Thermal Contact I : Symmetries ruled by Exchange Entropy Variations

F. Cornu
Laboratoire de Physique Théorique, UMR 8627 du CNRS
Université Paris-Sud, Bât. 210
F-91405 Orsay, France

M. Bauer
Institut de Physique Théorique de Saclay*, CEA Saclay
F-91191 Gif-sur-Yvette Cedex, France

May 2, 2015

Most results contained in the joined papers arXiv:1302.4538 and arXiv:1302.4540 put on cond-mat.stat-mech on February 19 2013 have now appeared in print as

- *[a]* Thermal Contact through a Diathermal Wall: A Solvable Toy Model,
  F. Cornu and M. Bauer, J. Stat. Mech. (2013) P10009
- *[b] Part of* Affinity and Fluctuations in a Mesoscopic Noria,
  M. Bauer and F. Cornu, J. Stat. Phys. **155** *(2014) 703, [arXiv:1402.2422]*
- *[c]* Local detailed balance : a microscopic derivation,
  M. Bauer and F. Cornu, J. Phys. A: Math. Theor. **48** *(2015) 015008, [arXiv:1412.8179]*

The present revised version takes into account the minor corrections made in the published articles but the bibliography is that of the first version (not updated). In reference [c] the denomination local detailed balance *has been used in place of* modified detailed balance *in order to fit the terminology that seems to be most commonly used nowadays.*

Abstract

Thermal contact is the archetype of non-equilibrium processes driven by constant non-equilibrium constraints enforced by reservoirs exchanging conserved microscopic quantities. At a mesoscopic scale only the energies of the macroscopic bodies are accessible together with the configurations of the contact system. We consider a class of models where the contact system, as well as macroscopic bodies, have a finite number of possible configurations. The global system with only discrete degrees of freedom has no microscopic Hamiltonian dynamics, but it is shown that, if the microscopic dynamics is assumed to be deterministic and ergodic and to conserve energy according to some specific pattern, and if the mesoscopic evolution of the global system is approximated by a Markov process as closely as possible, then the mesoscopic transition rates obey three constraints. In the limit where macroscopic bodies can be considered as reservoirs at thermodynamic equilibrium (but with different intensive parameters) the mesoscopic transition rates turn into transition rates for the contact system and the third constraint becomes modified detailed balance (MDB) ; the latter is generically expressed in terms of the microscopic exchange entropy variation, namely the opposite of the variation

*CEA/DSM/IPhT, Unité de recherche associée au CNRS

of the thermodynamic entropy of the reservoir involved in a given microscopic jump of the contact system configuration. We investigate the generic statistical properties for measurable quantities that arise from the MDB constraint. For a finite-time evolution after the system prepared in an equilibrium state has been set in contactwith thermostats at different temperatures, we derive a detailed fluctuation relation for the excess exchange entropy variation and an associated integral fluctuation relation. In the non-equilibrium stationary state (long-time limit), the proper mathematical definition of a large deviation function is introduced together with alternative definitions, and fluctuation relations are derived. The fluctuation relation for the exchange entropy variation is merely a particular case of the Lebowitz-Spohn fluctuation relation for the action functional [1]. The generalization to systems exchanging energy, volume and matter with several reservoirs, with a possible conservative external force acting on the contact system, is given explicitly. In the case of several independent macroscopic currents, the infinite time limit of any odd cumulant per unit time of exchanged quantities is expressed in terms of a series involving higher even cumulants and powers of the thermodynamic forces associated to the currents. Every relation can be seen as a generalized Einstein-Green-Kubo relation valid far from equilibrium. It entails a relation between the nth-order non-linear response coefficients of any odd cumulant per unit time in the vicinity of equilibrium and a finite number of lower-order non-linear response coefficients of even cumulants per unit time. The latter relations, already known in the literature, can be seen as another kind of generalizations of the standard Einstein-Green-Kubo formula pertaining to the first-order coefficient.

PACS : 05.70.Ln, 02.50.Ga, 05.60.Cd

KEYWORDS : thermal contact; master equation; ergodicity; modified detailed balance; exchange entropy; large deviation function ; fluctuation relations; generalized Green-Kubo relations.

Corresponding author : CORNU Françoise, E-mail: Francoise.Cornu@u-psud.fr

Contents

1 Introduction

1.1 Issues at stake

Thermal contact between two energy reservoirs is one of the first issues addressed by early thermodynamics and its second principle about entropy variation. It has been revisited in the context of the description of transport phenomena either from the statistical approach by the Boltzmann equation and its BBGKY hierarchy generalization or from the phenomenological thermodynamics of irreversible processes [2, 3, 4] and the local equilibrium phenomenological approach for inhomogeneous continuous media [5, 6]. Until recently there have been repeated attempts to formulate thermodynamics for non-equilibrium steady states [7, 8]. Nowadays the statistical description for the fluctuations of heat exchanges between a small system and two thermal baths is part of the so-called " stochastic thermodynamics" of small systems incorporating the effects of fluctuations [9, 10]. In the last two decades the latter domain has been the subject of an increasingly intense activity, from the theoretical as well as from the experimental point of view with very fast technological improvements [11, 12]. A recent extended review about stochastic thermodynamics is to be found in Ref.[10], and its formulation in the specific class of Markov jump processes in continuous time is reviewed in Refs.[13]. (For brief introductions see also for instance Refs.[14, 15].)

First steps in out-of-equilibrium statistical mechanics have been the study of the linear (static or dynamic) response to an external constraint that drives the system out of its equilibrium state. This vast topic began with the Einstein fluctuation-dissipation relation for Brownian motion (see collected translated papers in [16]) and is still active (for a review see [17]). Nowadays, the question of linear response in the vicinity of a non-equilibrium steady state is under investigation (see for instance [18, 19, 20, 21, 22]), but the subject is beyond the scope of the present work.

Non-perturbative approaches for systems far from equilibrium have followed various ways. One trend has been the search for a unifying variational principle based on a large deviation point of view, which would generalize the entropy-probability relation introduced by Einstein in his theory of thermodynamic equilibrium fluctuations [23] in order to explain the opalescence phenomenon. Indeed, principles of equilibrium statistical mechanics can be seen as resulting from the maximization of the Shannon entropy under various constraints about macroscopic observables in the framework of information theory [24, 25]. From this point of view, statistical mechanics, which allows one to retrieve the laws of thermodynamics while describing mesoscopic fluctuations, may be interpreted as an example of a large deviation theory, which provides the probabilities for an observable to deviate from its most probable value [26]. In order to go beyond the phenomenological irreversible thermodynamics, Oono [27] has promoted the idea of applying the large deviation theory to non-equilibrium statistical physics seen as a statistical mechanics along the time axis. In this direction efforts have been devoted to the study of the maximization of the Shannon entropy for the system histories under some constraints, such as fixed values for the macroscopic out-of-equilibrium currents in a non-equilibrium steady state (see [28, 29] and references therein).

In the same spirit a second path, the quest for some relevant physical quantities that would obey some universal principle, is the subject of the fluctuation relations either at finite time (transient regimes) or in the infinite-time limit (stationary regimes). The active topic of fluctuation relations was initiated by works about a symmetry of the large deviation function of some entropy production rate in the non-equilibrium stationary states of some chaotic dynamical systems [30, 31, 32, 33], and about the Jarzynski identity for the finite-time cumulative work when a Hamiltonian system is driven out of an initial equilibrium state by the variation of some external parameter [34]. The first results pertained to classical systems but fluctuation relations in quantum systems have also been investigated (see for instance the reviews [35, 36]). A short list for successive steps on the narrower pathway of fluctuation relations for systems with Markovian stochastic dynamics is given by the following references [37, 38, 39, 40, 1, 41, 42, 9, 43]. Among these results, in the more specific case of systems with stochastic evolutions described by a master equation for the probability of the system configurations, the milestone is the fluctuation relation obeyed by the dimensionless action functional introduced by Lebowitz and Spohn [1].

In the even more restricted case where the system exchanges conserved microscopic quantities,

such as energy quanta or particles or elementary volumes, with infinite size reservoirs which stay in their equilibrium states during the considered evolution of the finite system, the transition rates must obey a relation, which will be referred to as the Modified Detailed Balance (MDB)[1]. When the transition rates obey the MDB, the action functional introduced by Lebowitz and Spohn coincides with the variation of the sum of the reservoirs thermodynamic entropies. Then, apart from the fluctuation relation obeyed by the latter entropy variation, the known results about large deviation functions which are the most relevant for the topic of the present paper deal with the following quantities in finite size systems : the heat current that goes through the finite system that sets up a thermal contact between two thermostats [44, 45] (also first addressed in the context of classical or quantum Hamiltonian dynamics in Ref.[46]); more generally the various macroscopic currents through a system in a non-equilibrium stationary state sustained by various kinds of reservoirs [47, 48].

Apart from general approaches, a third trend in the quest for some non-perturbative statistical theory of out-of-equilibrium phenomena has been the search for solvable models which could give some hints in the comprehension of these phenomena in the absence of any theoretical framework. We are interested in a class of models where the system has a finite number of possible configurations, the heat exchanges are described as changes in the populations of its energy levels, and the configurations stochastically evolve under a master equation with transition rates bound to obey the MDB relation arising from the existence of an underlying ergodic deterministic microscopic dynamics which conserves energy. In paper II [49] we perform explicit analytical calculations for the very simple case where the thermal contact system is reduced to two spins, each of which is flipped by a single thermostat, the two thermostats being at different temperatures. This model allows the description of a thin diathermal interface between two macroscopic bodies as a collection of independent spin pairs.

1.2 Main results valid beyond the thermal contact example

In the present section we point out the main ideas and generic results of our study when we consider a class of models that generalize the case of thermal contact in the sense that the contact system $\mathcal{S}$ can exchange not only energy but also matter or some portion of occupied volume with several macroscopic bodies according to some specific conservative exchange pattern where one and only one macroscopic body is involved in a microscopic jump of the whole system (see section 6). For the sake of pedagogy, we express the main ideas and generic results in the language of the thermal contact example. The general arguments and results in the more general situation are more detailed in section 6.

In the whole paper we put emphasis on the *exchange entropy variation* $\Delta_{\rm exch}S$ of the system[2], which is opposite to the well-defined entropy variation of the reservoirs described in the thermodynamic limit, because it seems to be the crucial quantity to consider in order to answer the key-point question of identifying the relevant measurable quantities which are to obey non-trivial universal properties (beyond long-time decorrelations and the subsequent existence of large deviation functions). We stress that the term "measurable quantities" excludes the stationary probability distribution of the system configurations, which is very hard to determine from experiments. This point of view slightly differs from the common picture where the focus is put on the variation of the system Shannon-Gibbs entropy which explicitly involves the probability distribution of the system configurations. The latter picture has arisen from the seminal works by Crooks [39, 40, 50] and has led to the notion of entropy production along a stochastic trajectory of the system microscopic configurations (see [42, 10] for a recent formulation). In the whole paper Boltzmann constant k_B is set to 1 and S denotes dimensionless entropies.

We point out that Appendix E contains the proper mathematical definition of the large deviation function, and some variants for the models at stake, together with demonstrations of

[1] The denomination "generalized" detailed balance is also to be found in the literature.

[2] Exchange entropy variation is an abbreviation for "variation of entropy due to exchanges" of energy with a thermal bath, or more generally, of some measurable conserved quantities with various reservoirs.

properties about large deviation functions which are not usually exhibited in the literature of the physicists community.

1.2.1 Constraints from ergodicity and energy conservation pattern

We consider models for thermal contact where the energy exchanges between several very large systems with indices a's are realized through interactions with a system $\mathcal{S}$ which has a finite number of configurations, namely which has discrete variables (also called degrees of freedom) in finite number, and whose configurations can changed only thanks to interactions with the large systems. More precisely, the domain $\mathcal{D}$ occupied by the so-called small system $\mathcal{S}$ can be divided in several disjoint areas $\mathcal{D}_a$, such that the value of every degree of freedom of $\mathcal{S}$ that sits inside $\mathcal{D}_a$ can change only thanks to an energy exchange with a macroscopic body $\mathcal{B}_a$. The small system has an internal energy which is bounded, since its configuration space is finite. When each large part with index a is described only at a macroscopic level, i.e. only by its energy E_a, we have to resort to a statistical description, where the knowledge of the system is specified by the following data : the configuration $\mathcal{C}$ of the small system and the energies of the large parts. Such a set of data defines what we call a mesoscopic configuration of the total system in the sequel. Then the stochastic evolution is determined by transition rates between these mesoscopic configurations. The key points are the following.

First we have to answer the question : which constraints must be obeyed by the choice for the transition rates ? Indeed, one cannot define a Hamiltonian dynamics for discrete variables. However, when the microscopic dynamics is deterministic, ergodic and conserves the global energy, then in a given energy level all ergodic microscopic dynamics have the same period, and the probability of a mesoscopic configuration $(\mathcal{C}, \{E_a\})$ calculated as a time average over a period of microscopic dynamics coincides with the microcanonical distribution. Moreover, the chosen interaction pattern entails that, during a period of microscopic dynamics, the number of deterministic jumps that occur from one mesoscopic configuration of the global system to another one is equal to the number of jumps in the opposite sense (subsection 2.1 and Appendix A).

As a consequence, if the mesoscopic dynamics is approximated by a Markov process according to the prescription derived in Appendix B, then the corresponding transition rates between mesoscopic configurations $(\mathcal{C}, \{E_a\})$ must satisfy three constraints (subsection 2.2). In particular these transition rates obey the microcanonical detailed balance (2.11). (For a comparison the derivation of the microcanonical detailed balance for mesoscopic variables defined from continuous microscopic variables evolving according to a Hamiltonian dynamics invariant under time reversal is presented in Appendix C). The interaction pattern further transforms the microcanonical detailed balance into the reduced expression (2.13) which only involves the variation of the Boltzmann entropy of the large part $\mathcal{B}_a$ that exchanges energy in the transition. Eventually, the three constraints for the transition rates between mesoscopic configurations $(\mathcal{C}, \{E_a\})$ ensure that in the infinite-time limit the stochastic evolution does lead to a unique stationary state which coincides with the microcanonical probability distribution $P_{\rm mc}(\mathcal{C}, \{E_a\})$.

In the limit where the sizes of the large parts go to infinity before the time evolution of the system is considered (subsection 2.3), the global system does not reach equilibrium but, in a time window where the variations of the macroscopic energy of every large part divided by its number of degrees of freedom are negligible with respect to some microscopic energy scale, the large parts can be described as if they were in the corresponding thermodynamic equilibrium state. The *thermodynamic entropy* of a macroscopic body at equilibrium characterized by its internal energy U_a and its number of degrees of freedom $\mathcal{N}_a$ (and possibly other extensive quantities such as volume which we omit) is an extensive quantity : $S_a^{TH}(U_a, \mathcal{N}_a) = \mathcal{N}_a s_a(U_a/\mathcal{N}_a)$ where $s_a(\epsilon)$ is a differentiable function of the mean energy per constituent ϵ. In equilibrium statistical mechanics, for an isolated system $s_a(\epsilon)$ can be calculated (on principle) in the microcanonical ensemble from the *Boltzmann entropy* $S^B(E_a, \mathcal{N}_a) \equiv \ln \Omega(E_a, \mathcal{N}_a)$, where Boltzmann constant k_B has been set equal to 1 and $\Omega(E_a, \mathcal{N}_a)$ is the configuration number when the system is characterized by the constant global extensive quantities E_a and $\mathcal{N}_a$. The relation reads $s_a(\epsilon) = \lim_{\rm TH} S^B(E_a, \mathcal{N}_a)/\mathcal{N}_a$, where the thermodynamic limit $\lim_{\rm TH}$ corresponds to the coupled limits $\mathcal{N}_a \to +\infty$ and $E_a \to +\infty$

with $\lim E_a/\mathcal{N}_a = \epsilon$. In the case of thermal contact the large parts behave as thermostats, each of which is characterized by the intensive thermodynamic parameter β_a conjugated with the extensive macroscopic quantity U_a through $ds_a(\epsilon) = \beta_a d\epsilon$. When Boltzmann constant is set equal to 1, β_a is the inverse temperature of thermostat a. Then the constraints for the transition rates between mesoscopic configurations $(\mathcal{C}, \{E_a\})$ of the global system become three conditions which must be obeyed by the transition rates between the small system configurations $\mathcal{C}$ in the corresponding transient regime. Let $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$ denote the transition rate from configuration $\mathcal{C}$ to configuration $\mathcal{C}'$ in the transient regime. The three conditions are the following ones. First, the set of transition rates must be such that any configuration $\mathcal{C}$ can be reached by a succession of jumps with non-zero transition rates from any configuration $\mathcal{C}'$, namely in the network representation of the stochastic evolution

the graph G associated with the transition rates is connected. (1.1)

In other words the Markov matrix $\mathbb{M}$ defined from the transition rates by (2.17) must be irreducible and the property (1.1) will be referred to as the *irreducibility* condition. Second, the transition rates must obey the *microscopic reversibility* condition for any couple of configurations $(\mathcal{C}, \mathcal{C}')$, namely

$$(\mathcal{C}'|\mathbb{W}|\mathcal{C}) \neq 0 \qquad \Leftrightarrow \qquad (\mathcal{C}|\mathbb{W}|\mathcal{C}') \neq 0. \tag{1.2}$$

Third, the transition rates have to satisfy the *modified detailed balance* which takes the form (2.18) in the case of pure thermal contact. According to the derivation, the modified detailed balance reads quite generally

$$\frac{(\mathcal{C}'|\mathbb{W}|C)}{(\mathcal{C}|\mathbb{W}|C')} = e^{-\delta_{\rm exch} S(\mathcal{C}' \leftarrow \mathcal{C})}, \tag{1.3}$$

where the exchange entropy variation $\delta_{\rm exch} S(\mathcal{C}' \leftarrow \mathcal{C})$ associated with a jump of the small system from a microscopic configuration $\mathcal{C}$ to another one $\mathcal{C}'$ is defined as the opposite of the infinitesimal variation of the thermodynamic entropy of the reservoir $\mathcal{B}_a$ that causes the jump of configuration from $\mathcal{C}$ to $\mathcal{C}'$ by exchanging energy and/or volume and/or matter with the constituents of the small system. The explicit expression for $\delta_{\rm exch} S(\mathcal{C}' \leftarrow \mathcal{C})$ in terms of configuration observables is given in (6.6); it includes the case where a conservative external force acts on some global coordinate of the contact system. In the case of thermal contact, namely energy transfers, $\delta_{\rm exch} S(\mathcal{C}' \leftarrow \mathcal{C}) = \beta_a \delta q_a(\mathcal{C}' \leftarrow \mathcal{C})$ where $\delta q_a(\mathcal{C}' \leftarrow \mathcal{C})$ is the opposite of the heat amount received by heat source $\mathcal{B}_a$, which may be associated with its heat capacity at fixed volume or its heat capacity at fixed pressure (when the volume of the interface and that of the heat reservoirs both vary by opposite amounts). The modified detailed balance has been used in various specific forms, for instance, for heat currents between two heat sources or particle currents between two particle reservoirs or for coupled exchanges of energy and particles in molecular motor models (see among others [51, 45, 52, 53]). We stress that $\delta_{\rm exch} S(\mathcal{C}' \leftarrow \mathcal{C})$ is defined unambiguously and involves experimentally measurable quantities (in the thermal contact case the inverse temperatures of the energy reservoirs β_a's as well as the heat amounts received by the system from the reservoirs).

1.2.2 Various entropy variations inherent to Markovian dynamics

In section 3 we stress some consequences of the fact that the configuration probability obeys a master equation. First, even if the transition rates do not obey the MDB, the irreducibility of the Markov matrix implies the uniqueness of the stationary state and the role played by the *relative entropy with respect to the stationary solution*, $S_{\rm rel}[P(t)|P_{\rm st}]$ is recalled (subsection 3.1). We introduce the generic currents (3.10)-(3.11) for a microscopic variation when the system goes out of a given configuration $\mathcal{C}$; the averages of such currents show off either in the time derivative of the average of an observable or in the flow of an exchange quantity (subsection 3.2). In order to make the comparison with the thermodynamics of irreversible processes, we recall how the time-derivative of the *Shannon-Gibbs entropy* $S^{SG}[P(t)]$ of the system is split into two contributions, an exchange (or external) part $d_{\rm exch} S/dt$ arising from exchanges with the external reservoirs, hereafter called the *exchange entropy flow*, and an internal (or irreversible) part $d_{\rm int} S^{SG}/dt$ due to the internal irreversible processes in the system, also called the *entropy production rate* (subsection

3.3). The notation is chosen so as to emphasize that neither the exchange entropy flow $d_{\rm exch}S/dt$ nor the entropy production rate are time derivatives of some function.

In the stationary state, the time derivative $dS^{SG}[P_{\rm st}]/dt$ vanishes and the entropy production rate becomes opposite to the exchange entropy flow. Henceforth, in the case where there are only two energy reservoirs, the stationary entropy production rate can be rewritten as in phenomenological irreversible thermodynamics, namely as

$$\left.\frac{d_{\rm int}S^{SG}}{dt}\right|_{\rm st} = -\left.\frac{d_{\rm exch}S}{dt}\right|_{\rm st} = \mathcal{F}J, \tag{1.4}$$

where J is the mean heat current that goes through the system from heat bath 2 to heat bath 1 and $\mathcal{F} = \beta_1 - \beta_2$ is the associated "thermodynamic force". In the transient regime, when the MDB is satisfied the expressions for $d_{\rm exch}S/dt$ and $d_{\rm int}S^{SG}/dt$ coincide with the splitting of dS^{SG}/dt into two functionals $\sigma_{\rm exch}[P(t)]$ and $\sigma_{\rm int}[P(t)]$ [54, 1]. As noticed in [43] for the more generic case where the transition rates are time-dependent, $\sigma_{\rm int}[P(t)]$ appears as the sum of two positive terms. In the case of time-homogeneous Markov evolution, the first positive term is the opposite of the time-derivative of the relative entropy of $P(t)$ with respect to $P_{\rm st}$ and we rewrite the second positive term as the average of the microscopic current associated with some stationary affinity variation when the system jumps out of a configuration $\delta A[P_{\rm st}]$, average which is calculated with respect to the probability $P(t)$. Then the entropy production rate can be written as the sum of the following two positive contributions,

$$\frac{d_{\rm int}S^{SG}}{dt} \underset{MDB}{=} -\frac{dS_{\rm rel}[P(t)|P_{\rm st}]}{dt} + \langle j_{\delta A[P_{\rm st}]}\rangle_t. \tag{1.5}$$

The current $j_{\delta A[P_{\rm st}]}(\mathcal{C})$ proves to be positive even before averaging over the configurations (see (3.40)). For the sake of completeness we recall how, in the graph theory where the master equation is represented by a network, a non equilibrium stationary state is characterized by the affinities and probability currents associated with a restricted number of cycles defined from the graph built with the transition rates. When the graph is a pure cycle, there is a single affinity which has a simple probabilistic interpretation given in paper II.

1.2.3 MDB and non-perturbative symmetries in transient regimes

In the form (1.3) where it involves the microscopic exchange entropy variation $\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C})$ associated with a jump from configuration $\mathcal{C}$ to configuration $\mathcal{C}'$, the MDB entails time-reversal symmetries for finite time intervals at more and more mesoscopic levels as follows. The ratio (4.6) of the probabilities for a microscopic history $\mathcal{H}ist$ and the time-reversed one is determined by the *cumulative exchange entropy variation along a history*, $\Delta_{\rm exch}S[\mathcal{H}ist]$, defined in (4.1). Then, the ratio (4.8) between, on the one hand, the probability for all evolutions with fixed initial and final configurations $\mathcal{C}_0$ and $\mathcal{C}_f$ and given heat amounts $\mathcal{Q}_1$ and $\mathcal{Q}_2$ received from the thermostats and, on the other hand, the probability for all backward evolutions with exchanged initial and final configurations and opposite heat amounts $-\mathcal{Q}_1$ and $-\mathcal{Q}_2$ is determined only by the exchange entropy variation

$$\Delta_{\rm exch}S(\mathcal{Q}_1, \mathcal{Q}_2) = \beta_1\mathcal{Q}_1 + \beta_2\mathcal{Q}_2. \tag{1.6}$$

The ratio of probabilities does not explicitly depend on the initial and final configurations $\mathcal{C}_0$ and $\mathcal{C}_f$. There is only an implicit dependence on these configurations through the energy conservation rule that the quantities $\mathcal{Q}_1$ and $\mathcal{Q}_2$ for a given history must satisfy, $\mathcal{Q}_1 + \mathcal{Q}_2 = \mathcal{E}(\mathcal{C}_f) - \mathcal{E}(\mathcal{C}_0)$.

When the system is prepared in an equilibrium state at the inverse temperature β_0 and suddenly put in thermal contact at the initial time of measurements with two heat baths at the inverse temperatures β_1 and β_2, the symmetry (4.8) together with the specific form of the equilibrium canonical distribution lead one to consider the following measurable quantity : the *excess exchange entropy variation* $\Delta_{\rm exch}^{{\rm excs},\beta_0}S(\mathcal{Q}_1, \mathcal{Q}_2)$ defined as the difference between the exchange entropy variation under the non-equilibrium external constraint $\beta_1 \neq \beta_2$ and that under the equilibrium

condition $\beta_1 = \beta_2 \equiv \beta_0$, namely

$$\Delta_{\rm exch}^{{\rm excs},\beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2) = \Delta_{\rm exch} S(\mathcal{Q}_1, \mathcal{Q}_2) - \beta_0(\mathcal{Q}_1 + \mathcal{Q}_2). \tag{1.7}$$

$\Delta_{\rm exch}^{{\rm excs},\beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2)$ obeys the symmetry relation at any finite time, or finite-time "detailed fluctuation relation",

$$\frac{P_{P_{\rm can}^{\beta_0}}\left(\Delta_{\rm exch}^{{\rm excs},\beta_0} S\right)}{P_{P_{\rm can}^{\beta_0}}\left(-\Delta_{\rm exch}^{{\rm excs},\beta_0} S\right)} = e^{-\Delta_{\rm exch}^{{\rm excs},\beta_0} S}. \tag{1.8}$$

The latter relation itself entails the identity, or finite-time "integral fluctuation relation"

$$\langle e^{\Delta_{\rm exch}^{{\rm excs},\beta_0} S(t)} \rangle_{P_{\rm can}^{\beta_0}} = 1 \tag{1.9}$$

in the spirit of Jarzynski identity [34]. The integral fluctuation relation (1.9) entails through Jensen's inequality $\langle e^x \rangle \geq e^{\langle x \rangle}$ that, the mean heat amounts that are exchanged during the time interval t from the initial time where the thermal contact is set on between the system initially at equilibrium at inverse temperature β_0 and the two heat sources must satisfy

$$(\beta_0 - \beta_1) Q_1(t) + (\beta_0 - \beta_2) Q_2(t) = -\langle \Delta_{\rm exch}^{{\rm excs},\beta_0} S \rangle \geq 0, \tag{1.10}$$

where $Q_a(t)$ denotes the expectation value of $\mathcal{Q}_a$ when the experiment is repeated a large number of times, $Q_a(t) = \langle \mathcal{Q}_a(t) \rangle$. To our knowledge the two relations (1.8) and (1.9) have not appeared explicitly in the literature, though the calculations involved in the derivation of the present finite-time fluctuation relations are analogous to those that lead to finite-time detailed fluctuation relations for protocols where the system is in thermal contact with only one heat bath and is driven out of equilibrium by a time-dependent external parameter (see the argument first exhibited by Crooks [40] for work fluctuations and then Seifert [42] for the entropy production along a stochastic trajectory (see also the review [9]). The latter class of protocols is very different as for the physical mechanisms that they involve : the changes in energy level populations are caused by energy exchanges with only one thermal bath and the system is driven out of equilibrium by the time dependence of the energy levels enforced by external time-dependent forces [55]. Moreover, in Jarzynski-like protocols the system evolves from an initial equilibrium state and measurements are performed until work ceases to be provided to the system, which then relaxes to another equilibrium state, whereas in Hatano-Sasa-like protocols the system evolves from an initial non-equilibrium steady state to another one (and then housekeeping heats and excess heats are introduced as in the steady state thermodynamics introduced by Oono and Paniconi [7]). In the finite-time protocol considered here the system starts in an equilibrium state and at time t it has not yet reached the steady state controlled by β_1 and β_2. The present protocol does not either involve the comparison of forward and backward evolutions corresponding to two different series of experiments. Moreover, the integral fluctuation relation (1.9) differs from the Hatano-Sasa relation [56] in the sense that the quantity to be averaged over repeated experiments does not involve the probability distribution of the system.

In a time window sufficient long so that β_1 times the maximal possible energy variation of the system that sets up thermal contact is negligible with respect to $(\beta_1 - \beta_2)\langle \mathcal{Q}_2 \rangle$, the definition (1.6) can be replaced by its typical value, $\Delta_{\rm exch} S(\mathcal{Q}_1, \mathcal{Q}_2) \simeq -(\beta_1 - \beta_2)\mathcal{Q}_2$. Moreover, if the time window is also such that β_0 times the maximal possible energy variation of the system is also negligible with respect to $(\beta_1 - \beta_2)\langle \mathcal{Q}_2 \rangle$, the definition (1.7) can be replaced by the approximation $\Delta_{\rm exch}^{{\rm excs},\beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2) \simeq \Delta_{\rm exch} S(\mathcal{Q}_1, \mathcal{Q}_2)$. Then the relations (1.8) and (1.9) are compatible with the finite-time equalities settled in [46] for a heat transfer $\mathcal{Q}_2$ between two finite bodies initially prepared at different inverse temperatures β_1 and β_2 and whose microscopic Hamiltonian dynamics involves a negligible interaction turned on at time 0 and switched off at time t (while the temperatures of both bodies may vary). We also notice that in the considered time window the inequality (1.10) can be approximated by $(\beta_1 - \beta_2) Q_2(t) \geq 0$. We retrieve the result derived from thermodynamics principles : on the average heat flows from the hotter heat bath to the colder one, in the absence of work given to the system that ensures contact between them.

We notice that if the configuration probability distribution in a stationary state of the system happens to be the canonical distribution at an effective inverse temperature $\beta_\star(\beta_1, \beta_2)$, as it is the case in the solvable model considered in paper II, there exist similar detailed and integral fluctuation relations for a protocol where the system is initially prepared in a non-equilibrium stationary state with two heat baths at the inverse temperatures β_1^0 and β_2^0 and suddenly put at the initial time of measurements in thermal contact with two heat baths at the inverse temperatures β_1 and β_2. Then $\Delta_{\rm exch}^{{\rm excs},\beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2)$ is to be replaced by $\Delta_{\rm exch}^{{\rm excs},\beta_\star^0} S(\mathcal{Q}_1, \mathcal{Q}_2)$ where $\beta_\star^0$ is the effective inverse temperature $\beta_\star^0(\beta_1^0, \beta_2^0)$.

1.2.4 MDB and fluctuation relations in the stationary regime

For a system with a finite number of configurations, when the Markov stochastic matrix for the continous-time evolution of the configuration probabilities is irreducible, (see definition (1.1)), the Perron-Frobenius theorem can be applied: the system has a single stationary state, and it is such that every configuration has a non-vanishing weight (see for instance Ref. [57]). Moreover the system reaches its stationary state in a exponentially-short time [58]. Then the symmetry relation (4.8) enforced by the MDB at finite time leads to the existence of lower and upper bounds for the ratio between the finite-time probability to measure heat amounts $\mathcal{Q}_1$ and $\mathcal{Q}_2$ and the corresponding probability for the opposite values, when the system is in its stationary state (see (5.1)). Similar bounds are exhibited in Ref.[45].

The Markovian property of the evolution implies that the cumulative heats $\mathcal{Q}_1$ and $\mathcal{Q}_2$, and subsequently the cumulative exchange entropy variation $\Delta_{\rm exch} S$, all grow linearly with time in the long-time limit and that there exist large deviation functions for each of them [59, 60]. The proper mathematical definition of a large deviation function is recalled in Appendix E.1, and other alternative definitions when exchanged quantities are discrete are introduced in Appendix E.2.

The finite-time inequalities (5.1) entail a symmetry in the long-time limit, according to the general results derived in Appendix E.3 : the dimensionless exchange entropy variation $\Delta_{\rm exch} S$ obeys the fluctuation relation

$$f_{\Delta_{\rm exch} S}(\mathcal{J}) - f_{\Delta_{\rm exch} S}(-\mathcal{J}) = -\mathcal{J}, \tag{1.11}$$

where $\mathcal{J}$ denotes the values taken by the cumulative current $\Delta_{\rm exch} S(t)/t$. The fluctuation relation for $\Delta_{\rm exch} S$ is a special case of the more general fluctuation relation for the action functional introduced by Lebowitz and Spohn [1] ; indeed when the transition rates obey the MDB, the action functional for a given history becomes equal to the entropy variation of the reservoirs, namely to the opposite of the exchange entropy variation.

For a system with a finite number of configurations, $\mathcal{Q}_1 + \mathcal{Q}_2$ is bounded. Then, according to the results of Appendix E.4, the large deviation function for $\mathcal{Q}_2$ is equal to that for $-\mathcal{Q}_1$, while the fluctuation relation (1.11) for $f_{\Delta_{\rm exch} S}$ also entails a fluctuation relation for the large deviation function $f_{\mathcal{Q}_2}$ for the cumulative heat $\mathcal{Q}_2$, because the difference between $\Delta_{\rm exch} S$ and $-(\beta_1 - \beta_2)\mathcal{Q}_2$ is finite at any time t. The fluctuation relation can be written in the generic form

$$f_{\mathcal{Q}_2}(\mathcal{J}) - f_{\mathcal{Q}_2}(-\mathcal{J}) = \mathcal{F}\mathcal{J}, \tag{1.12}$$

where $\mathcal{J}$ denotes the values taken by the cumulative current $\mathcal{Q}_2(t)/t$ and $\mathcal{F}$ is the thermodynamic force conjugated to the mean instantaneous heat current from heat bath 2 in the stationary state $J \equiv \langle j_2 \rangle_{\rm st}$ through the expression (1.4) of the exchange entropy flow in the stationary state. Indeed on the one hand $\lim_{t\to+\infty} \langle \Delta_{\rm exch} S \rangle / t = -(\beta_1 - \beta_2) \lim_{t\to+\infty} \langle \mathcal{Q}_2(t) \rangle / t$ and on the other hand $\lim_{t\to+\infty} \langle \Delta_{\rm exch} S \rangle / t = d_{\rm exch} S/dt|_{\rm st}$ and $\lim_{t\to+\infty} \langle \mathcal{Q}_2(t) \rangle / t = \langle j_2 \rangle_{\rm st}$; then comparison of $-d_{\rm exch} S/dt|_{\rm st} = (\beta_1 - \beta_2)\langle j_2 \rangle_{\rm st}$ with definition (1.4) leads to identify the coefficient $(\beta_1 - \beta_2)$ that arises in the fluctuation relation with the thermodynamic force $\mathcal{F}$.

1.2.5 MDB and generalized Einstein-Green-Kubo relations

Quite generally, when there exists a large deviation function for the cumulative current $\mathcal{J}_t \equiv X_t/t$ associated with the cumulative quantity X_t, the generic expression of the linear response in the non-equilibrium steady state far from equilibrium reads

$$\frac{\partial J}{\partial \mathcal{F}} = \left.\frac{\partial^2 f(\mathcal{J};\mathcal{F})}{\partial \mathcal{F}\partial \mathcal{J}}\right|_{\mathcal{J}=J} \times \lim_{t\to+\infty} \frac{\langle X_t^2\rangle_{\rm st} - \langle X_t\rangle_{\rm st}^2}{t}. \tag{1.13}$$

It relates on the one hand, the coefficient $\partial J/\partial \mathcal{F}$ of the linear response of the heat current J to a variation of the thermodynamic force $\mathcal{F}$ and, on the other hand, the infinite-time limit of the variance per unit time of the cumulative quantity X_t in the non-equilibrium steady state, with a coefficient whose expression depends on the system.

