TIME-REVERSAL AND ENTROPY

Christian Maes[1] and **Karel Netočný**
Instituut voor Theoretische Fysica
K.U.Leuven, Belgium.

Abstract: There is a relation between the irreversibility of thermodynamic processes as expressed by the breaking of time-reversal symmetry, and the entropy production in such processes. We explain on an elementary mathematical level the relations between entropy production, phase-space contraction and time-reversal starting from a deterministic dynamics. Both closed and open systems, in the transient and in the steady regime, are considered. The main result identifies under general conditions the statistical mechanical entropy production as the source term of time-reversal breaking in the path space measure for the evolution of reduced variables. This provides a general algorithm for computing the entropy production and to understand in a unified way a number of useful (in)equalities. We also discuss the Markov approximation. Important are a number of old theoretical ideas for connecting the microscopic dynamics with thermodynamic behavior.

1. Introduction

An essential characteristic of irreversible thermodynamic processes is that the time-reversal invariance of the microscopic dynamics is apparently broken. This means that out of equilibrium a particular sequence of macrostates and its time-reversal can have a very different plausibility. This, basically, must be the reason for the positivity of transport coefficients, or, more generally, for the positivity of entropy production. It has already been argued before in [16, 17, 18], mostly via examples, how there is a direct relation between entropy production and the ratio of probabilities for time-reversed trajectories. Most of this was however concentrated on finding a unifying framework for equalities and inequalities that have recently appeared in nonequilibrium statistical mechanics, generalizing, so it is hoped, close to equilibrium relations. Most prominent among those is the symmetry expressed in the Gallavotti-Cohen fluctuation theorem, [4, 7]. In the present paper, we turn to more fundamental issues for identifying the statistical mechanical definition of entropy production rate and to offer a possible answer for various interpretational problems that have remained. The emphasis is on the simplicity of the explanation avoiding technical issues.

[1]email: christian.maes@fys.kuleuven.ac.be

2. Results

Nonequilibrium statistical mechanics is to a large extent still under construction. Recently, there have been made various proposals for a definition of statistical mechanical entropy production going beyond the close to equilibrium regime and through which fluctuations in irreversible processes could be studied. In some cases, the theory of dynamical systems has been a source of inspiration and it was argued that phase space contraction can be identified with entropy production with nonequilibrium ensembles obtained as limits of ergodic averages, see e.g. [7, 20]. A somewhat different approach started from taking advantage of the Gibbsian structure of the distribution of space-time histories where entropy production appeared as the source term for time-reversal breaking in the space-time action functional. These two approaches have in fact much in common, at least concerning the mathematical analysis, see [16]. Since then, many examples have passed the test of verifying that the various algorithms indeed give rise to the physical entropy production. There has however not been a derivation from first principles to convince also the stubborn that the algorithm of [16, 17] applied to models in nonequilibrium dynamics to identify the entropy production, is entirely trustworthy. The main result of the present paper is to give such a derivation: that indeed under very general conditions, both for closed systems and for open systems, both in the transient regime and in the steady state regime, the entropy production can be obtained as the source term of time-reversal breaking in the action functional of the path space measure that gives the distribution of the histories (on some thermodynmaic scale) of the system. This representation is useful because it gives the entropy production as a function of the trajectories and it allows easy mathematical manipulations for taking the average (to prove that it is positive) and for studying the fluctuations (to understand symmetries under time-reversal).

This paper is more or less self-contained with a first Section 3 introducing the main actors. Sections 4 and 5 contain the main result. The difference is that 4 is entirely about the transient regime for closed systems, while Section 5 deals with open systems and discusses the steady state regime. Sections 6 and 7 discuss their consequences in the Markov approximation. Section 8 relates the approach to results inspired by the theory of chaotic dynamical systems, in particular how phase space contraction can play the role of entropy production. Along the way, we suggest interpretations that we think are helpful for starting nonequilibrium statistical mechanics.

3. Set-up

3.1. **Phase space and microscopic dynamics.** Let Ω be the phase space of a closed isolated mechanical system with $x \in \Omega$ representing the microstates, i.e., as described by canonical variables for a classical system of N particles, $x = (q_1, \ldots, q_N, p_1, \ldots, q_N)$. The Hamiltonian dynamics specifies the flow $x \mapsto \phi_t(x)$ on Ω under which x (at some initial time t_0) evolves into $\phi_t(x)$ at time $t_0 + t$. The dynamics is reversible in the sense that $\pi\phi_t\pi = \phi_t^{-1}$ where the time-reversal π on Ω is the involution that changes the sign of the momenta p_i. The flow preserves the phase space volume (Liouville's theorem); the Jacobian determinant equals unity, $|d\phi_t(x)/dx| = 1$ for each t and the Liouville measure dx is time-invariant.
We fix a time-interval δ and write $f \equiv \phi_\delta$. Of course, f preserves the phase space volume and $\pi f \pi = f^{-1}$.

3.2. **Reduced variables.** The time evolution preserves the total energy. We introduce therefore the state space $\Omega_E \equiv \Gamma$, the energy shell, corresponding to a fixed total energy E or better, some interval around it. We denote by $|A|$ the phase space volume of a region $A \subset \Gamma$ given by the projection ρ of the Liouville measure into Γ. Since Γ is thought of as containing a huge number of degrees of freedom, it is reasonable to divide it further. For comparison with experimental situations, we look at some special set of variables, suitably defined in each case, which give a more coarse-grained, contracted, or reduced description of the system, [19, 23, 13]. Depending on the context or on the application, their precise nature may vary. It could be that we look at macroscopic variables $\alpha(x)$ implying a subdivision of Γ by cutting it up in phase cells defined by $a < \alpha(x) < a + \Delta a$ (with some tolerance Δa), or that we split up $x = (y, z) \in \Gamma$ into an observable part y and a background part z. For example, the y might refer to the coordinates of the particles in a subsystem while the background is only monitored as the macrostate of reservoir(s).
At this moment, we do not commit ourselves to one picture but rather imagine somewhat abstractly a map $M : \Gamma \to \hat{\Gamma} : x \mapsto M(x)$ where $\hat{\Gamma}$ is the reduced phase space, a finite partition of Γ. When having in mind macrostates, this space $\hat{\Gamma}$ would correspond to the $\mu-$space of Gibbs. The fact that this partition is assumed finite is not realistic, it is more like $\mathbb{R}^d$, but it is convenient for the notation and it is not essential. With some abuse of notation, the elements of $\hat{\Gamma}$ are denoted by M (standing for all possible values of $M(x)$) and of course, every microscopic trajectory $\gamma = (x, fx, \ldots, f^n x)$ gives rise to a trajectory

$\omega = (M(x), M(fx), \ldots, M(f^n x))$ in $\hat{\Gamma}$. We also assume for simplicity that πM is well defined via $\pi M = M\pi$, that is $Mx = My$ implies $M\pi x = M\pi y$, for all $x, y \in \Gamma$.

