

## Fluctuation theorems and inequalities generalizing the second law of thermodynamics out of equilibrium

Gatien Verley and David Lacoste

*ESPCI, Laboratoire PCT, 10 rue Vauquelin, 75005 Paris, France*

(Received 20 June 2012; published 26 November 2012)

We present a general framework for systems which are prepared in a nonstationary nonequilibrium state in the absence of any perturbation and which are then further driven through the application of a time-dependent perturbation. By assumption, the evolution of the system must be described by Markovian dynamics. We distinguish two different situations depending on the way the nonequilibrium state is prepared; either it is created by some driving or it results from a relaxation following some initial nonstationary conditions. Our approach is based on a recent generalization of the Hatano-Sasa relation for nonstationary probability distributions. We also investigate whether a form of the second law holds for separate parts of the entropy production and for any nonstationary reference process, a question motivated by the work of M. Esposito *et al.* [*Phys. Rev. Lett.* **104**, 090601 (2010)]. We find that although the special structure of the theorems derived in this reference is not recovered in the general case, detailed fluctuation theorems still hold separately for parts of the entropy production. These detailed fluctuation theorems contain interesting generalizations of the second law of thermodynamics for nonequilibrium systems.

DOI: [10.1103/PhysRevE.86.051127](https://doi.org/10.1103/PhysRevE.86.051127)

PACS number(s): 05.40.–a, 05.70.Ln

### I. INTRODUCTION

In recent years, a broad number of works summarized under the name of fluctuations theorems have led to significant progress in our understanding of the second law of thermodynamics [1–3]. A central idea, namely the application of thermodynamics at the level of trajectories, has developed into a field of its own, called stochastic thermodynamics [2,4].

In a similar spirit as the Crooks relation [5], the total entropy production can be expressed as the relative entropy between the probability distributions of trajectories associated with a forward and backward experiment [6–8]. As a consequence, the entropy production quantifies the time-symmetry breaking and reversibility which means zero entropy production, only occurs when the forward and backward experiments are undistinguishable. While this statement for the entropy production encompasses the second law after averaging over many trajectories, it also provides additional implications at the trajectory level.

This particular idea has also played a central role in recent developments of the framework of fluctuation relations for systems operating under feedback control [9]. A generalization of the Jarzynski relation [10], including the transfer of information due to feedback predicted theoretically in this reference, has been tested experimentally [11]. With these concepts, it is possible to reinterpret Landauer’s principle linking information and thermodynamics [12] and devise new experiments to test it in a particularly elegant and direct way [13]. Besides providing new insights into the deep connection between thermodynamics and information, progresses in stochastic thermodynamics make it possible to address optimization problems which should be relevant for many applications [14].

In previous work, we have analyzed some consequences of a generalized Hatano-Sasa relation, in which the stationary distribution entering the original Hatano-Sasa relation [15] is replaced by a nonstationary one. In Ref. [16], we have

shown that this approach is particularly adapted to obtain a modified fluctuation-dissipation theorem valid near an arbitrary nonequilibrium state; in Ref. [17], we have also derived from it an interesting generalization of the second law of thermodynamics for nonstationary states. Such generalizations of the second law of thermodynamics and of the fluctuation-dissipation theorem are useful to describe the following situations: (i) the system is driven by at least two control parameters, so even when the driving of interest  $h$  is constant in time, the probability distribution remains nonstationary, and (ii) the system undergoes a transient regime due to the choice of initial conditions, and before the relaxation of this transient regime is finished, the system is further driven. We note that the second situation is typical of systems with a slow relaxation time, such as aging systems, in which case the system never reaches a stationary state on any reasonable time. Therefore, it seems to us that this framework should be ideally suited to analyze aging systems.

In this paper, we provide a more detailed analysis of the results of Ref. [17], and we add some new applications. The first section contains preliminaries on fluctuation theorems. We then discuss a particular point concerning the symmetry property that the initial and final probability distributions should have for a detailed fluctuation theorem for the total entropy production to hold. Although this particular point is known in the literature [3], it has been overlooked in many other works in the field. For this reason, it seemed useful to us to revisit this rather subtle point. In the next section, we discuss extensions of the nonadiabatic and adiabatic entropy productions which were introduced in Ref. [18] for the case of a stationary reference process. We find that the special structure of the “three theorems” derived in this reference is not recovered in the general case of a nonstationary reference process. We interpret this as being due to the contribution of a new time-symmetric contribution in the dynamical action, which takes a form similar to the traffic introduced in Ref. [19]. We then discuss the second-law-like inequalities which follow from the integral

fluctuation theorems and which should be applicable to a broad class of nonequilibrium systems. In the last section, we present some illustrative examples of these ideas, using a two-state model or a particle in an harmonic potential submitted to Langevin dynamics.

## II. FLUCTUATION THEOREMS FROM GENERAL CONSIDERATIONS OF TIME-REVERSAL SYMMETRY

### A. Stochastic modeling and definitions

We consider a system which is assumed to evolve according to a continuous-time Markovian dynamics of a pure jump type [20]. Let us introduce the transition rate  $w_t(c, c')$  for the rate to jump from a state  $c$  to a state  $c'$  at time  $t$ . The subscript  $t$  in  $w_t(c, c')$  indicates that there are processes which are nonstationary even in the absence of explicit driving. The origin of such processes is arbitrary, and they can result from an additional underlying driving, which differs from explicit driving and does not need to be specified. Note that if the system is submitted to an initial quench and a constant driving (explicit or not), the rates are time independent but evolution is still nonstationary due to the initial quench. At time  $t = 0$ , an arbitrary explicit driving protocol  $h_t$  is applied to the system, and we denote by  $p_t(c, [h_t])$  the probability to observe the system in the state  $c$  at a time  $t$  in the presence of this driving. The evolution of the system for  $t > 0$  is controlled by the generator  $L_t^{h_t}$ , which is defined by

$$L_t^{h_t}(c', c) = w_t^{h_t}(c', c) - \delta(c, c') \sum_{c''} w_t^{h_t}(c', c''), \quad (1)$$

where  $w_t^{h_t}(c', c)$  is a transition rate in the presence of the driving  $[h_t]$ . Then  $p_t(c, [h_t])$  is the solution of

$$\frac{dp_t(c, [h_t])}{dt} = \sum_{c'} p_t(c', [h_t]) L_t^{h_t}(c', c). \quad (2)$$

The notation  $p_t(c, [h_t])$  emphasizes that this probability distribution depends functionally on the whole protocol history  $[h_t]$  up to time  $t$ . We assume that at  $t = 0$  there is no driving, so  $p_0(c, [h_0]) = p_0(c)$ . We also note that, in practice, the driving  $[h_t]$  may not start immediately at  $t = 0^+$  but may be turned on only later, after a certain time, called the waiting time in the context of aging systems.

We now introduce a different probability distribution denoted  $\pi_t(c, h)$ , which represents the probability to observe the system in the state  $c$  at a time  $t > 0$  in the presence of a constant (time independent) driving  $h$ . In other words,  $\pi_t(c, h)$  follows from  $p_t(c, [h_t])$  by freezing the time dependence in the driving  $[h_t]$ . This distribution, which will play a key role in the following, obeys the master equation

$$\left( \frac{\partial \pi_t}{\partial t} \right) (c, h) = \sum_{c'} \pi_t(c', h) L_t^h(c', c). \quad (3)$$

From the fact that  $\pi_t(c, h)$  and  $p_t(c, [h_t])$  should coincide for a constant protocol, we deduce the initial condition to be  $\pi_0(c, h_0) = p_0(c)$ .

In Ref. [16], we have shown that one can construct with this distribution the following functional:

$$\mathcal{Y}[c] = - \int_0^T d\tau \dot{h}_\tau \partial_h \ln \pi_\tau(c_\tau, h_\tau), \quad (4)$$

which has clear similarities with the functionals introduced by Jarzynski [10] and Hatano-Sasa [15]. We find from the analysis of this paper that the functional  $\mathcal{Y}$  has the interpretation of the driving part in the total entropy production. Using a Feynman-Kac approach, which has also played a central role for the Jarzynski relation [21], we have shown in Ref. [16] that this functional  $\mathcal{Y}$  obeys a generalized Hatano-Sasa relation,

$$\langle \exp(-\mathcal{Y}[c]) \rangle = 1. \quad (5)$$

This relation qualifies for a generalization of the Hatano-Sasa relation because the stationary probability distribution  $p_{st}(c, h)$  which enters in the functional  $Y[c] = - \int_0^T dt \dot{h}_t \partial_h \ln p_{st}(c_t, h_t)$  in the standard Hatano-Sasa relation is now replaced by the more general distribution  $\pi_t(c, h)$ . From a linear expansion of this generalized Hatano-Sasa, we have obtained modified fluctuation-dissipation theorems valid near an arbitrary nonequilibrium state [16,22]. In the next sections, we derive this generalized Hatano-Sasa relation in a different way and we investigate other consequences not contained in such a linear expansion.

### B. Path probability distributions and action functional

Let us consider a trajectory  $[c] = (c_0, c_1, \dots, c_N; \tau_1, \dots, \tau_N)$  where the  $c_i$  are the states which are visited by the system and  $\tau_i$  are the jumping times to go from  $c_{i-1}$  to  $c_i$ . The total time range of the trajectory is  $[0 \dots T]$ . We denote  $\mathcal{P}[c]$ , the probability to observe such a trajectory  $[c]$ , also called path probability, below:

$$\mathcal{P}[c] = p_0(c_0) \left\{ \prod_{j=1}^N \exp \left[ - \int_{\tau_{j-1}}^{\tau_j} d\tau \lambda_\tau^{h_\tau}(c_{j-1}) \right] w_{\tau_j}^{h_{\tau_j}}(c_{j-1}, c_j) \right\} \times \exp \left[ - \int_{\tau_N}^T d\tau \lambda_\tau^{h_\tau}(c_N) \right], \quad (6)$$

where  $\lambda_\tau^{h_\tau}(c') = \sum_{c \neq c'} w_\tau^{h_\tau}(c', c)$  represents the escape rate to leave the state  $c'$  and  $p_0(c_0) = p_0(c_0, h_0)$  represents the probability distribution of the initial condition.