In the limit where the thermodynamic force $\mathcal{F}$ vanishes, $\partial J/\partial \mathcal{F}$ tends to the linear-response coefficient near equilibrium, namely the kinetic Onsager coefficient, $L \equiv \partial J/\partial \mathcal{F}|_{\mathcal{F}=0}$. If the system obeys the fluctuation relation (1.12), with X in place of $\mathcal{Q}_2$, then the coefficient in the identity (1.13) takes the universal value $\frac{1}{2}$ and the identity becomes the fluctuation-dissipation relation (also referred to as the Einstein-Green-Kubo relation) between the kinetic Onsager coefficient and the infinite-time limit of the variance per unit time of X_t at equilibrium. In the case of thermal contact, in the limit where the difference $\beta_1 - \beta_2$ between the inverse temperatures vanishes, the ratio between the stationary heat current that goes through the system from heat bath 2 to heat bath 1 and the difference $\beta_1 - \beta_2$ becomes equal to $\frac{1}{2}$ times the variance per unit time of the heat amount exchanged with one thermal bath at equilibrium. (When the total system is at equilibrium, the net heat amount $\mathcal{Q}_1 + \mathcal{Q}_2$ received by the system remains finite at any time, but the variance of the heat amount received from one bath grows linearly in time in the long-time limit).

For a system with a finite number of configurations, the MDB entails a symmetry of the generating function for the infinite-time limit of the cumulants of X_t per unit time. This symmetry takes the generic form (5.58) in terms of the thermodynamic force $\mathcal{F}$. We then show that, in the infinite-time limit, any odd cumulant per unit time $\kappa^{[2n+1]}/t$ can be expressed in terms of even-order cumulants per unit time through the relation

$$\lim_{t\to+\infty} \frac{\kappa^{[2n+1]}(\mathcal{F})}{t} = \sum_{k=0}^{+\infty} d_k \mathcal{F}^{2k+1} \lim_{t\to+\infty} \frac{\kappa^{[2(n+k+1)]}(\mathcal{F})}{t} \qquad \text{for } n = 0, 1, \cdots, \tag{1.14}$$

where d_k is given in (5.68). The corresponding expressions for the ratios $(1/\mathcal{F}) \times \lim_{t\to+\infty} \kappa^{[2n+1]}/t$ may be viewed as generalized Einstein-Green-Kubo relations, where the term "generalized" refers to the fact that they are valid both far from equilibrium and for all cumulants.

We also express the response of any odd cumulant per unit time at any order in the thermodynamic force $\mathcal{F}$ near equilibrium in terms of non-linear response coefficients for even cumulants per unit time at lower orders in $\mathcal{F}$ near equilibrium (see (5.74)). At first order, namely at the level of linear response, one gets the generalized fluctuation-dissipation relations (5.75) where the term "generalized" refers to the fact that they are valid for any odd cumulant.

In the more general situation where there are several independent mean currents between reservoirs, we derive the corresponding generalized Einstein-Green-Kubo relations (6.29). From the latter equations one can derive relations between non-linear response coefficients. The latter relations have already been settled by another method by Andrieux and Gaspard [61]. As noticed by these authors some of them lead to symmetries which are generalizations of the Onsager reciprocity relation.

2 Constraints upon transitions rates

In the present section we review some of the constraints that ergodic deterministic energy-conserving microscopic dynamics puts on the statistical mesoscopic description of a finite system $\mathcal{S}$ which establishes thermal contact between energy reservoirs $\mathcal{B}_a$'s with $a = 1, \ldots, A$.

Indeed, the following situation occurs commonly : the interactions in the whole system allow to define one small part $\mathcal{S}$ in contact with otherwise independent large parts $\mathcal{B}_a$'s. The large parts, which involve a huge number of degrees of freedom, do not interact directly among each other (this gives a criterion to identify the distinct large parts), but are in contact with the small part, which involves only a few degrees of freedom. Moreover each degree of freedom in the small part is directly in contact with at most one large part and can vary only through its interaction with the latter large part. This results in a star-shaped interaction pattern. It is convenient then to forget about the microscopic description of the large parts, and turn to a statistical description of their interactions with the small part. Some general features of the statistical description can be inferred from microscopic ergodicity.

2.1 Ergodicity

2.1.1 Ergodicity in classical Hamiltonian dynamics

In classical mechanics, the time evolution of a system in phase space is described by a Hamiltonian H. If the system is made of several interacting parts, the Hamiltonian is then referred to as the total Hamiltonian, $H \equiv H_{\text{tot}}$, and it splits as $H_{\text{tot}} = H_{\text{dec}} + H_{\text{int}}$, where H_{dec} accounts for the dynamics if the different parts were decoupled and H_{int} accounts for interactions. The energy hypersurface $H_{\text{tot}} = E$, usually a compact set, is invariant under the time evolution, and in a generic situation, this will be the only conserved quantity.

The ergodic hypothesis states that a generic trajectory of the system will asymptotically cover the energy hypersurface uniformly. To be more precise, phase space is endowed with the Liouville measure (i.e. in most standard cases the Lebesgue measure for the product of couples made by every coordinate and its conjugate momentum), which induces a natural measure on the energy hypersurface, and ergodicity means that, in the long run, the time spent by the system in each open set of the energy hypersurface will be proportional to its Liouville measure. Ergodicity can sometimes be built in the dynamics, or proved, but this usually requires immense efforts.

Ergodicity depends crucially on the fact that the different parts are coupled : if $H_{\text{int}} = 0$, each part will have its energy conserved, and motion will take place on a lower-dimensional surface. If H_{int} is very small, the system will spend a long time very nearby this lower-dimensional surface, but, at even longer time scales, ergodicity can be restored. By taking limits in a suitable order (first infinite time and then vanishing coupling among the parts) one can argue that the consequences of ergodicity can be exploited by reasoning only on H_{dec}.

Notice that in this procedure, we have in fact some kind of dichotomy : H_{dec} defines the energy hypersurface, but cannot be used to define the ergodic motion, which is obtained from $H_{\text{tot}} = H_{\text{dec}} + H_{\text{int}}$ via a limiting procedure. So the dynamics conserves H_{dec} but is not determined by H_{dec}.

2.1.2 Ergodicity in deterministic dynamics for discrete variables

Our aim is to translate the above considerations in the context of a large but finite system described by discrete variables such as classical Ising spins.

In the case of discrete dynamical variables, one can still talk about the energy E_{tot} of a configuration, but there is no phase space and no Hamiltonian dynamics available. So there is no obvious canonical time evolution. This is where we exploit the previously mentioned dichotomy: we do not define the time evolution in terms of E_{tot}, but simply impose that the deterministic time evolution preserves E_{dec} and that it respects the star-shaped interaction pattern between the small part and the large parts. Besides we also impose that the time evolution is ergodic.

We consider that time is discrete as well, because ergodicity is most simply expressed in discrete time. Then deterministic dynamics is given by a bijective map, denoted by $\mathcal{T}$ in what follows, on configuration space, applied at each time step to get a new configuration from the previous one. As the configuration space is finite, the trajectories are bound to be closed. Then, for a given initial value E of $E_{\rm dec}$, a specific dynamics $\mathcal{T}$ conserving $E_{\rm dec}$ corresponds to a periodic evolution of the microscopic configuration of the full system inside the energy level $E_{\rm dec} = E$.

Ergodicity entails that the corresponding closed trajectory covers fully the energy level $E_{\rm dec} = E$, and then it must cover it exactly once during a period because the dynamics is one to one. As a consequence the period of the ergodic evolution inside a given energy level of $E_{\rm dec}$ is the same for all choices of ergodic dynamics $\mathcal{T}$ that conserves $E_{\rm dec}$. This period, denoted by N in time step units, depends only on the value E of $E_{\rm dec}$,

$$N = \Omega_{\rm dec}(E), \tag{2.1}$$

where $\Omega_{\rm dec}(E)$ is the total number of microscopic configurations in the level $E_{\rm dec} = E$. This is reminiscent of the microcanonical ensemble. Let us note that in classical mechanics, there is a time reversal symmetry, related to an involution of phase space, changing the momenta to their opposites while leaving the positions fixed (see (C.2)). In the discrete setting, involutions J such that $J\mathcal{T}J = \mathcal{T}^{-1}$ always exist, but there is no obvious candidate among them for representing time reversal and allowing to draw conclusions from it.

In the context of discrete variables the star-shaped interaction pattern is implemented as follows. We may naively assume that the energy conserved by the dynamics $\mathcal{T}$ is simply $E_{\rm dec}$, as if there were no energy for the interactions between the small part and the large ones, but $\mathcal{T}$ must reflect the fact that the large parts interact only indirectly: there is an internal interaction energy $\mathcal{E}(\mathcal{C})$ for every configuration $\mathcal{C}$ of the small part and each change in the small part can be associated with an elementary energy exchange with one of the large parts. If the small part can jump from configuration $\mathcal{C}$ to configuration $\mathcal{C}'$ in a single time step by exchanging energy with large part a, we use the notation $\mathcal{C}' \in \mathbb{F}_a(\mathcal{C})$. In this configuration jump the energy $E_{\rm dec}$ of the global system is conserved and the energy of large part a is changed from E_a to E'_a according to the conservation law

$$E'_a - E_a = \begin{cases} -\left[\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})\right] & \text{if} \quad \mathcal{C}' \in \mathbb{F}_a(\mathcal{C}) \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

while the energies of the other large parts are unchanged. Apart from these energy exchange constraints and from ergodicity, the deterministic dynamics $\mathcal{T}$ is supposed to obey some other natural physical constraints which will be specified later (see subsection 2.2).

As a final remark, we mention how, in a very simple case, some kind of deterministic map $\mathcal{T}$ that preserves $E_{\rm dec}$ and obeys the star-shaped interaction pattern can be associated with a deterministic map $\widetilde{\mathcal{T}}$ that conserves $E_{\rm tot}$. We consider the case where the energy exchange between every large part a and the small part is ensured by an interaction energy $E_{\rm int}^{(a)}$ between a classical spin $\sigma_a^\star$ in part a and a classical spin σ_a in the small part and we consider only maps $\widetilde{\mathcal{T}}$ that not only conserve $E_{\rm tot}$ but also satisfy the following rules for all large parts : (1) spins $\sigma_a^\star$ and σ_a are always flipped at successive time steps (in an order depending on the precise dynamics $\widetilde{\mathcal{T}}$) ; (2) the variations of the interaction energy $E_{\rm int}^{(a)}$ associated with these successive two flips are opposite to each other. Then the map $\mathcal{T}$ that conserves $E_{\rm dec}$ is defined from the map $\widetilde{\mathcal{T}}$ by merging every pair of time steps where $\sigma_a^\star$ and σ_a are successively flipped into a single time step where $\sigma_a^\star$ and σ_a are simultaneously flipped. Indeed, in the latter pair of time steps of $\widetilde{\mathcal{T}}$, by virtue of hypothesis (2), the successive two variations of $E_{\rm int}^{(a)}$ cancel each other and the variation of $E_{\rm tot}$ after these two time steps coincides with the variation of $E_{\rm dec}$, since by definition the latter variation is $\Delta E_{\rm dec} = \Delta E_{\rm tot} - \Delta E_{\rm int}^{(a)}$. Therefore, if map $\widetilde{\mathcal{T}}$ conserves $E_{\rm tot}$ at every time step, then the corresponding map $\mathcal{T}$ where the latter successive two flips occur in a single time step preserves $E_{\rm dec}$: the conservation rule (2.2) is indeed satisfied.. The suppression of time steps in the procedure that defines $\mathcal{T}$ from $\widetilde{\mathcal{T}}$ corresponds to a modification of the accessible configuration space that reflects the fact that the energy level $E_{\rm tot} = E$ and $E_{\rm dec} = E$ do not coincide.

2.1.3 Constraints from ergodicity and interaction pattern upon coarse-grained evolution

We start from the the familiar observation that keeping track of what happens in detail in the large parts is out of our abilities, and often not very interesting anyway. Our ultimate interest is in the evolution of the configuration $\mathcal{C}$ of the small part in an appropriate limit. As an intermediate step, we keep track also of the energies E_a's in the large parts, but not of the detailed configurations in the large parts.

The coarse graining that keeps track only of the time evolution of the configuration $\mathcal{C}$ of the small part and the energies E_a's of the large parts is defined as follows. With each microscopic configuration ξ of the full system we can associate the corresponding configuration $\mathcal{C} = \mathcal{C}(\xi)$ of the small part and the corresponding energy $E_a = E_a(\xi)$ carried by part a. To simplify the notation, we let $\underline{E}$ denote the collection of E_a's, so $\underline{E}$ is a vector with as many coordinates as there are large parts.

As shown in previous subsubsection, if the initial value of the energy $E_{\rm dec}$ is equal to E, then, over a period equal to $N = \Omega_{\rm dec}(E)$ in time step units, the trajectory of ξ under any microscopic ergodic dynamics $\mathcal{T}$ corresponds to a cyclic permutation of all the microscopic configurations in the energy level $E_{\rm dec} = E$. Thus ergodicity entails that at the coarse grained level, if $N_{(\mathcal{C},\underline{E})}$ denotes the occurrence number of $(\mathcal{C},\underline{E})$ during the period of N time steps, $N_{(\mathcal{C},\underline{E})}$ is nothing but $\Omega_{\rm dec}(\mathcal{C},\underline{E})$, the number of microscopic configurations of the full system when the small part is in configuration $\mathcal{C}$ and the large parts have energies E_a's in the energy level $E = E_{\rm dec}$, namely

$$N(\mathcal{C},\underline{E}) = \Omega_{\rm dec}(\mathcal{C},\underline{E}) \tag{2.3}$$

with

$$E = E_{\rm dec} \equiv \mathcal{E}(\mathcal{C}) + \sum_a E_a. \tag{2.4}$$

In the following we fix the value of E and N is called the period of the dynamics, while the energy constraint $E = \mathcal{E}(\mathcal{C}) + \sum_a E_a.$ is often implicit in the notations.

Another crucial point is that, since any specific dynamics $\mathcal{T}$ under consideration respects both the conservation of $E_{\rm dec}$ and the interaction pattern specified at the end of subsubsection 2.1.2, the number of jumps from $(\mathcal{C},\underline{E})$ to $(\mathcal{C}',\underline{E}')$ over the period of N time steps, denoted by $N_{(\mathcal{C},\underline{E}),(\mathcal{C}',\underline{E}')}$, is equal to the number of the reversed jumps from $(\mathcal{C}',\underline{E}')$ to $(\mathcal{C},\underline{E})$ during the same time interval, namely

$$N_{(\mathcal{C},\underline{E}),(\mathcal{C}',\underline{E}')} = N_{(\mathcal{C}',\underline{E}'),(\mathcal{C},\underline{E})}. \tag{2.5}$$

In Appendix A we give graph-theoretic conditions, not related to ergodicity, that ensure this property, and show that they are fulfilled in one relevant example, as a consequence of the star-shaped interaction pattern.

2.2 Markovian approximation for the mesoscopic dynamics

2.2.1 Definition of a Markovian approximation for the mesoscopic dynamics

For our purpose, we first rephrase the coarse-grained evolution as follows. As already noticed, over the period of $N = \Omega_{\rm dec}(E)$ time steps, a trajectory under any microscopic ergodic dynamics $\mathcal{T}$ in the energy level $E_{\rm dec} = E$ corresponds to a cyclic permutation of the N microscopic configurations ξ's in the energy level. Therefore, if the configuration at some initial time is denoted by ξ_1, then the trajectory is represented by the sequence $\omega = \xi_1\xi_2\cdots\xi_N$ where $\xi_{i+1} = \mathcal{T}\xi_i$ with $\xi_{N+1} = \xi_1$. By the coarse-graining procedure that retains only the mesoscopic variable $x \equiv (\mathcal{C},\underline{E})$, the succession of distinct microscopic configurations ω is replaced by $w = x_1x_2\cdots x_N$ where $x_i = x(\xi_i) = (\mathcal{C}(\xi_i),\underline{E}(\xi_i))$. In w various x_i's take the same value, and a so-called transition corresponds to the case $x_{i+1} \neq x_i$, namely the case where the configuration $\mathcal{C}$ of the small system is changed in the jump of the microscopic configuration of the full system from ξ_i to $\xi_{i+1} = \mathcal{T}\xi_i$.

Since the large parts involve many degrees of freedom, the number of times some given value $x = (\mathcal{C},\underline{E})$ appears in the coarse-grained sequence w is huge, and even if ω is given by the

deterministic rule $\xi_{i+1} = \mathcal{T}\xi_i$, there is no such rule to describe the sequence w. Moreover the microscopic configuration at the initial time, ξ_1, is not known so that the coarse-grained sequence that actually appears in the course of time is in in fact a sequence deduced from w by a translation of all indices.

As explained in Appendix B.1, one may associate to the sequence w a (discrete time) Markov chain such that the mean occurrence frequencies of the patterns x and xx' in a stationary stochastic sample are equal to the corresponding values, N_x/N and $N_{xx'}/N$, in the sequence w determined by the dynamics $\mathcal{T}$ (up to a translation of all indices corresponding to a different value of the initial microscopic configuration). Whether this Markovian effective description is accurate depends on several things: the choice of $\mathcal{T}$, the kind of statistical properties of w one wants to check, etc.

We may also argue (see Appendix B.2) that a continuous time description is enough if we restrict our attention to microscopic dynamical maps $\mathcal{T}$'s such that transitions, namely the patterns xx' with $x' \neq x$, are rare and of comparable mean occurrence frequencies over the period of N time steps. In other words, most of the steps in the dynamics amount to reshuffle the configurations of the large parts without changing their energies, leaving the configuration of the small part untouched. The latter physical constraint and the hypothesis of the validity of the Markovian approximation select a particular class of dynamics $\mathcal{T}$.

With these assumptions, we associate to the sequence w of coarse grained variables $(\mathcal{C}, \underline{E})$ a Markov process whose stationary measure shares some of the statistical properties of w, namely the values of the mean occurrence frequencies of length 1 and length 2 patterns. The transition rate from $(\mathcal{C}, \underline{E})$ to $(\mathcal{C}', \underline{E}')$ with $(\mathcal{C}, \underline{E}) \neq (\mathcal{C}', \underline{E}')$ in the approximated Markov process is given by (B.4), where we just have to make the substitutions $N_x = N_{(\mathcal{C},\underline{E})}$ and $N_{xx'} = N_{(\mathcal{C},\underline{E}),(\mathcal{C}',\underline{E}')}$, with the result

$$W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E}) = \frac{N_{(\mathcal{C},\underline{E}),(\mathcal{C}',\underline{E}')}}{\tau\, N_{(\mathcal{C},\underline{E})}} \quad \text{for } (\mathcal{C}, \underline{E}) \neq (\mathcal{C}', \underline{E}'), \tag{2.6}$$

where τ is a time scale such that $W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E})$ is of order unity. The corresponding stationary distribution by given by (B.2),

$$P^W_{\rm st}(\mathcal{C}, \underline{E}) = \frac{N_{(\mathcal{C},\underline{E})}}{N}. \tag{2.7}$$

We recall that $N = \sum_{(\mathcal{C},\underline{E})} N_{(\mathcal{C},\underline{E})}$ and $N_{(\mathcal{C},\underline{E})} = \sum_{(\mathcal{C}',\underline{E}')} N_{(\mathcal{C},\underline{E}),(\mathcal{C}',\underline{E}')}$.

2.2.2 Microcanonical detailed balance and other properties

By virtue of the ergodicity property (2.3) at the coarse-grained level, the transition rate in the approximated Markov process reads

$$W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E}) = \frac{N_{(\mathcal{C},\underline{E}),(\mathcal{C}',\underline{E}')}}{\tau\, \Omega_{\rm dec}(\mathcal{C}, \underline{E})} \quad \text{for } (\mathcal{C}, \underline{E}) \neq (\mathcal{C}', \underline{E}'). \tag{2.8}$$

Meanwhile, by virtue of the ergodicity properties (2.1) and (2.3), the corresponding stationary distribution (2.7) is merely the microcanonical distribution

$$P^W_{\rm st}(\mathcal{C}, \underline{E}) = \frac{\Omega_{\rm dec}(\mathcal{C}, \underline{E})}{\Omega_{\rm dec}(E)} \equiv P_{\rm mc}(\mathcal{C}, \underline{E}). \tag{2.9}$$

Ergodicity also entails that, since all microscopic configurations ξ's in the energy level appear in the sequence ω, all possible values of x also appear in the coarse-grained sequence w : so any mesoscopic state $(\mathcal{C}, \underline{E})$ can be reached from any other mesoscopic state $(\mathcal{C}', \underline{E}')$ by a succession of elementary transitions, even if they are not involved in an elementary transition (i.e. if $N_{(\mathcal{C},\underline{E}),(\mathcal{C}',\underline{E}')} = 0$); in other words the graph associated with the transition rates $W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E})$ is connected, or, equivalently, the Markov matrix defined from the transition rates (see definition below in (2.17)) is *irreducible*.

The constraint (2.5) imposed by the interaction pattern upon the coarse-grained evolution over a period of N time steps entails that the transition rates of the approximated Markov process

defined in (2.8) obey two properties. First,

$$W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E}) \quad \text{and} \quad W(\mathcal{C}, \underline{E} \leftarrow \mathcal{C}', \underline{E}') \tag{2.10}$$

are either both $= 0$ or both $\neq 0$. This property may be called *microreversibility*. Second, if the transition rates do not vanish, they obey the equality

$$\frac{W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E})}{W(\mathcal{C}, \underline{E} \leftarrow \mathcal{C}', \underline{E}')} = \frac{\Omega_{\rm dec}(\mathcal{C}', \underline{E}')}{\Omega_{\rm dec}(\mathcal{C}, \underline{E})} = \frac{P_{\rm mc}(\mathcal{C}', \underline{E}')}{P_{\rm mc}(\mathcal{C}, \underline{E})}. \tag{2.11}$$

Observe that the arbitrary time scale τ has disappeared in this equation. The equality between the ratio of transition rates and the ratio of probabilities in the corresponding stationary distribution is the so-called detailed balance relation. Here the stationary distribution is that of the microcanonical ensemble, and we will refer to relation (2.11) as the *microcanonical detailed balance*.

In appendix C we rederive a microcanonical detailed balance similar to (2.11) when the underlying microscopic dynamics is Hamiltonian and invariant under time reversal. The evolution of the probability distribution of the mesoscopic variables is approximated by a Markov process according to the same scheme as that introduced in subsubsection 2.2.1. Eventually the comparison between the ways in which the microcanonical detailed balance arises in that case and in our previous argument is the following.

- When the microscopic variables are continuous coordinates in phase space and evolve according to a Hamiltonian dynamics, in the framework of statistical ensemble theory the stationary measure for the mesoscopic variables is the measure that is preserved under the microscopic dynamics ; the fact that it coincides with the microcanonical distribution is enforced by the invariance of the Liouville measure in phase space under the Hamiltonian evolution ; the microcanonical detailed balance for mesoscopic variables that are even functions of microscopic momenta mainly arises from the invariance under time reversal of the trajectories in phase space (see (C.7)).

- When the microscopic dynamical variables are discrete and evolve under an energy-conserving map, the stationary measure for mesoscopic variables is defined as the average over a period of the microscopic dynamics ; the fact that it is equal to the microcanonical distribution arises from the ergodicity imposed on the microscopic map $\mathcal{T}$ (see (2.1) and (2.3)) ; the microcanonical detailed balance emerges from the equality between the frequencies of a given transition and the reversed one (over the period needed for the microscopic map to cover the energy level), equality which is enforced by the star-shaped interaction pattern (see (2.5)).

2.2.3 Further consequence of the interaction pattern

In the interaction pattern large parts do not interact directly with one another and $\Omega_{\rm dec}(\mathcal{C}, \underline{E}) = \prod_a \Omega_a(E_a)$, where $\Omega_a(E_a)$ denotes the number of configurations in large part a with energy E_a when it is isolated. Moreover the energy of a single large part is changed in a given transition, so if $(\mathcal{C}', \underline{E}')$ is obtained from $(\mathcal{C}, \underline{E})$ by an energy exchange with bath a that makes $\mathcal{C}$ jump to $\mathcal{C}'$ we have the result, with the notation introduced in (2.2),

$$\text{if } \mathcal{C}' \in \mathbb{F}_a(\mathcal{C}) \quad \frac{\Omega_{\rm dec}(\mathcal{C}', \underline{E}')}{\Omega_{\rm dec}(\mathcal{C}, \underline{E})} = \frac{\prod_b \Omega_b(E'_b)}{\prod_b \Omega_b(E_b)} = \frac{\Omega_a(E'_a)}{\Omega_a(E_a)}. \tag{2.12}$$

We have used the energy conservation rule $E'_b = E_b - \delta_{a,b}\,[\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})]$, so that for $b \neq a$ the multiplicity factors are unchanged in the transition.

The latter ratio of microstate numbers can be expressed in terms of the Boltzmann entropies when each part a is isolated. When Boltzmann constant k_B is set to 1, the dimensionless Boltzmann entropy $S_a^B(E_a)$ for the isolated part a when its energy is equal to E_a is defined by $\Omega_a(E_a) \equiv \exp S_a^B(E_a)$. With these notations, if the transition rate $W(\mathcal{C}', \underline{E}' \overset{\mathbb{F}_a}{\leftarrow} \mathcal{C}, \underline{E})$, where $\mathcal{C}' \in \mathbb{F}_a(\mathcal{C})$, is nonzero, then the transition rate for the reversed jump $W(\mathcal{C}, \underline{E} \overset{\mathbb{F}_a}{\leftarrow} \mathcal{C}', \underline{E}')$, where $\mathcal{C} \in \mathbb{F}_a(\mathcal{C}')$, is

also non zero (see (2.10)) and, by virtue of (2.12) the relation (2.11) is reduced to

$$\frac{W(\mathcal{C}', \underline{E}' \overset{\mathbb{F}_a}{\leftarrow} \mathcal{C}, \underline{E})}{W(\mathcal{C}, \underline{E} \overset{\mathbb{F}_a}{\leftarrow} \mathcal{C}', \underline{E}')} = \frac{\Omega_a(E'_a)}{\Omega_a(E_a)} \equiv e^{S_a^B(E'_a) - S_a^B(E_a)}. \tag{2.13}$$

The latter formula is the first important stage of the argument.

We stress that the present argument does not involve any kind of underlying time reversal. Here the time reversal symmetry arises only at the statistical level of description represented by the Markov evolution ruled by the transition rates.

Notice also that, as only certain ratios are fixed, different ergodic deterministic microscopic dynamics can lead to very different transition rates, a remnant of the fact that the coupling between a large part and the small part can take any value a priori.

Formula (2.13) is also a clue to understand a contrario what kind of physical input is needed for the homogeneous Markov approximation to be valid. Indeed, why didn't we do the homogeneous Markov approximation directly on the small part ? We could certainly imagine dynamics making this a valid choice. However, it is in general incompatible with the pattern of interactions (see the end of subsubsection 2.1.2) which is the basis of our argument. For instance, if in the coarse-graining procedure we had retained only the configurations $\mathcal{C}$'s of the small part, then the corresponding graph introduced in Appendix A would have been a cycle instead of a tree in the case of a small part made of two spins (see paper II), and the crucial property (2.5) would have been lost : over the period of N time steps of the microscopic dynamics $N_{\mathcal{C},\mathcal{C}'} \neq N_{\mathcal{C}',\mathcal{C}}$. In fact, we may expect, or impose on physical grounds, that Ω_a will be exponentially large in the size of large part a (i.e. its number of degrees of freedom $\mathcal{N}_a$), so that even the ratio $\Omega_a(E'_a)/\Omega_a(E_a) = \Omega_a(E_a - [\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})])/\Omega_a(E_a)$ will vary significantly over the trajectory, meaning that transition probabilities involving only the small part cannot be taken to be constant along the trajectory: the energies of the large parts are relevant variables.

2.3 Transient regime when large parts are described in the thermodynamic limit

2.3.1 Large parts in the thermodynamic limit

We now assume that the large parts are large enough that they are accurately described by a thermodynamic limit, which we take at the most naive level. To recall what we mean by that, we concentrate on one large part for a while, and suppress the index used to label it. Suppose this large part has $\mathcal{N}$ degrees of freedom, and suppose that energies are close to an energy E for which the Boltzmann entropy is $S^B(E)$. That the thermodynamic limit exists means that if one lets $\mathcal{N} \to +\infty$ while the ratio $E/\mathcal{N}$ goes to a finite limit ϵ, there is a differentiable function $s^B(\epsilon)$ such that the ratio $S^B(E)/\mathcal{N}$ goes to $s^B(\epsilon)$. The quantity

$$\frac{ds^B}{d\epsilon} \equiv \beta \tag{2.14}$$

is nothing but the inverse temperature. In that case, as long as $\Delta E \ll E$ (where E scales as $\mathcal{N}\Delta e$ with Δe some finite energy scale), $S^B(E + \Delta E) - S^B(E) \to \beta \Delta E$ when $\mathcal{N} \to +\infty$. For $\mathcal{N}$ large enough, the relation $S^B(E + \Delta E) - S^B(E) \sim \beta \Delta E$ is a good approximation.

2.3.2 Transient regime and modified detailed balance (MDB)

Notice that when transitions occur, which, by the definition of τ in (2.6), happens typically once on the macroscopic time scale, the changes in the energies of the large parts are finite, so that over long windows of time evolution, involving many changes in the small part, the relation

$$S_a^B(E'_a) - S_a^B(E_a) \sim \beta_a \, [E'_a - E_a] \tag{2.15}$$

is not spoiled, where E'_a and E_a are the energies in large part a at any moment within the window.

In fact the larger the large parts, the longer the time window for which (2.15) remains valid. The relation between the sizes $\mathcal{N}_a$'s of the large parts and of the length of the time window depends on the details of $\mathcal{T}$ (which still has to fulfill the imposed physical conditions). This relation also depends on the values of the energy per degree of freedom in every large part, $E_a/\mathcal{N}_a$, which are essentially constant in such a window.

Because of the ergodicity hypothesis, the largest window (of size comparable to the period of $\mathcal{T}$ to logarithmic precision) has the property that the energies E_a's in the large parts will be such that all β_a's are close to each other and the system will be at equilibrium. Indeed, inside the largest time window, the system remains in the region of the energy level where the energies E_a's are the most probable, and in the thermodynamic limit the most probable values for the E_a's in the microcanonical ensemble are the values $E_a^\star$'s that maximize the product $\prod_a \Omega_a(E_a)$ under the constraint $E = \sum_a E_a$ (since the system energies are negligible with respect to those of the large parts). The latter maximization condition is equivalent to the equalities $ds^B/d\epsilon_a(\epsilon_a^\star) = ds^B/d\epsilon_b(\epsilon_b^\star)$ for all pairs of large parts.

However, if the system starts in a configuration such that the β_a's are distinct, the time window over which $E_a/\mathcal{N}_a$ and β_a are constant (to a good approximation) will be short with respect to the period of the microscopic dynamics, but long enough that (2.15) still holds for a long time interval. Then by putting together the information on the ratio of transition rates in terms of Boltzmann entropies (2.13), the transient regime approximation (2.15) and the energy conservation (2.2), we get

$$\frac{W(\mathcal{C}', \underline{E}' \overset{\mathbb{F}_a}{\leftarrow} \mathcal{C}, \underline{E})}{W(\mathcal{C}, \underline{E} \overset{\mathbb{F}_a}{\leftarrow} \mathcal{C}', \underline{E}')} \sim e^{-\beta_a [\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})]}. \tag{2.16}$$

Now the right-hand side depends only on the configurations of the small system, and the parameters β_a's are constants. Letting the large parts get larger and larger while adjusting the physical properties adequately, we can ensure that the time over which (2.16) remains valid gets longer and longer, so, in the thermodynamic limit for the large parts, the transient regime lasts forever. This situation is our main interest in what follows.

The transition rates in the transient regime correspond to a Markov matrix $\mathbb{M}$ defined by

$$(\mathcal{C}'|\mathbb{M}|\mathcal{C}) = \begin{cases} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) & \text{if} \quad \mathcal{C}' \neq \mathcal{C} \\ -\sum_{\mathcal{C}''} (\mathcal{C}''|\mathbb{W}|\mathcal{C}) & \text{if} \quad \mathcal{C}' = \mathcal{C}. \end{cases} \tag{2.17}$$

As well as the transition rates $W(\mathcal{C}', \underline{E}' \overset{\mathbb{F}_a}{\leftarrow} \mathcal{C}, \underline{E})$ they must satisfy the three consequences derived from the properties of the underlying microscopic deterministic dynamics pointed out in subsubsection 2.1.2, namely ergodicity, energy conservation and specific interaction pattern. First, as shown in subsubsection 2.2.2, the Markov matrix is irreducible, or in other words the graph associated with the transition rates is connected (see (1.1)). Second, from (2.10) the transition rates must obey the microscopic reversibility condition (1.2) for any couple of configurations $(\mathcal{C}, \mathcal{C}')$. Third, from (2.16) one gets a constraint obeyed by the ratio of transition rates in the transient regime,

$$\text{for } \mathcal{C}' \in \mathbb{F}_a(\mathcal{C}) \quad \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')} = e^{-\beta_a [\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})]}. \tag{2.18}$$

The latter relation is the so-called modified detailed balance (MDB), which is also referred to in the literature as the "generalized detailed balance".

We stress that, by selecting a time window while taking the thermodynamic limit for the large parts, the microcanonical detailed balance (2.11) is replaced by the modified detailed balance (2.13), except in the case of the largest time window where all β_a's are equal. In the latter case, the microcanonical detailed balance (2.11) is replaced by the canonical detailed balance and the statistical time reversal symmetry is preserved. Indeed, the equilibrium thermodynamic regime is reached either if the we start from a situation in which $\prod_a \Omega_a(E_a)$ is close to its maximum along the trajectory in the energy level $E_{\text{dec}} = E$, or if we wait long enough so that $\prod_a \Omega_a(E_a)$ becomes close to this maximum. As recalled above, this is true for most of the period of the

microscopic dynamics, but reaching this situation may however take a huge number of time steps if the starting point was far from the maximum. By an argument similar to that used in the derivation of (2.15), when $\prod_a \Omega_a(E_a)$ is closed to its maximum and the large parts are considered in the thermodynamic limit, all β_a's are equal to the same value β and the relative weight of two configurations in the microcanonical ensemble, $P_{\rm mc}(\mathcal{C}', \underline{E}')/P_{\rm mc}(\mathcal{C}, \underline{E})$ given by (2.9), is shown to tend to $\exp\left(-\beta[\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})]\right)$. Then the equilibrium microcanonical distribution $P_{\rm mc}(\mathcal{C}, \underline{E})$ tends to the canonical distribution

$$P^{\beta}_{\rm can}(\mathcal{C}) \equiv \frac{e^{-\beta\mathcal{E}(\mathcal{C})}}{Z(\beta)}, \tag{2.19}$$

where $Z(\beta)$ is the canonical partition function at the inverse temperature β. Meanwhile, the detailed balance relation (2.11) in the microcanonical equilibrium ensemble for the transition rates $W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E})$ becomes a detailed balance relation in the canonical ensemble at the inverse temperature β of the whole system for the transition rates $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$, namely

$$\frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')} = e^{-\beta[\mathcal{E}(\mathcal{C}')-\mathcal{E}(\mathcal{C})]} = \frac{P^{\beta}_{\rm can}(\mathcal{C}')}{P^{\beta}_{\rm can}(\mathcal{C})}. \tag{2.20}$$

The modified detailed balance (2.18), valid in transient regimes, differs from the latter detailed balance in the canonical ensemble by two features : the various β_a's of the distinct large parts appear in place of the common equilibrium inverse temperature β, and the stationary distribution for the transition rates is not known a priori.