3.3. **Distributions.** Probabilities enter the description because the exact microstate of the system is not accessible to us. This is so when preparing the system and also later when we observe the system. Even when we know the reduced state, we still need to evaluate the plausibility of background configurations in order to predict the future development on the level of the reduced states. A natural choice here is to use the microcanonical ensemble. That is, we sample the reduced variables according to some probability distribution $\hat{\nu}$ on $\hat{\Gamma}$ and we impose the microcanonical distribution on each phase cell M. If $\hat{\nu}$ is a probability on $\hat{\Gamma}$, then $\hat{\nu} \times \rho\,(x) \equiv \hat{\nu}(M(x))/|M(x)|$ is the probability density on Γ obtained from $\hat{\nu}$ by uniform randomization (microcanonical ensemble) inside each $M \in \hat{\Gamma}$. It is uniquely determined from the two conditions (1) $\hat{\nu} \times \rho\,(M) = \hat{\nu}(M)$ and (2) $\hat{\nu} \times \rho\,(x|M) = 1/|M|$, $x \in M$, $M \in \hat{\Gamma}$. (Remark: the writing $\hat{\nu} \times \rho$ has no meaning in itself except that it is the notation we use for this probability density.) In words, the probability of a microstate x is the probability (under $\hat{\nu}$) of its corresponding reduced state Mx multiplied with the *a priori* probability (under the Liouville measure) of x given the reduced state Mx. So if we take $\hat{\nu} = \delta(M - \cdot)$ concentrated on the reduced state $M \in \hat{\Gamma}$, then $\hat{\nu} \times \rho$ is the initial probability density corresponding to an experiment where the system is started in equilibrium subject to constraints; that is a uniform (i.e., microcanonical) distribution of the phase points over the set M.
For the opposite direction, we note that every density ν on Γ gives rise to its projection $p(\nu)$, a probability on $\hat{\Gamma}$, via

$$p(\nu)(M) \equiv \nu(M) = \int dx \nu(x) \delta(M(x) - M)$$

and obviously, $p(\hat{\nu} \times \rho) = \hat{\nu}$. All this is very much like what enters in projection-operator techniques, [23].
It now makes sense to ask for the probabilities on Γ at time t, given that the system started at time zero in $M_0 \in \hat{\Gamma}$; we always mean by this that the microstates were uniformly sampled out of M_0. They are given by the ratio

$$\text{Prob}[\phi_t(x) \in A \,|\, M(x) = M_0] \equiv \frac{|\phi_t^{-1} A \cap M_0|}{|M_0|} \tag{3.1}$$

More generally, when, for all we know at time zero, the statistics of the reduced variables is given in terms of the probability $\hat{\nu}$ on $\hat{\Gamma}$, then, at time t, the statistics on $\hat{\Gamma}$ is obtained from

$$\hat{\nu}_t \equiv p((\hat{\nu} \times \rho)_t)$$

where $(\hat{\nu}\times\rho)_t$ gives the distribution at time t as solution of the Liouville equation with initial distribution $\hat{\nu} \times \rho$ on Γ.
Finally, given an initial probability $\hat{\nu}$ on $\hat{\Gamma}$, we can look at the collection of all paths $\omega = (M(x), M(fx), \ldots, M(f^n x)), (n$ fixed) where, first, a reduced state M_0 is drawn according to $\hat{\nu}$ and then, with uniform probability on M_0 a microstate $x \in M_0$ is drawn (so that $M(x) = M_0$). We denote the resulting distribution on these paths by $\boldsymbol{P}_{\hat{\nu}}$; it defines the path space measure on trajectories in $\hat{\Gamma}$.

3.4. **Entropies.** There will be various types of entropies appearing in what follows, each having their specific role and meaning. There is first the Shannon entropy $S(\mu)$, a functional on probability densities μ on Γ:

$$S(\mu) \equiv -\int dx\, \mu(x) \ln \mu(x) \tag{3.2}$$

We can also define the Shannon entropy $\hat{S}(\hat{\mu})$ for probability laws on $\hat{\Gamma}$ through

$$\hat{S}(\hat{\mu}) \equiv -\sum_M \hat{\mu}(M) \ln \hat{\mu}(M) \tag{3.3}$$

There is secondly the Boltzmann entropy S_B which is first defined on $M \in \hat{\Gamma}$, and then for each microstate $x \in \Gamma$ as

$$\hat{S}_B(M) \equiv \ln |M|; \quad S_B(x) \equiv \hat{S}_B(M(x)) \tag{3.4}$$

The dependence on the number of particles N is ignored here as it shall be of no concern. Most frequently, we have in mind here macroscopic variables (such as density and/or velocity profile(s)) for characterizing the reduced states. Any two microstates on Γ are *a priori* equivalent but if we randomly pick a microstate x from Γ, the chance that its reduced state $M(x)$ equals $M \in \hat{\Gamma}$ increases with greater Boltzmann entropy $\hat{S}_B(M)$. We can then expect, both for the forward evolution and for the backward evolution (positive or negative times) that the Boltzmann entropy should increase. This time-reflection invariance of the increase of entropy is an instance of the dynamic reversibility and it interprets the paradoxical words of Boltzmann when speaking about the increase of entropy (minus the $H-$ functional) “that every point of the $H-$curve is a maximum,” see [12]. From this, equilibrium is understood as the state of maximal entropy, given the constraints

in terms of macroscopic values that define the equilibrium conditions (such as energy, volume and number of particles). Upon varying the constraints, this maximal Boltzmann entropy will behave as the thermodynamic entropy (defined operationally). Yet, even out of equilibrium the Boltzmann entropy makes sense which is essential and needed even to discuss fluctuations around equilibrium. It then continues to correspond to the thermodynamic entropy in the close to equilibrium treatments of irreversible processes.
The Boltzmann entropy thus tells how typical a macroscopic appearance is from counting its possible microscopic realizations. Also the Shannon entropy has its origin in counting (for example in evaluating Stirling's formula or other combinatorial computations) and it is therefore not surprising that there are relations between the two. For our context, the following identity between Shannon and Boltzmann entropies holds:

$$S(\hat{\nu} \times \rho) - \hat{S}(\hat{\nu}) = \sum_M \hat{\nu}(M)\hat{S}_B(M) \tag{3.5}$$

Thirdly, we will need the Gibbs entropy $S_G(\hat{\nu})$ which is a functional on the statistics $\hat{\nu}$ of reduced states:

$$S_G(\hat{\nu}) \equiv \sup_{p(\mu)=\hat{\nu}} S(\mu) \tag{3.6}$$

Equivalently,

$$S_G(\hat{\nu}) = S(\hat{\nu} \times \rho) \tag{3.7}$$

because we always have $p(\hat{\nu} \times \rho) = \hat{\nu}$ and a standard computation shows that $S(\hat{\nu} \times \rho) \geq S(\mu)$ for every density μ on Γ for which also $p(\mu) = \hat{\nu}$ (Gibbs variational principle). At the same time, from (3.5), note that for $\hat{\nu} = \delta(M - \cdot)$,

$$\hat{S}_B(M) = S(\hat{\nu} \times \rho)$$

Combining this with (3.7), we observe that in case $\hat{\nu}$ concentrates on one $M \in \hat{\Omega}$, then the Boltzmann and the Gibbs entropies coincide:

$$\hat{S}_B(M) = S_G(\delta(M - \cdot)) \tag{3.8}$$

which indicates that the Gibbs entropy is, mathematically speaking, an extension of the Boltzmann entropy since it lifts the definition of $\hat{S}_B$ to the level of distributions on $\hat{\Gamma}$.