In the following, we consider several ratios of path probabilities of the form

$$\Delta \mathcal{A}[c] = \ln \frac{\mathcal{P}[c]}{\tilde{\mathcal{P}}[c^*]}, \quad (7)$$

where the tilde symbol ( $\tilde{\cdot}$ ) corresponds to a transformation of the original dynamics into a new dynamics. This new dynamics is defined by its own initial condition and by the transformed transition rates denoted by  $\tilde{w}$ . The  $(*)$  denotes a different transformation which acts on the trajectory itself. The transformed trajectory  $[c^*] = (c_0^*, c_1^*, \dots, c_N^*; \tau_1^*, \dots, \tau_N^*)$  results from the application of an involution on the trajectory  $[c]$ , which we assume to be either the identity ( $[c^*] = [c]$ ) or the time-reversal symmetry acting on the trajectories  $[[c^*] = [\bar{c}] = (c_N, c_{N-1}, \dots, c_0; T - \tau_N, \dots,$

$T - \tau_1$ ]. In other words, we have

$$c_i^* = \begin{cases} c_i & \text{if } * \text{ is identity} \\ c_{N-i} & \text{if } * \text{ is time reversal} \end{cases}, \quad (8)$$

$$\tau_i^* = \begin{cases} \tau_i & \text{if } * \text{ is identity} \\ \tau_{N-i+1} & \text{if } * \text{ is time reversal} \end{cases},$$

with the convention that  $\tau_0^*$  and  $\tau_{N+1}^*$  are, respectively, 0 and  $T$  when  $*$  is identity and are, respectively,  $T$  and 0 when  $*$  is the time-reversal symmetry. Substituting the trajectory  $[c^*]$  in replacement of  $[c]$ , and the rates of the modified dynamics  $\tilde{w}$  instead of the original rates  $w$  in Eq. (6), one obtains directly for the transformed path probability

$$\tilde{\mathcal{P}}[c^*] = \tilde{p}_0(c_0^*) \left\{ \prod_{j=1}^N \exp \left[ - \int_{\tau_{j-1}^*}^{\tau_j^*} dt \tilde{\lambda}_t^{h_t}(c_{j-1}^*) \right] \tilde{w}_{\tau_j^*}^{h_{\tau_j^*}}(c_{j-1}^*, c_j^*) \right\} \times \exp \left[ - \int_{\tau_N^*}^{\tau_{N+1}^*} dt \tilde{\lambda}_t^{h_t}(c_N^*) \right], \quad (9)$$

where  $\tilde{\lambda}_t^{h_t}(c') = \sum_{c \neq c'} \tilde{w}_t^{h_t}(c', c)$  represents the escape rate to leave the state  $c'$  in the dynamics modified via the operation tilde. From this, we see that  $\Delta\mathcal{A}[c]$  can be written as

$$\Delta\mathcal{A}[c] = \ln \frac{p_0(c_0)}{\tilde{p}_0(c_0^*)} - \int_0^T dt [\lambda_t^{h_t}(c_t) - \tilde{\lambda}_t^{h_t}(c_t)] + \sum_{j=1}^N \ln \frac{w_{\tau_j}^{h_{\tau_j}}(c_{j-1}, c_j)}{\tilde{w}_{\tau_j^*}^{h_{\tau_j^*}}(c_{j-1}^*, c_j^*)}, \quad (10)$$

with  $c_t = c_j$  if  $t \in [\tau_j, \tau_{j+1}[$  and  $\lambda_t^{h_t} = \lambda_{T-t}^{h_{T-t}}$  (or  $\lambda_t^{h_t} = \lambda_t^{h_t}$ ) if the involution  $*$  is the time reversal (or, respectively, if  $*$  is identity). Thus,  $\Delta\mathcal{A}[c]$  has three different contributions: The first term is a boundary term which depends only on the initial or final configurations, and the last term is a bulk term which depends on the whole trajectory. The second term is related to the notion of traffic [23], which represents the integral of the escape rate  $\lambda_t$  evaluated at the actual configuration  $c_t$  of the system at time  $t$ . In view of this property, the second term in Eq. (10) represents a difference of traffic between the original dynamics (which corresponds to  $\mathcal{P}$ ) and the transformed dynamics (which corresponds to  $\tilde{\mathcal{P}}$ ).

### C. Protocol-reversal symmetry and the probability distributions of the initial and final points

Fluctuations theorems can be derived from considerations of symmetry for an arbitrary observable and arbitrary initial and final probability distributions [24]. These choices of observables, of initial and final probability distributions, determine precisely which fluctuation theorem holds. In this construction, we emphasize that the fluctuation theorem takes a strong form if the initial and final probability distributions are related by a reversal of protocol and a weaker form if not [3]. Then two cases must be considered, either the initial and final path probabilities are not related by the reversal of the protocol and the transformation ( $\sim$ ) is not an involution or such a symmetry exists and the transformation is an involution. In the following, we discuss both cases separately:

(i) Let us, first, assume that  $\sim$  is not an involution. This occurs, for instance, when the initial condition does not satisfy  $\tilde{p}_0(c) = p_0(c)$ . Following Ref. [18], we consider

$$P(\Delta\mathcal{A}[c] = \Delta\mathcal{A}) = \sum_{[c]} \delta(\Delta\mathcal{A} - \Delta\mathcal{A}[c]) \mathcal{P}[c], \quad (11)$$

$$= \exp(\Delta\mathcal{A}) \sum_{[c]} \delta(\Delta\mathcal{A} - \Delta\mathcal{A}[c]) \tilde{\mathcal{P}}[c^*], \quad (12)$$

$$= \exp(\Delta\mathcal{A}) \tilde{P}(\Delta\mathcal{A}[c^*] = \Delta\mathcal{A}), \quad (13)$$

with

$$\tilde{P}(\Delta\mathcal{A}[c^*] = \Delta\mathcal{A}) = \sum_{[c]} \delta(\Delta\mathcal{A} - \Delta\mathcal{A}[c^*]) \tilde{\mathcal{P}}[c]. \quad (14)$$

With words,  $\tilde{P}(\Delta\mathcal{A}[c^*] = \Delta\mathcal{A})$  corresponds to the probability to have on a given trajectory  $[c^*]$ ,  $\Delta\mathcal{A}[c^*]$  equal to  $\Delta\mathcal{A}$  in the tilde experiment (or dynamics). When comparing with the expression of  $P(\Delta\mathcal{A}[c] = \Delta\mathcal{A})$ , it appears that the same function  $\Delta\mathcal{A}$  is evaluated on different trajectories ( $[c]$  or  $[c^*]$ ), which are themselves generated by different dynamics (the original dynamics or the tilde dynamics). Thus, the probability  $\tilde{P}(\Delta\mathcal{A}[c^*] = \Delta\mathcal{A})$  cannot be defined in itself, i.e., without reference to the quantity  $\Delta\mathcal{A}$  introduced in the original dynamics [3]. For this reason, we regard the detailed fluctuation theorem (DFT) of Eq. (13) as a weak version of the theorem.

(ii) Let us then assume that the operation ( $\sim$ ) is an involution acting on the path probabilities,  $\tilde{\tilde{\mathcal{P}}} = \mathcal{P}$ . This implies that the distribution of initial condition satisfies the condition  $\tilde{\tilde{p}}_0(c) = p_0(c)$  and that the transition rates satisfy  $\tilde{\tilde{w}}_t^{h_t}(c, c') = w_t^{h_t}(c, c')$ . From these two conditions or, equivalently, directly from the definition Eq. (7), it follows that

$$\Delta\mathcal{A}[c] = -\Delta\tilde{\mathcal{A}}[c^*], \quad (15)$$

where  $\Delta\tilde{\mathcal{A}}[c] = \ln \tilde{\mathcal{P}}[c] / \tilde{\tilde{\mathcal{P}}}[c^*]$ . With this symmetry property, the fluctuation relation for  $\Delta\mathcal{A}$  now takes the form

$$P(\Delta\mathcal{A}[c] = \Delta\mathcal{A}) = \exp(\Delta\mathcal{A}) \sum_{[c]} \delta(\Delta\mathcal{A} + \Delta\tilde{\mathcal{A}}[c^*]) \tilde{\mathcal{P}}[c^*], \quad (16)$$

$$= \exp(\Delta\mathcal{A}) \tilde{P}(\Delta\tilde{\mathcal{A}}[c] = -\Delta\mathcal{A}), \quad (17)$$

with

$$\tilde{P}(\Delta\tilde{\mathcal{A}}[c] = -\Delta\mathcal{A}) = \sum_{[c]} \delta(\Delta\mathcal{A} + \Delta\tilde{\mathcal{A}}[c]) \tilde{\mathcal{P}}[c], \quad (18)$$

which corresponds with words to the probability to have on a given trajectory  $[c]$ ,  $\Delta\tilde{\mathcal{A}}[c]$  equal to  $-\Delta\mathcal{A}$  in the tilde experiment or dynamics. As expected, one can obtain directly Eq. (18) from Eq. (14) using Eq. (15). The main difference with the previous case where the tilde was not an involution is that now it is not the same function which must be evaluated in the two experiments (or dynamics) characterized by  $P$  (respectively,  $\tilde{P}$ ); rather, it is two different functions, namely  $\Delta\mathcal{A}[c]$  and  $\Delta\tilde{\mathcal{A}}[c]$ , but they are related because they represent the same physical quantity which takes different forms on

each dynamics. This is similar to the Crooks relation [5,25], where the same physical concept, namely the dissipated work, must be evaluated in the direct and tilde dynamics (although the precise function which represents this physical concept takes a different form in both cases). The main point is that here, unlike in the previous case, the function which must be evaluated is linked to the process (direct or reversed) under consideration. We thus regard Eq. (17) as a strong form of the detailed fluctuation theorem.

As a particular important illustration of this point, we discuss below the detailed fluctuation theorem satisfied by the entropy production. To do so, we consider both involutions introduced above, namely  $(*)$  and  $(\sim)$ , to represent a reversal symmetry, respectively, the reversal of trajectories and of protocol, which we both denote with a bar  $(-)$ . We recall that the effect of this symmetry must be considered separately on the trajectories and on the dynamics. The rates which control the dynamics are transformed as

$$\bar{w}_\tau^{h_i}(c, c') = w_{T-\tau}^{h_i}(c, c'), \quad (19)$$

since the order in the visited configurations is not affected by the transformation while the time dependence of the driving is. Therefore, one can think of this transformation as basically a time reversal of all protocols (the driving  $[h_i]$  and the other protocols represented by the extra subscript in the rates). Note also that Eq. (19) represents a transformation for the rates, which is always an involution unlike the full reversal of the path probabilities, which may or may not be an involution, depending on the initial conditions. This point is very relevant for the existence of a detailed fluctuation theorem for the entropy production. Indeed, in order to identify  $\mathcal{A}$  as entropy production, the initial probability distribution of the reversed process must correspond to the final probability distribution reached by the direct process [24]. In other words, one must choose  $\bar{p}_0(\bar{c}_0) = p_T(c_T)$ , where  $p_T$  is the solution of the master equation or Fokker Planck equation at time  $T$ . From Eq. (10), due to the vanishing of the second term, one obtains the familiar result [6,24]

$$\Delta S_{\text{tot}}[c] = \ln \frac{\mathcal{P}[c]}{\bar{\mathcal{P}}[\bar{c}]} = \Delta S + \sum_{j=1}^N \ln \frac{w_{\tau_j}^{h_i}(c_{j-1}, c_j)}{w_{\tau_j}^{h_i}(c_j, c_{j-1})}, \quad (20)$$

where the first term  $\Delta S = \ln p_0(c_0) - \ln p_T(c_T)$  represents the change in system stochastic entropy while the second term represents the change in reservoir entropy  $\Delta S_r[c]$  along the specified trajectory  $[c]$ .

In view of the discussion above, it is not obvious that the transformation of the full path probability denoted  $(-)$  as defined above is an involution in the particular case of the entropy production. Only when additional assumptions are made, namely that the initial and final probability distributions are related by a reversal of the protocol, can this transformation be an involution. Incidentally, this condition means equivalently that the system stochastic entropy  $\Delta S$  is antisymmetric with respect to a reversal of the protocol. When this is the case, one obtains from Eq. (17) the following detailed fluctuation relation:

$$\ln \frac{P(\Delta S_{\text{tot}}[c] = \Delta S_{\text{tot}})}{\bar{P}(\Delta \bar{S}_{\text{tot}}[\bar{c}] = -\Delta S_{\text{tot}})} = \Delta S_{\text{tot}}, \quad (21)$$

which many authors as Ref. [18] have denoted using a simplified notation,

$$\ln \frac{P(\Delta S_{\text{tot}})}{\bar{P}(-\Delta S_{\text{tot}})} = \Delta S_{\text{tot}}. \quad (22)$$

Note that this relation takes the form of the Evans and Searles theorem [26] in the following particular cases: (i) for nonequilibrium stationary processes and (ii) for processes generated by time-symmetric driving protocols with the additional condition that the initial and final conditions are related by the reversal of the protocol.