We conclude this discussion with the following remarks. We have not tried to exhibit explicit physical descriptions of the large parts, or explicit formulæ for the dynamical map $\mathcal{T}$. Though it is not too difficult to give examples for fixed sizes of the large parts, it is harder to get a family of such descriptions sharing identical physical properties for varying large part sizes, a feature which is crucial to really make sense of the limits we took blindly in our derivation. It is certainly doable, but cumbersome, and we have not tried to pursue this idea. Let us note also that in principle, taking large parts of increasing sizes can be used to enhance the validity of the approximation of the (discrete-time) Markov chain by a (continuous-time) Markov process. As the physics of the continuous time limit does not seem to be related to the physics of convergence towards a heat bath description we have preferred to keep the discussion separate, taking a continuous-time description as starting point.

2.3.3 Expression of MDB in terms of exchange entropy variation

Observe that though we have given no detailed analysis of the dynamics or the statistical properties of the large parts, their influence on the effective Markov dynamics of the small system enters only through the inverse temperatures β_a defined in (2.14). So we can consistently assume that each large part becomes a thermal bath with its own temperature. The leading term in $S^B_a(E'_a) - S^B_a(E_a)$ is the variation $\delta S^{TH}_a(\mathcal{C}' \leftarrow \mathcal{C})$ of the thermodynamic entropy of bath a when it flips the small system from configuration $\mathcal{C}$ to configuration $\mathcal{C}'$,

$$\delta S^{TH}_a(\mathcal{C}' \leftarrow \mathcal{C}) = \begin{cases} \beta_a \left[E'_a - E_a\right] & \text{if} \quad \mathcal{C}' \in \mathbb{F}_a(\mathcal{C}) \\ 0 & \text{otherwise.} \end{cases} \tag{2.21}$$

Then we have an idealized description of a thermal contact between heat baths. This is the situation on which we concentrate in this paper.

By definition of a heat source, the variation $\delta S^{TH}_a(\mathcal{C}' \leftarrow \mathcal{C})$ of the thermodynamic entropy of bath a when it flips the system from configuration $\mathcal{C}$ to configuration $\mathcal{C}'$ reads

$$\delta S^{TH}_a(\mathcal{C}' \leftarrow \mathcal{C}) = -\beta_a \delta q_a(\mathcal{C}' \leftarrow \mathcal{C}), \tag{2.22}$$

where $\delta q_a(\mathcal{C}' \leftarrow \mathcal{C})$ is the heat received by the small system from part a. According the expression (2.21) and to the energy conservation relation (2.2),

$$\begin{cases} \delta q_a(\mathcal{C}' \leftarrow \mathcal{C}) = \mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C}) & \text{if} \quad \mathcal{C}' \in \mathbb{F}_a(\mathcal{C}) \\ \delta q_a(\mathcal{C}' \leftarrow \mathcal{C}) = 0 & \text{otherwise.} \end{cases} \tag{2.23}$$

Let us introduce $\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C})$ the exchange entropy variation of the small system (see footnote 2) that is associated with the heat exchanges with the thermostats when the small system goes from configuration $\mathcal{C}$ to configuration $\mathcal{C}'$. Thanks to the definition (2.21)

$$\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C}) \equiv -\sum_a \delta S_a^{TH}(\mathcal{C}' \leftarrow \mathcal{C}), \tag{2.24}$$

namely, in the case of a pure energy reservoir (which does not exchange particles)

$$\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C}) \equiv \sum_a \beta_a \delta q_a(\mathcal{C}' \leftarrow \mathcal{C}). \tag{2.25}$$

Then the modified detailed balance (2.18) can be rewritten in a form which does not involve explicitly the heat bath responsible for the transition from $\mathcal{C}$ to $\mathcal{C}'$,

$$\frac{(\mathcal{C}'|\mathbb{W}|C)}{(\mathcal{C}|\mathbb{W}|C')} = e^{-\delta_{\rm exch}S(\mathcal{C}'\leftarrow\mathcal{C})}. \tag{2.26}$$

3 Master equation, exchange entropy flow and various entropy variations

3.1 Evolution of the probability distribution without MDB (known results)

In this subsection we recall previously known results which are important milestones to our original results and which are consequences of the first two properties (1.1) and (1.2), among the three mesoscopic conditions derived from the ergodicity of the underlying conservative deterministic dynamics.

The starting point is that the evolution of the probability $P(\mathcal{C};t)$ that the system is in configuration $\mathcal{C}$ at time t is ruled by the master equation

$$\frac{dP(\mathcal{C};t)}{dt} = \sum_{\mathcal{C}'\neq\mathcal{C}} (\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}';t) - \sum_{\mathcal{C}'\neq\mathcal{C}} (\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t) = \sum_{\mathcal{C}'} (\mathcal{C}|\mathbb{M}|\mathcal{C}')P(\mathcal{C}';t), \tag{3.1}$$

where $\mathbb{M}$ is defined in (2.17). Since $P(\mathcal{C};t)$ is to be interpreted as a probability distribution, it has to satisfy the positivity and normalization conditions,

$$\forall \mathcal{C} \qquad P(\mathcal{C};t) \geq 0 \quad \text{and} \quad \sum_{\mathcal{C}} P(\mathcal{C};t) = 1. \tag{3.2}$$

3.1.1 Generic properties

According to the theory of systems of ordinary differential equations with constant coefficients, the solution $P(t)$ of the equation (3.1) exists and is unique for a given initial function $P(t=0)$. The Markov matrix $\mathbb{M}$ obeys the property $\sum_{\mathcal{C},\mathcal{C}'}(\mathcal{C}|\mathbb{M}|\mathcal{C}')P(\mathcal{C}';t) = 0$ for any P and this ensures that the normalization constraint is preserved under the time evolution,

$$\sum_{\mathcal{C}} P(C;t=0) = 1 \qquad \Rightarrow \qquad \forall t>0 \quad \sum_{\mathcal{C}} P(C;t) = 1. \tag{3.3}$$

The above property of $\mathbb{M}$ also ensures the existence of at least one stationary solution, but there is no argument for every stationary solution to obey the positivity constraint in (3.2) without further assumptions.

3.1.2 Properties arising from the irreducibility condition

When the transition rates satisfy the irreducibility condition (1.1), and if positivity and normalization (3.2) are satisfied at the initial time, the solution of the master equation (3.1) not only meets these conditions for being a probability distribution at any subsequent time, but it even has the more stringent property that any configuration has a strictly non-vanishing weight,

$$\forall t > 0 \quad \forall \mathcal{C} \qquad P(\mathcal{C}; t) > 0. \tag{3.4}$$

This result can be derived in the framework of the theory of ordinary differential equations with constant coefficients [54].

Moreover the irreducibility condition (1.1) allows one to build at least formally a stationary solution which fulfills the conditions (3.2) for being a probability (with the even more stringent property (3.4)). This solution is obtained in the framework of graph theory by using the network representation of the master equation; the corresponding expression is called Kirchhoff's theorem in Ref.[54].

The irreducibility condition (1.1) also entails that the stationary solution of the master equation is unique, and therefore coincides with the expression given by Kirchhoff's theorem. The uniqueness of the stationary solution can be derived either in the framework of matrix theory by using the Perron-Frobenius theorem (see for instance [57]) or in the framework of the theory of ordinary differential equations with constant coefficients [54].

The latter derivation uses the stability criterion introduced by Schögl [62]. As noticed by Schlögl a good candidate for the Liapunov function involved in a stability criterion is the relative entropy introduced by Kullback and Leibler [63] in the context of information theory, namely

$$S_{\rm rel}[P(t)|P_{\rm st}] \equiv \sum_{\mathcal{C}} P(\mathcal{C}; t) \ln \frac{P(\mathcal{C}; t)}{P_{\rm st}(\mathcal{C})}. \tag{3.5}$$

The relative entropy is well-defined at any time according to (3.4). (In the generic case $S_{\rm rel}[P|P_0]$ is well defined if $P_0(\mathcal{C}) = 0$ implies $P(\mathcal{C}) = 0$.) The definition of $S_{\rm rel}[P|P_0]$, with any given P_0, entails that $S_{\rm rel}[P|P_0]$ is positive for any P and vanishes only when P is equal to P_0 (by virtue of the inequality $\ln x < x - 1$ if $x > 0$ and $x \neq 1$). Therefore

$$S_{\rm rel}[P|P_{\rm st}] > 0 \quad \text{if } P \neq P_{\rm st} \text{ and } \quad S_{\rm rel}[P_{\rm st}|P_{\rm st}] = 0. \tag{3.6}$$

The definition also ensures that $S_{\rm rel}[P|P_0]$ is convex (i.e. concave upward) for any P, so that

$$\delta^{(2)} S_{\rm rel}[P|P_{\rm st}] > 0 \quad \text{for any } P, \tag{3.7}$$

with

$$\delta^{(2)} f[P] \equiv \frac{1}{2} \sum_{\mathcal{C},\mathcal{C}'} \left. \frac{\partial^2 f}{\partial \widetilde{P}(\mathcal{C}) \partial \widetilde{P}(\mathcal{C}')} \right|_{\widetilde{P}=P} \delta P(\mathcal{C}) \delta P(\mathcal{C}') \tag{3.8}$$

and $\delta P(\mathcal{C}) \equiv \widetilde{P}(\mathcal{C}) - P(\mathcal{C})$. Moreover, because of the structure of the master equation (3.1) combined with the properties $\ln x \leq x - 1$ for any $x > 0$ and $\sum_{\mathcal{C}'} (\mathcal{C}|\mathbb{M}|\mathcal{C}') P_{\rm st}(\mathcal{C}') = 0$, the time derivative of $S_{\rm rel}[P(t)|P_{\rm st}]$ is negative at any time [54],

$$\frac{dS_{\rm rel}[P(t)|P_{\rm st}]}{dt} \leq 0. \tag{3.9}$$

The properties (3.6) and (3.9) define a Liapunov function and ensure that for any initial distribution $P(t_0)$ *in the vicinity* of $P_{\rm st}$ $\lim_{t\to+\infty} P(t) = P_{\rm st}$ (because the property (3.6) ensures that $\delta^{(2)} S_{\rm rel}[P|P_{\rm st}] > 0$ for any P in the vicinity of $P_{\rm st}$). With the extra property (3.7) one can apply the stability theorem by Schlögl which states that $\lim_{t\to+\infty} P(t) = P_{\rm st}$ *for any* initial distribution $P(t = 0)$. The interpretation given by Schögl of the stability condition $dS_{\rm rel}[P(t)|P_{\rm st}]/dt \leq 0$ can be rephrased as follows. If the observer does not know more than that the system was initially

in some unknown state P_0 of the stability region, then an unbiased estimate for the state $P(t)$ at time t would be $P_{\rm st}$. But if the observer knows P_0 by measurements, then his excess knowledge at initial time is equal to $S_{\rm rel}[P_0|P_{\rm st}]$, and the property $dS_{\rm rel}[P(t)|P_{\rm st}]/dt \leq 0$ reflects the fact that the spontaneous development of the states after the last observation can only go such that this knowledge does not increase.

3.2 Microscopic currents

3.2.1 Definitions

Our results in the framework of Markovian stochastic dynamics described by a master equation can be written in compact forms if we introduce the generic current $j(\mathcal{C})$ for a microscopic variation when the system goes out of a given configuration $\mathcal{C}$. Such a current is associated either with the variation $\mathcal{O}(\mathcal{C}') - \mathcal{O}(\mathcal{C})$ of a configuration observable $\mathcal{O}(\mathcal{C})$ and then

$$j_{\mathcal{O}}(\mathcal{C}) \equiv \sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) \left[\mathcal{O}(\mathcal{C}') - \mathcal{O}(\mathcal{C})\right], \tag{3.10}$$

or more generally with some exchange quantity $\delta K(\mathcal{C}' \leftarrow \mathcal{C})$ such as a heat amount or the variation of the reservoir entropies when the system jumps from configuration $\mathcal{C}$ ot $\mathcal{C}'$, and then

$$j_{\delta K}(\mathcal{C}) \equiv \sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) \delta K(\mathcal{C}' \leftarrow \mathcal{C}). \tag{3.11}$$

We include in the definition of an exchange quantity that it obeys the antisymmetry property $\delta K(\mathcal{C}' \leftarrow \mathcal{C}) = -\delta K(\mathcal{C} \leftarrow \mathcal{C}')$. We notice that the above definitions are valid even if the observable $\mathcal{O}$ (or the variation $\delta K(\mathcal{C}' \leftarrow \mathcal{C})$) depends explicitly on time.

The average of a current $j(\mathcal{C})$ at time t is given by the generic formula for the mean value $\langle \mathcal{O} \rangle_t$ of a configuration observable $\mathcal{O}(\mathcal{C})$ at time t, namely

$$\langle \mathcal{O} \rangle_t \equiv \sum_{\mathcal{C}} \mathcal{O}(\mathcal{C}) P(\mathcal{C}; t). \tag{3.12}$$

According to the master equation (3.1) the time derivative of the mean value $\langle \mathcal{O} \rangle_t$ of a configuration observable $\mathcal{O}(\mathcal{C})$ which does not depend explicitly on time is equal to the mean value $\langle j_{\mathcal{O}} \rangle_t$ of the associated current $j_{\mathcal{O}}(\mathcal{C})$,

$$\frac{d\langle \mathcal{O} \rangle_t}{dt} = \langle j_{\mathcal{O}} \rangle_t, \tag{3.13}$$

where $j_{\mathcal{O}}(\mathcal{C})$ is the current (3.10) of observable $\mathcal{O}$ associated with the stochastic jumps going out of configuration $\mathcal{C}$.

3.2.2 Probabilistic interpretation of microscopic observable currents

With every observable i.e. with every real-valued function $\mathcal{O}(\mathcal{C})$ on the configuration space, one can associate the random process $\mathcal{O}(\mathcal{C}_t)$ where $\mathcal{C}_t$ denotes the configuration of the system at time t. Note that $P(\mathcal{C}; t)$ is nothing but $P(\mathcal{C}_t = \mathcal{C})$, so that $\langle \mathcal{O} \rangle_t$ as defined above could also be written $\langle \mathcal{O}(\mathcal{C}_t) \rangle$.

The process associated with the current observable $j_{\mathcal{O}}(\mathcal{C})$ at time t has a clear probabilistic meaning. Without explaining the details, let us just say that the decomposition

$$\mathcal{O}(\mathcal{C}_t) = \left(\mathcal{O}(\mathcal{C}_t) - \int_0^t ds j_{\mathcal{O}}(\mathcal{C}_s) \right) + \left(\int_0^t ds j_{\mathcal{O}}(\mathcal{C}_s) \right) \equiv M_t + N_t, \tag{3.14}$$

called the Doob-Meyer decomposition of the process $\mathcal{O}(\mathcal{C}_t)$, expresses $\mathcal{O}(\mathcal{C}_t)$ as the sum of a martingale M_t and a predictible process N_t vanishing at $t = 0$. Such a decomposition is unique. Informally a martingale is a process whose expectation in the future knowning all the past up to

now is equal to its present value. In particular, the expectation $\langle M_t \rangle$ is equal to 0. Taking this as a fact, one gets immediately (3.13). Informally again, N_t is predictible because its value at $t+dt$, N_{t+dt}, is not sensitive to the randomness (i.e. to a possible jump occuring) between t and $t+dt$.

Even if $\mathcal{O}(\mathcal{C}_t)$ depends only on the configuration at time t, this is not true anymore of M_t and N_t which in general depend on the history up to time t. Moreover, the decomposition does not behave trivially under nonlinear maps, so that N_t^2 is not the predictible process that appears in the Doob-Meyer decomposition of $[\mathcal{O}(\mathcal{C}_t)]^2$.

3.2.3 Mean energy time derivative

As an example of (3.13), the time derivative of the mean energy $\langle \mathcal{E} \rangle_t$ is equal to the mean value of the energy current $j_{\mathcal{E}}(\mathcal{C})$. In the present model, energy variations are only due to heat exchanges with the two thermostats according to the conservation rule

$$\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C}) = \delta q_1(\mathcal{C}' \leftarrow \mathcal{C}) + \delta q_2(\mathcal{C}' \leftarrow \mathcal{C}), \tag{3.15}$$

where $\delta q_a(\mathcal{C}' \leftarrow \mathcal{C})$ is the heat received from thermal bath a for a jump of configuration as defined in (2.23). Therefore the energy current $j_{\mathcal{E}}(\mathcal{C})$ can be split into two heat currents $j_{\delta q_1}(\mathcal{C})$ and $j_{\delta q_2}(\mathcal{C})$ received from the thermal baths 1 and 2 respectively. Then the evolution equation (3.13) applied to the energy observable $\mathcal{E}$ can be rewritten as

$$\frac{d\langle \mathcal{E} \rangle_t}{dt} = \langle j_1 \rangle_t + \langle j_2 \rangle_t, \tag{3.16}$$

where j_a is a short notation for the instantaneous heat current received from thermal bath a by the system when it leaves configuration $\mathcal{C}$:

$$j_a(\mathcal{C}) \equiv j_{\delta q_a}(\mathcal{C}) \equiv \sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C})\, \delta q_a(\mathcal{C}' \leftarrow \mathcal{C}). \tag{3.17}$$

Note that, unless $\beta_1 = \beta_2$, there is no observable $\mathcal{O}_a$ for which $j_{\delta q_a}(\mathcal{C})$ would be equal to $j_{\mathcal{O}_a}(\mathcal{C})$ with the definition of $j_{\mathcal{O}_a}(\mathcal{C})$ given in (3.10). Moreover, under the microscopic reversibility condition (1.2), if $\beta_1 = \beta_2 = \beta$ the modified detailed balance (2.18) becomes the canonical detailed balance (2.20). Then the current associated with any observable $\mathcal{O}_a$ (more generally any exchange quantity) has a zero mean in the stationary equilibrium state with distribution $P^{\beta}_{\rm can}$: $\langle j_{\mathcal{O}_a} \rangle_{P^{\beta}_{\rm can}} = 0$.

3.2.4 Exchange entropy flow

The heat currents are associated with an exchange entropy variation of the system (see footnote 2 for the meaning) according to the relation (2.25). Similarly to the definition (3.17) of the microscopic heat current $j_a(\mathcal{C})$ in terms of the heat amount $\delta q_a(\mathcal{C}' \leftarrow \mathcal{C})$, the microscopic exchange entropy current $j_{\delta_{\rm exch} S}(\mathcal{C})$ when the system goes out of configuration $\mathcal{C}$ is defined from the exchange entropy variation $\delta_{\rm exch} S(\mathcal{C}' \leftarrow \mathcal{C})$ as

$$j_{\delta_{\rm exch} S}(\mathcal{C}) \equiv \sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) \delta_{\rm exch} S(\mathcal{C}' \leftarrow \mathcal{C}). \tag{3.18}$$

The exchange entropy flow is defined as

$$\frac{d_{\rm exch} S}{dt} \equiv \sum_{\mathcal{C},\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) \delta_{\rm exch} S(\mathcal{C}' \leftarrow \mathcal{C}) P(\mathcal{C}; t). \tag{3.19}$$

It can be expressed as the average of the current $j_{\delta_{\rm exch} S}(\mathcal{C})$,

$$\frac{d_{\rm exch} S}{dt} = \langle j_{\delta_{\rm exch} S} \rangle_t. \tag{3.20}$$

The current definitions (3.17) and (3.18) together with the relation (2.25) between the exchange entropy variation and heat transfers imply that

$$\frac{d_{\rm exch}S}{dt} = \beta_1\langle j_1\rangle_t + \beta_2\langle j_2\rangle_t. \tag{3.21}$$

In the stationary state, the time derivative of the mean energy vanishes, namely $d < \mathcal{E} >_{\rm st}/dt = 0$, where $\langle\cdots\rangle_{\rm st}$ denotes an average with the stationary distribution $P_{\rm st}$. Henceforth, according to the evolution equation (3.16) of $< \mathcal{E} >_{\rm st}$,

$$\langle j_1\rangle_{\rm st} + \langle j_2\rangle_{\rm st} = 0. \tag{3.22}$$

By inserting the current balance (3.22) into the expression (3.21) of the exchange entropy flow, we get

$$\left.\frac{d_{\rm exch}S}{dt}\right|_{\rm st} = -(\beta_1 - \beta_2)\langle j_2\rangle_{\rm st}. \tag{3.23}$$

3.3 Evolution of the Shannon-Gibbs entropy

3.3.1 Definition of the entropy production rate

The dimensionless Shannon-Gibbs entropy (where the Boltzmann constant is set equal to 1) is defined from the configuration probability distribution $P(\mathcal{C};t)$ as

$$S^{SG}[P(t)] \equiv -\sum_{\mathcal{C}} P(\mathcal{C};t)\ln P(\mathcal{C};t) = -\langle \ln P(t)\rangle_t. \tag{3.24}$$

When the evolution is ruled by the master equation (3.1), the time derivative of $S^{SG}(t)$ takes the form

$$\frac{dS^{SG}}{dt} = -\sum_{\mathcal{C},\mathcal{C}'}(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t)\ln\frac{P(\mathcal{C}';t)}{P(\mathcal{C};t)} = -\langle j_{\ln P(t)}\rangle_t, \tag{3.25}$$

where we have used the definition (3.10) of an observable current $j_{\mathcal{O}}(\mathcal{C};t)$, also valid in the case of an observable which depends explicitly on time.

As it is done for the phenomenological entropy introduced in the thermodynamics of irreversible processes [5, 64] the time derivative of $S^{SG}(t)$ can be split into two contributions, an exchange (or external) part arising from exchanges with the external reservoirs, $d_{\rm exch}S/dt$, and an internal (or irreversible) part due to the internal irreversible processes in the system, $d_{\rm int}S^{SG}/dt$,

$$\frac{dS^{SG}}{dt} \equiv \frac{d_{\rm exch}S}{dt} + \frac{d_{\rm int}S^{SG}}{dt}. \tag{3.26}$$

By virtue of its definition (3.19), $d_{\rm exch}S/dt$ is expressed in terms of the exchange entropy variation $\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C})$, associated with a jump of the system from a microscopic configuration $\mathcal{C}$ to another one $\mathcal{C}'$. By virtue of the definition (3.26) $d_{\rm int}S^{SG}/dt$ is a functional of $P(\mathcal{C};t)$ determined from dS^{SG}/dt given in (3.25) and $d_{\rm exch}S/dt$ given in (3.19).

3.3.2 Comparison with the thermodynamics of irreversible processes

An implicit postulate in the modern literature is that, when the system is out of equilibrium and evolves on time scales far smaller than the reservoirs, one can still define some universe entropy with the following properties. It coincides with its equilibrium statistical expression when the system and the reservoirs are at equilibrium; when equilibrium conditions are not fulfilled, the equilibrium Gibbs entropy of the system is replaced by its instantaneous Shannon-Gibbs entropy while the reservoir entropies can be approximated by their thermodynamic entropies. In other words, the variation of the out-of-equilibrium universe entropy is the sum of the variation of the

Shannon-Gibbs entropy of the system and the variation of the total thermodynamic entropy of the reservoirs, namely

$$\frac{dS_{\rm Univ}}{dt} = \frac{dS^{SG}}{dt} + \frac{dS^{TH}_{\rm res}}{dt}. \tag{3.27}$$

On the other hand, by definition, the exchange entropy flow $d_{\rm exch}S/dt$ received by the system from the reservoirs is the opposite of the time-derivative of the thermodynamic entropy of the reservoirs $d_{\rm exch}S/dt \equiv -dS^{TH}_{\rm res}/dt$. As a consequence, by virtue of definition (3.26),

$$\frac{dS_{\rm Univ}}{dt} = \frac{d_{\rm int}S^{SG}}{dt}. \tag{3.28}$$

The interpretation of the latter equality is that, since the reservoirs are at thermodynamic equilibrium, the variation rate of the universe entropy, $dS_{\rm Univ}/dt$, is equal to the production rate of the Shannon-Gibbs entropy in the system, namely the internal part $d_{\rm int}S^{SG}/dt$ in the time-derivative dS^{SG}/dt of the system Shannon-Gibbs entropy. As shown below, if the modified detailed balance is obeyed, the universe entropy increases when the system is out-of-equilibrium, as the universe entropy increases between two equilibrium states, according to the second principle of thermodynamics.

In the stationary state the time derivative of the Shannon-Gibbs entropy (3.24) vanishes, $dS^{SG}[P_{\rm st}]/dt = 0$, and the decomposition (3.26) leads to

$$\left.\frac{d_{\rm int}S^{SG}}{dt}\right|_{\rm st} = -\left.\frac{d_{\rm exch}S}{dt}\right|_{\rm st}. \tag{3.29}$$

Then according to (3.23), $d_{\rm int}S^{SG}/dt|_{\rm st}$ can be written as the entropy production rate introduced in the framework of irreversible thermodynamics when there is only one independent mean instantaneous current J (see for instance Ref.[65]), namely in the form

$$\left.\frac{d_{\rm int}S^{SG}}{dt}\right|_{\rm st} = \mathcal{F}J, \tag{3.30}$$

where $\mathcal{F}$ is the so-called thermodynamic force associated with the heat current J. In the present case $\langle j_2\rangle_{\rm st} = -\langle j_1\rangle_{\rm st}$ and in (3.30) we can make the identification

$$J = \langle j_2\rangle_{\rm st} \quad \text{and} \quad \mathcal{F} = \beta_1 - \beta_2. \tag{3.31}$$

(Another identification might have been $J = \langle j_1\rangle_{\rm st}$ and $\mathcal{F} = \beta_2 - \beta_1$.) If $T_2 > T_1$, $\mathcal{F}$ is positive and the positivity of $d_{\rm int}S^{SG}_{\rm st}/dt$ (settled in next subsubsection when the modified detailed balance is satisfied) ensures that the mean current $\langle j_2\rangle_{\rm st}$ received from heat bath 2 is also positive.

3.3.3 Entropy production rate under MDB

When the transition rates obey the modified detailed balance (2.26), the microscopic exchange entropy current defined in (3.18) can be rewritten as

$$j_{\delta_{\rm exch}S}(\mathcal{C}) \underset{MDB}{=} -\sum_{\mathcal{C}'}(\mathcal{C}'|\mathbb{W}|\mathcal{C}) \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')}. \tag{3.32}$$

Then the exchange entropy flow defined in (3.19) becomes equal to

$$\frac{d_{\rm exch}S}{dt} \underset{MDB}{=} -\sum_{\mathcal{C},\mathcal{C}'}(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t) \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')} \equiv \sigma_{\rm exch}[P(t)], \tag{3.33}$$

while the entropy production rate determined by (3.25) and (3.26) becomes equal to

$$\frac{d_{\rm int}S^{SG}}{dt} \underset{MDB}{=} \sum_{\mathcal{C},\mathcal{C}'}(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t) \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}';t)} \equiv \sigma_{\rm int}[P(t)] \tag{3.34}$$

The latter expression takes the form of a relative entropy so that $d_{\rm int}S^{SG}/dt$ is positive under MDB.

We notice that, in the generic case where the modified detailed balance does not necessarily holds, Lebowitz and Spohn [1] have introduced the splitting of dS^{SG}/dt into $\sigma_{\rm exch}+\sigma_{\rm int}$, where $\sigma_{\rm exch}[P(t)]$ is defined in (3.33) and $\sigma_{\rm int}[P(t)]$ is defined as the symmetrized expression of the definition in (3.34),

$$\sigma_{\rm int}[P(t)]=\frac{1}{2}\sum_{\mathcal{C},\mathcal{C}'}[(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t)-(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}';t)]\ln\frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}';t)}, \tag{3.35}$$

which is obviously positive. Another rewritting for the definition of $\sigma_{\rm int}[P(t)]$ given in (3.34) is

$$\sigma_{\rm int}[P(t)]=\langle j_{\delta_{\rm int}S}(t)\rangle_t \quad \text{with} \quad \delta_{\rm int}S(\mathcal{C}'\leftarrow\mathcal{C})\equiv\ln\frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}';t)}. \tag{3.36}$$

Then the positivity of $\sigma_{\rm int}[P(t)]$ may be viewed as arising from the property : for $x>0$, $-\ln x\geq 1-x$. We notice that $j_{\delta_{\rm int}S}(\mathcal{C};t)$ is not the current introduced in [1]. However, the current $j_{\delta_{\rm exch}S}(\mathcal{C})$ is introduced through its expression (3.32) in the latter reference and it has been used in the literature, possibly under another denomination (see for instance [66]).

3.3.4 Decomposition of the entropy production rate into two positive contributions

The microscopic variation $\delta_{\rm int}S(\mathcal{C}'\leftarrow\mathcal{C})$ defined in (3.36) can be decomposed into two contributions as

$$\delta_{\rm int}S(\mathcal{C}'\leftarrow\mathcal{C})=-\left[\ln\frac{P(\mathcal{C}';t)}{P_{\rm st}(\mathcal{C}')}-\ln\frac{P(\mathcal{C};t)}{P_{\rm st}(\mathcal{C})}\right]+\delta A_{[P_{\rm st}]}(\mathcal{C}'\leftarrow\mathcal{C}), \tag{3.37}$$

where

$$\delta A_{[P_{\rm st}]}(\mathcal{C}'\leftarrow\mathcal{C})\equiv\ln\frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P_{\rm st}(\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P_{\rm st}(\mathcal{C}')}. \tag{3.38}$$

By analogy with chemical reaction kinetics (see for instance Ref.[54] and subsubsection 3.3.5 below), $\delta A_{[P_{\rm st}]}(\mathcal{C}'\leftarrow\mathcal{C})$ may be viewed as the affinity of the elementary reversible reaction (or phase change) $\mathcal{C}\rightleftharpoons\mathcal{C}'$ in the stationary state resulting from all possible reversible pair reactions between all configurations. We notice that $\delta_{\rm int}S(\mathcal{C}'\leftarrow\mathcal{C})$ depends explicitly on time, contrarily to $\delta A_{[P_{\rm st}]}(\mathcal{C}'\leftarrow\mathcal{C})$. According to the definitions (3.10) and (3.11) of the currents associated respectively with an observable $\mathcal{O}$ or with some exchange quantity $\delta K(\mathcal{C}'\leftarrow\mathcal{C})$, the decomposition (3.37) allows to rewrite the expression (3.36) for $\sigma_{\rm int}[P(t)]$ as

$$\sigma_{\rm int}[P(t)]=-\langle j_{\ln[P(t)/P_{\rm st}]}\rangle_t+\langle j_{\delta A[P_{\rm st}]}\rangle_t. \tag{3.39}$$

We stress that, by virtue of the already used inequality $-\ln x\geq 1-x$ for $x>0$ and according to the stationary condition $\sum_{\mathcal{C}'\neq\mathcal{C}}(\mathcal{C}|\mathbb{W}|\mathcal{C}')P_{\rm st}(\mathcal{C}')-\sum_{\mathcal{C}'\neq\mathcal{C}}(\mathcal{C}'|\mathbb{W}|\mathcal{C})P_{\rm st}(\mathcal{C})=0$ (see (3.1)), the microscopic current $j_{\delta A[P_{\rm st}]}(\mathcal{C})$ that we have introduced is positive,

$$j_{\delta A[P_{\rm st}]}(\mathcal{C})\geq 0 \quad \text{for any } \mathcal{C}. \tag{3.40}$$

The latter positivity may be viewed as a consequence of the fact that $j_{\delta A[P_{\rm st}]}(\mathcal{C})$ is the relative entropy (for a given $\mathcal{C}$) of the rate $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$ with respect to the rate $(\mathcal{C}'|\mathbb{W}^\dagger|\mathcal{C})\equiv[P_{\rm st}(\mathcal{C})]^{-1}(\mathcal{C}|\mathbb{W}|\mathcal{C}')P_{\rm st}(\mathcal{C}')$.

According to the master equation (3.1) and the definition (3.5), the first contribution in the r.h.s. of (3.39) is the opposite of the time derivative of the relative entropy of the probability distribution $P(t)$ with respect to the stationary solution $P_{\rm st}$. Therefore the expression (3.36) of the entropy production rate $\sigma_{\rm int}[P(t)]$ can be split into two contributions, both of which are positive,

$$\sigma_{\rm int}[P(t)]=-\frac{dS_{\rm rel}[P(t)|P_{\rm st}]}{dt}+\langle j_{\delta A[P_{\rm st}]}\rangle_t. \tag{3.41}$$

The positivity of the first term has played a role in the discussion of the uniqueness of the stationary state (see Eq.(3.9)) and the positivity of the second term arises from the positivity of the current pointed out in (3.40).

For the sake of completeness, we notice that, when the transition rates depend on time the master equation (3.1) remains unchanged, so that the decomposition of the time-derivative of the Shannon-Gibbs entropy into $\sigma_{\rm exch} + \sigma_{\rm int}$ as well as the decomposition (3.39) of $\sigma_{\rm int}$ remain unchanged, except that $P_{\rm st}$ is to be replaced by a function of time $P^{(0)}(t)$, namely the probability distribution that is the zero right-eigenvector of the Markov matrix at a given time t. Then the splitting (3.39) of $\sigma_{\rm int}$ corresponds to that introduced in Ref.[43] for an evolution where both the bath temperature and the energy levels may change with time. In the splitting of the latter reference $P_{\rm st}$ is to be replaced by $P^{(0)}(t)$, and then the so-called non-adiabatic term corresponds to minus the average of the current associated with the observable $\ln[P(t)/P^{(0)}(t)]$ while the so-called adiabatic term corresponds to the average of the affinity current $\langle j_{\delta A[P^{(0)}(t)]}\rangle_t$.

3.3.5 Comparison with NESS characterization from graph theory

The r.h.s. of the master equation (3.1) can be rewritten as a sum of probability currents between configurations,

$$\frac{dP(\mathcal{C};t)}{dt} = \sum_{\mathcal{C}'} J_{[P]}(\mathcal{C},\mathcal{C}';t), \tag{3.42}$$

where

$$J_{[P]}(\mathcal{C}',\mathcal{C};t) \equiv (\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t) - (\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}';t). \tag{3.43}$$

The probability current $J_{[P]}(\mathcal{C}',\mathcal{C};t)$ has the form of the chemical reaction rate associated with the reversible reaction (or phase change) $\mathcal{C} \rightleftharpoons \mathcal{C}'$ with the species concentrations replaced by the configuration probabilities $P(\mathcal{C};t)$ and $P(\mathcal{C}';t)$ and the reaction rate constant for $\mathcal{C} \rightharpoonup \mathcal{C}'$ replaced by the transition rate $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$. The entropy production rate $\sigma_{\rm int}[P(t)]$, introduced in Ref.[1] in the generic case (namely when the MDB is not necessarily satisfied) by the definition (3.35), can be rewritten (with notations similar to those of Ref.[54]) as

$$\sigma_{\rm int}[P(t)] = \frac{1}{2}\sum_{\mathcal{C},\mathcal{C}'} \delta A_{[P]}(\mathcal{C}',\mathcal{C};t) J_{[P]}(\mathcal{C}',\mathcal{C};t), \tag{3.44}$$

where the dimensionless affinity $\delta A_{[P]}(\mathcal{C}',\mathcal{C};t)$ of an oriented pair (already introduced in the case of the stationary distribution $P_{\rm st}$ in (3.38)) is defined as

$$\delta A_{[P]}(\mathcal{C}',\mathcal{C};t) \equiv \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}';t)}. \tag{3.45}$$

Indeed, in the irreversible thermodynamics of chemical reactions at a temperature T fixed by a thermostat (see for instance [67]), the affinity of the reaction $\mathcal{C} \rightleftharpoons \mathcal{C}'$ is equal to $k_{\scriptscriptstyle B}T$ times an expression similar to $\delta A_{[P]}(\mathcal{C}',\mathcal{C};t)$, with the concentration of species $\mathcal{C}$ ($\mathcal{C}'$) in place of $P(\mathcal{C};t)$ ($P(\mathcal{C}';t)$) and the reaction rate constant for $\mathcal{C} \rightharpoonup \mathcal{C}'$ in place of the transition rate $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$.