Another application of the Gibbs variational formula (3.6) concerns the change of entropy under the Hamiltonian dynamics. This was emphasized by Jaynes, see e.g. [10]. Suppose the system is initially (at time $t_0 = 0$) prepared with density $\hat{\nu} \times \rho$ (i.e., microcanonically with reduced states sampled from $\hat{\nu}$). Then, according to the Liouville equation, at time t we obtain the density $(\hat{\nu} \times \rho)_t$. But only the reduced state is monitored, i.e., its projection $p((\hat{\nu} \times \rho)_t) = \hat{\nu}_t$. This is for example obtained from the empirical distribution of the macrovariables. From (3.6) it follows that $S_G(\hat{\nu}_t) \geq S_G(\hat{\nu})$ because, by Liouville's theorem, $S((\hat{\nu} \times \rho)_t) = S(\hat{\nu} \times \rho) = S_G(\hat{\nu})$. We call the difference

$$S_G(\hat{\nu}_t) - S_G(\hat{\nu}) = \text{the (Gibbs-) entropy production}$$

It is always non-negative. From (3.8), if we initially prepare the system in some specific reduced state $M_0 \in \hat{\Gamma}$, then this (Gibbs-) entropy production equals, in fact, $S_G(\hat{\nu}_t) - S_B(x) \geq 0, x \in M_0$. If the set of reduced variables allows a hydrodynamic description in which, reproducibly for almost all $x \in M_0$, $M(\phi_t x) = M_t \in \hat{\Gamma}$, then the experimentalist will, for all practical purposes, identify $\hat{\nu}_t$ with $\delta(M_t - \cdot)$ and the (Gibbs-) entropy production is then given by the change $\hat{S}_B(M_t) - \hat{S}_B(M_0) = S_B(\phi_t x) - S_B(x)$ in Boltzmann entropy. In other words, the (Gibbs-) entropy production then coincides with the (Boltzmann-) entropy production. Yet, from this we see that, under second law conditions, the inequality $S_G(\hat{\nu}_t) \geq S_G(\hat{\nu})$ obtained from the Gibbs variational principle is doing great injustice to the actual difference between initial and final Boltzmann entropies: the value of $\hat{S}_B(M_0)$ will be very small compared to the equilibrium entropy $\ln|\Omega_E|$ ($\simeq \hat{S}_B(M_t)$ for $|M_t| \simeq |\Gamma|$) if M_0 corresponds to a preparation in a special nonequilibrium state. As it is often correctly emphasized, the stone-wall character of the second law derives from the great discrepancy in microscopic and macroscopic scales as a result of the huge number of degrees of freedom in a thermodynamic system. Moreover, a theoretical advantage of considering $S_B(\phi_t x) - S_B(x)$ is that this is directly defined on the phase space Γ and in fact allows a microscopic derivation of the second law based on statistical considerations concerning the initial state, see [8, 14] for recent discussions. Note however that this advantage also implies the challenge to relate the Boltzmann entropy with more operational definitions of entropy as practiced in thermodynamics of irreversible processes where entropy production appears as the products of fluxes and forces, as obtained from entropy balance equations. Yet, irreversible thermodynamics is restricted by the assumption of local equilibrium whose validity requires systems close to equilibrium.
Finally, there is the dynamical entropy S_K that is an immediate extension of the Boltzmann entropy but defined on trajectories: for a given trajectory $\omega = (M_0, M_1, \ldots, M_n)$ in $\hat{\Gamma}$, we put

$$S_K(\omega) \equiv \ln|\cap_{j=0}^{n} f^{-j} M_j| \tag{3.9}$$

counting the microstates $x \in \Gamma$ for which $M(f^j x) = M_j (\equiv \omega_j)$, $j = 0, \ldots, n$. In ergodic theory, this dynamical entropy (3.9) is related to the Kolmogorov-Sinai entropy via Breiman's theorem, [22]. In this way, the Kolmogorov-Sinai entropy gives the asymptotics of the number of different types of trajectories as time tends to infinity. Note however that in the physical case we have in mind, $\hat{\Gamma}$ does not correspond to some kind of mathematical coarse graining and there is no way in which it is assumed generating nor do we intend to let it shrink $\hat{\Gamma} \to \Gamma$.

3.5. **Transient versus steady state regime.** The family of nonequilibrium states is much more rich and varied than what is encountered in equilibrium. It is often instructive to divide nonequilibrium phenomena according to their appearance in the transient versus the steady state regime. The simplest example of a transient regime is when the total system starts from a nonequilibrium state and is allowed to relax to equilibrium. Steady state on the other hand refers to a maintained nonequilibrium state. For this we need an open system that is driven away from equilibrium by an environment. All of this however strongly depends on the type of variables that are considered and over what length and time scales. In this paper we take the point of view that the steady state is a special or limiting case of the transient situation. Since the fundamental dynamics is Hamiltonian for a closed system, any attempt to give a microscopic definition of entropy production must start there. We can then discuss the limiting cases or approximations through which we are able to define also the mean entropy production rate in the steady state. In any case, the identification of the statistical mechanical entropy production as function of the trajectory must be independent from the regime, be it transient or steady.

4. Entropy production: closed systems

Our main goal in the following two sections is to show how, via time-reversal, we can define a function on path space which will be recognized as the variable or statistical mechanical entropy production. It enables us to compute the time-derivative of the entropy production, the entropy production rate. Since this function is defined on trajectories, we can also study its fluctuations.
The situation to have in mind here is that of the transient regime in a closed system; entropy production is the change of entropy.

Recall (3.1). We start by observing that for any two reduced states $M_0, M_n \in \hat{\Gamma}$ and for every microscopic trajectory γ corresponding to a sequence of microstates starting in M_0 and ending in M_n,

$$\ln \frac{\text{Prob}[(x, fx, \ldots, f^n x) = \gamma \,|\, M(x) = M_0]}{\text{Prob}[(x, fx, \ldots, f^n x) = \gamma\Theta \,|\, M(x) = \pi M_n]} = \hat{S}_B(M_n) - \hat{S}_B(M_0) \quad (4.1)$$

where $\gamma\Theta$ is the time-reversed microscopic trajectory. That $\gamma\Theta$ is also a microscopic trajectory is an expression of dynamic reversibility. This identity (4.1) follows because, given the initial reduced state, the probability that a specific microscopic trajectory is realized only depends on the probability of the initial microstate. But since we know to what reduced state it belongs, that probability is just the exponential of minus the Boltzmann entropy.

While the previous relation indicates that time-reversal transformations are able to pick up the entropy production, we cannot in practice sample microscopic trajectories. In order to lift these relations to the level of trajectories on $\hat{\Gamma}$, we should also relax the condition that we start in a fixed reduced state; if we only know the reduced state the dynamics started from, we will not know in what specific reduced state we land after time t. For this additional uncertainty, there is a small price to be paid.

Let $\omega = (M_0, M_1, \ldots, M_n)$ be a possible trajectory on $\hat{\Gamma}$. Its time-reversal is $\omega\Theta = (\pi M_n, \ldots, \pi M_0)$. Let $\hat{\mu}$ and $\hat{\nu}$ be two probabilities on $\hat{\Gamma}$. We ask for the ratio of probabilities that the actual trajectory coincides with ω and with $\omega\Theta$, conditioned on starting the microscopic trajectory sampled from $\hat{\mu} \times \rho$ and from $\hat{\nu}\pi \times \rho$ respectively:

$$\frac{\text{Prob}_{\hat{\mu}\times\rho}[\text{trajectory } = \omega]}{\text{Prob}_{\hat{\nu}\pi\times\rho}[\text{trajectory } = \omega\Theta]} = \frac{\hat{\mu}(M_0)}{\hat{\nu}(M_n)} \frac{|M_n|}{|M_0|} \tag{4.2}$$

More precisely, this wants to say that the corresponding path space measures have a density with respect to each other, given by

$$\frac{d\boldsymbol{P}_{\hat{\mu}}}{d\boldsymbol{P}_{\hat{\nu}\pi}\Theta}(\omega) = \frac{\hat{\mu}(M_0)}{\hat{\nu}(M_n)} \frac{|M_n|}{|M_0|} \tag{4.3}$$

To prove this, it suffices to see that, on the one hand

$$\hat{\mu} \times \rho\,(\cap_{j=0}^n f^{-j} M_j) = \frac{\hat{\mu}(M_0)}{|M_0|} |\cap_{j=0}^n f^{-j} M_j|$$

and, on the other hand, for the denominator in the left-hand side of (4.2)-(4.3),

$$|\cap_{j=0}^n f^{-j} \pi M_{n-j}| = |\cap_{j=0}^n f^{j} M_{n-j}| = |\cap_{j=0}^n f^{j-n} M_{n-j}|$$

where we first used $\pi f^{-1}\pi = f$ and then the stationarity of the Liouville measure under f. Hence, the factor $|\cap_{j=0}^n f^{-j}M_j|$ will cancel when taking the ratio as in (4.2)-(4.3); this expresses the time-reversal invariance of the dynamical entropy, $S_K(\omega) = S_K(\Theta\omega)$, which excludes it as candidate for entropy production.