When  $p_0(c_0)$  and  $\bar{p}_0(\bar{c}_0)$  are not related by a protocol reversal, the detailed fluctuation theorem for the entropy production only holds in its weak form, namely Eq. (13). As explained above, this means that the quantity which enters this detailed fluctuation theorem for the reversed process is not the entropy production of that process.

#### D. Dual dynamics and difference of traffic

We now introduce a new transformation, called a duality transformation, which acts specifically on the dynamics of the process. In the following, this transformation is denoted with a hat  $(\hat{\ })$ . Analogously with the way this dual transformation has been introduced in the stationary case [15,18], we define the dual dynamics from the original dynamics by substituting the original rates  $w_\tau^h(c, c')$  by

$$\hat{w}_\tau^h(c, c') = \frac{w_\tau^h(c', c)\pi_\tau(c', h)}{\pi_\tau(c, h)}. \quad (23)$$

From this definition, it is not obvious that the duality transformation is an involution although it is indeed the case, as we show in Appendix. The basic idea is that this transformation essentially reverses the probability currents defined with respect to  $\pi_i(c, h)$ , and, because of this, it follows that this transformation is an involution. The proof also confirms that the dynamics constructed from the dual rates is Markovian. The generator of the dynamics still verify  $\sum_{c'} \hat{L}_i^{h_i}(c, c') = 0$ , where we have defined  $\hat{L}_i^{h_i}$  as in Eq. (1) substituting the rates  $w_\tau^{h_i}$  by the rates  $\hat{w}_\tau^{h_i}$ . The normalization of the probability distribution is, thus, conserved in time.

An important property of the probability distribution  $\pi_\tau(c, h)$ , justifying its use to define the duality transform, is that it is related to the difference between the escape rates of the direct and dual dynamics, because

$$\begin{aligned} \hat{\lambda}_\tau^h(c) - \lambda_\tau^h(c) &= \sum_{c' \neq c} [\hat{w}_\tau^h(c, c') - w_\tau^h(c, c')], \end{aligned} \quad (24)$$

$$= \sum_{c' \neq c} [\pi_\tau^{-1}(c, h)w_\tau(c', c)\pi_\tau(c', h) - w_\tau^h(c, c')], \quad (25)$$

$$= \sum_{c'} \pi_\tau^{-1}(c, h)w_\tau(c', c)\pi_\tau(c', h) - \sum_{c''} w_\tau^h(c, c''), \quad (26)$$

$$= \sum_{c'} \pi_\tau^{-1}(c, h) \left[ w_\tau(c', c) - \delta_{c'} \sum_{c''} w_\tau^h(c', c'') \right] \pi_\tau(c', h), \quad (27)$$

$$= \pi_\tau^{-1}(c, h) (\partial_\tau \pi_\tau)(c, h) = (\partial_\tau \ln \pi_\tau)(c, h), \quad (28)$$

where, in the last step, we used the evolution equation (3). We define the difference of traffic between the direct and dual dynamics as

$$\begin{aligned}\Delta\mathcal{T}[c] &= \int_0^T d\tau [\lambda_\tau^{h_\tau}(c_\tau) - \hat{\lambda}_\tau^{h_\tau}(c_\tau)] \\ &= - \int_0^T d\tau (\partial_\tau \ln \pi_\tau)(c_\tau, h_\tau).\end{aligned}\quad (29)$$

From this last expression, we note that the difference of traffic vanishes when the reference probability  $\pi_t$  is stationary and that this quantity is *antisymmetric* under the duality transformation  $\Delta\tilde{\mathcal{T}}[c] = -\Delta\mathcal{T}[c]$  but *symmetric* under the combined action of the reversal of the trajectories and of the full protocol (which we regard as the total-reversal symmetry),

$$\Delta\tilde{\mathcal{T}}[\bar{c}] = \int_0^T d\tau [\bar{\lambda}_\tau^{h_\tau}(\bar{c}_\tau) - \hat{\bar{\lambda}}_\tau^{h_\tau}(\bar{c}_\tau)], \quad (30)$$

$$= \int_0^T d\tau [\lambda_{T-\tau}^{h_{T-\tau}}(c_{T-\tau}) - \hat{\lambda}_{T-\tau}^{h_{T-\tau}}(c_{T-\tau})], \quad (31)$$

$$= \Delta\mathcal{T}[c]. \quad (32)$$

In the end, combining the total-reversal symmetry and the duality transform together, we obtain that  $\Delta\hat{\tilde{\mathcal{T}}}[\bar{c}] = -\Delta\mathcal{T}[c]$ .

### E. Adiabatic and nonadiabatic entropy productions

A system can fall into a nonequilibrium state by two mechanisms: either (i) detailed balance can be broken due to, for instance, boundary conditions or (ii) the system can be driven. Building on a number of works on steady-state thermodynamics [15,27–29], it was shown in Ref. [18] that these two different ways to put a system in a nonequilibrium state correspond to two separate contributions in the entropy production, called adiabatic for case (i) and nonadiabatic for case (ii). Note that this term *adiabatic* does not refer to the absence of heat exchange but rather to the fact that this contribution is the only one which remains in the adiabatic limit of very slow driving. In this reference, it was shown that, surprisingly, both terms can be expressed as logratios of probabilities, which implies that both quantities satisfy separately a DFT. This property is surprising because it is not expected to hold for a general splitting of the entropy production. Indeed, it does not hold, for instance, for the splitting of the entropy production into system entropy and reservoir entropy [24]. As a further consequence of these DFTs, both the adiabatic part and the nonadiabatic are positive on average, which means that the second law can be split into these two components.

In this section, we generalize the notions of adiabatic and nonadiabatic entropy productions defined as in Refs. [18,30] for the stationary case by replacing the stationary distribution by the distribution  $\pi_t(c, h)$  defined in Eq. (3). We obtain the following splitting:

$$\Delta S_{\text{na}}[c] = \ln \frac{p_0(c_0)}{p_T(c_T, [h_T])} + \sum_{j=1}^N \ln \frac{\pi_{\tau_j}(c_j, h_{\tau_j})}{\pi_{\tau_j}(c_{j-1}, h_{\tau_j})} \quad (33)$$

and

$$\Delta S_a[c] = \sum_{j=1}^N \ln \frac{w_{\tau_j}^{h_{\tau_j}}(c_{j-1}, c_j) \pi_{\tau_j}(c_{j-1}, h_{\tau_j})}{w_{\tau_j}^{h_{\tau_j}}(c_j, c_{j-1}) \pi_{\tau_j}(c_j, h_{\tau_j})}, \quad (34)$$

so we still have  $\Delta S_{\text{tot}}[c] = \Delta S_{\text{na}}[c] + \Delta S_a[c]$ .

We note that the adiabatic entropy production verifies  $\Delta\tilde{S}_a[\bar{c}] = -\Delta S_a[c]$ , which means that it is antisymmetric with respect to the combination of the protocol reversal and the time reversal of the trajectories, a transformation that we call total reversal. On the other side, the nonadiabatic entropy production is antisymmetric under the total reversal, i.e.,  $\Delta\tilde{S}_{\text{na}}[\bar{c}] = -\Delta S_{\text{na}}[c]$ , when the total entropy is (provided the appropriate condition on the initial and final states holds as explained in the previous section). We can also define an excess entropy production  $\Delta S_{\text{ex}}$  such that  $\Delta S_{\text{na}}[c] = \Delta S + \Delta S_{\text{ex}}[c]$  and  $\Delta S_a[c] = \Delta S_r[c] - \Delta S_{\text{ex}}[c]$ .

It is natural to ask at this point whether  $\Delta S_{\text{na}}$  and  $\Delta S_a$  separately satisfy a DFT. These quantities are not *a priori* of the form of Eq. (7), except for the particular case studied in Ref. [18], where the reference is stationary, so  $\Delta S_{\text{na}}$  and  $\Delta S_a$  should, thus, not in general satisfy separately a DFT. We thus loose, with the definition of Eqs. (33) and (34), the positivity of the mean adiabatic and nonadiabatic entropy productions. Despite this, we will see that their joint probability distribution still satisfies a DFT, as explained in Sec. III A.

### F. Nonadiabatic and adiabatic action functionals

In this section, we show that the difference of traffic  $\Delta\mathcal{T}$  introduced above is a key observable which can be used to construct quantities which satisfy a DFT. We start from the two possible decompositions of the entropy production as

$$(A) \quad \Delta S_{\text{tot}}[c] = \ln \frac{\mathcal{P}[c]}{\hat{\mathcal{P}}[\bar{c}]} + \ln \frac{\hat{\mathcal{P}}[\bar{c}]}{\bar{\mathcal{P}}[\bar{c}]} \quad (35)$$

$$\text{or} \quad (B) \quad \Delta S_{\text{tot}}[c] = \ln \frac{\mathcal{P}[c]}{\hat{\mathcal{P}}[c]} + \ln \frac{\hat{\mathcal{P}}[c]}{\bar{\mathcal{P}}[\bar{c}]}.$$

We first remark that, contrary to the case of Ref. [18], where the stationary probability distribution is chosen as a reference, the two decompositions are not equivalent. This is due to the fact that the two terms in the decomposition are no longer antisymmetric under total reversal, since they contain a nonzero difference of traffic term, defined above,

$$\ln \frac{\hat{\mathcal{P}}[\bar{c}]}{\bar{\mathcal{P}}[\bar{c}]} \neq \ln \frac{\mathcal{P}[c]}{\hat{\mathcal{P}}[c]} \quad \text{and} \quad \ln \frac{\hat{\mathcal{P}}[c]}{\bar{\mathcal{P}}[\bar{c}]} \neq \ln \frac{\mathcal{P}[c]}{\hat{\mathcal{P}}[\bar{c}]}.$$

#### I. Case A

We first focus on the first term in the right-hand side of Eq. (35A), which we call the nonadiabatic action  $\Delta A_{\text{na}}[c]$ . Using Eq. (10) with the choice  $\hat{\mathcal{P}} = \hat{\mathcal{P}}$  for the path probabilities and  $c^* = \bar{c}$  for the trajectories, we obtain

$$\begin{aligned}\Delta A_{\text{na}}[c] &= \ln \frac{\mathcal{P}[c]}{\hat{\mathcal{P}}[\bar{c}]} = \ln \frac{p_0(c_0)}{\hat{p}_0(\bar{c}_0)} - \int_0^T d\tau [\lambda_\tau^{h_\tau}(c_\tau) - \hat{\lambda}_\tau^{h_\tau}(c_\tau)] \\ &\quad + \sum_{j=1}^N \ln \frac{w_{\tau_j}^{h_{\tau_j}}(c_{j-1}, c_j)}{\hat{w}_{\tau_j}^{h_{\tau_j}}(c_j, c_{j-1})}.\end{aligned}\quad (36)$$