The previous rewritings are convenient to handle the master equation in the framework of network theory where the master equation is represented by a graph G as follows. Each vertex of the graph corresponds to a given configuration $\mathcal{C}$ and there exists an edge between two vertices if at least one of the transition rates $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$ or $(\mathcal{C}|\mathbb{W}|\mathcal{C}')$ does not vanish. Moreover an arbitrary orientation is chosen for every edge in the graph G so that the transition rate $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$ (or $(\mathcal{C}|\mathbb{W}|\mathcal{C}')$) can be shortly referred to as the transition rate along the edge in either the positive sense or the negative sense.

From a connected graph G one can define several possible fundamental sets of N_c circuits (or closed paths) on the graph. The number N_c of circuits is only determined by the edge number N_e and the vertex number N_v through the relation $N_c = N_e - N_v + 1$. (The final results are independent of the specific fundamental set used in intermediate algebraic calculations.) A fundamental set is built from one among the various possible maximal (or spanning) trees which are defined by

removing N_c edges of G. For a given maximal tree $T(G)$ the corresponding removed edges are called chords and indexed by $\alpha = 1, \ldots, N_c$. For each α the circuit C_α is obtained by first considering the graph made by adding the chord α to $T(G)$ and then removing all edges which are not part of the circuit closed by the insertion of the chord α. A cycle $\vec{C}_\alpha$ is associated with every circuit C_α by choosing an arbitrary orientation to go around the circuit, so "cycle" is a synonymous for "oriented circuit".

The affinity $A(\vec{C}_\alpha)$ associated with a cycle $\vec{C}_\alpha$ is defined as an algebraic sum of all edge affinities which is calculated as follows: each edge affinity, whose expression is given by (3.45) for a positive orientation from $\mathcal{C}$ to $\mathcal{C}'$ in the graph G, is multiplied by the sign of the relative orientation of the edge in the cycle $\vec{C}_\alpha$ and in the graph G. Because of the cyclic structure of $\vec{C}_\alpha$ the affinity $A(\vec{C}_\alpha)$ does not depend explicitly on the configuration probability distribution P and it is determined only from the transition rates $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$. Indeed, if $N(\vec{C}_\alpha)$ is the number of configurations involved in the cycle $\vec{C}_\alpha$, and if the configurations are labeled with indices increasing with one unit when one goes from one configuration to the next one in the sense chosen for the orientation of the cycle $\vec{C}_\alpha$, then

$$A(\vec{C}_\alpha) = \ln \prod_{i=1}^{N(\vec{C}_\alpha)} \frac{(\mathcal{C}_{i+1}|\mathbb{W}|\mathcal{C}_i)}{(\mathcal{C}_i|\mathbb{W}|\mathcal{C}_{i+1})}, \tag{3.46}$$

with the notational convention $\mathcal{C}_{N(\vec{C}_\alpha)+1} \equiv \mathcal{C}_1$. The probability current $J_{[P]}(\vec{C}_\alpha; t)$ associated with the cycle is defined as the probability current in the chord α in the sense of the cycle orientation : it is given by definition (3.43) where $\mathcal{C}$ and $\mathcal{C}'$ are respectively the initial and final configurations in the sense of the cycle orientation.

As shown in Ref.[54], the stationary state has the following properties. First, the vanishing of all cycle affinities $A(\vec{C}_\alpha)$ is equivalent to the vanishing of all cycle probability currents $J_{[P_{\rm st}]}(\vec{C}_\alpha)$ in the stationary state,

$$\forall \alpha \quad A(\vec{C}_\alpha) = 0 \qquad \Leftrightarrow \qquad \forall \alpha \quad J_{[P_{\rm st}]}(\vec{C}_\alpha) = 0. \tag{3.47}$$

Second, the vanishing of all stationary cycle probability currents $J_{[P_{\rm st}]}(\vec{C}_\alpha)$ is equivalent to the fact that the stationary state obeys the detailed balance condition, namely

$$\forall \alpha \quad J_{[P_{\rm st}]}(\vec{C}_\alpha) = 0 \qquad \Leftrightarrow \qquad \forall (\mathcal{C}, \mathcal{C}') \quad \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C}) P_{\rm st}(\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}') P_{\rm st}(\mathcal{C}')} = 1. \tag{3.48}$$

Schnakenberg specifies that it is in fact a "complete detailed balance" condition in the sense that if there exist several kinds of independent transitions between two configurations $\mathcal{C}$ and $\mathcal{C}'$, the detailed balance must be satisfied by every kind of transition. Since equilibrium is characterized at the mesoscopic level by the detailed balance (see the derivation of (2.20)), the properties (3.47) and (3.48) entail that the equilibrium is characterized by either the vanishing of all stationary cycle currents or the vanishing of all cycle affinities. Moreover, the entropy production rate (3.44) in the stationary state reads

$$\sigma_{\rm int}[P_{\rm st}] = \sum_{\alpha=1}^{N_c} A(\vec{C}_\alpha)) J_{[P_{\rm st}]}(\vec{C}_\alpha), \tag{3.49}$$

where $A(\vec{C}_\alpha)$ is given directly in terms of the transition rates by the formula (3.46). By virtue of the definitions (3.33) and (3.34), $\sigma_{\rm exch}[P(t)] + \sigma_{\rm int}[P(t)] = dS^{SG}/dt$ and in the stationary state $\sigma_{\rm int}[P_{\rm st}] = -\sigma_{\rm exch}[P_{\rm st}]$.

When the MDB is satisfied, according to (3.33), $\sigma_{\rm exch}[P_{\rm st}] \underset{MDB}{=} d_{\rm exch}S/dt|_{\rm st}$ so that $\sigma_{\rm int}[P_{\rm st}] \underset{MDB}{=} -d_{\rm exch}S/dt|_{\rm st}$ and (3.49) becomes

$$\left.\frac{d_{\rm exch}S}{dt}\right|_{\rm st} \underset{MDB}{=} -\sum_{\alpha=1}^{N_c} A(\vec{C}_\alpha)) J_{[P_{\rm st}]}(\vec{C}_\alpha). \tag{3.50}$$

Meanwhile, by virtue of (1.3) the affinity of a cycle $A(\vec{C}_\alpha)$ given by (3.46) becomes equal to

$$A(\vec{C}_\alpha) \underset{MDB}{=} -\sum_{i=1}^{N(\vec{C}_\alpha)} \delta_{\rm exch} S(\mathcal{C}_{i+1} \leftarrow \mathcal{C}_i), \tag{3.51}$$

where the latter expression depends on the thermodynamic parameters of the reservoirs and on the quanta of microscopic quantities that the system exchanges with the reservoirs. In the generic case there may be several cycles corresponding to the same current between two reservoirs and there is no straightforward correspondence between the expression of $-d_{\rm exch}S/dt|_{\rm st}$ given by (3.50) in terms of the $A(\vec{C}_\alpha)$'s and $J_{[P_{\rm st}]}(\vec{C}_\alpha)$'s and the expression of $-d_{\rm exch}S/dt|_{\rm st}$ given by the entropy production rate for the phenomenological entropy S in thermodynamics of irreversible processes $-d_{\rm exch}S/dt|_{\rm st} = d_{\rm int}S/dt|_{\rm st} = \sum_\gamma \mathcal{F}_\gamma J^\star_\gamma$, where the $J^\star_\gamma$'s are independent macroscopic currents.

However in the case of the simple solvable model considered in paper II, where the thermal contact between two thermostats is settled by a set of independent two-spin systems, where each spin σ_a is flipped by thermostat a ($a = 1, 2$), only one cycle is involved for every spin pair; then one can make the correspondence between, on the one hand, the affinity $A(\vec{C})$ of the cycle and the stationary probability current $J_{[P_{\rm st}]}(\vec{C})$ that goes through it, and, on the other hand, the thermodynamic force $\mathcal{F}$ and the mean heat current J. Indeed, in this model, the graph associated with the master equation is itself a cycle, which can be orientated to read

$$\begin{array}{ccc} (+,+) & \rightarrow & (-,+) \\ \uparrow & & \downarrow \\ (+,-) & \leftarrow & (-,-) \end{array} . \tag{3.52}$$

The cycle may be rewritten as $\mathcal{C}_1 \rightarrow \mathcal{C}_2 \rightarrow \mathcal{C}_3 \rightarrow \mathcal{C}_4 \rightarrow \mathcal{C}_1$ where the configurations are labeled in the positive sense of the cycle orientation. Then, since the model obeys the MDB, the affinity of the cycle is given by (3.51), and by virtue of the definitions (2.23) and (2.25) it reads $A(\vec{C}) = -\beta_1 q_1(\vec{C}) - \beta_2 q_2(\vec{C})$, where $q_a(\vec{C})$ is the heat received from the thermostat a when the system configuration performs the cycle once in the positive sense. Moreover, according to the energy conservation law, $q_1(\vec{C}) + q_2(\vec{C})$ is equal to the energy difference between the final and initial states when the cycle is performed once : this difference vanishes for a cycle so that

$$A(\vec{C}) = (\beta_1 - \beta_2) q_2(\vec{C}). \tag{3.53}$$

On the other hand, since the graph is exactly a cycle, the current along an edge defined in (3.43) has the same value for all edges in the stationary state (because, by virtue of (3.42), the stationary condition $dP(\mathcal{C};t)/dt|_{\rm st} = 0$ is equivalent to Kirchoff's current law at every vertex of the graph). As a consequence, the stationary current associated with the cycle reads

$$J_{[P_{\rm st}]}(\vec{C}) = (\mathcal{C}_{i+1}|\mathbb{W}|\mathcal{C}_i) P_{\rm st}(\mathcal{C}_i) - (\mathcal{C}_i|\mathbb{W}|\mathcal{C}_{i+1}) P_{\rm st}(\mathcal{C}_{i+1}), \tag{3.54}$$

where i is any label in $\{1, 2, 3, 4\}$. The exchange entropy flow given by (3.50) reads $d_{\rm exch}S/dt|_{\rm st} = -(\beta_1 - \beta_2) q_2(\vec{C}) J_{[P_{\rm st}]}(\vec{C})$. Comparison with the expression $d_{\rm exch}S/dt = -\mathcal{F}J$ in irreversible processes thermodynamics (see (3.29)-(3.30)-(3.31)) leads to the following identification of the macroscopic heat current $J = \langle j_2 \rangle_{\rm st}$ received from heat bath 2 (and given to heat bath 1),

$$J = q_2(\vec{C}) J_{[P_{\rm st}]}(\vec{C}), \tag{3.55}$$

while the thermodynamic force $\mathcal{F} = \beta_1 - \beta_2$ is to be identified with

$$\mathcal{F} = \frac{A(\vec{C})}{q_2(\vec{C})}. \tag{3.56}$$

The probabilistic interpretation of the affinity $A(\vec{C})$ is given in paper II.

4 Exchange entropy variation and symmetries at finite time under MDB

4.1 Exchange entropy variation for a history

For a history $\mathcal{H}ist$ where the system starts in configuration $\mathcal{C}_0$ at time $t_0 = 0$ and ends in configuration $\mathcal{C}_f$ at time t after going through successive configurations $\mathcal{C}_0$, $\mathcal{C}_1$,..., $\mathcal{C}_N = \mathcal{C}_f$, the exchange entropy variation $\Delta_{\rm exch}S[\mathcal{H}ist]$ corresponding to the history is defined from the heat amounts $\mathcal{Q}_1[\mathcal{H}ist]$ and $\mathcal{Q}_2[\mathcal{H}ist]$ received from two thermal baths as

$$\Delta_{\rm exch}S[\mathcal{H}ist] \equiv \beta_1 \mathcal{Q}_1[\mathcal{H}ist] + \beta_2 \mathcal{Q}_2[\mathcal{H}ist] \quad \text{with} \quad \mathcal{Q}_a[\mathcal{H}ist] \equiv \sum_{i=0}^{N-1} \delta q_a(\mathcal{C}_{i+1} \leftarrow \mathcal{C}_i). \tag{4.1}$$

The expectation value of $\Delta_{\rm exch}S[\mathcal{H}ist]$ with respect to the measure over all possible histories starting from configurations distributed according to some initial probability distribution (see appendix D) is equal to the time integral of the mean exchange entropy current calculated with the instantaneous configuration probability distribution,

$$\langle \Delta_{\rm exch}S \rangle = \int_0^t dt' \langle j_{\delta_{\rm exch}S} \rangle_{t'} = \int_0^t dt' \frac{d_{\rm exch}S}{dt'}. \tag{4.2}$$

The second equality arises from (3.20).

4.2 MDB and symmetry between time-reversed histories

Let $\mathbb{T}$ be the time reversal operator for histories. If $\mathcal{H}ist$ is a history that starts at time $t_0 = 0$ in $\mathcal{C}_0$ and ends at time t in $\mathcal{C}_f$ after N jumps from $\mathcal{C}_{i-1}$ to $\mathcal{C}_i$ at time T_i, $\mathbb{T}\mathcal{H}ist$ is a history that starts at time $t_0 = 0$ in $\mathcal{C}_f$ and ends at $\mathcal{C}_0$ at time t after N jumps from $\mathcal{C}'_{i-1}$ to $\mathcal{C}'_i$ at time T'_i with $\mathcal{C}'_i = \mathcal{C}_{N-i}$ and $T'_i = t - T_{N-i+1}$, namely

$$\begin{aligned} \mathcal{H}ist: \qquad & \mathcal{C}_0 \text{ at } t_0 = 0 \qquad \mathcal{C}_0 \xrightarrow{T_1} \mathcal{C}_1 \cdots \mathcal{C}_{N-1} \xrightarrow{T_N} \mathcal{C}_f \\ \mathbb{T}\mathcal{H}ist: \qquad & \mathcal{C}_f \text{ at } t_0 = 0 \qquad \mathcal{C}_f \xrightarrow{T'_1} \mathcal{C}_{N-1} \cdots \mathcal{C}_1 \xrightarrow{T'_N} \mathcal{C}_0. \end{aligned} \tag{4.3}$$

From the definition of the measure $dP_{\mathcal{C}_0,\mathcal{C}_f}$ over histories starting in configuration $\mathcal{C}_0$ and ending in configuration $\mathcal{C}_f$ (see Appendix D),

$$\frac{dP_{\mathcal{C}_f,\mathcal{C}_0}[\mathcal{H}ist]}{dP_{\mathcal{C}_0,\mathcal{C}_f}[\mathbb{T}\mathcal{H}ist]} = \prod_{i=0}^{N-1} \frac{(\mathcal{C}_{i+1}|\mathbb{W}|\mathcal{C}_i)}{(\mathcal{C}_i|\mathbb{W}|\mathcal{C}_{i+1})}. \tag{4.4}$$

When the transition rates obey the modified detailed balance (2.26) written in terms of $\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C})$, the exchange entropy variation for the history, defined in (4.1), can be rewritten as

$$\Delta_{\rm exch}S[\mathcal{H}ist] \underset{MDB}{=} -\ln \prod_{i=0}^{N-1} \frac{(\mathcal{C}_{i+1}|\mathbb{W}|\mathcal{C}_i)}{(\mathcal{C}_i|\mathbb{W}|\mathcal{C}_{i+1})}, \tag{4.5}$$

and equation (4.4) can be rewritten as

$$\frac{dP_{\mathcal{C}_f,\mathcal{C}_0}[\mathcal{H}ist]}{dP_{\mathcal{C}_0,\mathcal{C}_f}[\mathbb{T}\mathcal{H}ist]} \underset{MDB}{=} e^{-\Delta_{\rm exch}S[\mathcal{H}ist]}. \tag{4.6}$$

We stress that, according to (4.5), when the MDB is satisfied the expression of the exchange entropy variation for a history defined in (4.1) coincides with the opposite of the "action functional" introduced by Lebowitz and Spohn in Ref.[1] in the generic case where the MDB does not necessarily hold.

4.3 Symmetry between time-reversed evolutions with fixed heat amounts

The probability $P\left(\mathcal{C}_f|\mathcal{Q}_1,\mathcal{Q}_2,t|\mathcal{C}_0\right)$ that the system has evolved from configuration $\mathcal{C}_0$ at $t_0=0$ to configuration $\mathcal{C}_f$ at t while receiving the heat amounts $\mathcal{Q}_1$ and $\mathcal{Q}_2$ from the thermostats 1 and 2 reads

$$P\left(\mathcal{C}_f|\mathcal{Q}_1,\mathcal{Q}_2;t|\mathcal{C}_0\right)\equiv\int dP_{\mathcal{C}_f,\mathcal{C}_0}\left[\mathcal{H}ist\right]\delta\left(\mathcal{Q}_1[\mathcal{H}ist]-\mathcal{Q}_1\right)\delta\left(\mathcal{Q}_2[\mathcal{H}ist]-\mathcal{Q}_2\right), \tag{4.7}$$

where $\int dP_{\mathcal{C}_f,\mathcal{C}_0}$ denotes the "summation" over the histories from $\mathcal{C}_0$ to $\mathcal{C}_f$. The time-reversal symmetry property (4.6) for the history measure $dP_{\mathcal{C}_f,\mathcal{C}_0}\left[\mathcal{H}ist\right]$ implies the following relation between probabilities of forward and backward evolutions where initial and final configurations are exchanged (and heat amounts are changed into their opposite values),

$$\frac{P\left(\mathcal{C}_f|\mathcal{Q}_1,\mathcal{Q}_2;t|\mathcal{C}_0\right)}{P\left(\mathcal{C}_0|-\mathcal{Q}_1,-\mathcal{Q}_2;t|\mathcal{C}_f\right)}=e^{-\Delta_{\rm exch}S(\mathcal{Q}_1,\mathcal{Q}_2)}, \tag{4.8}$$

with the definition

$$\Delta_{\rm exch}S(\mathcal{Q}_1,\mathcal{Q}_2)\equiv\beta_1\mathcal{Q}_1+\beta_2\mathcal{Q}_2. \tag{4.9}$$

An analogous relation for $\Delta_{\rm exch}S$ in place of $(\mathcal{Q}_1,\mathcal{Q}_2)$ is derived in [68] in the case where the microscopic dynamics of the heat baths is assumed to be Hamiltonian.

4.4 Symmetries in protocols starting from an equilibrium state

We consider a protocol where the system is prepared in an equilibrium state at the inverse temperature β_0 and suddenly put at time $t_0=0$ in thermal contact with the two thermostats at the inverse temperatures β_1 and β_2 respectively. Then the system evolution is a relaxation from an equilibrium state to a stationary non-equilibrium state.

The initial equilibrium distribution at the inverse temperature β_0 is the canonical distribution (2.19). $Z(\beta_0)$ cancels in the ratio $P^{\beta_0}_{\rm can}(\mathcal{C}_0)/P^{\beta_0}_{\rm can}(\mathcal{C}_f)$ and

$$\ln\frac{P^{\beta_0}_{\rm can}(\mathcal{C}_0)}{P^{\beta_0}_{\rm can}(\mathcal{C}_f)}=\beta_0\left[\mathcal{E}(\mathcal{C}_f)-\mathcal{E}(\mathcal{C}_0)\right]=\beta_0(\mathcal{Q}_1+\mathcal{Q}_2), \tag{4.10}$$

where the last equality is enforced by energy conservation. Then the time-reversal symmetry (4.8) and the specific form (4.10) for $\ln[P^{\beta_0}_{\rm can}(\mathcal{C}_0)/P^{\beta_0}_{\rm can}(\mathcal{C}_f)]$ imply that

$$\frac{P\left(\mathcal{C}_f|\mathcal{Q}_1,\mathcal{Q}_2;t|\mathcal{C}_0\right)P^{\beta_0}_{\rm can}(\mathcal{C}_0)}{P\left(\mathcal{C}_0|-\mathcal{Q}_1,-\mathcal{Q}_2;t|\mathcal{C}_f\right)P^{\beta_0}_{\rm can}(\mathcal{C}_f)}=e^{-\Delta^{\rm excs,\beta_0}_{\rm exch}S(\mathcal{Q}_1,\mathcal{Q}_2)}, \tag{4.11}$$

where the excess exchange entropy variation $\Delta^{\rm excs,\beta_0}_{\rm exch}S(\mathcal{Q}_1,\mathcal{Q}_2)$ is defined as the difference between the exchange entropy variation in an evolution under the non-equilibrium constraint $\beta_1\neq\beta_2$ where the system receives heat amounts $\mathcal{Q}_1$ and $\mathcal{Q}_2$ and that in an evolution under the equilibrium condition $\beta_1=\beta_2=\beta_0$ where the system would received the same heat amounts. It reads

$$\Delta^{\rm excs,\beta_0}_{\rm exch}S(\mathcal{Q}_1,\mathcal{Q}_2)=\Delta_{\rm exch}S(\mathcal{Q}_1,\mathcal{Q}_2)-\beta_0(\mathcal{Q}_1+\mathcal{Q}_2)=(\beta_1-\beta_0)\mathcal{Q}_1+(\beta_2-\beta_0)\mathcal{Q}_2. \tag{4.12}$$

A crucial point is that $\Delta^{\rm excs,\beta_0}_{\rm exch}S(\mathcal{Q}_1,\mathcal{Q}_2)$ does not depend explicitly on the initial and final configurations and is only a function of the heat amounts received from the thermal baths,

As a consequence, the measurable joint distribution $P_{P^{\beta_0}_{\rm can}}\left(\mathcal{Q}_1,\mathcal{Q}_2;t\right)$ for the heat amounts $\mathcal{Q}_1$ and $\mathcal{Q}_2$ received between $t_0=0$ and t when the initial configuration of the system is distributed according to $P^{\beta_0}_{\rm can}$, namely $P_{P^{\beta_0}_{\rm can}}\left(\mathcal{Q}_1,\mathcal{Q}_2;t\right)=\sum_{\mathcal{C}_0,\mathcal{C}_f}P\left(\mathcal{C}_f|\mathcal{Q}_1,\mathcal{Q}_2,t|\mathcal{C}_0\right)P_{P^{\beta_0}_{\rm can}}(\mathcal{C}_0)$, satisfies the identity

$$\frac{P_{P^{\beta_0}_{\rm can}}\left(\mathcal{Q}_1,\mathcal{Q}_2;t\right)}{P_{P^{\beta_0}_{\rm can}}\left(-\mathcal{Q}_1,-\mathcal{Q}_2;t\right)}=e^{-\Delta^{\rm excs,\beta_0}_{\rm exch}S(\mathcal{Q}_1,\mathcal{Q}_2)}. \tag{4.13}$$

Subsequently the measurable quantity $\Delta_{\rm exch}^{{\rm excs},\beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2)$, with the distribution probability $P_{P_{\rm can}^{\beta_0}}\left(\Delta_{\rm exch}^{{\rm excs},\beta_0} S\right) = \sum_{\mathcal{Q}_1,\mathcal{Q}_2} \delta\left(\Delta_{\rm exch}^{{\rm excs},\beta_0} S - (\beta_1-\beta_0)\mathcal{Q}_1 - (\beta_2-\beta_0)\mathcal{Q}_2\right) P_{P_{\rm can}^{\beta_0}}(\mathcal{Q}_1, \mathcal{Q}_2; t)$ obeys the symmetry relation at any finite time, which may be referred to as a detailed fluctuation relation,

$$\frac{P_{P_{\rm can}^{\beta_0}}\left(\Delta_{\rm exch}^{{\rm excs},\beta_0} S\right)}{P_{P_{\rm can}^{\beta_0}}\left(-\Delta_{\rm exch}^{{\rm excs},\beta_0} S\right)} = e^{-\Delta_{\rm exch}^{{\rm excs},\beta_0} S}. \tag{4.14}$$

The latter relation itself entails the identity, which may be referred to as an integral fluctuation relation,

$$\langle e^{\Delta_{\rm exch}^{{\rm excs},\beta_0} S}\rangle_{P_{\rm can}^{\beta_0}} = 1. \tag{4.15}$$

To our knowledge these two relations have not appeared explicitly in the literature. We notice that, from a purely technical point of view, the derivation has similarities with the argument first exhibited by Crooks [40] and then Seifert [42, 9] for the entropy production along a stochastic trajectory when the system is in thermal contact with only one heat bath and is driven out of equilibrium by a time-dependent external parameter. (In Crooks' argument the initial configurations for the forward and backward evolutions are distributed with different equilibrium probabilities, $P_{\rm can}^{\beta_0}$ and $P_{\rm can}^{\beta_f}$, whereas forward and backward evolutions with the same initial distribution had already been considered in [69, 70]).

4.5 Symmetries in protocols starting from a stationary state with a canonical distribution

For some systems, such as the two-spin model studied in paper II, the stationary distribution when the thermostats are at the inverse temperatures β_1 and β_2 proves to be a canonical distribution at the effective inverse temperature $\beta_\star(\beta_1, \beta_2)$.

When the system is prepared in a stationary state between two heat baths at the inverse temperatures β_1^0 and β_2^0 and then put in thermal contact with two thermostats at the inverse temperatures β_1 and β_2 at time $t_0 = 0$, the protocol describes the relaxation from a given stationary state corresponding to (β_1^0, β_2^0) to another stationary state corresponding to (β_1, β_2). When the initial stationary state has the canonical distribution at the effective inverse temperature $\beta_\star^0 = \beta_\star(\beta_1^0, \beta_2^0)$, the argument of the previous subsection can be repeated and the equalities (4.14) and (4.15) still hold with β_0 replaced by $\beta_\star^0$ and $\Delta_{\rm exch}^{{\rm excs},\beta_0} S$ replaced by

$$\Delta_{\rm exch}^{{\rm excs},\beta_\star^0} S(\mathcal{Q}_1, \mathcal{Q}_2) = (\beta_1 - \beta_\star^0)\mathcal{Q}_1 + (\beta_2 - \beta_\star^0)\mathcal{Q}_2. \tag{4.16}$$

When the system is already in the stationary state corresponding to the inverse temperatures β_1 and β_2 at time $t_0 = 0$, the equalities (4.14) and (4.15) for $\Delta_{\rm exch}^{{\rm excs},\beta_0} S$ still hold with $\beta_\star(\beta_1, \beta_2)$ in place of β_0 :

$$\frac{P_{\rm st}\left(\Delta_{\rm exch}^{{\rm excs},\beta_\star} S\right)}{P_{\rm st}\left(-\Delta_{\rm exch}^{{\rm excs},\beta_\star} S\right)} = e^{-\Delta_{\rm exch}^{{\rm excs},\beta_\star} S}, \tag{4.17}$$

where the subscript "st" in the notation for the probability is a reminder of the fact that the initial configurations are distributed according to the stationary measure, which is equal to $P_{\rm can}^{\beta_\star}$ in the present case. Another detailed fluctuation relation involving the forward histories for the original dynamics and the backward histories for the dual reversed dynamics is derived in [71] for the case where the external parameters also vary during the time interval $]t_0, t]$; these considerations are out of the scope of the present paper.

5 Long-time symmetries : non-equilibrium stationary state with MDB

5.1 Fluctuation relations

5.1.1 Fluctuation relation for the cumulative exchange entropy variation

As recalled with some details in section 3, if the Markov matrix is irreducible and the system has a finite number of configurations, then the Perron-Frobenius theorem entails that there exists a single stationary state $P_{\rm st}$ and every configuration $\mathcal{C}$ has a non-zero probability $P_{\rm st}(\mathcal{C})$. Let us call $P_{\rm st}^{\rm min}$ and $P_{\rm st}^{\rm max}$ the minimum and maximum values taken by $P_{\rm st}$. Since $P_{\rm st}\left(\mathcal{Q}_1,\mathcal{Q}_2;t\right) = \sum_{\mathcal{C}_0,\mathcal{C}_f} P\left(\mathcal{C}_f|\mathcal{Q}_1,\mathcal{Q}_2;t|\mathcal{C}_0\right) P_{\rm st}(\mathcal{C}_0)$, and $\frac{P_{\rm st}^{\rm min}}{P_{\rm st}^{\rm max}} \leq \frac{P_{\rm st}(\mathcal{C}_f)}{P_{\rm st}(\mathcal{C}_0)} \leq \frac{P_{\rm st}^{\rm max}}{P_{\rm st}^{\rm min}}$, the time-reversal symmetry (4.8) for $P\left(\mathcal{C}_f|\mathcal{Q}_1,\mathcal{Q}_2;t|\mathcal{C}_0\right)$ entails that

$$\frac{P_{\rm st}^{\rm min}}{P_{\rm st}^{\rm max}} \leq \frac{P_{\rm st}\left(\mathcal{Q}_1,\mathcal{Q}_2;t\right)}{P_{\rm st}\left(-\mathcal{Q}_1,-\mathcal{Q}_2;t\right) e^{-\Delta_{\rm exch}S(\mathcal{Q}_1,\mathcal{Q}_2)}} \leq \frac{P_{\rm st}^{\rm max}}{P_{\rm st}^{\rm min}}. \tag{5.1}$$

The distribution probability for the exchange entropy variation $\Delta_{\rm exch}S$ can be determined from measurements of heat amounts through the relation

$$P_{\rm st}\left(\Delta_{\rm exch}S;t\right) = \sum_{\mathcal{Q}_1,\mathcal{Q}_2} \delta\left(\Delta_{\rm exch}S - \beta_1\mathcal{Q}_1 - \beta_2\mathcal{Q}_2\right) P_{\rm st}\left(\mathcal{Q}_1,\mathcal{Q}_2;t\right). \tag{5.2}$$

The inequalities (5.1) imply that

$$\frac{P_{\rm st}^{\rm min}}{P_{\rm st}^{\rm max}} \leq \frac{P_{\rm st}\left(\Delta_{\rm exch}S;t\right)}{P_{\rm st}\left(-\Delta_{\rm exch}S;t\right) e^{-\Delta_{\rm exch}S}} \leq \frac{P_{\rm st}^{\rm max}}{P_{\rm st}^{\rm min}}. \tag{5.3}$$

In the long-time limit the system reaches its non-equilibrium stationary state exponentially fast, so the existence and value of a large deviation function for the cumulative current $\mathcal{J} \equiv \Delta_{\rm exch}S/t$ are not expected to depend on the initial distribution. Appendix E contains several definitions of large deviation functions for a cumulative random variable X_t with the sign convention used in mathematical literature (see subsections E.1 and E.2). In the present paper we use the opposite sign convention and we consider the large deviation function $f_X(\mathcal{J}) \equiv -R_X(\mathcal{J})$ where R_X denotes the rate function introduced by the proper mathematical definition (E.3)-(E.4) for the values $\mathcal{J}$ taken by X_t/t. When $f_X(\mathcal{J})$ is continuous over some interval I, the property (E.8) reads

$$\lim_{t\to+\infty} \frac{1}{t} \ln P\left(\frac{X_t}{t} \in I\right) = \sup_{\mathcal{J}\in I} f_X(\mathcal{J}). \tag{5.4}$$

The presence of the upper bound reflects the fact that, if I is split into small intervals I_k, only the interval where $P\left(\frac{X_t}{t} \in I_k\right)$ is maximum contributes to the limit in the left-hand side of (5.4), and this limit can indeed be rewritten as the upper bound over I of a function $f_X(\mathcal{J})$ that depends only on a single variable (see the heuristic argument in subsection E.1). The supremum in (5.4) is also crucial in the derivation of the Gärtner-Ellis theorem which, in its simplified version, allows to compute $f_X(\mathcal{J})$ from the generating function for infinite-time cumulants of X_t per unit time, namely $\lim_{t\to+\infty} \frac{1}{t} \ln\langle e^{\lambda X_t}\rangle$, when the latter generating function exists and is differentiable for all λ in $\mathbb{R}$ (see subsection 5.1.3 below). A consequence of (5.4) is that

$$\lim_{\epsilon\to 0} \lim_{t\to+\infty} \frac{1}{t} \ln P\left(\frac{X_t}{t} \in [\mathcal{J}-\epsilon, \mathcal{J}+\epsilon]\right) = f_X(\mathcal{J}). \tag{5.5}$$

Eventually we recall an expression often encountered in the literature. In the case where $f_X(\mathcal{J})$ is strictly convex downward, with a maximum value at $J = \lim_{t\to+\infty} \frac{X_t}{t}$ where $f_X(\mathcal{J})$ vanishes, then

$f_X(\mathcal{J})$ can be expressed in terms of the cumulative distribution function and the complementary cumulative distribution function as

$$f_X(\mathcal{J}) \equiv \begin{cases} \lim_{t\to\infty} \frac{1}{t} \ln P(\frac{X_t}{t} > \mathcal{J}; t) & \text{for } \mathcal{J} > J \\ \lim_{t\to\infty} \frac{1}{t} \ln P(\frac{X_t}{t} < \mathcal{J}; t) & \text{for } \mathcal{J} < J \end{cases}. \tag{5.6}$$

We notice that the latter expression is not to be generalized to the case of several independent currents, contrarily to the definition (5.4).

Subsection E.3 contains a proof that, whatever the definition among the three ones discussed in Appendix E, a relation like (5.3) implies that, if $\Delta_{\rm exch} S$ has a large deviation function $f_{\Delta_{\rm exch} S}(\mathcal{J})$ – under $P_{\rm st}$ but then also under any initial probability distribution – then

$$f_{\Delta_{\rm exch} S}(\mathcal{J}) - f_{\Delta_{\rm exch} S}(-\mathcal{J}) = -\mathcal{J}. \tag{5.7}$$

Our derivation of the fluctuation relation is close to the argument given in Ref.[45] for the large deviation of the cumulative heat current (see (5.12)). It relies on the MDB obeyed by the transition rates. Since the opposite of the exchange entropy variation is the specific form that the Lebowitz-Spohn action functional [1] takes in the presence of MDB, the fluctuation relation (5.7) is in fact a special case of the fluctuation relation satisfied by the action functional for a system with a finite number of configurations under the assumption (1.2) of reversibility for configuration jumps and the assumption (1.1) that the Markov matrix is irreducible, without the extra assumption of MDB (1.3).

5.1.2 Constraints from the bound upon the system energy

In a system with a finite number of configurations, $\mathcal{Q}_1 + \mathcal{Q}_2 = \mathcal{E}(\mathcal{C}_f) - \mathcal{E}(\mathcal{C}_0)$ is bounded and this entails several properties upon the large deviation functions of the cumulative currents $\mathcal{J}_1 = \mathcal{Q}_1/t$ and $\mathcal{J}_2 = \mathcal{Q}_2/t$.