The most interesting case is obtained by taking $\hat{\nu} = \hat{\mu}_t$ for $t = n\delta$ in (4.2) or (4.3). Remember that $\hat{\mu}_t = p((\hat{\mu}\times\rho)_t)$ is the projection on the reduced states of the measure at time t when started from the constrained equilibrium $\hat{\mu}\times\rho$. We then get as a direct consequence of (4.3):

Proposition 4.1. *For every probability $\hat{\mu}$ on $\hat{\Gamma}$,*

$$\ln\frac{d\boldsymbol{P}_{\hat{\mu}}}{d\boldsymbol{P}_{\hat{\mu}_t\pi}\Theta}(\omega) = [S_B(\phi_t x) - S_B(x)] + [-\ln\hat{\mu}_t(M(\phi_t x)) + \ln\hat{\mu}(Mx)] \tag{4.4}$$

for all trajectories $\omega = (M(x), M(fx), \ldots, M(f^n x))$ in $\hat{\Gamma}$, $x\in\Gamma$, $t = n\delta$.

The right-hand side of (4.4) contains two contributions. The first difference of Boltzmann entropies has already appeared (alone) in (4.1) when the comparison was made between probabilities for microscopic trajectories. The second contribution to (4.4) can thus be viewed as originating from the 'stochasticity' of the reduced dynamics. Note in particular that even when $\hat{\mu}$ is concentrated on some $M\in\hat{\Gamma}$, then $\hat{\mu}_t$ is still smeared out over various possible reduced states. Yet, this second contribution can be expected to be very small under second law conditions. After all, if $\hat{\mu}(M) = \delta(M - M(x))$, then $\hat{\mu}(Mx) = 1$. And if $M(f^n x)$ is large in the sense that 'typically' almost all $z\in\Gamma$ get into $M(f^n x)$ after a sufficient time $t = n\delta$, then, we expect, $|M(\phi_t x)\cap\phi_t M(x)| \simeq |M(x)|$ so that by (3.1), also $\hat{\mu}_t(M(\phi_t x)) \simeq 1$ and hence only the first Boltzmann contribution survives. Of course, for smaller systems, there is hardly a notion of what is typical but we can see what to expect in general.

Let $\boldsymbol{E}_{\hat{\mu}}$ stand for the expectation with respect to the path space measure $\boldsymbol{P}_{\hat{\mu}}$.

Proposition 4.2. *Denote (4.4) by*

$$R^t_{\hat{\mu}} \equiv \ln\frac{d\boldsymbol{P}_{\hat{\mu}}}{d\boldsymbol{P}_{\hat{\mu}_t\pi}\Theta}$$

Then,

$$\boldsymbol{E}_{\hat{\mu}}[e^{-R^t_{\hat{\mu}}}] = 1 \tag{4.5}$$

In particular, its expectation equals the (Gibbs-) entropy production:

$$\boldsymbol{E}_{\hat\mu}[R^t_{\hat\mu}] = S_G(\hat\mu_t) - S_G(\hat\mu) \geq 0 \tag{4.6}$$

Proof. The identity (4.5) is the normalization

$$\boldsymbol{E}_{\hat\mu}[\frac{d\boldsymbol{P}_{\hat\mu_t\pi}\Theta}{d\boldsymbol{P}_{\hat\mu}}] = 1$$

The relation (4.6) follows from (3.5) and the definition (3.6)-(3.7) of Gibbs entropy from inspecting the expectation of the right-hand side of (4.4). The non-negativity of the (Gibbs-) entropy production was already explained under Section 3.4 but it can also be obtained from applying the Jensen (convexity) inequality to (4.5). □

Remark 1: The above calculations have relied heavily on the structure $\hat\mu \times \rho$ of the distributions. We will see in Section 5 what happens in the case where the distribution has the form $\bar\mu \times \nu$ where $\bar\mu$ will refer to the distribution of a subsystem and the ν takes into account the macrostate of the enviroment that further constraints the evolution.
Remark 2: One may wonder about the physical significance of the terms $[-\ln\hat\mu_t(M(\phi_t x)) + \ln\hat\mu(Mx)]$ appearing in (4.4). There is no general answer: they have *a priori* nothing to do with entropy production but their addition can become physically significant to the same extent that $\hat\mu$ and $\hat\mu_t$ are physically motivated. Nevertheless we continue to call the right-hand side of

$$[-\ln\hat\mu_t(\omega_n) + \ln\hat\mu(\omega_0)] + [\hat S_B(\omega_n) - \hat S_B(\omega_0)] = R^t_{\hat\mu}(\omega)$$

the total variable (or statistical mechanical) entropy production $R^t_{\hat\mu}$. It coincides with the (Boltzmann-) entropy production under second law conditions and its expectation is the (Gibbs-) entropy production. Even though $R^t_{\hat\mu}$ does not quite coincide with the change of Boltzmann entropy, it has a useful structure (as ratio of two probabilities) making the studies of its fluctuations much easier, see for example (4.5). The amazing point is that while $\boldsymbol{P}_{\hat\mu}(\omega)$ and $\boldsymbol{P}_{\hat\mu_t\pi}(\Theta\omega)$ both depend on the entire path $\omega = (\omega_0, \omega_1, \ldots, \omega_n)$, their ratio is a state function, only depending on the initial and final states ω_0 and ω_n. We will use it throughout the following.

5. Entropy production: open systems

As a general remark, in the set-up of the present paper, nonequilibrium steady states correspond in reality to a transient regime for the whole system over timescales short enough for the macrostates of the reservoirs not to have changed appreciably while long enough for the internal subsystem to have reached a stationary condition. We start however again from the Hamiltonian dynamics of the total system.

We consider the situation where the complete system consists of one smaller and one larger subsystem. The latter represents an environment to which the smaller subsystem is coupled and it can have the form of one or more thermal reservoirs, for instance. For short, the smaller subsystem will simply be called system. The total phase space is $\Omega = \Omega^0 \times \Omega^1$, where the superscripts 0 and 1 stand for the system and for the environment, respectively. The dynamics of the total system is Hamiltonian and we use the same notation as introduced in Section 3; namely f is the discretized flow and the time-reversal π on Ω is assumed to have the form $\pi = \pi^0 \otimes \pi^1$ (we will soon forget these superscripts). The reduced picture is given by the partition $\hat{\Omega}$ of Ω, with the product structure $\hat{\Omega} = \hat{\Omega}^0 \times \hat{\Omega}^1$. One choice could be taking $\hat{\Omega}^0 = \Omega^0$ in case the system has only a few microscopic degrees of freedom, and the elements of $\hat{\Omega}^1$ corresponding to fixed values of the energies of the individual reservoirs. Preparing the system in the initial state $\hat{\mu}^0$ and, independently, preparing the environment in $\hat{\mu}^1$, we construct the initial distribution on Ω, from which the microstates are sampled, as $(\hat{\mu}^0 \otimes \hat{\mu}^1) \times \rho$. At this moment the microscopic dynamics, conserving the total energy, takes over and the system gets coupled with the environment. We then get, as used in the previous section and as defined in Section 3.3, a path space measure $\boldsymbol{P}_{\hat{\mu}^0 \otimes \hat{\mu}^1}$ for the trajectories.