Given the initial condition  $\hat{p}_0(\bar{c}_0) = p_T(c_T)$  for the dual reversed experiment, the first term in this equation corresponds to what we have denoted before,  $\Delta S = \ln p_0(c_0) - \ln p_T(c_T)$ . Using Eqs. (23) and (33), we obtain

$$\begin{aligned} \Delta A_{\text{na}}[c] &= \Delta S + \int_0^T d\tau (\partial_\tau \ln \pi_\tau)(c_\tau, h_\tau) \\ &\quad + \sum_{j=1}^N \ln \frac{\pi_{\tau_j}(c_j, h_{\tau_j})}{\pi_{\tau_j}(c_{j-1}, h_{\tau_j})}, \\ &= \Delta S_{\text{na}}[c] - \Delta \mathcal{T}[c], \end{aligned} \quad (37)$$

which corresponds to a decomposition into two terms, where the first term,  $\Delta S_{\text{na}}[c]$ , is antisymmetric and the second term,  $\Delta \mathcal{T}[c]$ , is symmetric under total reversal. Alternatively, we can also write the same quantity as

$$\begin{aligned} \Delta A_{\text{na}}[c] &= \Delta S + \int_0^T d\tau (\partial_\tau \ln \pi_\tau)(c_\tau, h_\tau) \\ &\quad + \sum_{j=1}^N \ln \frac{\pi_{\tau_j}(c_j, h_{\tau_j})}{\pi_{\tau_j}(c_{j-1}, h_{\tau_j})}, \\ &= \Delta S - \Delta \psi + \int_0^T d\tau \partial_\tau (\psi_\tau(c_\tau, h_\tau)) \\ &\quad + \int_0^T d\tau (\partial_\tau \ln \pi_\tau)(c_\tau, h_\tau) + \sum_{j=1}^N \ln \frac{\pi_{\tau_j}(c_j, h_{\tau_j})}{\pi_{\tau_j}(c_{j-1}, h_{\tau_j})} \\ &= \Delta S_b + \mathcal{Y}[c], \end{aligned} \quad (38)$$

where  $\Delta S_b = \Delta S - \Delta \psi$  is a boundary term, with  $\Delta \psi = -\ln \pi_T(c_T, h_T) + \ln \pi_0(c_0, h_0)$ .

Therefore, since  $p_0(c_0) = \pi_0(c_0, h_0)$ , by construction,

$$\Delta S_b = \ln \frac{\pi_T(c_T, h_T)}{p_T(c_T)}. \quad (40)$$

As a result, the average of  $\Delta S_b$ , is related to the Kullback-Leibler divergence between the distributions  $\pi_T$  and  $p_T$ . Physically, this quantity can be viewed as a measure of the lag between the two distributions, in the same way that one can look at the dissipated work as a measure of the lag between the actual distribution at time  $t$  and the corresponding equilibrium distribution with the control parameter at the same value [31]. When there is no lag, either because  $p_T$  has relaxed towards  $\pi_T$  or because the initial probability distribution of the reversed protocol, namely  $p_T(c_T)$ , is chosen to be  $\pi_T(c_T, h_T)$ , then the two distributions are identical and  $\Delta S_b$  vanishes. In this case,  $\Delta A_{\text{na}}[c] = \mathcal{Y}[c]$ , which satisfies the symmetry condition  $\hat{\mathcal{Y}}[\bar{c}] = -\mathcal{Y}[c]$ . Therefore, from Eqs. (7) and (17), this quantity satisfies a DFT,

$$\ln \frac{P(\mathcal{Y}[c] = \mathcal{Y})}{\hat{P}(\hat{\mathcal{Y}}[\bar{c}] = -\mathcal{Y})} = \mathcal{Y}. \quad (41)$$

If we do not have a vanishing boundary term, then, unfortunately, only the weak fluctuation theorem of Eq. (13) is verified,

$$\ln \frac{P(\Delta A[c] = \Delta A_{\text{na}})}{\hat{P}(\Delta A[\bar{c}] = \Delta A_{\text{na}})} = \Delta A_{\text{na}}. \quad (42)$$

We now look at the second term in Eq. (35A), namely

$$\Delta A_a[c] = \Delta S_{\text{tot}}[c] - \Delta A_{\text{na}}[c] = \ln \frac{\hat{\mathcal{P}}[\bar{c}]}{\bar{\mathcal{P}}[\bar{c}]}. \quad (43)$$

We can rewrite Eq. (43) as

$$\begin{aligned} \Delta A_a[c] &= \int_0^T d\tau [\lambda_\tau^{h_\tau}(c_\tau) - \hat{\lambda}_\tau^{h_\tau}(c_\tau)] \\ &\quad + \sum_{j=1}^N \ln \frac{\hat{w}_{T-\tau_j}^{h_{T-\tau_j}}(c_j, c_{j-1})}{\bar{w}_{T-\tau_j}^{h_{T-\tau_j}}(c_j, c_{j-1})}, \end{aligned} \quad (44)$$

$$\begin{aligned} &= - \int_0^T d\tau (\partial_\tau \ln \pi_\tau)(c_\tau, h_\tau) \\ &\quad + \sum_{j=1}^N \ln \frac{w_{\tau_j}^{h_{\tau_j}}(c_{j-1}, c_j) \pi_{\tau_j}(c_{j-1}, h_{\tau_j})}{\bar{w}_{\tau_j}^{h_{\tau_j}}(c_j, c_{j-1}) \pi_{\tau_j}(c_j, h_{\tau_j})}, \end{aligned} \quad (45)$$

$$= \Delta S_a[c] + \Delta \mathcal{T}[c], \quad (46)$$

which corresponds again to a decomposition where the first term,  $\Delta S_a[c]$ , is antisymmetric and the second term,  $\Delta \mathcal{T}[c]$ , is symmetric under total reversal. As a self-consistent check, we see that the difference of traffic,  $\Delta \mathcal{T}$ , in  $\Delta A_a$  exactly compensates an opposite contribution in  $\Delta A_{\text{na}}$  so

$$\Delta S_{\text{tot}} = \Delta A_a + \Delta A_{\text{na}} = \Delta S_{\text{na}} + \Delta S_a. \quad (47)$$

One important point is to realize that  $\Delta A_a$  is not of the form of Eq. (7) because it involves a modified path probability both at the numerator and the denominator in its definition. Therefore, a detailed fluctuation relation of the form of Eq. (41) is *not* verified for this quantity.

## 2. Case B

One can, however, find another DFT by starting from the splitting of the entropy production of Eq. (35B). We define the first term on the right-hand side by

$$\Delta B_a[c] = \ln \frac{\mathcal{P}[c]}{\hat{\mathcal{P}}[c]} = -\Delta \bar{A}_a[\bar{c}] = -\Delta \mathcal{T}[c] + \Delta S_a[c] \quad (48)$$

and we also introduce the quantity

$$\Delta B_{\text{na}}[c] = \ln \frac{\hat{\mathcal{P}}[c]}{\bar{\mathcal{P}}[\bar{c}]} = \Delta \hat{A}_{\text{na}}[c] = \Delta \mathcal{T}[c] + \Delta S_{\text{na}}[c]. \quad (49)$$

Here  $\Delta B_a$  plays a role similar to  $\Delta A_{\text{na}}$  since it, too, has the required form to satisfy a detailed FT, which is

$$\ln \frac{P(\Delta B_a[c] = \Delta B_a)}{\hat{P}(\Delta \hat{B}_a[\bar{c}] = -\Delta B_a)} = \Delta B_a. \quad (50)$$

As before for  $\Delta A_a$ , the remaining part in the total entropy, namely  $\Delta B_{\text{na}}$ , does not satisfy a detailed FT.

## G. Some limiting cases of interest

In this section, we discuss some of the limiting cases for which the detailed fluctuation relations obtained above simplify. Let us assume that the driving starts at time  $t_{\text{di}} > 0$  and ends at time  $t_{\text{df}} < T$  for a total duration  $t_d = t_{\text{df}} - t_{\text{di}}$ .

(i) When  $\pi_\tau(c, h)$  relaxes very quickly to the stationary distribution (on a time scale  $\tau_{\text{st}}$  such that  $\tau_{\text{st}} \ll T$  and  $\tau_{\text{st}} \ll t_d$ ), one recovers from Eqs. (33) and (34) the usual definitions of

the nonadiabatic and adiabatic parts of the entropy production. In this case  $\Delta\mathcal{T} = 0$ , and, as a result, Eq. (41) and Eq. (50) become the usual DFTs satisfied by the nonadiabatic and adiabatic entropies, respectively [18].

(ii) In the limit of slow driving  $\dot{h}_t \simeq 0$ , which can happen without having  $\pi_t(c, h)$  relaxed to a stationary distribution, the driving part of the entropy production,  $\mathcal{Y}[c]$ , vanishes. Furthermore, the boundary term  $\Delta S_b$  also vanishes, because in this case  $p_t(c, [h_t])$  relaxes to  $\pi_t(c, h)$  since  $[h_t] \rightarrow h$ . In this limit  $\Delta A_{\text{na}} = 0$ , which justifies *a posteriori* the name nonadiabatic action for  $\Delta A_{\text{na}}$ . The vanishing of  $\Delta A_{\text{na}}$  has two further consequences; the first is that Eq. (36) implies  $\mathcal{P}[c] = \hat{\mathcal{P}}[\bar{c}]$ , in other words, the duality and total reversal compensate each other exactly. Another consequence is that  $\Delta A_a = \Delta B_a = \Delta S_{\text{tot}}$ , which implies that  $\Delta\mathcal{T} = 0$  although  $\pi_t(c, h)$  is time dependent. Furthermore, the fluctuation theorem of entropy production, namely Eq. (21), coincides with that for  $\Delta B_a$ , namely Eq. (50).

#### H. Modified second law for transition between nonstationary states

The observables introduced in Sec. IIF verify an integral fluctuation theorem and, therefore, are submitted to second-law-like inequalities, which are valid for an arbitrary nonequilibrium reference process. This results from the positivity of the Kullback-Leibler divergence between the path probabilities  $\mathcal{P}[c]$  and  $\hat{\mathcal{P}}[\bar{c}]$ ,

$$D(\mathcal{P}[c]||\hat{\mathcal{P}}[\bar{c}]) \equiv \sum_{[c]} \mathcal{P}[c] \ln \frac{\mathcal{P}[c]}{\hat{\mathcal{P}}[\bar{c}]} = \langle \Delta A_{\text{na}} \rangle \geq 0, \quad (51)$$

and between the distributions  $\mathcal{P}[c]$  and  $\hat{\mathcal{P}}[c]$ ,

$$D(\mathcal{P}[c]||\hat{\mathcal{P}}[c]) = \sum_{[c]} \mathcal{P}[c] \ln \frac{\mathcal{P}[c]}{\hat{\mathcal{P}}[c]} = \langle \Delta B_a \rangle \geq 0. \quad (52)$$

In view of Eqs. (37), (48), and (47), this implies

$$\langle \Delta S_{\text{na}} \rangle \geq \langle \Delta\mathcal{T} \rangle. \quad (53)$$

Furthermore, one also has  $\langle \Delta S_a \rangle \geq \langle \Delta\mathcal{T} \rangle$  and  $\langle \Delta S_{\text{tot}} \rangle \geq 0$ , which, taken together, imply  $\langle \Delta S_{\text{tot}} \rangle \geq \max(2\langle \Delta\mathcal{T} \rangle, 0)$ . However, all these inequalities are less binding than Eq. (53), because the adiabatic entropy production and the total entropy production are generally increasing functions of time, whereas the inequality Eq. (53) becomes an equality in the long time limit, as explained below. We note, furthermore, the following.