When $\mathcal{Q}_1 + \mathcal{Q}_2$ is bounded, the consequences for the cumulative heat currents are conveniently investigated if one considers the couple of variables $\left(\mathcal{Q}_1^{\rm d}, \mathcal{Q}_2\right)$ where $\mathcal{Q}_1^{\rm d} = -\mathcal{Q}_1$ is the heat amount dissipated towards thermal bath 1. The fact that the difference $\mathcal{Q}_1^{\rm d} - \mathcal{Q}_2$ is bounded means that there exists some (time-independent) cosntant $M > 0$ such that

$$|\mathcal{Q}_1^{\rm d} - \mathcal{Q}_2| < M. \tag{5.8}$$

The first straightforward consequence upon the cumulative heat currents $\mathcal{J}_1^d = \mathcal{Q}_1^{\rm d}/t$ and $\mathcal{J}_2 = \mathcal{Q}_2/t$ is that $\lim_{t\to+\infty} \langle \mathcal{J}_1^d \rangle = \lim_{t\to+\infty} \langle \mathcal{J}_2 \rangle \equiv J$, namely

$$\lim_{t\to+\infty} \frac{-\langle \mathcal{Q}_1 \rangle}{t} = \lim_{t\to+\infty} \frac{\langle \mathcal{Q}_2 \rangle}{t} = J. \tag{5.9}$$

The second consequence of the fact that $\mathcal{Q}_1^{\rm d} - \mathcal{Q}_2$ is sub-extensive is that, according to (E.32), $\mathcal{Q}_1^{\rm d}$ and $\mathcal{Q}_2$ have the same large deviation function, $f_{\mathcal{Q}_1^{\rm d}}(\mathcal{J}) = f_{\mathcal{Q}_2}(\mathcal{J})$, namely

$$f_{\mathcal{Q}_1}(\mathcal{J}) = f_{\mathcal{Q}_2}(-\mathcal{J}). \tag{5.10}$$

Similarly the difference between $\Delta_{\rm exch} S$ and $-(\beta_1 - \beta_2)\mathcal{Q}_2$, which is equal to $\beta_1[\mathcal{E}(\mathcal{C}_f) - \mathcal{E}(\mathcal{C}_0)]$, is bounded. Therefore the difference is sub-extensive, so that according to (E.32) $f_{\Delta_{\rm exch} S}(\mathcal{J}) = f_{-(\beta_1-\beta_2)\mathcal{Q}_2}(\mathcal{J})$, namely

$$f_{\Delta_{\rm exch} S}(\mathcal{J}) = f_{\mathcal{Q}_2}\left(-\frac{\mathcal{J}}{\beta_1 - \beta_2}\right). \tag{5.11}$$

As a consequence, the fact that the exchange entropy variation $\Delta_{\rm exch} S$ obeys the fluctuation relation (5.7) is equivalent to the fact that the cumulative heat current received from heat bath 2 obeys the fluctuation relation

$$f_{\mathcal{Q}_2}(\mathcal{J}) - f_{\mathcal{Q}_2}(-\mathcal{J}) = (\beta_1 - \beta_2)\,\mathcal{J}. \tag{5.12}$$

The latter equation is the long-time limit of the relation first exhibited by Jarzynski and Wojcik [46] for the thermal contact between two bodies initially prepared at different inverse temperatures β_1 and β_2 and whose microscopic Hamiltonian dynamics involves an interaction turned on at time $t_0 = 0$ and turned off at time t and which is assumed to be negligible with respect to the heat quantity that goes from one body to the other during time t. The fluctuation relation (5.12) is also derived in the framework of the master equation approach in Refs.[44, 45].

We notice that, since the function $\Delta_{\rm exch}^{{\rm excs},\beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2)$ defined in (4.12) is equal to $\Delta_{\rm exch} S$ plus a term $-\beta_0(\mathcal{Q}_1 + \mathcal{Q}_2)$ which is bounded by virtue of energy conservation, the large deviation functions for $\Delta_{\rm exch}^{{\rm excs},\beta_0} S$ and $\Delta_{\rm exch} S$ coincide (see (E.32)). As a consequence, (5.7) entails that $\Delta_{\rm exch}^{{\rm excs},\beta_0}(\mathcal{Q}_1, \mathcal{Q}_2)$ obeys the fluctuation relation

$$f_{\Delta_{\rm exch}^{{\rm excs},\beta_0}}(\mathcal{J}) - f_{\Delta_{\rm exch}^{{\rm excs},\beta_0}}(-\mathcal{J}) = -\mathcal{J}. \tag{5.13}$$

The relation (5.12) can be written in a more generic form by the following argument. The fact that the difference between $\Delta_{\rm exch} S$ and $-(\beta_1 - \beta_2)\mathcal{Q}_2$ is bounded yields the following relation between the infinite-time mean values of the corresponding cumulative currents,

$$\lim_{t\to+\infty} \frac{\langle \Delta_{\rm exch} S(t)\rangle}{t} = -(\beta_1 - \beta_2) \lim_{t\to+\infty} \frac{\langle \mathcal{Q}_2(t)\rangle}{t}. \tag{5.14}$$

Since the infinite-time limit of the mean value of a cumulative current $\mathcal{J}_t$ measured during the interval $[0, t]$ is equal to the mean value of the corresponding instantaneous current in the stationary state, namely

$$\lim_{t\to\infty} \langle \mathcal{J}_t\rangle = \langle j\rangle_{\rm st} \equiv J, \tag{5.15}$$

we get $\lim_{t\to+\infty}\langle \Delta_{\rm exch} S\rangle/t = \langle j_{\delta_{\rm exch} S}\rangle_{\rm st} = d_{\rm exch} S/dt|_{\rm st}$ as well as $\lim_{t\to+\infty}\langle \mathcal{Q}_2\rangle/t = \langle j_2\rangle_{\rm st}$ (with the instantaneous current definitions (3.18) and (3.17) respectively). With these identifications the comparison of (5.14) with the property (1.4) (which exhibits the analogy with the thermodynamics of irreversible processes) shows that $\beta_1 - \beta_2$ in (5.14) is to be interpreted as the thermodynamic force $\mathcal{F}$. Therefore the fact that the fluctuation relation (5.12) arises from the boundedness of the difference between $\Delta_{\rm exch} S$ and $-(\beta_1 - \beta_2)\mathcal{Q}_2$, as the relation between the mean values (5.14) does, implies that the fluctuation relation (5.12) for $f_{\mathcal{Q}_2}$ is a special case of the more generic fluctuation relation

$$f(\mathcal{J}; \mathcal{F}) - f(-\mathcal{J}; \mathcal{F}) = \mathcal{F}\mathcal{J}, \tag{5.16}$$

where $\mathcal{F}$ is the thermodynamic force that appears in the stationary exchange entropy flow $d_{\rm exch} S/dt|_{\rm st} = -\mathcal{F}J$. (We recall that in the case where $f(\mathcal{J}; \mathcal{F}) = -\infty$, the relation (5.16) makes sense when written as $f(\mathcal{J}; \mathcal{F}) = f(-\mathcal{J}; \mathcal{F}) - \mathcal{F}\mathcal{J}$.)

5.1.3 Gärtner-Ellis theorem and some of its consequences

As exemplified in the previous subsubsection, large deviation functions may depend on auxiliary parameters. For certain questions, they are simply spectators. This is the case in the following discussion, so we do not mention possible auxiliary parameters explicitly. However, we nevertheless write all derivatives as partial derivatives.

In this short subsubsection, we introduce an important tool to study the existence and properties of a large deviation function $f(\mathcal{J})$: the infinite-time limit of the generating function for the cumulants of $X_t = t\mathcal{J}_t$ per unit time, namely $\alpha(\lambda) \equiv \lim_{t\to+\infty}(1/t)\ln\langle e^{\lambda X_t}\rangle$. $\alpha(\lambda)$ is known as the scaled cumulant generating function in the literature about large deviations. This function, if it exists, is automatically convex (downward).

According to a simplified version of the Gärtner-Ellis theorem (see e.g. the review for physicists [72] or the mathematical point of view [73]), if $\alpha(\lambda)$ exists and is differentiable for all λ in $\mathbb{R}$, then the large deviation function f of the current $\mathcal{J}$ exists and it can be calculated as the Legendre-Fenchel transform of $\alpha(\lambda)$, namely, with the signs chosen in the definitions used in the present paper,

$$f(\mathcal{J}) = \min_{\lambda\in\mathbb{R}}\{\alpha(\lambda) - \lambda\mathcal{J}\}. \tag{5.17}$$

If $\alpha(\lambda)$ is strictly convex and continuously differentiable, then for each $\mathcal{J}$ the minimum is achieved for a single value of λ, which is the unique solution λ_c of $\frac{\partial \alpha}{\partial \lambda} = \mathcal{J}$ i.e. the Legendre-Fenchel transform reduces to the usual Legendre transform, and the duality relation, $\lambda_c(\mathcal{J}) = -\frac{\partial f}{\partial \mathcal{J}}$, holds, i.e.

$$\left.\frac{\partial \alpha}{\partial \lambda}\right|_{\lambda=-\frac{\partial f}{\partial \mathcal{J}}} = \mathcal{J}. \tag{5.18}$$

If moreover $\alpha(\lambda)$ is differentiable twice, taking the derivative of this relation with respect to $\mathcal{J}$ one obtains

$$\left.\frac{\partial^2 f}{(\partial \mathcal{J})^2}\,\frac{\partial^2 \alpha}{(\partial \lambda)^2}\right|_{\lambda=\lambda_c(\mathcal{J})} = -1. \tag{5.19}$$

From the fundamental properties of any large deviation function, $f(\mathcal{J})$ is maximum at $\mathcal{J} = J$ defined in (5.15), so by construction $\lambda_c(J) = 0$. But derivatives of $\alpha(\lambda)$ at $\lambda = 0$ are related to cumulants of X_t at large t. For instance

$$\left.\frac{\partial^2 \alpha}{(\partial \lambda)^2}\right|_{\lambda=0} = \lim_{t\to+\infty} \frac{\langle X_t^2\rangle_{\rm st} - \langle X_t\rangle_{\rm st}^2}{t}, \tag{5.20}$$

and one gets:

$$\left.\frac{\partial^2 f}{(\partial \mathcal{J})^2}\right|_{\mathcal{J}=J} = -\left[\lim_{t\to+\infty} \frac{\langle X_t^2\rangle_{\rm st} - \langle X_t\rangle_{\rm st}^2}{t}\right]^{-1}. \tag{5.21}$$

We now apply these considerations to linear response.

5.1.4 Linear response far from equilibrium

In the case at hand, the large deviation function f depends on other parameters. The one relevant for the discussion of linear response is the thermodynamic force $\mathcal{F}$, and we write $f(\mathcal{J};\mathcal{F})$. In general f depends on other variables which we do not mention explicitly, and which are supposed to be kept constant whenever a partial derivative is taken in the sequel.

We start from the property recalled in the previous subsubsection:

$$\left.\frac{\partial f(\mathcal{J};\mathcal{F})}{\partial \mathcal{J}}\right|_{\mathcal{J}=J} = 0. \tag{5.22}$$

Taking the derivative with respect to $\mathcal{F}$ leads to

$$\left.\frac{\partial J}{\partial \mathcal{F}}\frac{\partial^2 f(\mathcal{J};\mathcal{F})}{(\partial \mathcal{J})^2}\right|_{\mathcal{J}=J} = -\left.\frac{\partial^2 f(\mathcal{J};\mathcal{F})}{\partial \mathcal{F}\partial \mathcal{J}}\right|_{\mathcal{J}=J}. \tag{5.23}$$

According to (5.21), $\left.\frac{\partial^2 f(\mathcal{J};\mathcal{F})}{(\partial \mathcal{J})^2}\right|_{\mathcal{J}=J}$ can be expressed in terms of the second cumulant of X_t in the infinite-time limit, and we can rewrite the equality (5.23) as:

$$\frac{\partial J}{\partial \mathcal{F}} = \left.\frac{\partial^2 f(\mathcal{J};\mathcal{F})}{\partial \mathcal{F}\partial \mathcal{J}}\right|_{\mathcal{J}=J} \times \lim_{t\to+\infty} \frac{\langle X_t^2\rangle_{\rm st} - \langle X_t\rangle_{\rm st}^2}{t}. \tag{5.24}$$

This is the generic expression of the linear response in the non-equilibrium state far from equilibrium. In the case of thermal contact X_t is the cumulative heat $\mathcal{Q}_2(t)$ received from heat bath 2.

5.1.5 Einstein-Green-Kubo relation as a consequence of the fluctuation relation for the large deviation function

When the large deviation function obeys the fluctuation relation (5.16) arising from the MDB, successive partial derivatives of the fluctuation relation entail that

$$\frac{\partial^2 f(\mathcal{J};\mathcal{F})}{\partial\mathcal{F}\partial\mathcal{J}} + \frac{\partial^2 f(\mathcal{J}';\mathcal{F})}{\partial\mathcal{F}\partial\mathcal{J}'}\bigg|_{\mathcal{J}'=-\mathcal{J}} = 1. \tag{5.25}$$

Then the relation (5.25) for $\mathcal{J}=0$ yields $\frac{\partial^2 f(\mathcal{J};\mathcal{F})}{\partial\mathcal{F}\partial\mathcal{J}}\Big|_{\mathcal{J}=0} = \frac{1}{2}$ and, since the stationary state with $\mathcal{F}=0$ is in fact the equilibrium state, the linear response relation (5.24) becomes

$$\frac{\partial J}{\partial\mathcal{F}}\bigg|_{\mathcal{F}=0} = \frac{1}{2}\times\lim_{t\to+\infty}\frac{\langle X_t^2\rangle_{\rm eq}-\langle X_t\rangle_{\rm eq}^2}{t}. \tag{5.26}$$

In the following the latter fluctuation-dissipation relation will be referred to as the Einstein-Green-Kubo relation. The precise terminology has been pointed out in Ref.[61]. When the equilibrium fluctuations are written in terms of a second-order cumulant, namely the mean value of the squared Helfand moment $X_t-\langle X_t\rangle$, as it is the case for Einstein relation, the fluctuation-dissipation relations are called Einstein-Helfand formulae [16, 74]. When the equilibrium fluctuations are written in terms of the time-correlation function of the instantaneous current, they are are known as the Green-Kubo relations [75, 76, 77, 78] or the Yamamoto-Zwanzig formulae in the context of chemical relations [79, 80].

The Einstein-Green-Kubo relation (5.26) can be rephrased in terms of the Onsager coefficient L defined in the framework of thermodynamics of irreversible processes near equilibrium as

$$L \equiv \frac{\partial J}{\partial\mathcal{F}}\bigg|_{\mathcal{F}=0} = \lim_{\mathcal{F}\to0}\frac{J}{\mathcal{F}}, \tag{5.27}$$

where the last equality is valid when there is only one nonzero mean current in the stationary state, in which case this single current J vanishes as $\mathcal{F}$ goes to zero.

In the case of thermal contact, the limit $\mathcal{F}\to0$ in (5.27) can be stated more precisely according to the following argument. If $\beta_1=\beta_2$ the stationary state is the equilibrium state and $J(\beta_1,\beta_2)$ becomes $J(\beta,\beta)=\langle j_2\rangle_{\rm eq}=0$ (see the comment after (3.17)). As a consequence $J(\beta+d\beta,\beta+d\beta)-J(\beta,\beta)=0$ for any infinitesimal $d\beta$, namely

$$\frac{\partial J}{\partial\beta_1}\bigg|_{\beta_1=\beta_2=\beta} = -\frac{\partial J}{\partial\beta_2}\bigg|_{\beta_1=\beta_2=\beta}, \tag{5.28}$$

so that

$$J(\beta_1,\beta_2) \underset{\beta_1,\beta_2\to\beta}{\sim} (\beta_1-\beta_2)\,\frac{\partial J}{\partial\beta_1}\bigg|_{\beta_1=\beta_2=\beta} \tag{5.29}$$

independently of the way β_1 and β_2 go to β. Subsequently the linear response coefficient L defined in (5.27) also reads

$$L = \frac{\partial J}{\partial\beta_1}\bigg|_{\beta_1=\beta_2=\beta}, \tag{5.30}$$

while the Einstein-Green-Kubo relation (5.26) can also be stated as

$$\lim_{\beta_1,\beta_2\to\beta}\frac{J(\beta_1,\beta_2)}{\beta_1-\beta_2} = \frac{1}{2}\lim_{t\to+\infty}\frac{\langle\mathcal{Q}_2^2(t)\rangle_{\rm eq}^{\beta}-\big(\langle\mathcal{Q}_2(t)\rangle_{\rm eq}^{\beta}\big)^2}{t}. \tag{5.31}$$

where $J(\beta_1,\beta_2)=\langle j_2\rangle_{\rm st}^{(\beta_1,\beta_2)}$ and the limit is independent of the way $\beta_1-\beta_2$ vanishes. Eventually in the case of thermal contact the Einstein-Green-Kubo relation relates the proportionality coefficient between the stationary heat current and the difference between the bath inverse temperatures when the system is weakly out of equilibrium to the fluctuations of the heat amount received from one thermal bath in the equilibrium situation where both thermostats are at the same temperature.

5.2 Symmetry of the generating function for cumulants per unit time under MDB

Contrarily to the case of the Einstein-Green-Kubo relation (5.26) between the mean current and the equilibrium infinite-time second cumulant per unit time, relations between higher cumulants per unit time cannot be readily obtained from the fluctuation relation (5.16). In order to obtain such relations one has to resort to the cumulant generating function.

5.2.1 Infinite-time cumulants per unit time for $\mathcal{Q}_1$ and $\mathcal{Q}_2$

The kth cumulant $\kappa^{[p]}_{\mathcal{Q}_a}$ for the heat amount $\mathcal{Q}_a$ received from bath a or the joint cumulant $\kappa^{[p,q]}_{\mathcal{Q}_a,\mathcal{Q}_b}$ for the heat amounts $\mathcal{Q}_a$ and $\mathcal{Q}_b$ can be computed from the "characteristic function"

$$\langle e^{\lambda_1 \mathcal{Q}_1(t)+\lambda_2 \mathcal{Q}_2(t)}\rangle = \sum_{\mathcal{Q}_1}\sum_{\mathcal{Q}_2} e^{\lambda_1 \mathcal{Q}_1+\lambda_2 \mathcal{Q}_2} \sum_{\mathcal{C}_f}\sum_{\mathcal{C}_0} P\left(\mathcal{C}_f|\mathcal{Q}_1,\mathcal{Q}_2;t|\mathcal{C}_0\right) P(\mathcal{C}_0,t_0=0) \tag{5.32}$$

through the following derivatives

$$\kappa^{[p]}_{\mathcal{Q}_a}(t) = \left.\frac{\partial^p \ln\langle e^{\lambda_1 \mathcal{Q}_1(t)+\lambda_2 \mathcal{Q}_2(t)}\rangle}{\partial \lambda_a^p}\right|_{\lambda_1=\lambda_2=0} \quad \text{for } a=\{1,2\} \tag{5.33}$$

$$\kappa^{[p,q]}_{\mathcal{Q}_1,\mathcal{Q}_2}(t) = \left.\frac{\partial^{p+q} \ln\langle e^{\lambda_1 \mathcal{Q}_1(t)+\lambda_2 \mathcal{Q}_2(t)}\rangle}{\partial \lambda_1^p \partial \lambda_2^q}\right|_{\lambda_1=\lambda_2=0}. \tag{5.34}$$

For a Markov process, the leading long-time behaviors of these cumulants are proportional to the time t elapsed from the beginning of the measurements. The asymptotic behavior of the cumulants per unit time are given by the derivatives of

$$\alpha_{1,2}(\lambda_1,\lambda_2) \equiv \lim_{t\to+\infty} \frac{1}{t} \ln\langle e^{\lambda_1 \mathcal{Q}_1(t)+\lambda_2 \mathcal{Q}_2(t)}\rangle \tag{5.35}$$

with respect to λ_1 and λ_2 at $\lambda_1=\lambda_2=0$. In other words,

$$\lim_{t\to+\infty} \frac{\kappa^{[p,q]}_{\mathcal{Q}_1,\mathcal{Q}_2}}{t} = \left.\frac{\partial^{p+q}\alpha_{1,2}}{\partial \lambda_1^p \partial \lambda_2^q}\right|_{\lambda_1=\lambda_2=0}. \tag{5.36}$$

Similarly

$$\lim_{t\to+\infty} \frac{\kappa^{[p]}_{\mathcal{Q}_a}}{t} = \left.\frac{\partial^p \alpha_a}{\partial \lambda_a^p}\right|_{\lambda_a=0} \quad \text{for } a=\{1,2\} \tag{5.37}$$

where

$$\alpha_a(\lambda) \equiv \lim_{t\to+\infty} \frac{1}{t} \ln\langle e^{\lambda \mathcal{Q}_a(t)}\rangle. \tag{5.38}$$

The generating function for infinite-time cumulants per unit time, $\alpha_a(\lambda)$, may also be referred to as the scaled cumulant generating function.

We notice that, since for any cumulant $\lim_{t\to\infty} \kappa^{[p]}_{\mathcal{Q}_2}/t$ is finite, under some technical conditions, the probability distribution of the variable $\left[\mathcal{Q}_2(t)-\langle\mathcal{Q}_2(t)\rangle\right]/\sqrt{t}$ becomes Gaussian in the long-time limit. Indeed the logarithm of the characteristic function for the variable $Y_2(t) = \left[\mathcal{Q}_2(t)-\langle\mathcal{Q}_2(t)\rangle\right]/\sqrt{t}$ reads $\ln\langle e^{\lambda Y_2(t)}\rangle = \sum_{p=2}^{+\infty} (1/p!)\left(\lambda/\sqrt{t}\right)^p \kappa^{[p]}_{\mathcal{Q}_2}(t)$. If the sum and the $t\to+\infty$ limit can be interchanged, $\ln\langle e^{\lambda Y_2(t)}\rangle$ becomes proportional to λ^2 in the limit where t goes to infinity : only the second cumulant of $Y_2(t)$ survives in the long-time limit.

5.2.2 Constraints from the bound upon the system energy

In the case of a system with a finite number of configurations, $\mathcal{Q}_1+\mathcal{Q}_2=\mathcal{E}(\mathcal{C}_f)-\mathcal{E}(\mathcal{C}_0)$ is restricted to some interval $[-|\Delta\mathcal{E}|_{\max},+|\Delta\mathcal{E}|_{\max}]$, and the definition (5.32) entails the inequalities

$$e^{-\lambda_1|\Delta\mathcal{E}|_{\max}}\leq\frac{\langle e^{\lambda_1\mathcal{Q}_1(t)+\lambda_2\mathcal{Q}_2(t)}\rangle}{\langle e^{(\lambda_2-\lambda_1)\mathcal{Q}_2(t)}\rangle}\leq e^{\lambda_1|\Delta\mathcal{E}|_{\max}}. \tag{5.39}$$

As a consequence,

$$\alpha_{1,2}(\lambda_1,\lambda_2)=\alpha_a(\lambda_a-\lambda_b)\quad\text{with }\{a,b\}=\{1,2\}. \tag{5.40}$$

The relation (5.40) can be rewritten as

$$\alpha_{1,2}(\lambda_1,\lambda_2)=\alpha_2(\lambda_2-\lambda_1)\quad\text{and}\quad\alpha_1(\lambda)=\alpha_2(-\lambda). \tag{5.41}$$

The specific dependence of $\alpha_{12}(\lambda_1,\lambda_2)$ upon $\lambda_2-\lambda_1$ together with the generic formulæ (5.36) -(5.37) imply the following relations between the infinite-time cumulants per unit time,

$$\lim_{t\to\infty}\frac{\kappa^{[p]}_{\mathcal{Q}_1}}{t}=(-1)^p\lim_{t\to\infty}\frac{\kappa^{[p]}_{\mathcal{Q}_2}}{t} \tag{5.42}$$

and

$$\lim_{t\to\infty}\frac{\kappa^{[p,q]}_{\mathcal{Q}_1,\mathcal{Q}_2}}{t}=(-1)^p\lim_{t\to\infty}\frac{\kappa^{[p+q]}_{\mathcal{Q}_2}}{t}. \tag{5.43}$$

5.2.3 MDB and symmetry of the generating function for infinite-time cumulants per unit time

The modified detailed balance entails the time-reversal symmetry (4.8) for $P\left(\mathcal{C}_f|\mathcal{Q}_1,\mathcal{Q}_2;t|\mathcal{C}_0\right)$ at finite time. Henceforth, according to its definition (5.32), the characteristic function can be rewritten as

$$\langle e^{\lambda_1\mathcal{Q}_1(t)+\lambda_2\mathcal{Q}_2(t)}\rangle=\sum_{\mathcal{Q}_1}\sum_{\mathcal{Q}_2}e^{(\beta_1-\lambda_1)\mathcal{Q}_1+(\beta_2-\lambda_2)\mathcal{Q}_2}\sum_{\mathcal{C}_f}\sum_{\mathcal{C}_0}P\left(\mathcal{C}_0|\mathcal{Q}_1,\mathcal{Q}_2;t|\mathcal{C}_f\right)P(\mathcal{C}_0,t_0=0). \tag{5.44}$$

When the Markov matrix is irreducible and the system has a finite number of configurations, the Perron-Frobenius theorem entails that there exists a single stationary state $P_{\rm st}$ and that every configuration $\mathcal{C}$ has a non-zero probability $P_{\rm st}(\mathcal{C})$. Henceforth, in the case where the initial distribution is the stationary one, the relation (5.44) leads to the inequalities

$$\frac{P^{\min}_{\rm st}}{P^{\max}_{\rm st}}\leq\frac{\langle e^{(\beta_1-\lambda_1)\mathcal{Q}_1+(\beta_2-\lambda_2)\mathcal{Q}_2}\rangle_{\rm st}}{\langle e^{\lambda_1\mathcal{Q}_1+\lambda_2\mathcal{Q}_2}\rangle_{\rm st}}\leq\frac{P^{\max}_{\rm st}}{P^{\min}_{\rm st}}. \tag{5.45}$$

(We recall that $\langle\cdots\rangle_{\rm st}$ denotes an average when the initial configurations are distributed according to the stationary measure $P_{\rm st}(\mathcal{C})$, the maximum and minimum values of which are $P^{\max}_{\rm st}$ and $P^{\min}_{\rm st}$ respectively (see (5.1)). These inequalities entail that the generating function for the infinite-time limits of the joint cumulants per unit time defined in (5.35) obeys the symmetry

$$\alpha_{1,2}(\lambda_1,\lambda_2)=\alpha_{1,2}(\beta_1-\lambda_1,\beta_2-\lambda_2). \tag{5.46}$$

We notice that the symmetry (5.46) can also be derived by considering the evolution of the Laplace transform of $P(\mathcal{Q}_1,\mathcal{Q}_2;t)$. With the notations of paper II, $\langle e^{\lambda_1\mathcal{Q}_1(t)+\lambda_2\mathcal{Q}_2(t)}\rangle_{\rm st}=\sum_{\mathcal{C}_f}\sum_{\mathcal{C}_0}(\mathcal{C}_f|\widehat{\mathbb{U}}(e^{\lambda_1},e^{\lambda_2};t)|\mathcal{C}_0)P_{\rm st}(\mathcal{C}_0)$ and $\alpha_{12}(\lambda_1,\lambda_2)$ coincides with the largest eigenvalue of the operator $\widetilde{\mathbb{A}}(\lambda_1,\lambda_2)$ which rules the evolution of $\widehat{\mathbb{U}}(e^{\lambda_1},e^{\lambda_2};t)$ according to $d\widehat{\mathbb{U}}/dt=\widetilde{\mathbb{A}}\widehat{\mathbb{U}}$. The MDB implies that the operator $\widetilde{\mathbb{A}}(\lambda_1,\lambda_2)$ obeys the symmetry $\widetilde{\mathbb{A}}(\lambda_1,\lambda_2)=\widetilde{\mathbb{A}}^T(\beta_1-\lambda_1,\beta_2-\lambda_2)$, where $\widetilde{\mathbb{A}}^T$ denotes the transposed matrix of $\widetilde{\mathbb{A}}$ in the configuration basis. Then, by using the Perron-Frobenius theorem, one can prove that the largest eigenvalue of $\widetilde{\mathbb{A}}(\lambda_1,\lambda_2)$ satisfies the symmetry

relation (5.46) (see the argument in [1] where an analogous symmetry is exhibited and then used for the derivation of the long-time fluctuation relation obeyed by the action functional defined in the comment after(4.5)-(4.6)).

Moreover, since the system has a finite number of configurations, $\alpha_{12}(\lambda_1, \lambda_2) = \alpha_a(\lambda_a - \lambda_b)$ by virtue of (5.40), and the symmetry property (5.46) of $\alpha_{1,2}(\lambda_1, \lambda_2)$ becomes

$$\alpha_a(\lambda) = \alpha_a(\beta_a - \beta_b - \lambda) \quad \text{for } \{a, b\} = \{1, 2\}. \tag{5.47}$$

We can apply the Gärtner-Ellis theorem (see subsubsection 5.1.3 and references there): if $\alpha_2(\lambda)$ exists and is differentiable for all λ in $\mathbb{R}$, then the large deviation function of the current $\mathcal{J} = \mathcal{Q}_2/t$ exists and it can be calculated as the Legendre-Fenchel transform of $\alpha_2(\lambda)$, namely, with the signs chosen in the definitions used in the present paper,

$$f_{\mathcal{Q}_2}(\mathcal{J}) = \min_{\lambda \in \mathbb{R}} \{\alpha_2(\lambda) - \lambda \mathcal{J}\}. \tag{5.48}$$

Then, since $\alpha_2(\lambda)$ obeys the symmetry (5.47), $f_{\mathcal{Q}_2}(\mathcal{J})$ obeys the fluctuation relation (5.12). We notice that we can similarly retrieve the fluctuation relation obeyed by $f_{\Delta_{\rm exch} S}$. Indeed, $\alpha_{\Delta_{\rm exch} S}(\lambda) \equiv \lim_{t \to +\infty} (1/t) \ln \langle e^{\lambda \Delta_{\rm exch} S(t)} \rangle$ is equal to $\alpha_{12}(\lambda\beta_1, \lambda\beta_2)$ with $\alpha_{12}(\lambda_1, \lambda_2)$ defined in (5.35), and, according to (5.40), $\alpha_{\Delta_{\rm exch} S}(\lambda) = \alpha_a\left(\lambda(\beta_a - \beta_b)\right)$. Therefore the symmetry (5.47) implies that

$$\alpha_{\Delta_{\rm exch} S}(1 - \lambda) = \alpha_{\Delta_{\rm exch} S}(\lambda). \tag{5.49}$$

Under the above assumptions, the large deviation function of the exchange entropy cumulative current can be computed as the Legendre-Fenchel transform of $\alpha_{\Delta_{\rm exch} S}$, namely $f_{\Delta_{\rm exch} S}(\mathcal{J}) = \min_{\lambda \in \mathbb{R}} \{\alpha_{\Delta_{\rm exch} S}(\lambda) - \lambda \mathcal{J}\}$. Henceforth the symmetry (5.49) allows to retrieve the fluctuation relation (5.7).

Another consequence of the symmetry (5.47) is that the Einstein-Green-Kubo relation near equilibrium, which can be derived from the fluctuation relation (5.12) for $f_{\mathcal{Q}_2}(\mathcal{J})$ as shown in section 5.1.5, can equivalently be directly derived from the symmetry (5.47) of its Legendre-Fenchel transform $\alpha_2(\lambda)$. This can be checked as follows. Relation (5.47) can be rewritten for $a = 2$ as

$$\alpha_2(\lambda; \beta_1, \beta_2) = \alpha_2(-\mathcal{F} - \lambda; \beta_1, \beta_2), \tag{5.50}$$

where $\mathcal{F} = \beta_1 - \beta_2$ is the thermodynamic force. The Einstein-Green-Kubo relation involves the variation of the mean current $J(\beta_1, \beta_2) \equiv \frac{\partial \alpha_2}{\partial \lambda}\big|_{(\lambda=0;\beta_1,\beta_2)}$ with respect to $\mathcal{F}$, so we need to compute mixed derivatives of α_2 with respect to λ and $\mathcal{F}$. Notice that in the identity (5.50), the second and the third arguments are untouched. So if we make any invertible (λ-independent) change of variables $(\beta_1, \beta_2) \leftrightarrow (\mathcal{F}, \rho)$ and set $\alpha_2(\lambda; \beta_1, \beta_2) \equiv \alpha(\lambda; \mathcal{F}, \rho)$, we have the following identity

$$\alpha(\lambda; \mathcal{F}, \rho) = \alpha(-\mathcal{F} - \lambda; \mathcal{F}, \rho), \tag{5.51}$$

where the third variable ρ is purely a spectator. Keeping it fixed, we get from the symmetry (5.51):

$$\left.\frac{\partial^2 \alpha}{\partial \mathcal{F} \partial \lambda}\right|_{(\lambda;\mathcal{F})} = -\left.\frac{\partial^2 \alpha}{\partial \mathcal{F} \partial \lambda}\right|_{(-\mathcal{F}-\lambda;\mathcal{F})} + \left.\frac{\partial^2 \alpha}{\partial \lambda^2}\right|_{(-\mathcal{F}-\lambda;\mathcal{F})}. \tag{5.52}$$

Taking this relation at $\lambda = \mathcal{F} = 0$, we obtain

$$2 \left.\frac{\partial^2 \alpha}{\partial \mathcal{F} \partial \lambda}\right|_{(\lambda=0;\mathcal{F}=0)} = \left.\frac{\partial^2 \alpha}{\partial \lambda^2}\right|_{(\lambda=0;\mathcal{F}=0)}. \tag{5.53}$$

From now on, the discussion parallels the one in section 5.1.5. We repeat the argument in a slightly different form to stress again the slight subtlety involved in variations with respect to $\mathcal{F}$ in the context of thermal contact. The condition $\mathcal{F} = 0$ means $\beta_1 = \beta_2 \equiv \beta$. So the right-hand side of (5.53) is simply $\left.\frac{\partial^2 \alpha}{\partial \lambda^2}\right|_{(\lambda=0;\mathcal{F}=0)} = \left.\frac{\partial^2 \alpha_2}{\partial \lambda^2}\right|_{(\lambda=0;\beta_1=\beta_2=\beta)}$. The left-hand side of (5.53) can be

dealt with as follows. The change of variables $(\beta_1, \beta_2) \leftrightarrow (\mathcal{F}, \rho)$ is λ-independent, so we can set $\lambda = 0$ once the derivative with respect to λ is taken

$$\left.\frac{\partial \alpha}{\partial \lambda}\right|_{(\lambda=0;\mathcal{F},\rho)} = \left.\frac{\partial \alpha_2}{\partial \lambda}\right|_{(\lambda=0;\beta_1,\beta_2)} = J(\beta_1, \beta_2). \tag{5.54}$$

Of course, taking the partial derivative of $J(\beta_1, \beta_2)$ with respect to $\mathcal{F}$ is ambiguous, as it depends on the choice of the variable ρ. However, $\frac{\partial J}{\partial \mathcal{F}}\big|_{\mathcal{F}=0}$ has an intrinsic meaning, independent of the choice of the variable ρ. Indeed, by construction $J(\beta, \beta) = 0$ (equilibrium) so $\left.\frac{\partial J}{\partial \beta_1}\right|_{\beta_1=\beta_2=\beta} + \left.\frac{\partial J}{\partial \beta_2}\right|_{\beta_1=\beta_2=\beta} = 0$. Moreover, from $\mathcal{F} = \beta_1 - \beta_2$ we get $1 = \frac{\partial \beta_1}{\partial \mathcal{F}} - \frac{\partial \beta_2}{\partial \mathcal{F}}$. Now $\frac{\partial J}{\partial \mathcal{F}} = \frac{\partial J}{\partial \beta_1}\frac{\partial \beta_1}{\partial \mathcal{F}} + \frac{\partial J}{\partial \beta_2}\frac{\partial \beta_2}{\partial \mathcal{F}}$. This implies that $\frac{\partial J}{\partial \mathcal{F}}\big|_{\mathcal{F}=0} = \left.\frac{\partial J}{\partial \beta_1}\right|_{\beta_1=\beta_2=\beta}$ is intrinsic and we can write (5.53) as

$$\left.\frac{\partial J}{\partial \mathcal{F}}\right|_{\mathcal{F}=0} = \frac{1}{2}\left.\frac{\partial^2 \alpha_2}{\partial \lambda^2}\right|_{(\lambda=0;\beta_1=\beta_2=\beta)}. \tag{5.55}$$

By using the relation (5.37) for $p = 2$ between $\alpha_2(\lambda)$ and the infinite-time cumulant per unit time together with the fact that, when $\beta_1 = \beta_2$, the stationary state is the equilibrium state, we get the Einstein-Green-Kubo relation (5.26). We notice that the present derivation of the Einstein-Green-Kubo relation is very similar to the argument developed by Lebowitz and Spohn in Ref.[1] in the case where there are several independent currents corresponding to several parameters which drive the system out of equilibrium.