It is convenient to rephrase the above construction in the following, more formal, way to make a connection with the scenario of the previous section. We introduce yet another coarse-graining in which only the system is observed, while the macrostate of the environment is ignored. It is defined via the map $\bar{p} : \hat{\Omega} \mapsto \hat{\Omega}^0$ which assigns to every $(M, E) \in \hat{\Omega}$ its first coordinate $M \in \hat{\Omega}^0$. That means that we actually deal with two successive partitions: $\hat{\Omega}$ is a partition of the original phase space Ω (involving both system and environment) and $\bar{\Omega}$ is taken as a further partition of $\hat{\Omega}$. $\bar{\Omega}$ can be identified with $\hat{\Omega}^0$ involving only the system's degrees of freedom. The elements of $\bar{\Omega}$ are written as M and those of $\hat{\Omega}$ as (M, E). To every $(y, z) \in \Omega = \Omega^0 \times \Omega^1$ we thus associate an element $My \in \bar{\Omega}$ and an element $(My, Ez) \in \hat{\Omega}$.
On $\bar{\Omega}$ we put the distribution $\bar{\mu} \equiv \hat{\mu}^0$ which stands for the initial statistics of the small subsystem. On $\hat{\Omega}$ we put the distribution $\hat{\mu}$ for which we only ask that $\hat{\mu}(M, E) = \hat{\mu}(M)\hat{\mu}^1(E)$ thus representing the preparation of the environment. The first and crucial observation we

make is that

$$(\hat{\mu}^0 \otimes \hat{\mu}^1) \times \rho = \bar{\mu} \times (\hat{\mu} \times \rho) \tag{5.1}$$

The right-hand side is defined as the generalization of the randomization we introduced in Section 3.3: given a distribution ν on Ω and a distribution $\bar{\mu}$ on $\bar{\Omega}$ we let

$$\bar{\mu} \times \nu \ (y, z) \equiv \bar{\mu}(My) \frac{\nu(y, z)}{\nu(My)}$$

Take here $\nu = \hat{\mu} \times \rho$, then $\nu(y, z) = \hat{\mu}(My, Ez)/|My||Ez|$ and $\nu(My) = \hat{\mu}(My)$. Therefore,

$$\bar{\mu} \times (\hat{\mu} \times \rho) \ (y, z) = \frac{\hat{\mu}^0(My)\hat{\mu}^1(Ez)}{|My||Ez|}$$

which proves the identity (5.1).

The representation (5.1) enables us to consider the initial distribution as constructed directly from the $\bar{\mu}$ by randomizing it with the *a priori* distribution $\hat{\mu} \times \rho$ which is not time-invariant under the dynamics and depends on the initial state of the environment, cf. the first remark at the end of Section 4. The probability of a trajectory $\bar{\omega} = (\bar{\omega}_0, \dots, \bar{\omega}_n)$ in $\bar{\Omega}$, i.e., of the system, may then be evaluated as follows:

$$\begin{aligned} \boldsymbol{P}_{\hat{\mu}^0 \otimes \hat{\mu}^1}(\bar{\omega}) &\equiv (\bar{\mu} \times (\hat{\mu} \times \rho))(\cap_{j=0}^n f^{-j} \bar{\omega}_j) = \frac{\bar{\mu}(\bar{\omega}_0)}{\hat{\mu}(\bar{\omega}_0)} (\hat{\mu} \times \rho)(\cap_{j=0}^n f^{-j} \bar{\omega}_j) \\ &= \frac{\bar{\mu}(\bar{\omega}_0)}{\hat{\mu}(\bar{\omega}_0)} \sum_{\bar{p}(\hat{\omega}) = \bar{\omega}} \boldsymbol{P}_{\hat{\mu}}(\hat{\omega}) \end{aligned} \tag{5.2}$$

where $\boldsymbol{P}_{\hat{\mu}}(\hat{\omega})$ is the probability of the trajectory $\hat{\omega}$ on $\hat{\Omega}$ started from $\hat{\mu}$:

$$\boldsymbol{P}_{\hat{\mu}}(\hat{\omega}) \equiv (\hat{\mu} \times \rho)(\cap_{j=0}^n f^{-j} \hat{\omega}_j)$$

and we sum in (5.2) over all trajectories on $\hat{\Omega}$ (i.e., for environment and system together) that coincide in their first coordinate with the given trajectory $\bar{\omega}$. Similarly, for the time-reversed trajectory $\Theta\bar{\omega}$ with the system's initial distribution $\bar{\nu}\pi$ we have the probability (the initial microstates sampled from the distribution $\bar{\nu}\pi \times (\hat{\mu}\pi \times \rho) = (\hat{\nu}^0\pi \otimes \hat{\mu}^1\pi) \times \rho$)

$$\begin{aligned} \boldsymbol{P}_{\hat{\nu}^0\pi \otimes \hat{\mu}^1\pi}(\Theta\bar{\omega}) &= \frac{\bar{\nu}(\bar{\omega}_n)}{\hat{\mu}(\bar{\omega}_n)} \sum_{\bar{p}(\hat{\omega}) = \Theta\bar{\omega}} \boldsymbol{P}_{\hat{\mu}\pi}(\hat{\omega}) \\ &= \frac{\bar{\nu}(\bar{\omega}_n)}{\hat{\mu}(\bar{\omega}_n)} \sum_{\bar{p}(\hat{\omega}) = \bar{\omega}} \frac{\hat{\mu}(\hat{\omega}_n)}{\hat{\mu}(\omega_0)} \frac{|\hat{\omega}_0|}{|\hat{\omega}_n|} \boldsymbol{P}_{\hat{\mu}}(\hat{\omega}) \end{aligned} \tag{5.3}$$

where we used a version of (4.3). As always, we want to take the ratio of (5.2) and (5.3). We write this out in the most explicit form:

$$\frac{\boldsymbol{P}_{\hat{\mu}^0\otimes\hat{\mu}^1}(\bar{\omega})}{\boldsymbol{P}_{\hat{\nu}^0\pi\otimes\hat{\mu}^1\pi}(\Theta\bar{\omega})} = \frac{\hat{\mu}^0(M_0)}{\hat{\nu}^0(M_n)} r_n^{-1}(\bar{\omega}) \tag{5.4}$$

where

$$r_n(\bar{\omega}) \equiv \frac{\sum_{E_0,\ldots,E_n} \frac{\hat{\mu}^1(E_n)}{\hat{\mu}^1(E_0)} \frac{|M_0||E_0|}{|M_n||E_n|} \boldsymbol{P}_{\hat{\mu}}[(M_0,E_0),\ldots,(M_n,E_n)]}{\sum_{E_0,\ldots,E_n} \boldsymbol{P}_{\hat{\mu}}[(M_0,E_0),\ldots,(M_n,E_n)]}$$

for a trajectory $\bar{\omega} = (M_0,\ldots,M_n)$ of the system and the sums are over trajectories $(E_0,\ldots,E_n)$ of the environment.

The identity (5.4) is still exact and general. To proceed, we choose first to be more specific about the nature of the environment. As example, we suppose that the environment consists of m heat baths which are taken very large and which are prepared at inverse temperatures $\beta_1,\ldots,\beta_m$. This means that $\hat{\Omega}^1$ and $\hat{\mu}^1$ are split further as $m-$fold products and that the trajectories of the environment are obtained from the successive values of the energies E_i^k at times $i\delta$ in all heat baths, $k=1,\ldots,m$. We also suppose that these reservoirs are spatially separated, each being in direct contact only with the system. Through these system-heat bath interfaces a heat current will flow changing the energy contents of each reservoir. It implies that even though the initial energies E_0^k are sampled from $\hat{\mu}^1$ giving inverse temperatures β_k to each of the heat baths, *a priori* it need not be that the final energies E_n^k can be considered as sampled corresponding to the same temperatures. At this moment we need a steady state assumption for the reservoirs: that they maintain the same temperature during the evolution, or more precisely,

A1: The path space measure $\boldsymbol{P}_{\hat{\mu}}$ gives full measure to those trajectories for which

$$|E_n^k - E_0^k| \leq o(\sqrt{V})$$

of the order of less than the square root of the size of the volume V of the environment.