(i) Although we have shown that  $\langle \Delta A_{\text{na}} \rangle \geq 0$  and  $\langle \Delta B_a \rangle \geq 0$ , the corresponding conjugate quantities of  $\Delta A_{\text{na}}$  and  $\Delta B_a$  with respect to the total entropy production, namely  $\Delta A_a$  and  $\Delta B_{\text{na}}$ , do not have likewise a positive mean in general. That this should be the case can be understood from the consideration of a particular case, namely the case where the initial nonequilibrium condition has been prepared by the application of a protocol, which is exactly compensated by the second protocol (the perturbation) denoted  $[h_t]$  in this paper. In this case, the system is in equilibrium at all times in the presence of the perturbation. Since the rates satisfy the detailed balance condition, one can check explicitly that this implies  $\Delta A_a = -\Delta A_{\text{na}}$  as expected since the system is in equilibrium and  $\Delta S_{\text{tot}} = 0$ . It follows from this that,

$\langle \Delta A_a \rangle \leq 0$  and  $\langle \Delta B_{\text{na}} \rangle \leq 0$ , so in this case  $\Delta A_a$  and  $\Delta B_{\text{na}}$  do not have positive means.

(ii) The first inequality in Eq. (53) can be written

$$\langle \mathcal{Y} \rangle \geq D(p_T || \pi_T) \geq 0. \quad (54)$$

In other words, the average of the functional  $\mathcal{Y}$  is bounded by  $-\Delta S_b = D(p_T || \pi_T)$ , which is a measure of the lag between the distributions  $p_T$  and  $\pi_T$ . As noted before, a similar result holds for the dissipated work for the case of an initial equilibrium probability distribution [31]. Recalling the definition of the excess entropy,  $\Delta S_{\text{na}}[c] = \Delta S + \Delta S_{\text{ex}}[c]$ , one can also express this inequality as a Clausius-type inequality of the form

$$\langle \Delta S \rangle \geq -\langle \Delta S_{\text{ex}} \rangle + \langle \Delta\mathcal{T} \rangle, \quad (55)$$

which contains, as particular cases, the Clausius form of the second law for transitions between equilibrium states and a modified version of the second law for transitions between NESS [15].

(iii) As noted above, the equality in the inequality of Eq. (53) holds in the adiabatic limit for infinitely slow driving. In this limit the right-hand side of Eq. (54) is zero because there is no lag between the distribution  $p_T$  and  $\pi_T$ . The fact that the inequality can be saturated is essential for identifying Eq. (55) as a generalization of the second law of thermodynamics.

### III. FLUCTUATION THEOREMS FROM CONSIDERATION OF GENERATING FUNCTIONS

Generating functions provide an alternate way to understand fluctuation relations without considering trajectories explicitly. Let us introduce the generating functions of  $(\Delta\mathcal{T}, \Delta S_a)$  and of  $(\Delta\mathcal{T}, \Delta S_{\text{na}})$ , namely

$$g_T^{(a)}(c, \gamma, \epsilon) = \langle \delta(c - c_T) e^{-\gamma \Delta\mathcal{T}[c] - \epsilon \Delta S_a[c]} \rangle, \quad (56)$$

$$g_T^{(\text{na})}(c, \gamma, \epsilon) = \langle \delta(c - c_T) e^{-\gamma \Delta\mathcal{T}[c] - \epsilon \Delta S_{\text{na}}[c]} \rangle. \quad (57)$$

These quantities satisfy deformed master equations of the form

$$\begin{aligned} \partial_t g_t^{(a)}(c, \gamma, \epsilon) &= \sum_{c'} g_t^{(a)}(c', \gamma, \epsilon) \left[ \frac{w_t^{h_t}(c, c')}{\hat{w}_t^{h_t}(c, c')} \right]^\epsilon L_t^{h_t}(c', c) \\ &\quad + \gamma (\partial_t \ln \pi_t)(c, h_t) g_t^{(a)}(c, \gamma, \epsilon), \\ \partial_t g_t^{(\text{na})}(c, \gamma, \epsilon) &= \sum_{c'} g_t^{(\text{na})}(c', \gamma, \epsilon) \left[ \frac{\pi_t(c', h_t) p_t(c)}{\pi_t(c, h_t) p_t(c')} \right]^\epsilon L_t^{h_t}(c', c) \\ &\quad + [\gamma (\partial_t \ln \pi_t)(c, h_t) + \epsilon \partial_t \ln p_t(c)] g_t^{(\text{na})}(c, \gamma, \epsilon). \end{aligned} \quad (58)$$

$$+ [\gamma (\partial_t \ln \pi_t)(c, h_t) + \epsilon \partial_t \ln p_t(c)] g_t^{(\text{na})}(c, \gamma, \epsilon). \quad (59)$$

We can check that, in the special case where  $\gamma = -1$  and  $\epsilon = 1$ , the solutions are

$$g_T^{(a)}(c, -1, 1) = \hat{p}_T(c) = \langle \delta(c - c_T) e^{-\Delta B_a[c]} \rangle, \quad (60)$$

$$g_T^{(\text{na})}(c, -1, 1) = p_T(c) = \langle \delta(c - c_T) e^{-\Delta A_{\text{na}}[c]} \rangle, \quad (61)$$

where  $\hat{p}_t(c)$  is the solution of the master equation with generator  $\hat{L}_t^{h_t}$  as defined in Appendix. Note that Eq. (61) can be transformed to remove the boundary term in  $\Delta A_{\text{na}}[c]$  in the

following way:

$$p_T(c) = \langle \delta(c - c_T) e^{-\Delta S_b - \mathcal{Y}[c]} \rangle, \quad (62)$$

$$= \sum_{[c]} \mathcal{P}[c|c_0] p_0(c_0) \delta(c - c_T) \times \left[ \frac{p_T(c_T) \pi_0(c_0, h_0)}{p_0(c_0) \pi_T(c_T, h_T)} \right] e^{-\mathcal{Y}[c]}, \quad (63)$$

$$= \frac{p_T(c)}{\pi_T(c, h_T)} \sum_{[c]} \mathcal{P}[c] \delta(c - c_T) e^{-\mathcal{Y}[c]}, \quad (64)$$

so we finally get the result for the generating function of  $\mathcal{Y}[c]$  already obtained in Ref. [16],

$$\pi_T(c, h_T) = \langle \delta(c - c_T) e^{-\mathcal{Y}[c]} \rangle. \quad (65)$$

Through integration over  $c$ , we immediately obtain from Eq. (62) the integrated fluctuation theorem

$$\langle e^{-\Delta A_{na}[c]} \rangle = 1, \quad (66)$$

given that  $\Delta A_{na} = \Delta S_b + \mathcal{Y}$ . This integrated fluctuation theorem also follows directly from Eq. (42). Similarly, by integrating over  $c$  in Eq. (65), we have

$$\langle e^{-\mathcal{Y}[c]} \rangle = 1, \quad (67)$$

which is nothing but the generalized Hatano-Sasa relation given in Eq. (5). Using the Jensen inequality, we recover from these relations the second-law-like inequalities of the previous section.

#### A. Fluctuation theorems for joint probability distributions

As shown in Refs. [32,33], it is possible to derive fluctuation theorems for joint probability distributions of variables which form parts of the total entropy production, even when each variable does not satisfy separately a fluctuation theorem. This approach has many advantages as it offers a unifying principle to recover many fluctuation theorems. It is straightforward to apply this idea to the general case of an observable  $\Delta \mathcal{A}$  of the form of Eq. (7). We assume that this observable can be decomposed into a sum of  $m$  observables which are antisymmetric with respect to the combined action of the tilde and of the star involutions:  $\Delta \mathcal{A} = \sum_{i=1}^m \Delta \mathcal{A}_i$ , with  $\Delta \tilde{\mathcal{A}}_i[c^*] = -\Delta \mathcal{A}_i[c]$  for  $i = 1 \dots m$ . We now have

$$\begin{aligned} P(\Delta \mathcal{A}_1[c] = \Delta \mathcal{A}_1, \dots, \Delta \mathcal{A}_m[c] = \Delta \mathcal{A}_m) \\ &= \sum_{[c]} \prod_{i=1}^m \delta(\Delta \mathcal{A}_i - \Delta \mathcal{A}_i[c]) \mathcal{P}[c], \\ &= \sum_{[c^*]} \prod_{i=1}^m \delta(\Delta \mathcal{A}_i + \Delta \tilde{\mathcal{A}}_i[c^*]) \exp(\Delta \mathcal{A}) \tilde{\mathcal{P}}[c^*], \\ &= e^{\Delta \mathcal{A}} \tilde{P}(\Delta \tilde{\mathcal{A}}_1[c] = -\Delta \mathcal{A}_1, \dots, \Delta \tilde{\mathcal{A}}_m[c] \\ &= -\Delta \mathcal{A}_m), \end{aligned} \quad (68)$$

where the probability  $\tilde{P}$  is defined by

$$\begin{aligned} \tilde{P}(\Delta \tilde{\mathcal{A}}_1[c] = \Delta \mathcal{A}_1, \dots, \Delta \tilde{\mathcal{A}}_m[c] = \Delta \mathcal{A}_m) \\ &= \sum_{[c]} \tilde{\mathcal{P}}[c] \prod_{i=1}^m \delta(\Delta \mathcal{A}_i - \Delta \tilde{\mathcal{A}}_i[c]). \end{aligned}$$

Note that if the observables  $\Delta \mathcal{A}_i$  do not satisfy the antisymmetry property with respect to the combined action of the tilde and of the star involutions, we still have a weak form of the fluctuation theorem similar to Eq. (14).

In the particular case of the decomposition of  $\Delta S_{\text{tot}}$  into  $\Delta S_{na}$  and  $\Delta S_a$  that are antisymmetric by total reversal, we have

$$\ln \frac{P(\Delta S_a[c] = \Delta S_a, \Delta S_{na}[c] = \Delta S_{na})}{\tilde{P}(\Delta \tilde{S}_a[c] = -\Delta S_a, \Delta \tilde{S}_{na}[c] = -\Delta S_{na})} = \Delta S_a + \Delta S_{na}. \quad (69)$$

We can also apply the same idea on the decomposition  $\Delta A_{na} = \Delta S_{na} - \Delta \mathcal{T}$  obtained in Eq. (37). Since  $\Delta S_{na}$  and  $\Delta \mathcal{T}$  are both antisymmetric under the combination of the duality and the total reversal ( $\Delta \hat{S}_{na}[\bar{c}] = -\Delta S_{na}[c]$  and  $\Delta \hat{\mathcal{T}}[\bar{c}] = -\Delta \mathcal{T}[c]$ ), we have

$$\ln \frac{P(\Delta \mathcal{T}[c] = \Delta \mathcal{T}, \Delta S_{na}[c] = \Delta S_{na})}{\hat{P}(\Delta \hat{\mathcal{T}}[c] = -\Delta \mathcal{T}, \Delta \hat{S}_{na}[c] = -\Delta S_{na})} = \Delta S_{na} - \Delta \mathcal{T}. \quad (70)$$

In the same way, there is a DFT associated with the decomposition  $\Delta B_a = \Delta S_a - \Delta \mathcal{T}$  because both  $\Delta S_a$  and  $\Delta \mathcal{T}$  are also antisymmetric under the duality transformation ( $\Delta \hat{S}_a[c] = -\Delta S_a[c]$  and  $\Delta \hat{\mathcal{T}}[c] = -\Delta \mathcal{T}[c]$ ). Thus, we have

$$\ln \frac{P(\Delta \mathcal{T}[c] = \Delta \mathcal{T}, \Delta S_a[c] = \Delta S_a)}{\hat{P}(\Delta \hat{\mathcal{T}}[c] = -\Delta \mathcal{T}, \Delta \hat{S}_a[c] = -\Delta S_a)} = \Delta S_a - \Delta \mathcal{T}. \quad (71)$$

## IV. ILLUSTRATIVE EXAMPLES

In the following, we illustrate, using simple analytical models, the fluctuation relations and the modified second law discussed above. There are two main ways to create nonstationary reference distributions. These nonstationary distributions can be created due to either the choice of initial conditions or a driving force. We illustrate both cases with a driven two-states model, and we focus particularly on the case of sinusoidal driving. Besides this two state model, we also study a model for a particle in an harmonic potential and obeying Langevin dynamics.