5.3 Far from equilibrium relations for infinite-time heat cumulants per unit time

In the present subsection we settle generic results ; thus we replace the cumulative heat $\mathcal{Q}_2(t)$ by a generic cumulative quantity X_t. According to the Einstein-Green-Kubo relation (5.26) where, by virtue of (5.27), $\partial J/\partial \mathcal{F}|_{\mathcal{F}=0}$ can be replaced by $\lim_{\mathcal{F}\to 0} J/\mathcal{F}$, in the vicinity of equilibrium the ratio $J/\mathcal{F}$ is equal to the equilibrium value of the second cumulant of the cumulative quantity X_t per unit time. On the contrary, when the system is far from equilibrium, $\partial J/\partial \mathcal{F}$ and $J/\mathcal{F}$ are expected not to be equal to $1/2$ times the second cumulant of the cumulative heat X_t per unit time in the stationary state. The property

$$\frac{\partial J}{\partial \mathcal{F}} \neq \frac{1}{2}\lim_{t\to+\infty} \frac{\langle X_t^2\rangle_{\rm st} - \langle X_t\rangle_{\rm st}^2}{t} \tag{5.56}$$

arises from (5.24), while the discrepancy

$$\frac{J}{\mathcal{F}} \neq \frac{1}{2}\lim_{t\to+\infty} \frac{\langle X_t^2\rangle_{\rm st} - \langle X_t\rangle_{\rm st}^2}{t}. \tag{5.57}$$

is already mentioned in Ref.[1]. This discrepancy implies that the generating function $\alpha(\lambda)$ for the infinite-time cumulants of X_t per unit time is not a quadratic function of its argument λ, namely the large deviation function of the current $\mathcal{J}_t = X_t/t$ is not a quadratic in the generic case, i.e. the probability distribution of X_t is not asymptotically Gaussian. Indeed, if the generating function $\alpha(\lambda)$ were quadratic in λ, i.e. $\alpha(\lambda) = \alpha^{(1)}\lambda + \frac{1}{2}\alpha^{(2)}\lambda^2$, then the symmetry (5.51) would lead to $\alpha^{(1)} = \frac{1}{2}\alpha^{(2)}\mathcal{F}$. In the present section we derive an equation hierarchy for the infinite-time cumulants per unit time far from equilibrium, which in particular gives how $J/\mathcal{F}$ is related to all even cumulants per unit time in the infinite-time limit.

5.3.1 Generalized Einstein-Green-Kubo relations for cumulants per unit time far from equilibrium

The starting point is (5.50) which we rewrite for convenience in the generic form

$$\alpha(\lambda;\mathcal{F}) = \alpha(-\mathcal{F}-\lambda;\mathcal{F}), \tag{5.58}$$

where $\mathcal{F}$ is the thermodynamical force (an inverse temperature difference in (5.47)). The viewpoint is that the coefficients in the expansion of $\alpha(\lambda;\mathcal{F})$ in powers of λ at $\lambda = 0$ are related to the heat cumulants per unit time in the infinite-time limit,

$$\overline{\kappa}^{[p]} \equiv \lim_{t\to+\infty} \frac{1}{t}\kappa^{[p]}(t), \tag{5.59}$$

and these cumulants per unit time are experimentally measurable. This is true at least for the first few ones.

There are a number of ways to rewrite the symmetry (5.58). Formally it is equivalent to

$$e^{-\mathcal{F}\frac{\partial}{\partial\lambda}}\alpha(\lambda;\mathcal{F}) = \alpha(-\lambda;\mathcal{F}). \tag{5.60}$$

Expanding both sides in powers of λ and comparing we get that the cumulants per unit time, given by $\alpha(\lambda;\mathcal{F}) \equiv \sum_{p=0}^{+\infty}\frac{1}{p!}\lambda^p\overline{\kappa}^{[p]}(\mathcal{F})$ obey the equation hierarchy

$$\overline{\kappa}^{[p]}(\mathcal{F}) = \sum_{q=0}^{+\infty}(-1)^{p+q}\frac{\mathcal{F}^q}{q!}\overline{\kappa}^{[p+q]}(\mathcal{F}) \qquad \text{for } p = 0,1,\cdots. \tag{5.61}$$

In (5.61) $\overline{\kappa}^{[0]}(\mathcal{F}) = 0$, because $\alpha(\lambda;\mathcal{F})$ is a cumulant generating function. The relation (5.61) can be explicitly checked in the case of the solvable model studied in paper II in a kinetic regime where the probability distribution of the heat received from the slow thermostat is that of an asymmetric random walk.

A further expansion of the equations (5.61) in powers of $\mathcal{F}$, with $\overline{\kappa}^{[p]}(\mathcal{F}) \equiv \sum_{q=0}^{+\infty}\frac{1}{q!}\mathcal{F}^q\overline{\kappa}^{[p;q]}$, yields another hierarchy of relations, each of which involves only a finite number of derivatives of cumulants per unit time with respect to $\mathcal{F}$,

$$\overline{\kappa}^{[p;q]} = \sum_{r=0}^{q}(-1)^{p+r}\binom{q}{r}\overline{\kappa}^{[p+r;q-r]} \qquad \text{for } p,q = 0,1,\cdots, \tag{5.62}$$

where $\binom{q}{r} \equiv q!/[r!(q-r)!]$, and $\overline{\kappa}^{[p;0]} = \overline{\kappa}^{[p]}(\mathcal{F}=0)$ while for $q \geq 1$

$$\overline{\kappa}^{[p;q]} \equiv \left.\frac{\partial^q\overline{\kappa}^{[p]}}{\partial\mathcal{F}^q}\right|_{\mathcal{F}=0}. \tag{5.63}$$

From the cumulant generating function interpretation, in (5.62) it is to be understood that $\overline{\kappa}^{[0;q]} = 0$ for $q = 0,1,\cdots$, since $\overline{\kappa}^{[0]}(\mathcal{F}) = 0$.

The relation hierarchies (5.61) and (5.62) are pure consequences of the symmetry (5.58) without using any further properties of α. Anyway, they express nothing but a parity relation around $\lambda = \mathcal{F}/2$, which fixes half of the coefficients in terms of the other ones, so the equations are not expected to be all independent. Indeed, this is particularly transparent at small order in $\mathcal{F}$: for $q = 0$ (5.62) yields nothing if p is even, but if p is odd it leads to $\overline{\kappa}^{[2n+1;0]} = 0$. As for the hierarchy (5.61) obtained before expansions in powers of $\mathcal{F}$, the dependence between its equations can be exhibited as follows. By specifying the equations to $p = 2n$ or $p = 2n+1$, the infinite system of equations (5.61) can be rewritten as the equivalent hierarchy,

$$\left.\begin{aligned}\overline{\kappa}^{[2n+1]}(\mathcal{F}) &= \sum_{q=1}^{+\infty}(-1)^{q+1}\frac{\mathcal{F}^q}{(q+1)!}\overline{\kappa}^{[2n+q+1]}(\mathcal{F})\\ \overline{\kappa}^{[2n+1]}(\mathcal{F}) &= \frac{1}{2}\sum_{q=1}^{+\infty}(-1)^{q+1}\frac{\mathcal{F}^q}{q!}\overline{\kappa}^{[2n+q+1]}(\mathcal{F})\end{aligned}\right\} \qquad \text{for } n = 0,1,\cdots, \tag{5.64}$$

where the right-hand sides of both relations involves both odd and even cumulants per unit time. The difference between the two expressions for $\overline{\kappa}^{[2n+1]}(\mathcal{F})$ gives the infinite sum rule $0 = \sum_{q=2}^{+\infty}(-1)^{q+1}\left[\frac{1}{(q+1)!} - \frac{1}{2}\frac{1}{q!}\right]\mathcal{F}^q\overline{\kappa}^{[2n+q+1]}(\mathcal{F})$.

It is not difficult to obtain independent relations which express odd cumulants in terms of even cumulants. We rewrite (5.60) as

$$\left(1 + e^{-\mathcal{F}\frac{\partial}{\partial\lambda}}\right)\alpha(\lambda;\mathcal{F}) = \alpha(\lambda;\mathcal{F}) + \alpha(-\lambda;\mathcal{F}), \tag{5.65}$$

i.e.

$$\alpha(\lambda;\mathcal{F}) = \frac{1}{1 + e^{-\mathcal{F}\frac{\partial}{\partial\lambda}}}\left[\alpha(\lambda;\mathcal{F}) + \alpha(-\lambda;\mathcal{F})\right]. \tag{5.66}$$

Note that $\left[1+e^{-x}\right]^{-1} + \left[1+e^{x}\right]^{-1} = 1$ so that $\left[1+e^{-x}\right]^{-1} - \frac{1}{2}$ is an odd function of x. Hence the Taylor expansion in x can be written

$$\frac{1}{1+e^{-x}} = \frac{1}{2}\left(1 + \sum_{k=0}^{+\infty} d_k x^{2k+1}\right) = \frac{1}{2}\left(1 + \frac{x}{2} - \frac{x^3}{24} + \frac{x^5}{240} - \frac{17x^7}{40320} + \cdots\right), \tag{5.67}$$

where the coefficients d_k are related to the classical Bernoulli numbers by

$$d_k = 2\frac{4^{k+1}-1}{(2k+2)!}B_{2k+2}. \tag{5.68}$$

From the identity

$$1 = \frac{1}{2}(1+e^{-x})\left(1 + \sum_{k=0}^{+\infty} d_k x^{2k+1}\right) \tag{5.69}$$

one infers the recursion relation

$$d_k = \frac{1}{2}\left(\frac{1}{(2k+1)!} - \sum_{0\leq l<k}\frac{d_l}{(2(k-l))!}\right). \tag{5.70}$$

Finally, in the expansion of (5.66) the even terms cancel out and one is left with

$$\overline{\kappa}^{[2n+1]}(\mathcal{F}) = \sum_{k=0}^{+\infty} d_k \mathcal{F}^{2k+1}\overline{\kappa}^{[2(n+k+1)]}(\mathcal{F}) \qquad \text{for } n = 0, 1, \cdots, \tag{5.71}$$

which expresses systematically odd cumulants per unit time in terms of even cumulants per unit time. For instance, in the case $n = 0$, the ratio of the out-of-equilibrium current $J(\mathcal{F}) = \overline{\kappa}^{[1]}$ and the thermodynamic force $\mathcal{F}$ is determined by all even cumulants as

$$\frac{J(\mathcal{F})}{\mathcal{F}} = \sum_{q=0}^{+\infty} d_q \mathcal{F}^{2q} \lim_{t\to+\infty}\frac{\kappa^{[2(q+1)]}(\mathcal{F})}{t}. \tag{5.72}$$

The latter relation may be viewed as a a far-from-equilibrium generalization of the Einstein-Green-Kubo relation (5.26)-(5.27).

5.3.2 Relations between non-linear responses of cumulants per unit time near equilibrium

It is to be noted that the two equivalent hierarchies (5.61) and (5.71) for relations between cumulants make sense only if the cumulants satisfy some growth conditions, the second one being most stringent (note that $d_q \simeq 4(-1)^q \pi^{-2q-2}$). Such problem do not arise in the double expansions in both λ and $\mathcal{F}$. Expanding (5.71) in powers of $\mathcal{F}$ yields for $n = 0, 1, \cdots$

$$\begin{aligned} \overline{\kappa}^{[2n+1;0]} &= 0 \\ \overline{\kappa}^{[2n+1;q]} &= \sum_{k=0}^{\lfloor (q-1)/2 \rfloor} \frac{d_k q!}{[q-(2k+1)]!} \overline{\kappa}^{[2(n+k+1);q-(2k+1)]} \qquad q = 1, 2, \cdots \end{aligned} \tag{5.73}$$

where $\lfloor y \rfloor$ denotes the lower integer part of y. It is easy to check, at least for small values of q, that the contents of (5.62) and (5.73) are the same. The physical consequences of the latter relations are more readily inferred by explicitly rewriting the relations in the case where q is odd or even and in terms of either equilibrium cumulants per unit time or the partial derivatives of out-of-equilibrium cumulants per unit time with respect to the thermodynamic force $\mathcal{F}$. Then the formulae (5.73) read

$$\overline{\kappa}_{\rm eq}^{[2n+1]} = 0 \tag{5.74a}$$

$$\left. \frac{\partial \overline{\kappa}^{[2n+1]}}{\partial \mathcal{F}} \right|_{\mathcal{F}=0} = \frac{1}{2} \overline{\kappa}_{\rm eq}^{[2(n+1)]} \tag{5.74b}$$

$$\left. \frac{\partial^{2m} \overline{\kappa}^{[2n+1]}}{\partial \mathcal{F}^{2m}} \right|_{\mathcal{F}=0} \underset{m \geq 1}{=} \sum_{r=1}^{m} d_{m-r} \frac{(2m)!}{(2r-1)!} \left. \frac{\partial^{2r-1} \overline{\kappa}^{[2(n+m+1-r)]}}{\partial \mathcal{F}^{2r-1}} \right|_{\mathcal{F}=0} \tag{5.74c}$$

$$\left. \frac{\partial^{2m+1} \overline{\kappa}^{[2n+1]}}{\partial \mathcal{F}^{2m+1}} \right|_{\mathcal{F}=0} \underset{m \geq 1}{=} d_m (2m+1)! \, \overline{\kappa}_{\rm eq}^{[2(n+m+1)]} + \sum_{r=1}^{m} d_{m-r} \frac{(2m+1)!}{(2r)!} \left. \frac{\partial^{2r} \overline{\kappa}^{[2(n+m+1-r)]}}{\partial \mathcal{F}^{2r}} \right|_{\mathcal{F}=0} \tag{5.74d}$$

where the values of the d_k's are given in (5.68).

The equilibrium statements (5.74a) are exemplified in the case of thermal contact as follows. The first cumulant per unit time $\overline{\kappa}^{[1]}$ for the cumulative heat $\mathcal{Q}_2$ coincides with the average of the instantaneous current of j_2, $\langle j_2 \rangle_{\rm st}^{(\beta_1,\beta_2)} \equiv J$, and from the first equation (5.74a) for $n = 0$ we retrieve that at equilibrium the mean instantaneous current vanishes, $J_{\rm eq} = 0$. More generally, (5.74a) states that at equilibrium all odd cumulants per unit time of the heat amounts received from one of the two heat baths vanish.

The relations (5.74b) are generalized fluctuation-dissipation relations in the vicinity of equilibrium, which are called "generalized" in the sense that they express the linear response of any cumulant per unit time. Indeed, in the case of thermal contact, from (5.74b) for $n = 0$ one retrieves the fluctuation-dissipation relation (also called Einstein-Green-Kubo relation) in the form (5.26) where X_t is the heat amount received from one of the two heat baths. More generally, from (5.74b) for any value of n one gets the generalized fluctuation-dissipation relations

$$\lim_{(\beta_1,\beta_2) \to (\beta,\beta)} \frac{1}{\beta_1 - \beta_2} \left(\lim_{t \to +\infty} \frac{\kappa^{[2n+1]}}{t} \right) = \frac{1}{2} \lim_{t \to +\infty} \frac{\kappa_{\rm eq}^{[2n+2]}}{t}. \tag{5.75}$$

Analogous relations have been derived in the case of several independent out-of-equilibrium steady currents by Andrieux and Gaspard [61] (see also subsection 6.4) We notice that for $n = 1$, (5.75) means that, in the limit of vanishing $\beta_1 - \beta_2$, the ratio between the out-of-equilibrium third centered moment (non-normalized skewness) $\langle (\mathcal{Q}_2 - \langle \mathcal{Q}_2 \rangle)^3 \rangle$ per unit time and the thermodynamic force $\beta_1 - \beta_2$ tends to half the equilibrium fourth cumulant (kurtosis multiplied by the square of the variance) $\langle (\mathcal{Q}_2 - \langle \mathcal{Q}_2 \rangle_{\rm eq})^4 \rangle_{\rm eq} - 3 \langle (\mathcal{Q}_2 - \langle \mathcal{Q}_2 \rangle_{\rm eq})^2 \rangle_{\rm eq}$ per unit time. These cumulants are expected to be experimentally measurable.

The relations (5.74c) and (5.74d) deal with higher-order response coefficients. The relation (5.74c) in the case $m = 1$ reads $\partial^2 \overline{\kappa}^{[2n+1]}/\partial \mathcal{F}^2|_{\mathcal{F}=0} = \partial \overline{\kappa}^{[2n+2]}/\partial \mathcal{F}|_{\mathcal{F}=0}$. For instance, for $n = 0$ and $n = 1$ it leads respectively to

$$\left.\frac{\partial^2 J}{\partial \mathcal{F}^2}\right|_{\mathcal{F}=0} = \left.\frac{\partial \overline{\kappa}^{[2]}}{\partial \mathcal{F}}\right|_{\mathcal{F}=0}, \tag{5.76}$$

and

$$\left.\frac{\partial^2 \overline{\kappa}^{[3]}}{\partial \mathcal{F}^2}\right|_{\mathcal{F}=0} = \left.\frac{\partial \overline{\kappa}^{[4]}}{\partial \mathcal{F}}\right|_{\mathcal{F}=0}. \tag{5.77}$$

The relation (5.74d) in the case $m = 1$ gives $\partial^3 \overline{\kappa}^{[2n+1]} \partial \mathcal{F}^3|_{\mathcal{F}=0} = \frac{3}{2} \partial^2 \overline{\kappa}^{[2n+2]}/\partial \mathcal{F}^2|_{\mathcal{F}=0} - \frac{1}{4} \overline{\kappa}_{\rm eq}^{[2n+4]}$. For instance, in the case $n = 0$ it reads

$$\left.\frac{\partial^3 J}{\partial \mathcal{F}^3}\right|_{\mathcal{F}=0} = \frac{3}{2} \left.\frac{\partial^2 \overline{\kappa}^{[2]}}{\partial \mathcal{F}^2}\right|_{\mathcal{F}=0} - \frac{1}{4} \overline{\kappa}_{\rm eq}^{[4]}. \tag{5.78}$$

6 Extension of previous results to a larger class of models

6.1 Generic expression of exchange entropy variation

In this section we consider the generic case where the finite-size system $\mathcal{S}$ is made of ν_s species of mobile elementary constituents and can occupy a domain whose boundaries may be mobile interfaces. The degrees of freedom of every elementary constituent involve the site where it sits in discretized space and some possible internal degrees of freedom. In the following we call degrees of freedom of a configuration $\mathcal{C}$ of the system $\mathcal{S}$ the degrees of freedom of the elementary constituents in this configuration. (The number of constituents of every species may vary from one configuration to another.) When some boundaries are mobile interfaces, the description of any configuration $\mathcal{C}$ of the system not only involves the values of its degrees of freedom but it is also specified by the positions of the interfaces that surround the domain, called $\mathcal{D}(\mathcal{C})$, that the system $\mathcal{S}$ can occupy. For each configuration $\mathcal{C}$ one can define the following global quantities : the energy $\mathcal{E}(\mathcal{C})$, the volume $v(\mathcal{C})$ of $\mathcal{D}(\mathcal{C})$ and the total number of elementary constituents $n(\mathcal{C}) = \sum_{s=1}^{\nu_s} n_s(\mathcal{C})$, where $n_s(\mathcal{C})$ is the number of elementary constituents of species s that sit in $\mathcal{D}(\mathcal{C})$. All these quantities are assumed to take a finite number of values. The system $\mathcal{S}$ is in contact with several macroscopic bodies $\mathcal{B}_a$'s.

A crucial assumption is that in the course of the ergodic deterministic microscopic dynamics of the whole system ($\mathcal{S}$ and the large parts $\mathcal{B}_a$), for a given configuration $\mathcal{C}$, the domain $\mathcal{D}(\mathcal{C})$ can be divided in several disjoint subdomains $\mathcal{D}_a(\mathcal{C})$'s such that some boundary portion of $\mathcal{D}_a(\mathcal{C})$ can move only thanks to a corresponding volume variation of large part $\mathcal{B}_a$ and the values of the degrees of freedom that sit inside $\mathcal{D}_a(\mathcal{C})$ can vary only by exchanging microscopically conserved quantities (energy and/or matter) with the corresponding large part $\mathcal{B}_a$. Then a jump of system $\mathcal{S}$ from configuration $\mathcal{C}$ to another one $\mathcal{C}'$ is allowed only if $\mathcal{D}(\mathcal{C})$ and $\mathcal{D}(\mathcal{C}')$ differ by a displacement of some boundary portion of only one $\mathcal{D}_a(\mathcal{C})$ and by different values of the degrees of freedom inside $\mathcal{D}_a(\mathcal{C})$ and $\mathcal{D}_a(\mathcal{C}')$; then we use the notation $\mathcal{C}' \in \mathbb{F}_a(\mathcal{C})$. Moreover the corresponding jump of the microscopic configuration of $\mathcal{B}_a$ is such that conservation rules hold for the sum of the energies of $\mathcal{S}$ and $\mathcal{B}_a$, $\mathcal{E}(\mathcal{C}') + E'_a = \mathcal{E}(\mathcal{C}) + E_a$, for the sum of the volumes that they occupy, $v(\mathcal{C}') + V'_a = v(\mathcal{C}) + V_a$, and for the sum of the numbers of elementary constituents of species s that they contain, $n_s(\mathcal{C}') + N'_{a,s} = n_s(\mathcal{C}) + N_{a,s}$, where E_a, V_a and the $N_{a,s}$ are the values of the extensive parameters that characterize body $\mathcal{B}_a$ (and the prime denotes their values after the configuration jump). For instance if system $\mathcal{S}$ is a mobile diathermal thin solid wall separating a vessel in two parts filled with gases kept at different temperatures and pressures, then $\mathcal{S}$ can be viewed as made of two layers of constituents, each of which interacts respectively with body $\mathcal{B}_1$ and $\mathcal{B}_2$. Then an infinitesimal displacement of $\mathcal{S}$ such that the volume of $\mathcal{B}_1$ increases while that of

$\mathcal{B}_2$ decreases by the same absolute amount can be decomposed into two microscopic configuration jumps of the global system : in the first (second) jump only the layer in contact with $\mathcal{B}_2$ ($\mathcal{B}_1$) moves and the volume of $\mathcal{S}$ increases (decreases), and after two jumps the volume of $\mathcal{S}$ has retrieved its initial value.

Then the mesoscopic Markovian dynamics defined according to the prescription of subsection 2.2 is such that, when energy, volume and matter are exchanged, the transition rates obey the microcanonical detailed balance

$$\frac{W(\mathcal{C}', E'_a, V'_a, \{N'_{a,s}\} \leftarrow \mathcal{C}, E_a, V_a, \{N_{a,s}\})}{W(\mathcal{C}, E_a, V_a, \{N_{a,s}\} \leftarrow \mathcal{C}', E'_a, V'_a, \{N'_{a,s}\})} = \frac{\Omega_a(E'_a, V'_a, \{N'_{a,s}\})}{\Omega_a(E_a, V_a, \{N_{a,s}\})} \tag{6.1}$$

with

$$\frac{\Omega_a(E'_a, V'_a, \{N'_{a,s}\})}{\Omega_a(E_a, V_a, \{N_{a,s}\})} \equiv e^{S_a^B(E'_a, V'_a, \{N'_{a,s}\}) - S_a^B(E_a, V_a, \{N_{a,s}\})}, \tag{6.2}$$

where $\Omega_a(E_a, V_a, \{N_{a,s}\})$ is the number of configurations (or microstates) of large part $\mathcal{B}_a$ when it is isolated. In the following, in the spirit of the notations used by Callen [4] we denote by $X_i^{(a)}$ the extensive macroscopic parameters of $\mathcal{B}_a$, namely in the generic case

$$X_0^{(a)} \equiv E_a, \quad X_1^{(a)} \equiv V_a, \quad X_{1+s}^{(a)} = N_{a,s}. \tag{6.3}$$

The total number of elementary constituents in $\mathcal{B}_a$ is $N_a = \sum_s N_{a,s}$. The microscopic conservation rules that are associated with (6.1) read

$$X_i'^{(a)} - X_i^{(a)} = -\left[x_i(\mathcal{C}') - x_i(\mathcal{C})\right] \quad \text{if} \quad \mathcal{C}' \in \mathbb{F}_a(\mathcal{C}), \tag{6.4}$$

with the same notations for system $\mathcal{S}$ as those introduced in (6.3) for the macroscopic bodies.

If the bodies $\mathcal{B}_a$ are so large that they can be described by a thermodynamic limit, then, in a transient regime where the macroscopic extensive parameters $X_i^{(a)}$'s have negligible relative variations, $\mathcal{B}_a$ remains at thermodynamic equilibrium. The thermodynamic entropy per elementary constituent, S_a^{TH}/N_a, coincides with the thermodynamic limit of the Boltzmann entropy per elementary constituent, S_a^B/N_a. The intensive thermodynamic parameter $F_i^{(a)}$ conjugate to the extensive quantity $X_i^{(a)}$ by $F_i^{(a)} \equiv \partial S_a^{TH}/\partial X_i^{(a)}$ (with $X_i^{(a)}$ defined in (6.3)) is given by

$$F_0^{(a)} \equiv \beta_a, \quad F_1^{(a)} \equiv \beta_a P_a, \quad F_{1+s}^{(a)} = -\beta_a \mu_{a,s}, \tag{6.5}$$

where β_a is the inverse thermodynamic temperature, P_a the thermodynamic pressure and $\mu_{a,s}$ the chemical potential of species s in $\mathcal{B}_a$. Then from the relations (6.1) and (6.2), one can show, as in subsection 2.3, that in the transient regime the transition rates obey the modified detailed balance (2.26) where $\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C})$ is opposite to the infinitesimal variation at fixed intensive parameters of the thermodynamic entropy of the reservoir $\mathcal{B}_a$ that causes the jump from $\mathcal{C}$ to $\mathcal{C}'$ under the conservation rules (6.4). Therefore, if $\mathcal{C}' \in \mathbb{F}_a(\mathcal{C})$

$$\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C}) = \beta_a[\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})] + \beta_a P_a[v(\mathcal{C}') - v(\mathcal{C})] - \beta_a \sum_s \mu_{a,s}[n_s(\mathcal{C}') - n_s(\mathcal{C})]. \tag{6.6}$$

In the case of pure thermal contact, the volume of system $\mathcal{S}$ does not vary, $v(\mathcal{C}) = v(\mathcal{C}')$, and there is no matter exchange ; then $\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C}) = \beta_a[\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})]$ coincides with β_a times the opposite of the variation of the internal energy of $\mathcal{B}_a$, which is equal in that case to the heat given by $\mathcal{B}_a$ at constant volume. If system $\mathcal{S}$ and macroscopic body $\mathcal{B}_a$ are compressible then, during energy exchanges such that the thermodynamic pressure P_a of $\mathcal{B}_a$ remains fixed, the variation of the internal energy of the macroscopic body $\mathcal{B}_a$ involves both heat and pressure work ; then $\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C}) = \beta_a[\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})] + \beta_a P_a[v(\mathcal{C}') - v(\mathcal{C})]$ coincides with β_a times the opposite of the variation of the enthalpy of $\mathcal{B}_a$, which is equal in that case to the heat given by $\mathcal{B}_a$ at constant pressure. Moreover system $\mathcal{S}$ may receive work from some conservative external forces $f_b^{\rm ext}$ (such as gravitational or electrical fields), each of which causes the variation of some global coordinate

$Z_b(\mathcal{C})$ of the system (such as its mass center or its electrical barycenter) but does not act upon the macroscopic bodies. In that case the mechanical energy $\mathcal{E}(\mathcal{C})$ of system $\mathcal{S}$ in configuration $\mathcal{C}$ is the sum of its internal energy $\mathcal{E}_{\rm int}(\mathcal{C})$ and an external potential energy $\mathcal{E}_{\rm ext}(\mathcal{C})$. In all these situations the distances between the possible energy levels $\mathcal{E}$ are time-independent, and during the evolution only the occupation of the energy levels is modified, contrarily to the case where a time-dependent force acts on the system by changing the spacing between energy levels.

In order to handle compact notations for the successive variations $\delta_{\rm exch}S$ in the course of a history of system $\mathcal{S}$, let us introduce the notation $\delta\chi_i^{(a)}(\mathcal{C}' \leftarrow \mathcal{C})$ for the quantity with index i received by the system from reservoir $\mathcal{B}_a$ when the system jumps from $\mathcal{C}$ to $\mathcal{C}'$, with the same convention as in definition (2.23), namely

$$\begin{cases} \delta\chi_i^{(a)}(\mathcal{C}' \leftarrow \mathcal{C}) = x_i(\mathcal{C}') - x_i(\mathcal{C}) & \text{if} \quad \mathcal{C}' \in \mathbb{F}_a(\mathcal{C}) \\ \delta\chi_i^{(a)}(\mathcal{C}' \leftarrow \mathcal{C}) = 0 & \text{otherwise} \end{cases} . \tag{6.7}$$

(In the case where $\mathcal{C}' \in \mathbb{F}_a(\mathcal{C})$ but where in fact reservoir $\mathcal{B}_a$ does not exchange quantity with index i in the jump from $\mathcal{C}$ to $\mathcal{C}'$, $x_i(\mathcal{C}') - x_i(\mathcal{C}) = 0$). Then the exchange entropy variation in a jump takes the form

$$\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C}) = \sum_i \sum_{a \in R(i)} F_i^{(a)} \delta\chi_i^{(a)}(\mathcal{C}' \leftarrow \mathcal{C}), \tag{6.8}$$

where the first sum runs over the indices i of the extensive quantities defined in (6.3), the second sum runs over the indices of the reservoirs that indeed can exchange quantity i with the system and $R(i)$ denotes the set of the latter reservoirs.

6.2 Consequences of MDB at finite time

In the form (2.26) where it involves the microscopic exchange entropy variation $\delta_{\rm exch}S(\mathcal{C}' \leftarrow \mathcal{C})$ associated with a jump from configuration $\mathcal{C}$ to configuration $\mathcal{C}'$, the modified detailed balance entails the symmetry (4.6) between the probabilities for time-reversed histories. At a more mesoscopic level, let us compare the probability of all evolutions from configuration $\mathcal{C}_0$ to configuration $\mathcal{C}_f$, in the course of which the system receives given cumulative quantities with index i $\mathcal{X}_i^{(a)} = \sum \delta\chi_i^{(a)}$ from each reservoir $\mathcal{B}_a$, and the probability of the reversed evolutions, namely evolutions from $\mathcal{C}_f$ to $\mathcal{C}_0$ where the cumulative quantities are $-\mathcal{X}_i^{(a)}$'s. The symmetry (4.8) written in the case of thermal contact takes the following form in the generic case

$$\frac{P\left(\mathcal{C}_f | \{\mathcal{X}_i^{(a)}\}; t | \mathcal{C}_0\right)}{P\left(\mathcal{C}_0 | \{-\mathcal{X}_i^{(a)}\}; t | \mathcal{C}_f\right)} = e^{-\Delta_{\rm exch}S(\{\mathcal{X}_i^{(a)}\})}, \tag{6.9}$$

with, according to (6.8),

$$\Delta_{\rm exch}S(\{\mathcal{X}_i^{(a)}\}) = \sum_i \sum_{a \in R(i)} F_i^{(a)} \mathcal{X}_i^{(a)}. \tag{6.10}$$

In (6.9) and in the formulæ derived from it in the following, $\mathcal{X}_i^{(a)}$ occurs only if reservoir $\mathcal{B}_a$ indeed exchanges quantity with index i. The ratio of probabilities in (6.9) does not explicitly depend on the initial and final configurations $\mathcal{C}_0$ and $\mathcal{C}_f$. There is only an implicit dependence on these configurations through the conservation rules that the cumulative exchange quantities $\mathcal{X}_i^{(a)}$ for a given history must satisfy, namely $\sum_{a \in R(i)} \mathcal{X}_i^{(a)} = x_i(\mathcal{C}_f) - x_i(\mathcal{C}_0)$.

At the macroscopic level, namely when only the exchanges of extensive quantities with the reservoirs are measured, there appears a symmetry for transient regimes where the system is initially prepared in some equilibrium state with a fixed intensive parameter F_i^0 for each configuration observable x_i and then is suddenly put into contact with reservoirs with thermodynamic parameters $F_i^{(a)}$'s that drive the system into a non-equilibrium state. The symmetry involves the excess exchange entropy variation $\Delta_{\rm exch}^{{\rm excs},\{F_i^0\}}S$, defined as the difference between the exchange

entropy variation under the non-equilibrium external constraints $F_i^{(a)}$'s, defined in (6.10), and the corresponding variation under the equilibrium conditions where for all reservoirs that exchange quantity with index i the thermodynamic parameter has the same value F_i^0,

$$\Delta_{\rm exch}^{{\rm excs},\{F_i^0\}}(\{\mathcal{X}_i^{(a)}\}) \equiv \Delta_{\rm exch} S(\{\mathcal{X}_i^{(a)}\}) - \sum_i F_i^0 \sum_{a\in R(i)} \mathcal{X}_i^{(a)}. \tag{6.11}$$

When the system is prepared in the equilibrium state with probability distribution $P_{\rm eq}(\{F_i^0\})$ and put into contact with reservoirs with thermodynamic parameters $F_i^{(a)}$'s at the initial time of the measurements of exchanged quantities, the joint probability for the cumulative exchange quantities $\mathcal{X}_i^{(a)}$ obeys the symmetry relation at any finite time,

$$\frac{P_{P_{\rm eq}(\{F_i^0\})}\left(\{\mathcal{X}_i^{(a)}\}\right)}{P_{P_{\rm eq}(\{F_i^0\})}\left(\{-\mathcal{X}_i^{(a)}\}\right)} = e^{-\Delta_{\rm exch}^{{\rm excs},\{F_i^0\}} S(\{\mathcal{X}_i^{(a)}\})}. \tag{6.12}$$

As a consequence, the excess exchange entropy variation $\Delta_{\rm exch}^{{\rm excs},\{F_i^0\}} S$ obeys the symmetry relation at any finite time, or "detailed fluctuation relation",

$$\frac{P_{P_{\rm eq}(\{F_i^0\})}\left(\Delta_{\rm exch}^{{\rm excs},\{F_i^0\}} S\right)}{P_{P_{\rm eq}(\{F_i^0\})}\left(-\Delta_{\rm exch}^{{\rm excs},\{F_i^0\}} S\right)} = e^{-\Delta_{\rm exch}^{{\rm excs},\{F_i^0\}} S}. \tag{6.13}$$

The latter relation itself entails the identity, or "integral fluctuation relation",

$$\langle e^{\Delta_{\rm exch}^{{\rm excs},\{F_i^0\}} S}\rangle_{P_{\rm eq}(\{F_i^0\})} = 1. \tag{6.14}$$

6.3 Consequences of MDB in the infinite-time limit

A generalization of the argument in subsubsection 5.1.1 shows that the symmetry relation (6.9) enforced by the MDB at finite time leads to the existence of lower and upper bounds for the ratio between the finite-time probability to measure cumulative quantities with values $\{\mathcal{X}_i^{(a)}\}$ and the corresponding probability to measure the opposite values $\{-\mathcal{X}_i^{(a)}\}$, when the system is in its stationary state. The finite-time inequalities (5.1) become in the generic case

$$\frac{P_{\rm st}^{\rm min}}{P_{\rm st}^{\rm max}} \leq \frac{P_{\rm st}\left(\{\mathcal{X}_i^{(a)}\};t\right)}{P_{\rm st}\left(\{-\mathcal{X}_i^{(a)}\};t\right) e^{-\Delta_{\rm exch} S(\{\mathcal{X}_i^{(a)}\})}} \leq \frac{P_{\rm st}^{\rm max}}{P_{\rm st}^{\rm min}}. \tag{6.15}$$

As a consequence, the dimensionless exchange entropy variation $\Delta_{\rm exch} S$ given by (6.10) obeys the fluctuation relation (1.11).