For simplicity, we have used here $o(\sqrt{V})$ as a safe estimate. One could assume much less: the energy differences $|E_n^k - E_0^k|$ will be of the order of the product of the surface area $\partial_k\Lambda$ through which heat will flow between the $k-$th heat bath and the system, and the heat current (i.e., the flow of energy per unit area and per unit time) and the time $t = n\delta$. One can recognize this from the calculations in Appendix B. We need this assumption A1 to get rid of the ratio $\hat{\mu}^1(E_n)/\hat{\mu}^1(E_0)$

in (5.4). Under A1, this ratio is essentially equal to one, when the initial dispersion of the energy values under $\hat{\mu}^1$ is much larger than the changes of energy. For simplicity (but without loss of generality) we can suppose that for each heat bath of spatial size V, $\hat{\mu}^1$ gives the uniform distribution over an interval $[\mathcal{E}^k - \varepsilon, \mathcal{E}^k + \varepsilon]$ where we take $\mathcal{E}^k = \mathcal{O}(V)$ and $\varepsilon = o(V)$. This is just the simplest representation for an initial distribution that is peaked at energy value $\mathcal{E}$ with deviations of order $\varepsilon = \sqrt{V}$ say. Within such an interval, the temperatures $1/\beta_k$ are essentially constant if the size of the interval is large compared with the possible changes of the energy due to the flows from the system. These statements become sharp in the limit $V \to \infty$ but for finite reservoirs this steady state assumption can be expected realized over times $t \leq t^\star$ with $t^\star = t^\star(V)$ growing with V less than as $\sqrt{V}$. As conclusion, we set $\hat{\mu}^1(E_n)/\hat{\mu}^1(E_0) = 1$ in (5.4).

For notation, we denote by

$$\hat{S}^0_B(M_n) - \hat{S}^0_B(M_0) = \ln \frac{|M_n|}{|M_0|} \tag{5.5}$$

the change of Boltzmann entropy of the system. Usually, at least in close to equilibrium treatments, this change is divided into two parts: one corresponding to the entropy production properly speaking and one term corresponding to the entropy current through the surface of the system; one refers to it as an entropy balance equation, see Appendix D for a short discussion. The latter, the entropy current, is responsible for the change of entropy in the reservoirs, here

$$\ln \frac{|E_n|}{|E_0|} = \sum_{k=1}^{m} [\hat{S}^k_B(E^k_n) - \hat{S}^k_B(E^k_0)] \tag{5.6}$$

the sum of changes of the Boltzmann entropy in each bath. The sum of (5.5) and (5.6) is the total change of entropy (where total refers to the closed system consisting of the (smaller) system and all the reservoirs) and thus equals the entropy production. We write it as

$$\begin{aligned}\Delta S_B(\bar{\omega}) &\equiv \hat{S}_B(M_n) - \hat{S}_B(M_0) \qquad\qquad (5.7)\\ &- \ln \frac{\sum_{E_0,\ldots,E_n} e^{-\sum_{k=1}^m [\hat{S}^k_B(E^k_n) - \hat{S}^k_B(E^k_0)]} \boldsymbol{P}_{\hat{\mu}}[(M_0,E_0),\ldots,(M_n,E_n)]}{\sum_{E_0,\ldots,E_n} \boldsymbol{P}_{\hat{\mu}}[(M_0,E_0),\ldots,(M_n,E_n)]}\end{aligned}$$

As in (4.4), it is natural to take $\hat{\nu}^0 = \hat{\mu}^0_t \equiv \bar{p}(\bar{\mu} \times (\hat{\mu} \times \rho))_t$ in (5.4) which corresponds to the projection at time $t = n\delta$ on the system, and we obtain a first

Analogue of Proposition 4.1:

$$\ln \frac{\boldsymbol{P}_{\hat{\mu}^0\otimes\hat{\mu}^1}(\bar{\omega})}{\boldsymbol{P}_{\hat{\mu}^0_t\pi\otimes\hat{\mu}^1\pi}(\Theta\bar{\omega})} = \Delta S_B(\bar{\omega}) + [-\ln \hat{\mu}^0_t(\bar{\omega}_n) + \ln \hat{\mu}^0(\bar{\omega}_0)] \tag{5.8}$$

We can still do better by reconsidering (5.7), in particular the expectation over the path-space measure of the exponential change of Boltzmann entropies in the heat baths. These changes are caused by the heat dissipated in each of the reservoirs and it therefore corresponds to the entropy current. Since this is the energy current divided by the temperature of the reservoir, it should be possible to express it directly in terms of the microscopic trajectory over the surface separating the heat bath from the system. The basic question is now in what sense the trajectory $\bar{\omega}$ of the system determines the trajectory $\hat{\omega}$ of the total system. In the context of Hamiltonian dynamics it is not hard to see that (again, provided that the reservoirs are coupled to different parts of the system) the trajectory $\hat{\omega}$ is uniquely determined by its projection onto the system, $\bar{\omega}$, and by the initial energies of the reservoirs, E^k_0. We formulate this in the form of another assumption:

A2: Let $\hat{\omega}$ and $\hat{\omega}'$ be two trajectories of the total system such that $\bar{p}(\hat{\omega}) = \bar{p}(\hat{\omega}') = \bar{\omega}$. Then, for typical trajectories obtained as successive reduced states from the Hamiltonian dynamics, the energy changes $E^k_n - E^k_0$ depend only on $\bar{\omega}$.

This should be understood in the sense of all allowed trajectories, typical here referring to the path-space measure $\boldsymbol{P}_{\hat{\mu}}$. This assumption is too strong. The problem is that we also consider reduced states on the level of the system itself and that the $\bar{\omega}$ is not the continuous time microscopic trajectory of the system. It would for example have been better to use a finer time-scale for the evolution of the reduced states of the system (compared with that of the reservoirs). One could remedy that but we prefer to stick to A2 for simplicity, see Appendix B.
Assuming A2, we already know that $\Delta E^k \equiv E^k_n - E^k_0$ only depends on $\bar{\omega} : \Delta E^k = \Delta E^k(\bar{\omega})$. So, the change of Boltzmann entropy in each reservoir, $\hat{S}^k_B(E^k_n) - \hat{S}^k_B(E^k_0) = \hat{S}^k_B(E^k_0 + \Delta E^k(\bar{\omega})) - \hat{S}^k_B(E^k_0)$, depends on $\bar{\omega}$ and E^k_0. In order to get rid of the dependence on the initial state of the reservoirs, we must again use the distribution $\hat{\mu}^1$ and that the reservoirs are large. After all, thermodynamic behavior implies that the $E^k_i = \mathcal{O}(V)$ and $\hat{S}^k_B(E^k_i) = \mathcal{O}(V)$, all of the order of the volume V of the reservoirs, while $\beta_k \equiv \partial\hat{S}^k_B(E^k_i)/\partial E^k_i$ is kept fixed. This is again

our steady state assumption A1 for the environment; the reservoirs are heat baths at a fixed temperature. The change of Boltzmann entropy in each of the reservoirs is then

$$\hat{S}^k_B(E^k_n) - \hat{S}^k_B(E^k_0) = \beta_k \Delta E^k(\bar{\omega}) + \mathcal{O}(\frac{1}{\sqrt{V}})$$

for all trajectories of the system $\bar{\omega}$ and essentially all initial energies $E^k_0 \in (\mathcal{E}^k - \varepsilon, \mathcal{E}^k + \varepsilon)$ (most of trajectories started inside the interval will not leave it). This implies that the total change of entropy appearing in (5.7) and in (5.8) is, in good approximation,

$$\Delta S_B(\bar{\omega}) = \hat{S}^0_B(\bar{\omega}_n) - \hat{S}^0_B(\bar{\omega}_0) + \sum_k \beta_k \Delta E^k(\bar{\omega}) \tag{5.9}$$