### A. Two-states model dynamics

#### 1. Nonstationarity from relaxation due to the initial conditions

We consider a two-states model described by the following master equation:

$$\partial_t p_t(a) = -w^{h_t}(a, b) p_t(a) + w^{h_t}(b, a) p_t(b), \quad (72)$$

where the jump rate from state  $a$  to state  $b$  is denoted  $w^{h_t}(a, b)$  in the presence of the driving  $h_t$  and  $w(a, b)$  in the absence of this driving. We arbitrarily parametrize the rates as

$$w^{h_t}(a, b) = w(a, b) e^{-h_t/2} \quad \text{and} \quad w^{h_t}(b, a) = w(b, a) e^{h_t/2}, \quad (73)$$

where  $h_t$  can be thought of as a force which introduces a bias in the transitions rates. Note also that these rates depend on time only through  $h_t$ . In order to create a nonstationary distribution, we choose the initial probability distribution to be in state  $b$ ,  $p_0(b)$ , at an arbitrary value that differs from the steady-state



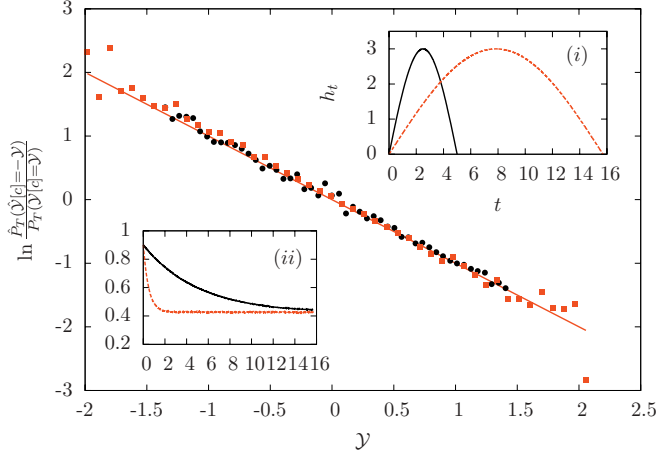


FIG. 1. (Color online) Illustration of the detailed fluctuation relation obeyed by the quantity  $\mathcal{Y}$ . Orange squares correspond to a long protocol with fast relaxation of the  $\pi_t$  function towards equilibrium (orange dashed line of the insets), whereas black dots are for a short protocol with slow relaxation towards equilibrium (black solid lines of the insets). Inset (i) shows the half sinusoidal protocols and inset (ii) shows the relaxation, as a function of time  $t$ , of the distribution  $\pi_t(b, h)$  towards equilibrium distribution for a given  $h$  value.

value  $\{p_{st}(b) = w(a, b)/[w(a, b) + w(b, a)]\}$ . As a result, even in the absence of driving, the system will relax in time.

For simplicity, we assume that the driving follows a half-sinusoidal protocol depicted in inset (i) of Fig. 1, which implies that both the driving and the transition rates are symmetric with respect to time. In inset (ii) of Fig. 1, we show the relaxation of  $\pi_t(b, h)$  towards the equilibrium distribution at a given value of  $h$  for two choices of the unperturbed rates. One relaxation is faster than the other one because the unperturbed rates are chosen to be larger.

In order to illustrate the DFT of Eq. (41), we have evaluated numerically the functions  $\pi_t(c, h)$  for different constant force protocols  $h$  with a kinetic Monte Carlo algorithm [34]. Using this data, we have generated an ensemble  $\hat{\mathcal{C}}$  of trajectories with the dual reversed dynamics. We have also built separately an ensemble  $\mathcal{C}$  of trajectories corresponding to the original dynamics. We then have measured the probability of  $\mathcal{Y}[c]$  with  $[c] \in \mathcal{C}$  and the probability of  $\hat{\mathcal{Y}}_t[c] = -\mathcal{Y}[\bar{c}]$  with  $[c] \in \hat{\mathcal{C}}$ , counting the number of times that the value of these functionals were in a given range  $[\mathcal{Y}, \mathcal{Y} + \delta\mathcal{Y}]$ . In Fig. 1, we verify the detailed fluctuation relation for  $\mathcal{Y}$  for the slow and the fast protocol of inset (i). In both cases, the initial condition of the dual reversed experiment was chosen to be  $\hat{p}_T(c) = \pi_T(c, h_T)$ , which means that, by construction,  $\Delta S_b = 0$  and  $\Delta A_{na}[c] = \mathcal{Y}[c]$ . We find that the probability distributions obtained from these simulations follow the expected symmetry. There is one practical difficulty in these simulations, which is also frequently encountered with other numerical tests of fluctuation theorems. One needs to find conditions such that the system is not too far from equilibrium to get a good overlap between  $P(\mathcal{Y})$  and  $\hat{P}(-\mathcal{Y})$  and, at the same time, sufficiently out of equilibrium so the probability distributions are distinct despite numerical errors.

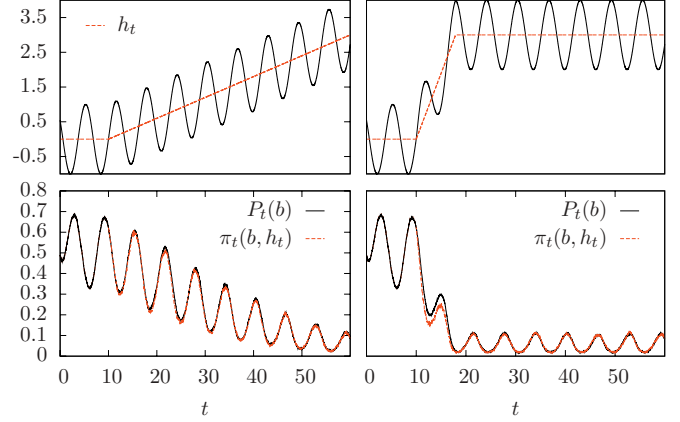


FIG. 2. (Color online) Transition between nonstationary oscillating states for different times of driving: (from left to right)  $t_d = 50$  and  $t_d = 7.91$ . (Top) Protocol imposed (orange dashed lines) and the total force applied on the system (black solid line). (Bottom) Exact (black) and accompanying probability distribution (orange) of state  $b$  as a function of time. For the shortest protocol, we see that the accompanying distribution differs from the exact solution while the protocol is changing.

## 2. Nonstationarity from periodic driving

To illustrate the inequalities generalizing the second law for transitions between nonstationary states obtained in Eq. (53), we use again the same two-states model but now with a different protocol. The shape of the protocols for the driving protocol  $h_t$  and the relaxation of  $p_t$  towards  $\pi_t$  is shown in Fig. 2: The protocol oscillates around an average value  $h_t^{\text{avg}}$  which evolves in time following a piecewise protocol of duration  $t_d = t_{\text{df}} - t_{\text{di}}$ . This average of the protocol  $h_t^{\text{avg}}$  represents the real driving which induces a transition from one nonstationary state to another one, while the oscillations around the average create these nonstationary states. As before, we have used kinetic Monte Carlo simulations to measure the functions  $\pi_t(a, h^{\text{avg}})$  for different values of  $h^{\text{avg}}$ . We then have carried out simulations with the time-dependent driving to obtain the quantities  $\langle \mathcal{Y} \rangle$ ,  $\langle \Delta \mathcal{T} \rangle$ , and  $\langle \Delta S_{na} \rangle$  at fixed final time  $T$  for different values of the duration of the driving  $t_d$ . As expected, we observe in Fig. 3 that  $\langle \Delta S_{na} \rangle \geq \langle \Delta \mathcal{T} \rangle$ . When we calculate  $\langle \Delta S_b \rangle$  at the final time  $T$ , we find a value close to zero irrespective of the duration of the protocol  $t_d$  because the system either has been driven so slowly that  $p_t(c, [h])$  has relaxed to  $\pi_t(c, h_t)$  already at the end of the protocol at  $t_{\text{df}}$  or the system has relaxed afterwards between the times  $t_{\text{df}}$  and  $T$ . This is compatible with  $\langle \mathcal{Y} \rangle \geq -\langle \Delta S_b \rangle$ , but in fact, since  $\langle \mathcal{Y} \rangle$  does not change between the time  $t_{\text{df}}$  and  $T$ , one can obtain a closer bound for  $\langle \mathcal{Y} \rangle$  by evaluating  $\langle \Delta S_b \rangle$  at the final time of the driving  $t_{\text{df}}$  instead of  $T$  as shown in Fig. 3. For this reason, in Fig. 3, we show  $\langle \mathcal{Y} \rangle$ ,  $\langle \Delta S_{na} \rangle$ , and  $\langle \Delta \mathcal{T} \rangle$  evaluated between time  $t_{\text{di}}$  and time  $T$ , whereas  $\langle \Delta S_b \rangle$  is evaluated at time  $t_{\text{df}}$ .

Furthermore, the behavior of  $\langle \mathcal{Y} \rangle$  is simple in two particular limits: when the driving is (i) very short ( $t_d \rightarrow 0$ ) or (ii) very slow. In case (i), one can show easily that  $\mathcal{Y} = \ln \pi_{t_{\text{di}}}(c_{t_{\text{di}}}, h_{t_{\text{di}}}) - \ln \pi_{t_{\text{df}}}(c_{t_{\text{df}}}, h_{t_{\text{df}}})$ . Since in this limit,  $\pi_{t_{\text{di}}} = p_{t_{\text{di}}}$ , one finds, after averaging over the distribution  $p_{t_{\text{di}}}$ , that  $\langle \mathcal{Y} \rangle \rightarrow D(p_{t_{\text{di}}} | \pi_{t_{\text{df}}})$ .