As for the "characteristic function" of the extensive exchanged quantities $\mathcal{X}_i^{(a)}$'s in the stationary state, namely $\langle e^{\sum_a \lambda_i^{(a)} \mathcal{X}_i^{(a)}(t)}\rangle_{\rm st}$, the symmetry (6.9)-(6.10) arising from the MDB entails that the characteristic function obeys an inequality similar to (5.45). As a consequence, the generating function of the infinite-time limit of the joint cumulants per unit time obeys a symmetry which is a generalization of (5.46)

$$\alpha_{\{\mathcal{X}_i^{(a)}\}}(\{\lambda_i^{(a)}\}) = \alpha_{\{\mathcal{X}_i^{(a)}\}}(\{-F_i^{(a)} - \lambda_i^{(a)}\}). \tag{6.16}$$

However, because of the conservation laws for the microscopic quantities exchanged with the reservoirs, the joint cumulants of the $\mathcal{X}_i^{(a)}$'s per unit time are not independent in the infinite-time limit. For instance, if $R(i)$ denotes the set of reservoirs which exchange the quantity with index i, because of the conservation law for every species of exchanged quantity, $\sum_{a\in R(i)} \mathcal{X}_i^{(a)}$

is equal to the difference between the values of some observable of the system in the final and initial configurations of the evolution. When the system has a finite number of configurations, $\sum_{a\in R(i)} \mathcal{X}_i^{(a)}$ is bounded and the mean currents $J_i^{(a)}$'s defined as $J_i^{(a)} = \lim_{t\to+\infty}\langle \mathcal{X}_i^{(a)}(t)\rangle/t$ are related by

$$\sum_{a\in R(i)} J_i^{(a)} = 0. \tag{6.17}$$

Other relations may arise from other microscopic conservation rules determined by the specific forms of the transitions rates. We consider the generic (but not universal) case where there exists a set of cumulative quantities Y_γ's, which are linear combinations of the $\mathcal{X}_i^{(a)}$'s, but less numerous than the $\mathcal{X}_i^{(a)}$'s, and a set of parameters $\mathcal{F}_\gamma$'s such that $\sum_i \sum_{a\in R(i)} F_i^{(a)} \mathcal{X}_i^{(a)} + \sum_\gamma \mathcal{F}_\gamma Y_\gamma$ is bounded by a constant independent of time. With these assumptions the expression for the exchange entropy flow in the stationary state takes the generic form

$$\left.\frac{d_{\rm exch}S}{dt}\right|_{\rm st} = \sum_i \sum_{a\in R(i)} F_i^{(a)} J_i^{(a)} = -\sum_\gamma \mathcal{F}_\gamma J_\gamma^\star, \tag{6.18}$$

where the mean currents $J_\gamma^\star$'s, defined as $J_\gamma^\star \equiv \lim_{t\to+\infty}\langle Y_\gamma(t)\rangle/t$, are less numerous than the $J_i^{(a)}$'s. Moreover, these assumptions entail that the symmetry (6.9)-(6.10) leads to the inequality

$$m \leq \frac{P\left(\mathcal{C}_f|\{Y_\gamma\};t|\mathcal{C}_0\right)}{P\left(\mathcal{C}_0|\{-Y_\gamma\};t|\mathcal{C}_f\right) e^{\sum_\gamma \mathcal{F}_\gamma Y_\gamma}}, \leq M, \tag{6.19}$$

where m and M are constants. As a consequence, the characteristic function for the Y_γ's obeys an inequality similar to (5.45) and subsequently the generating function of the infinite-time limits of their joint cumulants per unit time obeys a symmetry which is a generalization of (5.50),

$$\alpha_{\{Y_\gamma\}}(\{\lambda_\gamma\};\{\mathcal{F}_\gamma\}) = \alpha_{\{Y_\gamma\}}(\{-\mathcal{F}_\gamma - \lambda_\gamma\};\{\mathcal{F}_\gamma\}). \tag{6.20}$$

In the framework of graph theory (see subsubsection 3.3.5) Andrieux and Gaspard [47] have shown an analogous symmetry for dimensionless cumulative quantities Y_γ's, where each Y_γ is defined as the sum of the cumulative probability currents through the chords of all cycles in the graph that have the same affinity A_γ and where the $\mathcal{F}_\gamma$'s are replaced by the cycle affinities A_γ's. A similar result is also obtained in Ref.[81]. We also notice that an example of the symmetry relation (6.20) is derived in another context in Ref.[82].

6.4 Generalized Einstein-Green-Kubo relations for several independent currents

In standard fluctuation-dissipation Einstein-Green-Kubo formulae, valid near equilibrium, the relevant out-of-equilibrium quantity that is related to equilibrium fluctuations is the Onsager coefficient $L_{\alpha\gamma}$ introduced in the phenomenological thermodynamics of irreversible phenomena : "thermodynamic fluxes" $J_\gamma^\star$'s and "thermodynamic forces" $\mathcal{F}_\gamma$'s are defined by pairs from the expression of the entropy production rate $d_{\rm int}S/dt|_{\rm st} = -d_{\rm exch}S/dt|_{\rm st}$ given by (6.18), where the $J_\gamma^\star$'s are independent currents between reservoirs, and the Onsager coefficient is defined as

$$L_{\alpha\gamma} \equiv \left.\frac{\partial J_\alpha^\star}{\partial \mathcal{F}_\gamma}\right|_{\{\mathcal{F}_{\gamma'}=0\}}. \tag{6.21}$$

The generic statement of Einstein-Green-Kubo relations reads

$$L_{\alpha\gamma} = \frac{1}{2}\lim_{t\to+\infty} \frac{\langle Y_\alpha(t)Y_\gamma(t)\rangle_{\rm eq} - \langle Y_\alpha(t)\rangle_{\rm eq}\langle Y_\gamma(t)\rangle_{\rm eq}}{t}. \tag{6.22}$$

Therefore in the case where the non equilibrium stationary state involves several independent stationary currents, the symmetry $\langle Y_\alpha(t)Y_\gamma(t)\rangle_{\rm eq} = \langle Y_\gamma(t)Y_\alpha(t)\rangle_{\rm eq}$ allows to retrieve the phenomenological Onsager symmetry for the off-diagonal Onsager coefficients, namely $L_{\alpha\gamma} = L_{\gamma\alpha}$.

Far from equilibrium generalized Einstein-Green-Kubo relations can be derived from the symmetry (6.20) of the generating function for the infinite-time cumulants per unit time. The generalization of the combinatorics considerations of subsection 5.3 is straightforward. Suppose the independent currents are indexed by $\gamma \in \Gamma$ so that instead of a single parameter λ one deals with a collection $\underline{\lambda} \equiv (\lambda_\gamma)_{\gamma\in\Gamma}$. In the same way, one has a collection $\underline{\mathcal{F}} \equiv (\mathcal{F}_\gamma)_{\gamma\in\Gamma}$. However, one considers a single function α and we rewrite (6.20) as

$$\alpha(\underline{\lambda};\underline{\mathcal{F}}) = \alpha(-\underline{\mathcal{F}} - \underline{\lambda};\underline{\mathcal{F}}). \tag{6.23}$$

One could redo all the derivations performed in subsection 5.3. We content to express all odd cumulants in terms of the even ones. Write the expansion in powers of the λ_γ's in a compact way as

$$\alpha(\underline{\lambda};\underline{\mathcal{F}}) \equiv \sum_{\underline{p}\geq\underline{0}} \frac{1}{\underline{p}!}\underline{\lambda}^{\underline{p}}\overline{\kappa}^{[\underline{p}]}(\underline{\mathcal{F}}), \tag{6.24}$$

where the summation is over Γ-tuples $\underline{p} \equiv (p_\gamma)_{\gamma\in\Gamma}$ of non-negative integers, $\underline{p}! \equiv \prod_{\gamma\in\Gamma} p_\gamma!$ and $\underline{\lambda}^{\underline{p}} \equiv \prod_{\gamma\in\Gamma} \lambda_\gamma^{p_\gamma}$. Note that

$$\overline{\kappa}^{[\underline{p}]}(\underline{\mathcal{F}}) \equiv \left.\frac{\partial^{|\underline{p}|}\alpha(\underline{\lambda};\underline{\mathcal{F}})}{\partial\underline{\lambda}^{\underline{p}}}\right|_{\underline{\lambda}=\underline{0}}, \tag{6.25}$$

where $|\underline{p}| \equiv \sum_{\gamma\in\Gamma} p_\gamma$ and $\partial\underline{\lambda}^{\underline{p}} \equiv \prod_{\gamma\in\Gamma} \partial\lambda_\gamma^{p_\gamma}$. We expand the symmetry relation written in the form

$$\alpha(\underline{\lambda};\underline{\mathcal{F}}) = \frac{1}{1+e^{-\underline{\mathcal{F}}\frac{\partial}{\partial\underline{\lambda}}}}\left[\alpha(\underline{\lambda};\underline{\mathcal{F}}) + \alpha(-\underline{\lambda};\underline{\mathcal{F}})\right], \tag{6.26}$$

where

$$\underline{\mathcal{F}}\frac{\partial}{\partial\underline{\lambda}} \equiv \sum_{\gamma\in\Gamma} \mathcal{F}_\gamma \frac{\partial}{\partial\lambda_\gamma}. \tag{6.27}$$

From

$$\frac{1}{1+e^{-\underline{\mathcal{F}}\frac{\partial}{\partial\underline{\lambda}}}} = \frac{1}{2}\left(1 + \sum_{k=0}^{+\infty} d_k \sum_{\underline{p}\geq\underline{0},|\underline{p}|=2k+1} \frac{(2k+1)!}{\underline{p}!}\underline{\mathcal{F}}^{\underline{p}}\frac{\partial^{2k+1}}{\partial\underline{\lambda}^{\underline{p}}}\right) \tag{6.28}$$

one infers that, for $\underline{n}$ such that $|\underline{n}|$ is odd

$$\overline{\kappa}^{[\underline{n}]}(\underline{\mathcal{F}}) = \sum_{k=0}^{+\infty} d_k \sum_{\underline{p}\geq\underline{0},|\underline{p}|=2k+1} \frac{(2k+1)!}{\underline{p}!}\underline{\mathcal{F}}^{\underline{p}}\,\overline{\kappa}^{[\underline{n}+\underline{p}]}(\underline{\mathcal{F}}). \tag{6.29}$$

Again, these relations can be expanded in powers of $\underline{\mathcal{F}}$. The corresponding coefficients of the powers in $\mathcal{F}$ are the non-linear response coefficients. The relations between the latter non-linear response coefficients are derived by another method in [61]. For instance relations analogous to (5.74d) relate various non-linear response coefficients of the cumulants per unit time caused by a variation in thermodynamic forces $\mathcal{F}_\alpha$'s to some equilibrium cumulant per unit time, which is symmetric under permutations of the associated cumulative quantities. The latter symmetry is at the origin of the Onsager reciprocity relation in the case of $L_{\alpha\gamma}$ [1] (see definition (6.21)) and of generalized symmetry relations for higher-order response coefficients of cumulants per unit time as noted by Andrieux and Gaspard [83, 61].

A Property of the coarse grained dynamics

In the present appendix we derive the property (2.5) valid over a period of the ergodic deterministic microscopic dynamics $\mathcal{T}$ when $\mathcal{T}$ respects both the conservation of $E_{\rm dec}$ and the interaction pattern specified when the conservation law (2.2) was introduced.

Though we believe that a general development could be pursued, we prefer to concentrate on a specific model at this point. The system is made of two large parts and a small one, which is reduced to two Ising spins $\sigma_1, \sigma_2 = \pm 1$, each one directly in contact with one of the large parts. So a configuration C can be written as $C = (C_1, \sigma_1, \sigma_2, C_2)$. We assume that, when the small part is isolated, its energy $\mathcal{E}(\sigma_1, \sigma_2)$ does not describe independent spins. Moreover, the microscopic dynamics $\mathcal{T}$ conserves the energy $E_{\rm dec}(C)$ where the interactions between parts is neglected, namely $E_{\rm dec}(C) \equiv E_1(C_1) + E_2(C_2) + \mathcal{E}(\sigma_1, \sigma_2)$. The remnant of interactions between the parts is embodied in the following restrictions. For any C, $\mathcal{T}(C)$ is obtained from C by one of the following operations :

- (I) a flip of spin σ_1 together with a change in C_1 and a possible change in C_2 such that $E_1(C_1) + \mathcal{E}(\sigma_1, \sigma_2)$ and $E_2(C_2)$ both remain constant (i.e. the energy needed to flip the spin σ_1 entirely comes from or goes to large part 1).

- (II) a flip of spin σ_2 together with a change in C_2 and a possible change in C_1 such that $E_1(C_1)$ and $E_2(C_2) + \mathcal{E}(\sigma_1, \sigma_2)$ both remain constant (i.e. the energy needed to flip the spin σ_2 entirely comes from or goes to large part 2).

- (III) a change in C_1 and/or C_2 but no flip of σ_1 or σ_2, such that $E_1(C_1)$ and $E_2(C_2)$ both remain constant, as well as $\mathcal{E}(\sigma_1, \sigma_2)$.

In order to build an effective mesoscopic dynamics we just keep track of $(E_1, \sigma_1, \sigma_2, E_2)$ as a function of time. Therefore during the time evolution of a given configuration C of the whole system we concentrate only on time steps at which a change of type (I) or (II) occurs, namely when either spin σ_1 or spin σ_2 is flipped with known corresponding variations in E_1 and E_2. We do not follow precisely the changes (III) that modifies the configurations of large parts without changing the energy of any part (either E_1, E_2 or $\mathcal{E}(\sigma_1, \sigma_2)$). The possible changes are

$$(E_1, \sigma_1, \sigma_2, E_2) \to (E_1', -\sigma_1, \sigma_2, E_2) \quad \text{with} \quad E_1' = E_1 + \mathcal{E}(\sigma_1, \sigma_2) - \mathcal{E}(-\sigma_1, \sigma_2) \quad \text{type (I)} \tag{A.1}$$

and

$$(E_1, \sigma_1, \sigma_2, E_2) \to (E_1, \sigma_1, -\sigma_2, E_2') \quad \text{with} \quad E_2' = E_2 + \mathcal{E}(\sigma_1, \sigma_2) - \mathcal{E}(\sigma_1, -\sigma_2) \quad \text{type (II)} \tag{A.2}$$

Starting from some initial configuration $(E_1^0, \sigma_1^0, \sigma_2^0, E_2^0)$, some set of possible $(E_1, \sigma_1, \sigma_2, E_2)$ will be visited during the time evolution over a period of $\mathcal{T}$, and it is useful to view this set as the vertices of a graph, whose edges connect two vertices if the system can jump from one to the other by a single change of type (I) or (II). This graph can be chosen to be unoriented because a transformation of type (I) or of type (II) is its own inverse. Then a trajectory over a period of the underlying deterministic dynamics corresponds to a closed walk on this graph, which summarizes the coarse graining due to the macroscopic description of the large parts. During a period of the microscopic dynamics the closed walk on the graph goes through each edge a number of times.

The main observation is that, since energy $\mathcal{E}(\sigma_1, \sigma_2)$ does not describe independent spins, the graph has the topology of a segment. Indeed, the graph is connected by construction, but each vertex has at most two neighbors. So the graph is either a segment or a circle. Let us suppose that it is a circle. Then one can go from $(E_1^0, \sigma_1^0, \sigma_2^0, E_2^0)$ to itself by visiting each edge exactly once, i.e. by an alternation of moves of type (I) and (II). After two steps both spins in the small part are flipped, so the total length of the circle is a multiple of 4. But after 4 steps a definite amount of energy has been transferred between the two large parts, namely the energy in the first large part has changed by $\mathcal{E}(\sigma_1, \sigma_2) - \mathcal{E}(-\sigma_1, \sigma_2) + \mathcal{E}(-\sigma_1, -\sigma_2) - \mathcal{E}(\sigma_1, -\sigma_2)$ and a trivial computation shows that this cannot vanish unless $\mathcal{E}(\sigma_1, \sigma_2)$ describes two independent spins, a possibility which has been discarded. This excludes the circle topology.

Since the graph is a segment, any closed walk on the graph traverses a given edge the same number of times in one direction and in the other one. As a consequence, during a period of

the microscopic dynamics $\mathcal{T}$, the motion induced by $\mathcal{T}$ on the graph is such that the number of transitions of type (I) $(E_1, \sigma_1, \sigma_2, E_2) \to (E_1', -\sigma_1, \sigma_2, E_2)$ is equal to the number of their inverse transitions $(E_1', -\sigma_1, \sigma_2, E_2) \to (E_1, \sigma_1, \sigma_2, E_2)$, where the relation between E_1' and E_1 is given in (A.1). The same considerations apply for transition rates associated with the flipping of σ_2. The result is summarized in (2.5).

For a more general discrete system, the mesoscopic time evolution can again be characterized by the time steps where the small system variables are flipped while the changes in the large parts are only macroscopically described by their net energies. An analogous graph can be constructed but it can be much harder to analyze its topology, which is the crucial knowledge needed to exploit the consequences of ergodicity. These consequences are most stringent for a tree. There is no reason a priori why the graph should be a tree, but what this implies, namely that ratios for the transition rates involving the small part and a large part are given by ratios of energy level degeneracies in the large part, is physically quite appealing.

B The Markovian approximation

This appendix is a short digression on mathematics. The aim is to briefly recall a trick allowing to replace a fixed sequence of symbols by a random one with "analogous" statistical properties, and then to combine this trick with a coarse graining procedure.

B.1 Discrete time stochastic approximation

Suppose $w \equiv x_1 x_2 \cdots x_N$ is a given finite sequence of elements of a finite set S. It is convenient to assume that this sequence is periodic, i.e. that $x_{N+1} \equiv x_1$. For $x \in S$, set $N_x \equiv \#\{i \in [1, N], x_i = x\}$, i.e. N_x is the number of occurrences of x in the sequence w. There is no loss of generality in assuming that $N_x \neq 0$ for every $x \in S$, should it lead to consider a smaller S. For $x, x' \in S$ set $N_{xx'} \equiv \#\{i \in [1, N], x_i = x, x_{i+1} = x'\}$, i.e. $N_{xx'}$ is the number of occurrences of the pattern xx' in the sequence w. Notice that we accept that $x = x'$ in this definition. Of course, we could look at the occurrence of more general patterns. By definition, we have $\sum_{x' \in S} N_{xx'} = \sum_{x' \in S} N_{x'x} = N_x$. The result we want to recall is the following.

There is a single time-homogeneous irreducible Markov matrix on S that fulfills the following two requirements. First, in the stationary state of the corresponding discrete-time stochastic evolution, the probability $P_{\rm st}(x', i+1; x, i)$ that a sample $\widehat{x}$ takes the value x at time i and the value x' at time $i+1$[3] is equal to the frequency of the pattern xx' in the sequence w, namely

$$P_{\rm st}(x', i+1; x, i) = \frac{N_{xx'}}{N}. \tag{B.1}$$

Second, the stationary probability that the random variable $\widehat{x}$ takes the value x at any time i is equal to the frequency of x in the sequence w, namely

$$P_{\rm st}(x) = \frac{N_x}{N}. \tag{B.2}$$

In fact (B.2) is a consequence of (B.1), because of the relations $P_{\rm st}(x) = \sum_{x' \in S} P_{\rm st}(x', i+1; x, i)$ and $\sum_{x' \in S} N_{xx'} = N_x$. To say it in words, there is a unique time-homogeneous irreducible Markov chain whose stationary statistics for patterns of length 1 or 2 is the same as the corresponding statistics in w.

The proof is elementary. By the Markov property, the Markov matrix element from x to x', denoted by $T(x' \leftarrow x)$, must satisfy the relation $P_{\rm st}(x', i+1; x, i) = T(x' \leftarrow x) P_{\rm st}(x)$, so the only candidate is

$$T(x' \leftarrow x) = \frac{N_{xx'}}{N_x}. \tag{B.3}$$

[3] In the whole paper we use the convention that the evolution (here given by the joint probability $P_{\rm st}(x', i+1; x, i)$ or the transition matrix T) is written from the right to the left as in quantum mechanics.

Conversely, the corresponding matrix is obviously a Markov matrix ($\sum_{x'\in S} T(x' \leftarrow x) = 1$ since $\sum_{x'\in S} N_{xx'} = N_x$), and it is irreducible, because, as w contains all elements of S, transitions within w allow to go from every element of S to every other one. Checking that for this Markov matrix the stationary measure, namely the solution of $\sum_{x\in S} T(x' \leftarrow x)P_{\rm st}(x) = P_{\rm st}(x')$, is given by (B.2) boils down to the identity $\sum_{x\in S} N_{xx'} = N_{x'}$ recalled above, and then the two-point property (B.1) follows. This finishes the proof.

In the very specific case where $x_1, x_2, \cdots, x_N$ are all distinct, i.e. $|S|$, the cardinal of S, is equal to N, $N_x = 1$ for all $x \in S$ and $N_{xx'} = \delta_{x',x_{i+1}}$ where i is such that $x_i = x$. Then the only randomness lies in the choice of the initial distribution, and each trajectory of the Markov chain reproduces w up to a translation of all indices. If N is large and $|S|$ is $\sim N$, slightly weaker but analogous conclusions survive. A more interesting case is when $|S| \ll N$ by many orders of magnitude as discussed in next subsection.

This trick has been used for instance to write a random text "the Shakespeare way" by computing the statistics of sequences of two words in one of his books.

By definition, the Markovian approximation reproduces the statistics of the original sequence only for length 1 and length 2 patterns. It is a delicate issue to decide whether or not it also does a reasonnable job for other patterns. For instance, the random Shakespeare book certainly looks queer. Various physical but heuristic arguments suggest that, for the kind of sequences w relevant for this work, the Markov approximation is quite good, but we shall not embark on that.

B.2 Continuous time approximation

By the very same argument, there is a time-homogeneous irreducible Markov transition matrix on S (unique up to a time scale τ) such that, in the stationary state of the corresponding continuous-time stochastic evolution, the expected number of transitions from x to $x' \neq x$ per unit time is $N_{xx'}/\tau$. The formula one finds for the transition rate is

$$W(x' \leftarrow x) = \frac{N_{xx'}}{\tau N_x} \text{ for } x \neq x'. \tag{B.4}$$

The continuous-time approximation becomes more natural in the case when $|S|$, the cardinal of S, is such that $|S| \ll N$, while for all x $N_{xx} \sim N_x$ and for all $x \neq x'$ $N_{xx'} \ll N_x$, which means that transitions are rare and most of the time x follows x in the sequence w. Again, one can argue that for the kind of sequences w relevant for this work, this is guaranteed by physics. Then taking τ (in some macroscopic time unit) of the order of the largest value of the $N_{xx'}/N_x$, $x \neq x'$, one ends with a Markov transition matrix with elements of order unity (in some macroscopic inverse time unit), and $t_i = \tau i$ can be taken as the physical macroscopic time.

To summarize, in this work, we shall systematically associate to certain sequences w a continuous-time Markov process and exploit properties of w to constraint the structure of $W(x' \leftarrow x)$.

A natural application of the above ideas is to the case where the sequence w arises from some coarse graining procedure. One starts from a sequence $\omega = \xi_1\xi_2\cdots\xi_N$ where $\xi_1, \xi_2, \cdots, \xi_N$ belong to a set Σ that is so large that ω cannot be stored, and that only some of its features can be kept. Say we partition $\Sigma \equiv \cup_{x\in S}\Sigma_x$ where S is of reasonable size. Then all we keep of ω is $w = x_1x_2\cdots x_N$ where, for $i \in [1, N]$, x_i is substituted for ξ_i when $\xi_i \in \Sigma_{x_i}$. In the applications we have in mind, Σ is an N-element set, i.e. all terms in the sequence ω are distinct. In that case, even if ω is constructed in a perfectly deterministic way, by saying who follows who in the sequence, such a description is unavailable on w, and w may well look quite random, so the Markov chain approximation is worth a try. In fact, $|S| \ll N$ and we shall take as a physical input that transitions are rare, so that the (continuous-time) Markov process is an excellent approximation to the (discrete time) Markov chain.

C Microcanonical detailed balance and time reversal invariance in Hamiltonian dynamics

In view of comparison with the approach developed for discrete variables, we rederive the microcanonical detailed balance (2.11) for the probability distribution of some mesoscopic variables, within the framework of statistical ensemble theory, in the case where a microscopic configuration of the full system is described by continuous variables and the system has a microscopic deterministic dynamics whose Hamiltonian H, independent of time, is an even function of momenta. Moreover the coordinate of the system in phase space is denoted by ξ (by analogy with the notation in subsubsection 2.2.1), and $x[\xi]$ is a set of mesoscopic variables defined from the microscopic coordinate ξ and which are even functions of momenta. (Similar arguments can be found in derivations which rely on different assumptions in Refs.[84, 58].)

From the viewpoint of the statistical ensemble theory, the initial position of the system in phase space is not known. It is assumed to be uniformly randomly distributed in the energy shell $E = H[\xi]$, where E is the value of the energy of the full system at some initial time t_0. In other words, the initial probability distribution of ξ, $P_0(\xi)$, is such that

$$d\xi\, P_0(\xi) \equiv \frac{d\xi\, \delta\left(H[\xi] - E\right)}{\int d\xi'\, \delta\left(H[\xi'] - E\right)}, \tag{C.1}$$

where $d\xi$ is the Liouville measure in phase space. The choice of the uniform probability distribution in the energy shell for the initial position of the system in phase space is motivated by the fact that, at the microscopic level, it is the measure that is invariant under Hamiltonian dynamics. We recall the derivation of the latte property for the sake of completeness and in order to introduce notations used below.

The dynamics is invariant by time translation and we simply denote by $f_t(\xi)$ the position of the system in phase space at time $t_0 + t$ knowing that it is at position ξ at time t_0. (For the same reason, in the following the initial time t_0 is set equal to 0). Then the invariance of the Hamiltonian under the operation R that changes every momentum into its opposite implies that

$$[Rf_tRf_t](\xi) = \xi, \tag{C.2}$$

namely the microscopic dynamics is invariant under time reversal. Since the Hamiltonian is a constant of motion and the equations of motion conserve the infinitesimal volume in phase space, the probability $d\xi P_0(\xi)$ defined in (C.1) is also conserved under the microscopic evolution,

$$d(f_t[\xi])\, P_0(f_t[\xi]) = d\xi\, P_0(\xi). \tag{C.3}$$

When the initial position of the system in phase space is distributed according to P_0, the probability that at time t the set of mesoscopic variable x takes the value x_1, which is defined as

$$P_{P_0}(x_1, t) \equiv \int d\xi\, P_0(\xi)\, \delta\left(x[f_t(\xi)] - x_1\right), \tag{C.4}$$

is in fact independent of time (i.e. conserved by the microscopic dynamics). Indeed, let us consider the change of variable $\xi' = f_t(\xi)$. The conservation of $P_0(\xi)d\xi$ by the dynamics (C.3) implies that $d\xi\, P_0(\xi) = d\xi'\, P_0(\xi')$. Then the integral in (C.4) reads $\int d\xi'\, P_0(\xi')\, \delta\left(x[\xi'] - x_1\right)$. It is independent of time and, by virtue of the definition (C.1) of the distribution $P_0(\xi)$ in phase space, it is in fact equal to the microcanonical distribution $P_{\rm mc}(x_1)$,

$$P_{P_0}(x_1, t) = P_{\rm mc}(x_1) \equiv \frac{\int d\xi\, \delta\left(H[\xi] - E\right) \delta\left(x[\xi] - x_1\right)}{\int d\xi'\, \delta\left(H[\xi'] - E\right)}. \tag{C.5}$$

On the other hand, the probability that the set of mesoscopic variables x takes the value x_1 at time $t = 0$ and the value x_2 at time t, when the initial position of the system in phase space is distributed according to $P_0(\xi)$, reads

$$P_{P_0}(x_2, t; x_1, 0) = \int d\xi\, P_0(\xi)\, \delta\left(x[f_t(\xi)] - x_2\right) \delta\left(x[\xi] - x_1\right). \tag{C.6}$$

Since the variables denoted by $x(\xi)$ are even functions of momenta, we can write $x[f_t(\xi)] = x[[Rf_t](\xi)]$. Let us consider the change of variable $\xi' = [Rf_t](\xi)$. The invariance under time reversal of the microscopic dynamics (C.2) entails that $\xi = [Rf_t](\xi')$, so that $x[\xi] = x[[Rf_t](\xi')] = x[f_t(\xi')]$. By virtue of its definition (C.1) $d\xi\, P_0[\xi]$ is invariant under the operation R, so $d\xi'\, P_0(\xi') = d(f_t[\xi])\, P_0(f_t[\xi])$ and the conservation of $d\xi\, P_0(\xi)$ by the dynamics (C.3) implies that $d\xi'\, P_0(\xi') = d\xi\, P_0(\xi)$. Eventually the integral in (C.6) reads $\int d\xi'\, P_0(\xi')\delta\left(x[\xi'] - x_2\right)\delta\left(x[f_t(\xi')] - x_1\right)$ and we get

$$P_{P_0}(x_2, t; x_1, 0) = P_{P_0}(x_1, t; x_2, 0). \tag{C.7}$$

The joint probability is invariant by exchanging the times at which x_1 and x_2 occur.

We now assume that the evolution of the mesoscopic variable x, which is invariant by time translation, can be described by an homogeneous Markovian stochastic process whose stationary distribution $P_{\rm st}(x)$ is the time-independent probability $P_{P_0}(x)$ defined in (C.4) and whose transition rates, denoted by $W(x' \leftarrow x)$, are determined by the identification

$$P_{P_0}(x', dt; x, 0) \equiv W(x' \leftarrow x) P_{P_0}(x) \times dt, \tag{C.8}$$

where the time-displaced joint probability for x' and x and the probability of the single variable x are calculated with the same initial distribution P_0 for the microscopic variables ξ's. Then, the time reversal symmetry property (C.7) for the joint probability and the fact that $P_{P_0}(x)$ coincides with the microcanonical distribution $P_{\rm mc}(x)$ by virtue of (C.5) lead to the microcanonical detailed balance

$$W(x' \leftarrow x) P_{\rm mc}(x) = W(x \leftarrow x') P_{\rm mc}(x'). \tag{C.9}$$

D Definitions for statistics over histories

Consider a history where the system starts in configuration $\mathcal{C}_0$ at time $t_0 = 0$ and ends in configuration $\mathcal{C}_f$ at time t after going through successive configurations $\mathcal{C}_0$, $\mathcal{C}_1$,..., $\mathcal{C}_N = \mathcal{C}_f$. The N instantaneous jumps from one configuration to another occur at N successive intermediate times T_i which are continuous stochastic variables : the system jumps from $\mathcal{C}_{i-1}$ to $\mathcal{C}_i$ at time T_i $(i = 1, \ldots, N)$ in the time interval $[t_i, t_i + dt_i[$, with $t_0 < t_1 < t_2 < \cdots < t_N < t$. The probability measure for such a history is related to the probability density $\Pi_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist]$ by

$$dP_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist] \equiv dt_1 \ldots dt_N \Pi_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist] \tag{D.1}$$

where, for a time-translational invariant process,

$$\begin{aligned} \Pi_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist] \;=\; & e^{-(t-t_N)\Lambda(\mathcal{C}_N)} (\mathcal{C}_N|\mathbb{W}|\mathcal{C}_{N-1}) e^{-(t_N - t_{N-1})\Lambda(\mathcal{C}_{N-1})} \\ & \times \cdots e^{-(t_2 - t_1)\Lambda(\mathcal{C}_1)} (\mathcal{C}_1|\mathbb{W}|\mathcal{C}_0) e^{-(t_1 - t_0)\Lambda(\mathcal{C}_0)} \end{aligned} \tag{D.2}$$

and $\Lambda(\mathcal{C})$ is the total exit rate (also called escape rate) from configuration $\mathcal{C}$,

$$\Lambda(\mathcal{C}) \equiv \sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}'|\mathbb{W}|\mathcal{C}). \tag{D.3}$$

The average of a functional $F[\mathcal{H}ist]$ over the histories that start in configuration $\mathcal{C}_0$ and end in configuration $\mathcal{C}_f$ is computed as

$$\langle F \rangle_{\mathcal{C}_f, \mathcal{C}_0} = \int dP_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist]\, F[\mathcal{H}ist] \tag{D.4}$$

with

$$\int dP_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist] = \sum_{N=0}^{+\infty} \sum_{\mathcal{C}_1} \cdots \sum_{\mathcal{C}_{N-1}} \int_{t_0 < t_1 < \ldots < t_N} dP_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist]\,. \tag{D.5}$$

Then the average of a functional when the initial distribution of configurations is P_0 reads

$$\langle F \rangle_{P_0} = \sum_{\mathcal{C}_f} \sum_{\mathcal{C}_0} P_0(\mathcal{C}_0) \int dP_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist]\, F[\mathcal{H}ist]. \tag{D.6}$$

E Some remarks on large deviations

Suppose that X_t is some time-dependent random quantity. Typically, what we have in mind is a variation of some physical quantity (energy, matter) exchanged between a reservoir and a system during the interval $[0,t]$. So X_t is expected to scale like t at large times, with an average $\langle X_t\rangle/t$ (or even X_t/t almost surely) going to a constant J as t goes to $+\infty$.

There are a number of definitions to quantify the probability that X_t/t differs significantly from J at some large time. It often happens that this probability is exponentially small, and a general mathematical theory, large deviation theory, has emerged to describe this situation.

E.1 The mathematical definition of large deviations

The general definition is a bit abstract, dealing with general probability measures depending on t. We restrict here to the situation when the probability measure is the distribution of a real random variable X_t. If $j_- < j_+$ are two real numbers, observe that $P\left(\frac{X_t}{t} \in]j_-, j_+[\right) \leq P\left(\frac{X_t}{t} \in [j_-, j_+]\right)$ so that

$$-\frac{1}{t}\ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) \leq -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in]j_-, j_+[\right). \tag{E.1}$$

For large t, the two members may not converge but at least

$$\liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) \leq \limsup_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in]j_-, j_+[\right), \tag{E.2}$$

where both sides of the inequality belong to $[0,+\infty]$ with the definitions $\liminf_{t\to+\infty} g_t \equiv \lim_{t\to+\infty} \inf_{s\geq t} g_s$ and $\limsup_{t\to+\infty} g_t = \lim_{t\to+\infty} \sup_{s\geq t} g_s$. A nontrivial lower bound for the left-hand side or upper bound for the right-hand side gives information on the exponential rate of decrease of $P\left(\frac{X_t}{t} \in [j_-, j_+]\right)$ and $P\left(\frac{X_t}{t} \in]j_-, j_+[\right)$, the best situation being when the two exist and coincide.