What we have gained with respect to (5.7)-(5.8) is that this variable entropy production only depends on $\hat{\mu}^1$ through the initial temperatures of the reservoirs.
We denote

$$R^{\hat{\mu}^1,t}_{\hat{\mu}^0}(\bar{\omega}) \equiv \ln \frac{d\boldsymbol{P}_{\hat{\mu}^0 \otimes \hat{\mu}^1}}{d\boldsymbol{P}_{\hat{\mu}^0_t \pi \otimes \hat{\mu}^1 \pi}\Theta}(\bar{\omega}) \tag{5.10}$$

We conclude with the final result obtained under the assumptions above:

Analogue of Proposition 4.1:

$$R^{\hat{\mu}^1,t}_{\hat{\mu}^0}(\bar{\omega}) = \hat{S}^0_B(\bar{\omega}_n) - \hat{S}^0_B(\bar{\omega}_0) + \sum_k \beta_k \Delta E^k(\bar{\omega}) + [-\ln \hat{\mu}^0_t(\bar{\omega}_n) + \ln \hat{\mu}^0(\bar{\omega}_0)] \tag{5.11}$$

There are two big modifications with respect to (4.4) for closed systems. First, it is important to remember that the $\Delta E^k(\bar{\omega})$ in (5.11) are in general not differences of the form $E^k_n(\bar{\omega}_n) - E^k_0(\bar{\omega}_0)$ but they represent the heat flow depending on the complete path $\bar{\omega}$. So the right-hand side of (5.11) is not a difference. Secondly, here it is very well possible to take $\hat{\mu}^0_t = \hat{\mu}^0$ at least for small enough times t (compared to $\sqrt{V}$) what refers to the full steady state. In other words, we can study the stationary regime where the distribution $\hat{\mu}^0$ of the system is time-invariant.

Remark 1: The same construction applies of course also when the system is coupled to only one reservoir or to various reservoirs at the same temperature β^{-1}. We take $\hat{\Omega}^0 = \Omega^0 = \bar{\Omega}$ so that the system is described via its microscopic states y. The trajectory $\bar{\omega}$ gives successive microscopic states $\bar{\omega} = (y_0, \ldots, y_n)$ and the first two terms on the right-hand side of (5.11) are identically zero. By energy conservation, the total change of energy $\sum_k(E_n^k - E_0^k) = \Delta E$ in the reservoirs is always of the form $\Delta E = H(y_0) - H(y_n)$, the difference of the initial and final energies of the system. Therefore, in (5.9),

$$\Delta S_B(\bar{\omega}) = \beta[H(y_0) - H(y_n)]$$

It is interesting to see that then, when taking $\hat{\mu}^0(y) = \hat{\mu}_t^0(y) \sim \exp[-\beta H(y)]$ a Gibbs measure at inverse temperature β, the expression (5.11) becomes zero.

Remark 2: The same construction applies also to other scenario's (instead of via heat reservoirs) but it needs some change in notation. As an example of another physical mechanism we can consider the system coupled to a heat bath at constant temperature β^{-1} where some parameters (e.g. interaction coefficients) in the interaction of the components of the system are changed. This means that the effective Hamiltonian $H(\tau) \equiv H(\lambda(\tau), y), \tau \in [0, t]$, of the system is time-dependent with final value $H_f(y) \equiv H(\lambda(t), y)$ and initial value $H_i(y) \equiv H(\lambda(0), y)$. To change the parameter λ from $\lambda(0)$ to $\lambda(t)$ some heat must flow from the bath into the system so that the change of entropy of the bath equals $\Delta S_B = -\beta[H_f(y_n) - H_i(y_0) - W_t]$ where W_t is the work done over the time $[0, t]$. If we assume that the initial distribution $\hat{\mu}^0 = \exp[-\beta H_i]/Z_i$ and the final distribution $\hat{\mu}_t^0 = \exp[-\beta H_f]/Z_f$ are describing equilibrium with respect to the Hamiltonians H_i and H_f respectively, then (5.11) becomes

$$R_{\hat{\mu}^0}^{\hat{\mu}^1,t}(\bar{\omega}) = \beta W_t(\bar{\omega}) - \beta \Delta F \tag{5.12}$$

where $\Delta F \equiv -\beta^{-1}[\ln Z_f - \ln Z_i]$ is the change of (equilibrium) Helmholtz free energy.

This and the previous remark also indicate that the physical significance of the terms $-\ln \hat{\mu}_t^0(\bar{\omega}_n) + \ln \hat{\mu}^0(\bar{\omega}_0)$ in (5.11) depends on what can physically be assumed or said about $\hat{\mu}^0$ and $\hat{\mu}_t^0$. This can be different from case to case. Yet here again, as already said in Remark 2 of the previous section, while these terms have *a priori* nothing to do with entropy production, adding them gives rise to a more convenient form, both for the properties of the average (mean entropy production) and for the fluctuations of the entropy production.

The mean entropy production rate in the steady state where $\hat{\mu}_t^0 = \hat{\mu}^0 \equiv \bar{\mu}$ is time-invariant is obtained from taking the average of (5.11) with respect to $\boldsymbol{P}_{\hat{\mu}^0 \otimes \hat{\mu}^1}$. Let $\boldsymbol{E}_{\bar{\mu}}$ stand for the expectation (we do not indicate the dependence on $\hat{\mu}^1$). As is the case for our example, we suppose for simplicity for the rest of this section that $\hat{\mu}^1 = \hat{\mu}^1 \pi$. We have then the

Steady state analogue of Proposition 4.2: The entropy production ΔS_B of (5.9) satisfies

$$\boldsymbol{E}_{\bar{\mu}}[e^{-z\Delta S_B}] = \boldsymbol{E}_{\bar{\mu}}[e^{-(1-z)\Delta S_B}\frac{\bar{\mu}(\pi\omega_n)}{\bar{\mu}(\omega_0)}] \tag{5.13}$$

for all complex numbers z. In particular, its expectation equals the mean entropy current = mean entropy production =

$$\boldsymbol{E}_{\bar{\mu}}[\Delta S_B] = \sum_k \beta_k \boldsymbol{E}_{\bar{\mu}}[\Delta E^k] \geq 0 \tag{5.14}$$

The relation (5.13) expresses a symmetry in the fluctuations of ΔS_B. Modulo some technicalities that amount to estimating space-time boundary terms, as explained in [16, 18], it reproduces almost immediately the Gallavotti-Cohen symmetry, [7]. While it is the theory of smooth dynamical systems that has guided us to it, in our analysis, nothing has remained of a chaoticity hypothesis.
The relation (5.14) states the positivity of the mean entropy production. From its proof (below) we can understand under what (nonequilibrium) conditions, it is in fact strictly positive.