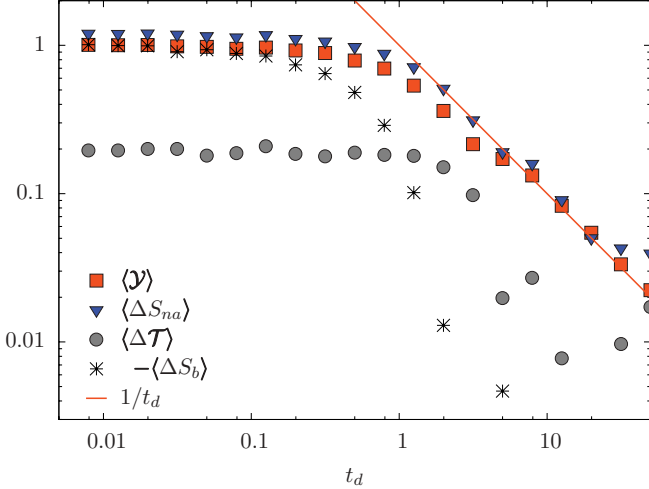


FIG. 3. (Color online) Illustration of the modified second law for nonstationary systems created by periodic driving. Symbols represent the driving entropy production  $\langle \mathcal{Y} \rangle$  (filled squares), the nonadiabatic entropy production  $\langle \Delta S_{na} \rangle$  (filled triangles), the difference of traffic  $\langle \Delta \mathcal{T} \rangle$  (bullets), and  $\langle \Delta S_b \rangle$  (stars) as a function of the duration of the driving  $t_d$ . Note that  $\langle \mathcal{Y} \rangle$ ,  $\langle \Delta S_{na} \rangle$ , and  $\langle \Delta \mathcal{T} \rangle$  are evaluated between the times  $t_{di}$  and the final time  $T$ , while  $\langle \Delta S_b \rangle$  is evaluated at time  $t_{df}$ . The solid line is simply  $1/t_d$  and shows that  $\langle \mathcal{Y} \rangle$  and  $\langle \Delta S_{na} \rangle$  behave as the inverse of the duration of the driving  $t_d$  in the limit of large  $t_d$ .

This is indeed the case as shown in Fig. 3 in the limit of small  $t_d$ . In the opposite limit of long  $t_d$  for case (ii), we find that  $\langle \mathcal{Y} \rangle$  approaches the adiabatic limit for large  $t_d$  as in  $1/t_d$ . Remarkably, this is precisely the asymptotic behavior of the dissipated work for a process starting in equilibrium. This dependence was first theoretically predicted in Ref. [35] and recently confirmed in an experiment aimed at testing experimentally the Landauer principle [13].

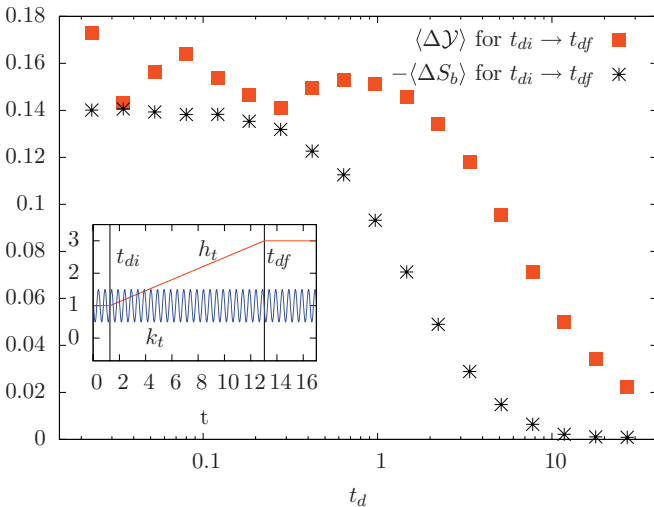


FIG. 4. (Color online) (Inset) Transitions between two nonstationary states corresponding to the forces  $h_{t_{di}} = 1.3$  and  $h_{t_{df}} = 3$ . The spring constant is oscillating with a period of 0.5 around the value 1 and with an amplitude of 0.5. The inverse temperature is taken at  $\beta = 0.2$ ; the friction is  $\gamma = 1$ .

## B. Overdamped Langevin dynamics

In the previous section, we used the same control parameter to create the nonstationary state and to induce a transition between the initial and the final nonstationary states. On the contrary, in this last section, we consider a particle obeying an overdamped Langevin dynamics in an harmonic potential with two different driving forces, a time-dependent spring constant  $k_t$  which oscillates and creates a nonstationary periodic state and a piecewise nonconservative time-dependent force  $h_t$  as a driving force inducing transitions. The position  $x_t$  of the particle is given by the following stochastic differential equation

$$\gamma \dot{x}_t = -k_t x_t + h_t + \eta_t \sqrt{2\gamma/\beta}, \quad (74)$$

with  $\gamma$  the friction coefficient,  $\beta$  the inverse temperature, and  $\eta_t$  a Gaussian white noise of mean zero and variance unity. In this case,  $x_t$  is a Gaussian process, which means that the probability distributions  $p_t(x, [h_t])$  and  $\pi_t(x, h)$  are known from the variance and the average value of the position [16]. From these quantities, we obtain  $\mathcal{Y}$  directly through Eq. (4) and the boundary term  $\Delta S_b(t_{df})$  from  $\ln \pi_{t_{df}}(x_{t_{df}}, h_{t_{df}}) - \ln p_{t_{df}}(x_{t_{df}}, [h])$ . As in the other example, we confirm that  $\langle \mathcal{Y} \rangle \geq -\langle \Delta S_b \rangle$  for all values of the duration of the driving  $t_d$ . Such a behavior is illustrated in Fig. 4, where these quantities are shown as a function of the duration of the driving  $t_d$ .

In a recent experiment, the heat fluctuations of a Brownian particle have been measured in an aging gel, created by a sudden temperature quench [36]. This aging gel plays the role of a nonequilibrium bath for the probe particle. With the same experimental setup, the deviation from the fluctuation-theorem has been measured by evaluating separately the correlations and the response function [37]. A complete discussion of these interesting results is out of place here, but instead we show that the detailed fluctuation for the heat exchange obtained in this reference can be obtained using the framework developed in previous sections. The dynamics followed by the probe particle in the experiment is similar to that described by Eq. (74) but differs from it in that, in the experiment, there is no driving force and the spring constant is not time dependent. We can adapt the formalism developed in Sec. II B to the experimental situation by choosing a similar log ratio of probabilities as in Eq. (7) with the tilde operation taken to be the identity and the star to represent time reversal. We therefore consider the quantity

$$\Delta\rho[x] = \ln \frac{\mathcal{P}[x]}{\mathcal{P}[\bar{x}]}. \quad (75)$$

Since the quench, which occurs at time 0, is very fast and there is no subsequent driving in the experiment, the dynamics occurring at time  $t > 0$  is described by time-independent rates denoted simply  $w(c, c')$ . If we consider now a path probability ratio with trajectories starting at time  $t > 0$  and finishing at time  $T$ , we obtain, from Eq. (10), that

$$\Delta\rho[x] = \ln \frac{p_t(x_t)}{p_t(x_T)} + \sum_{j \geq j_0}^N \ln \frac{w(x_{j-1}, x_j)}{w(x_j, x_{j-1})}, \quad (76)$$

where the index  $j_0$  corresponds to time  $t$  and index  $N$  corresponds to time  $T$ . Since the tilde was chosen to be the identity, it is obvious that the corresponding transformation

of the full path probability is an involution. It follows from Sec. II B that, in this case,

$$\ln \frac{P(\Delta\rho[x] = \Delta\rho)}{P(\Delta\rho[x] = -\Delta\rho)} = \Delta\rho. \quad (77)$$

Now, the second term on the right-hand side of Eq. (76) corresponds to what is called medium entropy,  $\Delta S_m$ . The temperature of the medium surrounding the probe particle equilibrates very fast (unlike the degrees of freedom associated with the polymers which constitute the gel), so we can consider that  $\Delta S_m = -\beta q$ , where  $\beta$  is the inverse equilibrium temperature of the surrounding medium times  $k_B$ , with  $q$  is the heat exchanged by the particle and the medium. Since there is no work, this heat is simply the variation of internal energy, so  $q = k(x_{t+T}^2 - x_t^2)/2$ . Furthermore, since the quench is fast and the system was prepared in an equilibrium state before the quench with a Gaussian distribution, the distribution of the initial condition at time  $t$  is still a Gaussian of variance denoted  $\sigma_x^2(t)$  in Ref. [36]. At time  $t + T$ , it is assumed that the system is equilibrated so  $k\sigma_x^2(t + T) = \beta^{-1}$ . In view of this, we obtain from Eq. (77) the fluctuation relation satisfied by the heat  $q$  obtained in this reference with  $\Delta\rho = -q\beta_{\text{eff}}/\beta$  and

$$\beta_{\text{eff}} = \frac{k_B}{k} \left[ \frac{1}{\sigma_x(t+T)^2} - \frac{1}{\sigma_x(t)^2} \right], \quad (78)$$

which is interpreted as an effective temperature [36]. This property, however, only holds for the case of linear Langevin dynamics with a time-independent spring constant.

We thus see in this example that the detailed fluctuation relation satisfied by the heat exchange in Eq. (77) follows from general considerations of a log ratio of probabilities of the form of Eq. (75). That this should be the case was also apparent in the derivation of a related fluctuation theorem satisfied by the heat exchange between a system and two thermostats [38].

## V. DISCUSSION

In this paper, we have presented a general framework for systems which are prepared in a nonstationary nonequilibrium state in the absence of any perturbation and which are then further driven through the application of a time-dependent perturbation. Below, we discuss the practical questions which arise when applying our framework to physical systems.

Typically for applications, we are often interested in reconstructing some properties of the unperturbed non-equilibrium state from its response to a perturbation. We can formally distinguish two different situations depending on the way the nonequilibrium state is prepared. In the first category, the nonequilibrium state is created by some driving, and thus the perturbation which will be applied to it after some time should be viewed as a second driving. As a particular simple example of this category, one can create the initial state by a periodic driving. In these conditions, our approach predicts a modified second law of thermodynamics for transitions between periodically driven states. Such periodically driven states are achievable in a number of experimental systems such as vibrated granular medium, electronic circuits, manipulated colloidal systems, or quantum optics for instance. In the second category, the initial nonstationary state is a transient state produced by the choice of initial conditions. For instance,

the system has been prepared by a quench of some parameter which can be the temperature or the concentration for instance, and the dynamics which follows involves relaxation or coarsening. This is typically what happens in a glassy system, where the slow relaxation following a quench leads to aging.

In this paper, it has been assumed that the dynamics of the system under consideration is Markovian. This condition limits the applicability of the theory, but it is important to appreciate that it does not impose a restriction on a specific physical system. As is well known, aging systems in particular are accessible to a Markovian description. On the theoretical side, the search for extension of fluctuations theorems to non-Markovian systems is an active field of research, which still awaits a complete understanding. Therefore, we believe that a complete discussion of this crucial point is beyond the scope of the present paper.

The practical application of our framework to experiments requires the determination of a specific nonstationary distribution  $\pi_t(c, h)$ . As is already the case with the standard Hatano-Sasa relation for nonequilibrium steady states, the corresponding distribution  $p_{\text{stat}}(c, h)$  is difficult to determine analytically and may not be a smooth function [39]. This state of affairs naturally persists for the nonstationary case. The distributions which are needed can only be calculated analytically in a few simple cases, such as in discrete models involving only a few states or for a particle in an harmonic trap obeying overdamped Langevin dynamics [36,37]. This is one reason for which there are only a handful of experimental verifications of the Hatano-Sasa relation. The first experimental verification used a colloidal particle which is manipulated by an optical tweezer [40]. In this case, the steady state is created by a nonconservative force in a simple 1D geometry. More recently, a new interesting and promising verification of this relation has been carried out [41], where a nonequilibrium steady state is created in a very different way, through the excitation of a granular gas.

In complex systems, the relevant distribution for such tests will not be accessible analytically. However, if the system (or subsystem) of interest is of small size, the numerical determination of this distribution is possible through extensive simulations as we have shown in an example based on the Glauber-Ising model [16]. Among the various other strategies which can facilitate this numerical determination, one recent interesting suggestion is to determine the distribution iteratively by starting from an approximate ansatz function [39].