One says that the random variable X_t obeys a large deviation principle (LDP) if there is a lower semi-continuous function R_X, called the rate function, such that

$$\inf_{j\in[j_-,j_+]} R_X(j) \leq \liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) \tag{E.3}$$

and

$$\limsup_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in]j_-, j_+[\right) \leq \inf_{j\in]j_-,j_+[} R_X(j). \tag{E.4}$$

The property that R_X is *lower semicontinuous* means that

$$\lim_{\varepsilon\searrow 0} \inf_{j'\in[j-\varepsilon,j+\varepsilon]} R_X(j') = R_X(j) \qquad \text{for every } j. \tag{E.5}$$

More generally a lower semicontinuous function achieves its minimum on any closed interval. Notice that using an open interval $]j-\varepsilon, j+\varepsilon[$ instead of $[j-\varepsilon, j+\varepsilon]$ in definition (E.5) leads to the same notion. The definitions (E.3) and (E.4) involve lim sup and lim inf and make a difference between $]j_-, j_+[$ and $[j_-, j_+]$ to take care of slightly pathological situations. According to the Portmanteau theorem (see Sec. D.2. of Ref.[73]) the latter definitions are equivalent to the weak convergence (or convergence in law) of the measure $-\frac{1}{t}\ln P\left(\frac{X_t}{t} \in I\right)$ to the measure $\inf_{j\in I} R_X(j)$.

By taking $j_+ = j+\varepsilon$, $j_- = j-\varepsilon$ and letting $\varepsilon \searrow 0$, we see that if X_t obeys a LDP then the rate function is given by

$$\begin{aligned} R_X(j) &= \lim_{\varepsilon\searrow 0} \liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in [j-\varepsilon, j+\varepsilon]\right) \\ &= \lim_{\varepsilon\searrow 0} \limsup_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in]j-\varepsilon, j+\varepsilon[\right). \end{aligned} \tag{E.6}$$

Note that we do not claim that the existence of the above limits, equal to some lower semi-continuous function $R_X(j)$, guarantees that X_t satisfies a LDP.

However, if X_t satisfies a LDP with a *continuous* rate function, then $\inf_{j\in]j_-,j_+[} R_X(j) = \inf_{j\in[j_-,j_+]} R_X(j)$, the lim sup and lim inf become standard limits and the inequalities become equalities :

$$\lim_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in]j_-,j_+[\right) = \lim_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in [j_-,j_+]\right) = \inf_{j\in[j_-,j_+]} R_X(j). \tag{E.7}$$

The latter property can be summarized as

$$\lim_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I\right) = \inf_{j\in I} R_X(j) \tag{E.8}$$

where I is some interval.

Before we turn to other possible characterizations of large deviations, let us explain briefly and heuristically the presence of the $\inf R_X(j)$ in these formulae. If the interval I is split into a finite number of intervals, say $I = \cup_{k=1}^K I_k$, we have

$$\max_k P\left(\frac{X_t}{t}\in I_k\right) \le P\left(\frac{X_t}{t}\in I\right) \le \sum_k P\left(\frac{X_t}{t}\in I_k\right) \le K\max_k P\left(\frac{X_t}{t}\in I_k\right). \tag{E.9}$$

We infer that

$$-\frac{\ln K}{t} + \min_k -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I_k\right) \le -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I\right) \le \min_k -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I_k\right) \tag{E.10}$$

so that, for $s \ge t$,

$$-\frac{\ln K}{t} + \min_k -\frac{1}{s}\ln P\left(\frac{X_s}{s}\in I_k\right) \le -\frac{1}{s}\ln P\left(\frac{X_s}{s}\in I\right) \le \min_k -\frac{1}{s}\ln P\left(\frac{X_s}{s}\in I_k\right). \tag{E.11}$$

First we take the inf over $s \ge t$ and interchange freely the inf and the min to get

$$-\frac{\ln K}{t} + \min_k \inf_{s\ge t} -\frac{1}{s}\ln P\left(\frac{X_s}{s}\in I_k\right) \le \inf_{s\ge t} -\frac{1}{s}\ln P\left(\frac{X_s}{s}\in I\right) \le \min_k \inf_{s\ge t} -\frac{1}{s}\ln P\left(\frac{X_s}{s}\in I_k\right). \tag{E.12}$$

Letting $t\to+\infty$ yields

$$\liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I\right) = \min_k \liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I_k\right). \tag{E.13}$$

Second we take the sup over $s \ge t$. This time only one of the inequalities survives the interchange of sup and min

$$\sup_{s\ge t} -\frac{1}{s}\ln P\left(\frac{X_s}{s}\in I\right) \le \sup_{s\ge t}\min_k -\frac{1}{s}\ln P\left(\frac{X_s}{s}\in I_k\right) \le \min_k \sup_{s\ge t} -\frac{1}{s}\ln P\left(\frac{X_s}{s}\in I_k\right) \tag{E.14}$$

and letting $t\to+\infty$ yields

$$\limsup_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I\right) \le \min_k \limsup_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I_k\right). \tag{E.15}$$

To summarize

$$\begin{aligned} \min_k \liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I_k\right) &= \liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I\right) \\ \limsup_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I\right) &\le \min_k \limsup_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t}\in I_k\right). \end{aligned} \tag{E.16}$$

Informally, this implies indeed that the large deviation estimates for the interval I come solely from some arbitrary small subinterval of I characterized by a minimizing property.

As a conclusion we list some informal rewritings often used in the literature of physicists community. There (E.8) can be found to be rewritten as

$$P\left(\frac{X_t}{t} \in I\right) \underset{t\to+\infty}{\sim} e^{-t \inf_{j\in I} R_X(j)}, \tag{E.17}$$

a notation which neglects the presence of a possibly unbounded t-dependent prefactor. The same notation convention is used when writing the consequence (5.5) of (E.8) for a continuous rate function $R_X(j)$ as

$$\Pi_t(j) \underset{t\to+\infty}{\sim} e^{-tR_X(j)}. \tag{E.18}$$

where $\Pi_t(j)$ is the probability density defined (if it exists) as $P\left(\frac{X_t}{t} \in [j, j+dj]\right) = \Pi_t(j)\, dj$. Sometimes $\Pi_t(j)$ in (E.18) is replaced by the notation $P\left(\frac{X_t}{t} = j\right)$, which is quite abusive, since $P\left(\frac{X_t}{t} = j\right) = 0$ whenever $\Pi_t(j)$ exists as a function (in other words $P\left(\frac{X_t}{t} = j\right) = 0$ unless X_t/t has a distribution made of δ peaks). An even more informal, and possibly misleading, is

$$P\left(\frac{X_t}{t} \in [j, j+dj]\right) \underset{t\to+\infty}{\sim} e^{-tR_X(j)} \tag{E.19}$$

which holds only under the convention that $\lim_{t\to+\infty} \frac{1}{t} \ln dj = 0$ for an infinitesimal dj.

E.2 Alternative definitions

There are a number of situations of physical interest when X_t takes values in a (time independent) discrete set $\mathbb{X}$. Typically $\mathbb{X}$ is $\mathbb{Z}$, the set of integers, or $a + \mathbb{Z}$, the set of integers shifted by a, or $a\mathbb{Z}$, the set of integers dilated by a.

Though this discreteness by no means prevents from using the general large deviation theory, other reasonable but ad hoc definitions come to mind.

For instance, one may ask whether there is a function $g_X(j)$ from $\mathbb{R}$ to $[0, +\infty]$ such that, for any map $t \to \xi_t$ from $[0, +\infty[$ to $\mathbb{X}$ such that $\lim_{t\to+\infty} \frac{\xi_t}{t} = j$,

$$\lim_{t\to+\infty} -\frac{1}{t} \ln P(X_t = \xi_t) \equiv g_X(j). \tag{E.20}$$

A weaker condition comes from asking whether there is a function $h_X(j)$ from $\mathbb{R}$ to $[0, +\infty]$ such that

$$\lim_{\substack{t\to+\infty \\ tj\in\mathbb{X}}} -\frac{1}{t} \ln P(X_t = tj) \equiv h_X(j). \tag{E.21}$$

Of course same care is needed when choosing $\mathbb{X}$, which would better be in some way minimal. For instance, if one takes $\mathbb{X} = \mathbb{Z}$ while X_t takes only even values, h_X or g_X have no chance to exist. However, when $\mathbb{X}$ is chosen appropriately and $P(X_t = x)$ behaves nicely for large t and x, it can be expected on physical grounds that all three functions R_X, g_X and h_X exist and are the same. Alas, we are not aware of practical and general enough mathematical criteria that guarantee this fact. Of the three, h_X is probably the easiest to tackle with bare hands. However, to deal with R_X one can often rely on powerful theorems of the general theory of large deviations. But whether they exist (and coincide) or not, the three functions are meaningful characteristics of the large deviations of X_t. With some imagination, we could probably invent others.

E.3 Inequalities at finite time and fluctuation relation between large deviation functions

One important feature of large deviation functions in out-of-equilibrium statistical mechanics is that they satisfy, under appropriate circumstances, symmetry relations. Usually, these relations are consequences of a, possibly generalized, time reversal symmetry. Typically, for some $0 < m < M < +\infty$ and for some real γ one has

$$mP(X_t = -x)e^{\gamma x} \leq P(X_t = x) \leq MP(X_t = -x)e^{\gamma x} \tag{E.22}$$

for every x. To be fair, this way of writing things applies, strictly speaking, only when X_t is atomic, i.e. takes values in a discrete set. A more correct statement is that the laws of X_t and $-X_t$ are absolutely continuous and the Radon-Nikodym derivative[4]

$$me^{\gamma x} \leq \frac{dP_{X_t}(x)}{dP_{-X_t}(x)} \leq Me^{\gamma x}, \tag{E.23}$$

of which the previous equation is a special case.

The point we want to make here is that, with any of the three definitions of large deviation functions given above (when they do exist), the above equation implies a symmetry relation. For g_X or h_X we can use the simplified equation evaluated when $x \in \mathbb{X}$. It implies

$$-\frac{1}{t}\ln M - \gamma\frac{x}{t} - \frac{1}{t}\ln P(X_t = -x) \leq -\frac{1}{t}\ln P(X_t = x) \leq -\frac{1}{t}\ln m - \gamma\frac{x}{t} - \frac{1}{t}\ln P(X_t = -x). \tag{E.24}$$

Assuming the existence of h_X one can take the limit $\lim_{\substack{t\to+\infty \\ tj\in\mathbb{X}}}$ to get $h_X(-j) - \gamma j \leq h_X(j) \leq h_X(-j) - \gamma j$, i.e.

$$h_X(j) = h_X(-j) - \gamma j, \tag{E.25}$$

the announced symmetry relation. If g_X exists, then so does h_X and they are equal, so the symmetry property for g_X (when it exists) is a consequence of the above bounds.

In the case when R_X exists, the proof of its symmetry based on the inequalities for the Radon-Nikodym derivative is only slightly more elaborate : if B is a Borel subset of $\mathbb{R}$ with $tj_- \leq \inf B$ and $\sup B \leq tj_+$ then, assuming $\gamma \geq 0$ for definiteness,

$$P(-X_t \in B)e^{\gamma t j_-} \leq \int_B dP_{-X_t}(x)e^{\gamma x} \leq P(-X_t \in B)e^{\gamma t j_+} \tag{E.26}$$

so, as a consequence of (E.23),

$$-\frac{1}{t}\ln M - \gamma j_+ - \frac{1}{t}\ln P(X_t \in -B) \leq -\frac{1}{t}\ln P(X_t \in B) \leq -\frac{1}{t}\ln m - \gamma j_- - \frac{1}{t}\ln P(X_t \in -B), \tag{E.27}$$

and taking $B =]tj_-, tj_+[$ and the $\limsup_{t\to+\infty}$ in the second inequality we get

$$\limsup_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in]j_-, j_+[\right) \leq -\gamma j_- + \limsup_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in]-j_+, -j_-[\right). \tag{E.28}$$

Similarly, taking $B = [tj_-, tj_+]$ and the $\liminf_{t\to+\infty}$ of the first inequality we get

$$-\gamma j_+ + \liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in [-j_+, -j_-]\right) \leq \liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right). \tag{E.29}$$

Taking $j_\pm = j \pm \varepsilon$, j fixed, $\varepsilon \searrow 0$ we find from the characterization (E.6) of R_X

$$R_X(j) \leq -\gamma j + R_X(-j) \text{ and } -\gamma j + R_X(-j) \leq R_X(j) \tag{E.30}$$

[4] One says that the Radon-Nikodym derivative of a probability measure μ with respect to a probability measure ν exists if there is an ν-integrable function f such that for every Borel set B one has $\int_B d\mu = \int_B f d\nu$. Then f is called the Radon-Nikodym derivative of μ with respect to ν and one writes $f = d\mu/d\nu$.

so

$$R_X(j) = R_X(-j) - \gamma j, \tag{E.31}$$

the announced symmetry. So with respect to fluctuation relations, our three definitions of large deviations are on the same footing.

We notice that the above relation is usually written in the more symmetric form $R_X(j) - R_X(-j) = -\gamma j$, which is however ambiguous when $R_X(j)$ is infinite.

E.4 Cumulated quantities with sub-extensive difference

We would like to stress that some natural and desirable properties of large deviation functions are not automatic with some of our definitions.

For instance, suppose that X_t and Y_t are random processes with a sub-extensive difference, i.e. there is a non-random function C_t such that $\lim_{t\to+\infty} C_t = 0$ and $|X_t - Y_t| \leq tC_t$. Then it should be expected that the following alternative holds :

$$\begin{cases} \text{- } X_t \text{ and } Y_t \text{ have the same large deviation function,} \\ \text{- neither } X_t \text{ nor } Y_t \text{ has a large deviation function.} \end{cases} \tag{E.32}$$

We shall show that this holds true within the general theory of large deviations, namely for R_X and R_Y. We notice that, when $R_X(j)$ is convex upward, then $-R_X(j)$ can be expressed as in (5.6) and the property (E.32) can also be rederived from the explicit formulæ (5.6). With our ad-hoc definitions, we can see that even when X_t and Y_t have the required properties, the discrete sets on which X_t and Y_t take their values can be very different. And even if the relation is simple, some problems remain at least with the weakest definition of large deviation functions, namely it may happen that $h_X \neq h_Y$.

To make a proof within the general theory of large deviations, we note that, if $j_- < j_+$, any real function R has the property that

$$\lim_{\varepsilon\searrow 0} \inf_{j\in]j_-+\varepsilon, j_+-\varepsilon[} R(j) = \inf_{j\in]j_-, j_+[} R(j). \tag{E.33}$$

Moreover, if R is lower semi-continuous then

$$\lim_{\varepsilon\searrow 0} \inf_{j\in [j_--\varepsilon, j_++\varepsilon]} R(j) = \inf_{j\in [j_-, j_+]} R(j). \tag{E.34}$$

Now, let $\varepsilon > 0$. For s large enough, $C_s < \varepsilon$ by hypothesis, so that

$$P\left(\frac{Y_s}{s} \in [j_-, j_+]\right) \leq P\left(\frac{X_s}{s} \in [j_- - \varepsilon, j_+ + \varepsilon]\right) \text{ for } s \geq t \tag{E.35}$$

whenever t is such that $C_s \leq \varepsilon$ for $s \geq t$. Then

$$\inf_{s\geq t} -\frac{1}{s}\ln P\left(\frac{X_s}{s} \in [j_- - \varepsilon, j_+ + \varepsilon]\right) \leq \inf_{s\geq t} -\frac{1}{s}\ln P\left(\frac{Y_s}{s} \in [j_-, j_+]\right), \tag{E.36}$$

so taking $t \to +\infty$ one gets

$$\liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{X_t}{t} \in [j_- - \varepsilon, j_+ + \varepsilon]\right) \leq \liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{Y_t}{t} \in [j_-, j_+]\right). \tag{E.37}$$

Assuming X_t has a large deviation function, the left-hand side is $\geq \inf_{j\in[j_--\varepsilon, j_++\varepsilon]} R_X(j)$ by (E.3), so we have proved that for any $\varepsilon > 0$ one has

$$\inf_{j\in[j_--\varepsilon, j_++\varepsilon]} R_X(j) \leq \liminf_{t\to+\infty} -\frac{1}{t}\ln P\left(\frac{Y_t}{t} \in [j_-, j_+]\right). \tag{E.38}$$

In the same vein, assuming $j_- < j_+$, one has, for ε small enough ($\varepsilon < \frac{j_+ - j_-}{2}$)

$$P\left(\frac{X_s}{s} \in]j_- + \varepsilon, j_+ - \varepsilon]\right) \leq P\left(\frac{Y_s}{s} \in]j_-, j_+[\right) \text{ for } s \geq t \tag{E.39}$$

whenever t is such that $C_s \leq \varepsilon$ for $s \geq t$. As above, we infer that

$$\limsup_{t\to+\infty} -\frac{1}{t} \ln P\left(\frac{Y_t}{t} \in]j_-, j_+[\right) \leq \limsup_{t\to+\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in]j_- + \varepsilon, j_+ - \varepsilon[\right). \tag{E.40}$$

Assuming X_t has a large deviation function, the right-hand side is $\leq \inf_{j\in]j_-+\varepsilon, j_+-\varepsilon]} R_X(j)$ by (E.4), so for ε small enough one has

$$\limsup_{t\to+\infty} -\frac{1}{t} \ln P\left(\frac{Y_t}{t} \in]j_-, j_+[\right) \leq \inf_{j\in]j_-+\varepsilon, j_+-\varepsilon[} R_X(j) \tag{E.41}$$

Taking the limit $\varepsilon \searrow 0$ in (E.38) and (E.41) (using the lower semi-continuity of R_X in (E.38)) we obtain precisely that Y_t satisfies an LDP (E.3)–(E.4) with large deviation function R_X. This proves that the alternative (E.32) holds.

We conclude this discussion with a (counter)-example which has the advantage of being extremely simple, but the drawback that it is artificial (in particular, it is not continuous in probability, it even has non-random discontinuity times).

The example is $X_t \equiv \lfloor t \rfloor$ the integer part of t, i.e. $\lfloor t \rfloor$ is an integer and $\lfloor t \rfloor \leq t < \lfloor t \rfloor + 1$, and $Y_t \equiv X_t + 1$. As shown below,

- X_t has a large deviation function R_X and a large deviation function h_X which coincide, namely

$$R_X(j) = h_X(j) = \begin{cases} 0 & \text{if } j = 1 \\ +\infty & \text{if } j \neq 1 \end{cases}. \tag{E.42}$$

- $Y_t \equiv X_t + 1$ has a large deviation function R_Y ($= R_X$ by the previous result) but h_Y, while existing, differs from h_X, namely

$$h_Y(j) = +\infty \text{ whatever } j. \tag{E.43}$$

- X_t (resp. Y_t) has no large deviation function g_X (resp. g_Y).

Here are proofs for the latter properties.

• The case of h_X : we take $\mathbb{X} = \mathbb{N}$

$$\lim_{\substack{t\to+\infty \\ tj\in\mathbb{N}}} -\frac{1}{t} \ln P(X_t = tj) = \lim_{\substack{t\to+\infty \\ tj\in\mathbb{N}}} -\frac{1}{t} \ln \mathbb{1}_{\lfloor t \rfloor = tj}. \tag{E.44}$$

If $j > 1$ $\lfloor t \rfloor \leq t < tj$ so $\mathbb{1}_{\lfloor t \rfloor = tj} = 0$. If $j < 1$ $\lfloor t \rfloor > t - 1 > tj$ whenever $t > \frac{1}{1-j}$, so $\mathbb{1}_{\lfloor t \rfloor = tj} = 0$ for t large enough. If $j = 1$ $\mathbb{1}_{\lfloor t \rfloor = t} = 1$ for $t \in \mathbb{N}$. This gives the announced formula (E.42) for h_X.

• The case of R_X.

First we consider $P\left(\frac{X_s}{s} \in [j_-, j_+]\right) = \mathbb{1}_{\lfloor s \rfloor \in [sj_-, sj_+]}$.

- If $1 \in [j_-, j_+]$ and $s = n \in \mathbb{N}$ then $P\left(\frac{X_n}{n} \in [j_-, j_+]\right) = 1$, so for any t $\inf_{s\geq t} -\frac{1}{s} \ln P\left(\frac{X_s}{s} \in [j_-, j_+]\right) = 0$ and $\liminf_{t\to+\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) = 0$.

- If $j_- > 1$ then $\lfloor s \rfloor < sj_-$ whenever $s > 0$ so for any $t > 0$ $\inf_{s>t} -\frac{1}{s} \ln P\left(\frac{X_s}{s} \in [j_-, j_+]\right) = +\infty$ and $\liminf_{t\to+\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) = +\infty$.

- If $j_+ < 1$ then $\lfloor s \rfloor > s - 1 \geq sj_+$ whenever $s \geq \frac{1}{1-j_+}$, so $\inf_{s>t} -\frac{1}{s} \ln P\left(\frac{X_s}{s} \in [j_-, j_+]\right) = +\infty$ whenever $t \geq \frac{1}{1-j_+}$ and $\liminf_{t\to+\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) = +\infty$.

To summarize

$$\liminf_{t\to+\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) = \begin{cases} 0 & \text{if } 1 \in [j_-, j_+] \\ +\infty & \text{else} \end{cases}. \tag{E.45}$$

Second we turn to $P\left(\frac{X_s}{s} \in]j_-, j_+[\right) = \mathbb{1}_{\lfloor s \rfloor \in]sj_-, sj_+[}$.

- If $1 \notin]j_-, j_+[$ and $s \neq 0$, then $s \notin]sj_-, sj_+[$ so that for $s = n \in \mathbb{N}$ we have $P\left(\frac{X_n}{n} \in]j_-, j_+[\right) = 0$. So for any t $\sup_{s \geq t} -\frac{1}{s} \ln P\left(\frac{X_s}{s} \in [j_-, j_+]\right) = +\infty$ and $\limsup_{t \to +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) = +\infty$.

- If $1 \in]j_-, j_+[$ then $\lfloor s \rfloor \in]sj_-, sj_+[$ as soon as $s \geq \frac{1}{1-j_-}$ so $\limsup_{t \to +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in]j_-, j_+[\right) = 0$. To summarize

$$\limsup_{t \to +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in]j_-, j_+[\right) = \begin{cases} 0 & \text{if } 1 \in]j_-, j_+[\\ +\infty & \text{else} \end{cases}. \tag{E.46}$$

From (E.46), (E.45) and the definition of an LDP (E.3)-(E.4) we infer that X_t has a large deviation function

$$R_X(j) = \begin{cases} 0 & \text{if } j = 1 \\ +\infty & \text{else} \end{cases} \tag{E.47}$$

and $R_X = h_X$.

- The case of h_Y : we take $\mathbb{X} = \mathbb{N}^\star$

$$\lim_{\substack{t \to +\infty \\ tj \in \mathbb{N}^\star}} -\frac{1}{t} \ln P(Y_t = tj) = \lim_{\substack{t \to +\infty \\ tj \in \mathbb{N}^\star}} -\frac{1}{t} \ln \mathbb{1}_{\lfloor t \rfloor + 1 = tj}. \tag{E.48}$$

If $j > 1$ $\lfloor t \rfloor \leq t+1 < tj$ whenever $t > \frac{1}{j-1}$ and then $\mathbb{1}_{\lfloor t \rfloor + 1 = tj} = 0$. If $j \leq 1$ then $\lfloor t \rfloor + 1 > t \geq tj$ for every t and then $\mathbb{1}_{\lfloor t \rfloor + 1 = tj} = 0$. So $\lim_{\substack{t \to +\infty \\ tj \in \mathbb{N}^\star}} -\frac{1}{t} \ln P(Y_t = tj) = +\infty$ for every j and $h_Y(j) = +\infty$ for every j as announced in (E.43).

- The case of g_X and g_Y.

For $j = 1$ we have to consider maps ξ_t such that $\lim_{t \to +\infty} \xi_t / t = 1$ and the behavior of $-\frac{1}{t} \ln P(X_t = \xi_t)$ when t goes to $+\infty$. Among these maps some are such that the corresponding limit is $0 = h_X(j = 1)$ and some other ones are such that the limit is $+\infty = h_Y(j = 1)$: for instance the map $\xi_t = \lfloor t \rfloor$ is such that $P(X_t = \xi_t) = 1$, whereas the map $\xi_t = \lfloor t \rfloor + a$ with $a > 0$ is such that $P(X_t = \xi_t) = 0$. The same argument can be applied to $Y_t \equiv X_t + 1$. As a result, neither g_X nor g_Y exist.

References

[1] J. L. Lebowitz and H. Spohn. A Gallavotti-Cohen Type Symmetry in the Large Deviation Functional for Stochastic Dynamics. *J. Stat. Phys.*, 95:333, 1999.

[2] L. Onsager. Reciprocal relations in irreversible processes. I. *Phys. Rev.*, 37:405, 1931.

[3] L. Onsager. Reciprocal relations in irreversible processes. II. *Phys. Rev.*, 38:2265, 1931.

[4] H.B. Callen. *Thermodynamics*. John Wiley and Sons, 1960.

[5] S. R. De Groot. *Thermodynamics of Irreversible Processes*. Interscience Publishers, Inc., New York, 1952.

[6] S. R. De Groot and P. Mazur. *Non-equilibrium thermodynamics*. North Holland, 1962.

[7] Y. Oono and M. Paniconi. Steady State Thermodynamics. *Prog. Theor. Phys. Suppl.*, 130:29, 1998.

[8] S. Sasa and H. Tasaki. Steady state thermodynamics. *J. Stat. Phys.*, 125:125, 2006.

[9] U. Seifert. Stochastic dynamics : principles and perspectives. *Eur. Phys. J. B*, 64:423, 2008.

[10] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.*, 75:126001, 2012.

[11] C. Bustamante, J. Liphardt, and F. Ritort. The Nonequilibrium Thermodynamics of Small Systems. *Phys. Today*, 58(7):43, 2005.

[12] F. Ritort. Single-molecule experiments in biological physics : methods and applications. *J. Phys. Condens. Matter*, 18:R531, 2006.

[13] R.J. Harris and G.M. Schütz. Fluctuation theorems for stochastic dynamics. *J. Stat. Mech.*, page P07020, 2007.

[14] K. Mallick. Some recent developments in non-equilibrium statistical physics. *Pramana - J. Phys.*, 73:417, 2009.

[15] C. Van den Broeck. Stochastic thermodynamics : A brief introduction. In C. Bechinger, F. Sciortino, and P. Ziherl, editors, *Physics of Complex Colloids*, volume 184 of *Proceedings of the International School of Physics "Enrico Fermi*, page 155. IOS Press, Amsterdam, 2013.

[16] A. Einstein. *Investigations on the Theory of the Brownian Movement*. New York : Dover, 1956.

[17] U.M.B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani. Fluctuation-dissipation : Response theory in statistical physics. *Phys. Rep.*, 461:111, 2008.

[18] R. Chétrite, G. Falkovich, and K. Gawędzki. Fluctuation relations in simple examples of non-equilibrium steady states. *J. Stat. Mech.*, page P08005, 2008.

[19] M. Baiesi, C. Maes, and B. Wynants. Fluctuations and response of nonequilibrium states. *Phys. Rev. Lett.*, 103:010602, 2009.

[20] M. Baiesi, C. Maes, and B. Wynants. Nonequilibrium linear response for Markov dynamics, I. Jump processes and overdamped diffusions. *J. Stat. Phys.*, 137:1094, 2009.

[21] M. Baiesi, E. Boksenbojm, C. Maes, and B. Wynants. Nonequilibrium linear response for Markov dynamics, II : Inertial dynamics. *J. Stat. Phys.*, 139:492, 2010.

[22] G. Verley, R. Chétrite, and D. Lacoste. Modified fluctuation-dissipation theorem for general non-stationary states and application to the Glauber-Ising chain. *J. Stat. Mech.*, page P10025, 2011.

[23] A. Einstein. Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes. *Ann. d. Phys.*, 33:1275, 1910.

[24] E.T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620, 1957.

[25] E.T. Jaynes. Information theory and statistical mechanics. II. *Phys. Rev.*, 108:171, 1957.

[26] R. S. Ellis. *Entropy, large deviations, and statistical mechanics.* Springer, 1985.

[27] Y. Oono. Large deviation and statistical physics. *Prog. Theor. Phys. Suppl.*, 99:165–205, 1989.

[28] A. A. Filyokov and V. Ya Karpov. Method of the most probable path of evolution in the theory of stationary irreversible processes. *Journal of engineering physics and thermophysics*, 13:416, 1967.

[29] C. Monthus. Non-equilibrium steady states: maximization of the Shannon entropy associated with the distribution of dynamical trajectories in the presence of constraints. *J. Stat. Mech.*, page P03008, 2011.

[30] D.J. Evans, E.G.D. Cohen, and G.P. Moriss. Probability of second law violations in shearing steady states. *Phys. Rev. Lett.*, 71:2401, 1993.

[31] D.J. Evans and D.J. Searles. Equilibrium microstates which generate second law violating steady states. *Phys. Rev. E*, 50:1645, 1994.

[32] G. Gallavotti and E.G.D. Cohen. Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.*, 74:2694, 1995.

[33] G. Gallavotti and E.G.D. Cohen. Dynamical ensembles in stationary states. *J. Stat. Phys.*, page 931, 1995.

[34] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, 78:2690, 1997.

[35] M. Esposito, U. Harbola, and S. Mukamel. Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems. *Rev. Mod. Phys.*, 81:1665, 2009.

[36] M. Campisi, Hänggi P., and P. Talkner. Colloquium : Quantum fluctuation relations : Foundations and applications. *Rev. Mod. Phys.*, 83:771, 2011.

[37] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements :a master-equation approach. *Phys. Rev. E*, 56:5018, 1997.

[38] J. Kurchan. Fluctuation theorem for stochastic dynamics. *J. Phys. A: Math. Gen.*, 31:3719, 1998.

[39] G. E. Crooks. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *J. Stat. Phys.*, 90:1481, 1998.

[40] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60:2721, 1999.

[41] C. Maes. The fluctuation theorem as a Gibbs property. *J. Stat. Phys.*, 95:367, 1999.

[42] U. Seifert. Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem. *Phys. Rev. Lett.*, 95:040602, 2005.

[43] M. Esposito and C. Van den Broeck. Three faces of the second law. I. Master equation formalism. *Phys. Rev. E*, 82:011143, 2010.

[44] B. Derrida. Non-equilibrium steady states: fluctuations and large deviations of the density and of the current. *J. Stat. Mech.*, page P07023, 2007.

[45] T. Bodineau and B. Derrida. Cumulants and large deviations of the current through non-equilibrium steady states. *C. R. Physique*, 8:540, 2007.

[46] C. Jarzynski and D. K. Wójcik. Classical and Quantum Fluctuation Theorems for Heat Exchange. *Phys. Rev. Lett.*, 92:230602, 2004.

[47] D. Andrieux and P. Gaspard. Fluctuation Theorem for Currents and Schnakenberg Network Theory. *J. Stat. Phys.*, 127:107, 2007.

[48] D. Andrieux and P. Gaspard. Network and thermodynamic conditions for a single macroscopic current fluctuation theorem. *C. R. Physique*, 8:579, 2007.

[49] F. Cornu and M. Bauer. Thermal Contact II. A Solvable Toy Model. Referred to as Paper II. *arXiv:1302.4540 [cond-mat.stat-mech]*, 2013.

[50] G. E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, 61:2361, 2000.

[51] D. Andrieux and P. Gaspard. Fluctuation theorems and the nonequilibrium thermodynamics of molecular motors. *Phys. Rev. E*, 74:011906, 2006.

[52] T. Schmiedl, T. Speck, and U. Seifert. Entropy Production for Mechanically or Chemically Driven Biomolecules. *J. Stat. Phys.*, 128:77, 2007.

[53] S. Liepelt and R. Lipowsky. Steady-state balance conditions for molecular motor cycles and stochastic nonequilibrium processes. *EPL*, 77:50002, 2007.

[54] J. Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Mod. Phys.*, 48:571, 1976.

[55] G. E. Crooks. *Excursions in Statistical Dynamics*. PhD thesis, University of California, 1999.

[56] T. Hatano and S. Sasa. Steady-State Thermodynamics of Langevin Systems. *Phys. Rev. Lett.*, 86:3463, 2001.

[57] D. R. Cox and H. D. Miller. *The Theory of Stochastic Processes*. Chapman and Hall, London, 1965.

[58] N.G. van Kampen. *Stochastic processes in physics and chemistry*. North Holland, 1992.

[59] J. T. Lewis and R. Russell. An introduction to large deviations for teletraffic engineers. *Unpublished, to be found on the web*, 1997.

[60] S. R. S. Varadhan. Large deviations. *The Annals of Probability*, 36:307, 2008.

[61] D. Andrieux and P. Gaspard. A fluctuation theorem for currents and non-linear response coefficients. *J. Stat. Mech.*, page P02006, 2007.

[62] F. Schlögl. On Stability of Steady States. *Z. Phys.*, 243:303, 1971.

[63] S. Kullbach and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79, 1951.

[64] N. Pottier. *Physique statistique hors équilibre, Processus irréversibles linéaires*. EDP Sciences, CNRS ÉDITIONS, 2007.

[65] E.T. Jaynes. The minimum entropy production principle. *Ann. Rev. Phys. Chem.*, 31:579, 1980.

[66] V. Lecomte, C. Appert-Rolland, and F. van Wijland. Chaotic properties of systems with Markov dynamics. *Phys. Rev. Lett.*, page 010601, 2005.

[67] I. Prigogine. *Introduction to thermodynamics of irreversible processes*. Wiley, New York, 1968.

[68] C. Jarzynski. Hamiltonian Derivation of a Detailed Fluctuation Theorem. *J. Stat. Phys.*, 98:77, 2000.

[69] G.N. Bochkov and Yu.E. Kuzovlev. Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics I. Generalized fluctuation-dissipation theorem. *Physica A*, 106:443, 1981.

[70] G.N. Bochkov and Yu.E. Kuzovlev. Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics II. Kinetic potential and variational principles for nonlinear irreversible processes. *Physica A*, 106:480, 1981.

[71] M. Esposito and C. Van den Broeck. Three Detailed Fluctuation Theorems. *Phys. Rev. Lett.*, 104:090601, 2010.

[72] H. Touchette. The large deviation approach to statistical mechanics. *Phys. Rep.*, 478:1, 2009.

[73] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, New York, 2nd edition, 1998.

[74] E. Helfand. Transport coefficients from dissipation in a canonical ensemble. *Phys. Rev.*, 119:1, 1960.

[75] M. S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. *J. Chem. Phys.*, 20:1281, 1952.

[76] M.S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. II irreversible processes in fluids. *J. Chem. Phys.*, 22:398, 1954.

[77] R. Kubo. Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Japan*, 12:570, 1957.

[78] R. Kubo, M. Yokota, and S. Nakajima. Statistical-mechanical theory of irreversible processes. II. Response to thermal disturbance. *J. Phys. Soc. Japan*, 12:1203, 1957.

[79] T. Yamamoto. Quantum statistical mechanical theory of the rate of exchange chemical reactions in the gas phase. *J. Chem. Phys.*, 33:281, 1960.

[80] R. Zwanzig. Time-correlation functions and transport coefficients in statistical mechanics. *Ann. Rev. Phys. Chem.*, 16:67, 1965.

[81] A. Faggionato and D. Di Pietro. Gallavotti-Cohen-type symmetry related to cycle decompositions for Markov chains and biochemical applications. *J. Stat. Phys.*, 143:11, 2011.

[82] B. Cleuren, C. Van den Broeck, and R. Kawai. Fluctuation theorem for the effusion of an ideal gas. *Phys. Rev. E*, 74:021117, 2006.

[83] D. Andrieux and P. Gaspard. Fluctuation theorem and Onsager reciprocity relations. *J. Chem. Phys.*, 121:6167, 2004.

[84] P. E. Wigner. Derivations of Onsager's Reciprocal Relations. *J. Chem. Phys.*, 22:1912, 1954.