## VI. CONCLUSION

In the first section of this paper, we have discussed a rather technical but important point concerning the importance of boundary conditions for fluctuations theorems. We have found that a detailed fluctuation theorem can be of strong or weak form depending on whether the initial and final probability distributions have a symmetry under protocol reversal. In the case of the entropy production, this property means that the system stochastic entropy may or may not be antisymmetric with respect to protocol reversal.

In the rest of the paper, we have derived a detailed fluctuation theorem for a quantity  $\langle \mathcal{J} \rangle$ , which can be viewed as a generalization of the average dissipated work for

nonthermodynamic systems. As a consequence of this detailed fluctuation theorem, this quantity satisfies a form of modified second law of thermodynamics. As a refinement of this modified second law, we have found that this quantity  $\langle \mathcal{Y} \rangle$  quantifies the lag between the actual probability distribution and the  $\pi_t$  distribution evaluated at the current value of the control parameter in the same way the dissipated work quantifies the lag with respect to the equilibrium distribution. Furthermore, we have found that  $\langle \mathcal{Y} \rangle$  approaches the adiabatic limit in a similar way as the dissipated work, i.e., in a manner which is proportional to the inverse of the duration of the driving.

We hope that this work can trigger experimental and theoretical interest in studying or testing fluctuation theorems for systems in contact with a nonequilibrium bath.

#### ACKNOWLEDGMENTS

We thank U. Seifert, M. Esposito, C. van den Broeck, R. Ch  trite, and A. Kundu for many insightful discussions in connection with this work.

#### APPENDIX: DEFINITION OF DUALITY FROM CURRENT REVERSAL

In the main text, we have introduced four dynamics with generators  $L_t^{h_t}$ ,  $\bar{L}_t^{h_t}$ ,  $\hat{L}_t^{h_t}$ , and  $\tilde{L}_t^{h_t}$ . For all these dynamics, with a generator that we write generally as  $\tilde{L}_t^{h_t}$  to encompass all cases, we define a probability distribution  $\tilde{p}_t(c)$  solution of the following master equation:

$$\frac{d\tilde{p}_t(c)}{dt} = \sum_{c'} \tilde{p}_t(c') \tilde{L}_t^{h_t}(c', c). \quad (\text{A1})$$

In the same spirit, we have several reference probability distributions,  $\pi_t(c, h)$ ,  $\bar{\pi}_t(c, h)$ ,  $\hat{\pi}_t(c, h)$ , and  $\tilde{\pi}_t(c, h)$ , associated to the generators  $L_t^{h_t}$ ,  $\bar{L}_t^{h_t}$ ,  $\hat{L}_t^{h_t}$ , and  $\tilde{L}_t^{h_t}$ , which we denote generally as  $\tilde{\pi}_t(c, h)$  with generator  $\tilde{L}_t^{h_t}$ . The corresponding general master equations is

$$\left( \frac{\partial \tilde{\pi}_t}{\partial t} \right) (c, h) = \sum_{c'} \tilde{\pi}_t(c', h) \tilde{L}_t^{h_t}(c', c) = \sum_{c'} \tilde{\mathcal{J}}_t^{h_t}(c, c'), \quad (\text{A2})$$

in which we have defined the reference probability current of the dynamics modified by the tilde transformation

$$\tilde{\mathcal{J}}_t^{h_t}(c, c') = \tilde{\pi}_t(c, h) \tilde{w}_t^{h_t}(c, c') - \tilde{\pi}_t(c', h) \tilde{w}_t^{h_t}(c', c). \quad (\text{A3})$$

We want to show in this appendix that the dual dynamics corresponds to the dynamics for which accompanying probability currents in the system at time  $t$  are opposite to accompanying probability currents in the system with reversed dynamics at time  $T - t$ , that is to say [5,42],

$$\hat{\mathcal{J}}_t^{h_t}(c, c') = -\tilde{\mathcal{J}}_{T-t}^{h_t}(c, c'). \quad (\text{A4})$$

To do so, we start from this definition of duality and find back the dual rates of Eq. (23). First, we remark that if all dynamics are connected, it is the same for the reference probability distributions. For instance, we can check that  $\hat{\pi}_t(c, h_t) = \pi_{T-t}(c, h_{T-t})$  by verifying that both quantities are the solution of the same differential equation,

$$\begin{aligned} (\partial_t \hat{\pi}_t)(c', h_t) &= \sum_c \hat{\mathcal{J}}_t^{h_t}(c, c') = - \sum_c \tilde{\mathcal{J}}_{T-t}^{h_t}(c, c') \\ &= -(\partial_{(T-t)} \bar{\pi}_{T-t})(c', h_{T-t}), \end{aligned} \quad (\text{A5})$$

with the same initial condition  $\hat{\pi}_0(c, h_0) = \pi_T(c, h_T)$ . Now, to obtain the dual rates, we use this symmetry  $\hat{\pi}_t(c, h_t) = \bar{\pi}_{T-t}(c, h_{T-t})$  and Eq. (A4) to get

$$\begin{aligned} \hat{\pi}_t(c, h_t) \hat{w}_t^{h_t}(c, c') - \hat{\pi}_t(c', h_t) \hat{w}_t^{h_t}(c', c) \\ = -[\hat{\pi}_t(c, h_t) \bar{w}_{T-t}^{h_t}(c, c') - \hat{\pi}_t(c', h_t) \bar{w}_{T-t}^{h_t}(c', c)], \\ = \hat{\pi}_t(c', h_t) w_t^{h_t}(c', c) - \hat{\pi}_t(c, h_t) w_t^{h_t}(c, c'). \end{aligned} \quad (\text{A6})$$

The simplest rates that verify this equality are

$$\hat{w}_t^h(c, c') = \frac{w_t^h(c', c) \hat{\pi}_t(c', h)}{\hat{\pi}_t(c, h)}. \quad (\text{A7})$$

The last step consists of using the fact that the duality obtained from Eq. (A4) is an involution so that  $\hat{w}_t^h(c, c') = w_t^h(c, c')$ . We then end with Eq. (23) as another definition of the duality transformation. Note that we have  $\tilde{\mathcal{J}}_{T-t}^{h_t}(c, c') \neq \tilde{\mathcal{J}}_t^{h_t}(c, c')$  because  $\bar{\pi}_{T-t}(c, h_{T-t}) \neq \pi_t(c, h)$ , as we can check using Eq. (A2), so duality is not a trivial reversal of the current as it was in the stationary reference framework.

- 
- [1] Christopher Jarzynski, *Annu. Rev. Condens. Matter Phys.* **2**, 329 (2011).  
[2] U. Seifert, [arXiv:1205.4176](https://arxiv.org/abs/1205.4176).  
[3] R. J. Harris and G. M. Sch  tz, *J. Stat. Mech.* (2007) P07020.  
[4] U. Seifert, *Eur. Phys. J. B* **64**, 423 (2008).  
[5] G. E. Crooks, *Phys. Rev. E* **61**, 2361 (2000).  
[6] C. Maes and K. Neto  n  y, *J. Stat. Phys.* **110**, 269 (2003).  
[7] P. Gaspard, *J. Stat. Phys.* **117**, 599 (2004).  
[8] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, *Phys. Rev. Lett.* **98**, 080602 (2007).  
[9] T. Sagawa and M. Ueda, *Phys. Rev. Lett.* **104**, 090602 (2010).  
[10] C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).  
[11] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, *Nat. Phys.* **6**, 988 (2010).  
[12] M. Esposito and C. Van den Broeck, *Europhys. Lett.* **95**, 40004 (2011).  
[13] A. Berut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, *Nature* **483**, 187–U1500 (2012).  
[14] E. Aurell, C. Mej  a-Monasterio, and P. Muratore-Ginanneschi, *Phys. Rev. Lett.* **106**, 250601 (2011).  
[15] T. Hatano and S. I. Sasa, *Phys. Rev. Lett.* **86**, 3463 (2001).  
[16] G. Verley, R. Ch  trite, and D. Lacoste, *J. Stat. Mech.* (2011) P10025.  
[17] G. Verley, R. Ch  trite, and D. Lacoste, *Phys. Rev. Lett.* **108**, 120601 (2012).  
[18] M. Esposito and C. Van den Broeck, *Phys. Rev. Lett.* **104**, 090601 (2010).

- [19] C. Maes and M. H. van Wieren, *Phys. Rev. Lett.* **96**, 240601 (2006).
- [20] W. Feller, *Trans. Am. Math. Soc.* **48**, 488 (1940).
- [21] G. Hummer and A. Szabo, *Proc. Natl. Acad. Sci. USA* **98**, 3658 (2001).
- [22] R. Chétrite, *Phys. Rev. E* **80**, 051107 (2009).
- [23] M. Baiesi, C. Maes, and B. Wynants, *Phys. Rev. Lett.* **103**, 010602 (2009).
- [24] U. Seifert, *Phys. Rev. Lett.* **95**, 040602 (2005).
- [25] J. Horowitz and C. Jarzynski, *J. Stat. Mech.* (2007) P11002.
- [26] D. J. Evans and D. J. Searles, *Adv. Phys.* **51**, 1529 (2002).
- [27] Y. Oono and M. Paniconi, *Prog. Theor. Phys. Suppl.* **130**, 29 (1998).
- [28] T. Speck and U. Seifert, *J. Phys. A: Math. Theor.* **38**, L581 (2005).
- [29] V. Y. Chernyak, M. Chertkov, and C. Jarzynski, *J. Stat. Mech.* (2006) P08001.
- [30] M. Esposito, U. Harbola, and S. Mukamel, *Phys. Rev. E* **76**, 031132 (2007).
- [31] S. Vaikuntanathan and C. Jarzynski, *Europhys. Lett.* **87**, 60005 (2009).
- [32] R. García-García, D. Domínguez, V. Lecomte, and A. B. Kolton, *Phys. Rev. E* **82**, 030104 (2010).
- [33] R. García-García, V. Lecomte, A. B. Kolton, and D. Domínguez, *J. Stat. Mech.* (2012) P02009.
- [34] D. T. Gillespie, *Abstr. Pap. Am. Chem. Soc.* **173**, 128 (1977).
- [35] K. Sekimoto and S.-I. Sasa, *J. Phys. Soc. Jpn.* **66**, 3326 (1997).
- [36] J. R. Gomez-Solano, A. Petrosyan, and S. Ciliberto, *Phys. Rev. Lett.* **106**, 200602 (2011).
- [37] J. R. Gomez-Solano, A. Petrosyan, and S. Ciliberto, *Europhys. Lett.* **98**, 10007 (2012).
- [38] C. Jarzynski and D. K. Wójcik, *Phys. Rev. Lett.* **92**, 230602 (2004).
- [39] C. Pérez-Espigares, A. B. Kolton, and J. Kurchan, *Phys. Rev. E* **85**, 031135, (2012).
- [40] E. H. Trepagnier, C. Jarzynski, F. Ritort, G. E. Crooks, C. J. Bustamante, and J. Liphardt, *Proc. Natl. Acad. Sci. USA* **101**, 15038 (2004).
- [41] A. Mounier and A. Naert, [arXiv:1208.4039](https://arxiv.org/abs/1208.4039).
- [42] R. Chétrite and K. Gawędzki, *Commun. Math. Phys.* **282**, 469 (2008).