The basic identity that drives fluctuation-symmetry relations is

$$\boldsymbol{E}_{\bar{\mu}}[e^{-z\mathcal{R}(\bar{\omega})}\psi(\omega)] = \boldsymbol{E}_{\bar{\mu}}[e^{-(1-z)\mathcal{R}(\bar{\omega})}\psi(\Theta\omega)] \tag{5.15}$$

for every function ψ of the trajectory of the system and with

$$\mathcal{R}(\omega) \equiv \ln \frac{d\boldsymbol{P}_{\bar{\mu}\otimes\hat{\mu}^1}}{d\boldsymbol{P}_{\bar{\mu}\otimes\hat{\mu}^1}\Theta}(\bar{\omega})$$

This identity (5.15) follows from the very definition of $\mathcal{R}$ as the logarithmic ratio of two probabilities from which also $\mathcal{R}(\Theta\bar{\omega}) = -\mathcal{R}(\bar{\omega})$. The equation (5.13) follows simply by taking for ψ in (5.15), $\psi(\bar{\omega}) = [\bar{\mu}(\bar{\omega}_0)/\bar{\mu}(\pi\bar{\omega}_n)]^z$.
Before we give the proof of (5.14), we give the version for the transient regime of the system (steady state for the reservoirs):
Transient regime analogue of Proposition 4.2: Recall the notation (5.10). Then,

$$\boldsymbol{E}_{\hat{\mu}^0\otimes\hat{\mu}^1}[e^{-R^{\hat{\mu}^1,t}_{\hat{\mu}^0}}] = 1 \tag{5.16}$$

Its expectation equals

$$\boldsymbol{E}_{\hat{\mu}^0\otimes\hat{\mu}^1}[R^{\hat{\mu}^1,t}_{\hat{\mu}^0}] = S_G(\hat{\mu}^0_t) - S_G(\hat{\mu}^0) + \sum_k \beta_k \boldsymbol{E}_{\hat{\mu}^0\otimes\hat{\mu}^1}[\Delta E^k] \geq 0 \tag{5.17}$$

The relation (5.16) for the example (5.12) gives the irreversible work - free energy relation of Jarzynski, [9].

Now to the proofs of (5.14)-(5.17). As before in (4.5), (5.16) just expresses a normalization of the probability measure $\boldsymbol{P}_{\hat{\mu}_t^0\pi\otimes\hat{\mu}^1}\Theta$. The equality in (5.17) follows as in (4.6) from taking the expectation of (5.11). From the Jensen inequality applied to (5.16), we obtain the inequality in (5.17). The relation (5.14) now follows from applying stationarity $\hat{\mu}^0 = \hat{\mu}_t = \bar{\mu}$.

We thus see that the positivity in (5.14) and in (5.17) follows from convexity. By the same argument, the strict positivity will express that the two path space measures $\boldsymbol{P}_{\hat{\mu}^0\otimes\hat{\mu}^1}$ and $\boldsymbol{P}_{\hat{\mu}^0\otimes\hat{\mu}^1}\Theta$ are really different, i.e., applying time-reversal really has an effect. (In [16] this is expressed via the relative entropy between these two path space measures.)

Remark 3: Note that the above fluctuation identities (5.13), (5.15) and (5.16) do not depend on Assumption A2. We can repeat them directly starting from (5.8).

6. Markov approximation

The stochastic processes of the previous sections give the statistics of trajectories for reduced states induced by the Hamiltonian dynamics. The stochasticity does not represent microscopic or intrinsic randomness, whatever that means, and is not an easy substitute for chaoticity. In the present section we make an approximation for this stochastic evolution that does go in the direction of assuming some chaoticity but again on the level of reduced states.

6.1. **Closed systems.**

We refer here to Section 4. Look at the time-evolved measure $(\hat{\mu}\times\rho)_t$ starting from $\hat{\mu}\times\rho$ at time zero: for $t = n\delta$,

$$(\hat{\mu}\times\rho)_t(x) = \hat{\mu}\times\rho(f^{-n}x)$$

Remember that we have used before its projection $\hat{\mu}_t$ on $\hat{\Gamma}$. Observe now that, quite generally, $(\hat{\mu}\times\rho)_t \neq \hat{\mu}_t\times\rho$. That is: the phase space distribution does not remain microcanonical; when two points $x, y \in \Gamma$ fall into the same reduced state ($Mx = My$), it need not be that $(\hat{\mu}\times\rho)_t(x) = (\hat{\mu}\times\rho)_t(y)$. This is an instance of so called memory-effects; the process $\boldsymbol{P}_{\hat{\mu}}$ does certainly not correspond to a Markov process on $\hat{\Gamma}$.

We can obtain a Markovian approximation by forcing uniformization at each step in the evolution. We then define the discrete time Markov approximation via the updating

$$\tilde{\mu}_n = p((\tilde{\mu}_{n-1}\times\rho)_\delta),\ n = 1, 2, \ldots \tag{6.1}$$

or more explicitly, from (3.1),

$$\tilde{\mu}_n(M) = \sum_{M'\in\hat{\Gamma}} \tilde{\mu}_{n-1}(M')\frac{|f^{-1}M \cap M'|}{|M'|}$$

corresponding to the Markov chain on $\hat{\Gamma}$ with transition probabilities $p(M', M) = |f^{-1}M \cap M'|/|M'|$. Naturally it satisfies the detailed balance condition

$$|M|\, p(M, M') = |M'|\, p(\pi M', \pi M) \tag{6.2}$$

or

$$\frac{p(M, M')}{p(\pi M', \pi M)} = e^{\hat{S}_B(M') - \hat{S}_B(M)} \tag{6.3}$$

It is an approximation in the sense that the evolution defined by (6.1) corresponds to a repeated randomization of the 'true' evolution. We expect it to be a good approximation in so far that $|M \cap f^j M'| \simeq |M||M'|/|\Gamma|$. That is to say, for δ large enough for the averaging over the reduced state to be valid. That is a mixing condition but for the evolution over the reduced states (as for Gibbs' inkdrop), see [2] for similar remarks. It also implies relaxation to equilibrium. Usually however this is combined with other limiting procedures through which the reduced variables (or their fluctuations) get an autonomous (stochastic) evolution. Most important in all this however remains the 'proper choice' of reduced states (or, thermodynamic variables).

We now have a Markov chain (X_k) on $\hat{\Gamma}$ with transition probabilities $p(M, M')$ and

$$\mathrm{Prob}_{\tilde{\mu}}[X_n = M_n, \ldots, X_0 = M_0] = \tilde{\mu}(M_0)p(M_0, M_1)\ldots p(M_{n-1}, M_n)$$

is the probability of a trajectory $\omega = (M_0, M_1, \ldots, M_n) \in \hat{\Gamma}^{n+1}$ when the Markov chain was started from the probability measure $\tilde{\mu}$. We have instead of (4.2):

$$\frac{\mathrm{Prob}_{\tilde{\mu}}(X_n = M_n, \ldots, X_0 = M_0)}{\mathrm{Prob}_{\tilde{\nu}\pi}(X_n = \pi M_0, \ldots, X_0 = \pi M_n)} = \frac{\tilde{\mu}(M_0)}{\tilde{\nu}(M_n)} \exp[\sum_{k=0}^{n-1} \ln \frac{p(M_k, M_{k+1})}{p(\pi M_{k+1}, \pi M_k)}] \tag{6.4}$$

Upon substituting (6.2), the exponential in (6.4) equals $|M_n|/|M_0|$ and, perhaps surprisingly, the identity (4.2)-(4.3) is unaffected in the Markov approximation:

$$\frac{\mathrm{Prob}_{\tilde{\mu}}(X_n = M_n, \ldots, X_0 = M_0)}{\mathrm{Prob}_{\tilde{\nu}\pi}(X_n = \pi M_0, \ldots, X_0 = \pi M_n)} = \frac{\tilde{\mu}(M_0)}{\tilde{\nu}(M_n)} \frac{|M_n|}{|M_0|} \tag{6.5}$$

Furthermore, take now $\tilde{\nu} = \tilde{\mu}_n$ of (6.1) and let us denote as in Proposition 4.2,

$$R^n_{\tilde{\mu}}(\omega) \equiv \ln \frac{\text{Prob}_{\tilde{\mu}}(X_n = M_n, \ldots, X_0 = M_0)}{\text{Prob}_{\tilde{\mu}_n \pi}(X_n = \pi M_0, \ldots, X_0 = \pi M_n)}$$

Its expectation, as in (4.6), under the now Markovian path space measure $\boldsymbol{P}_{\tilde{\mu}}$ is

$$\boldsymbol{E}_{\tilde{\mu}}[R^n_{\tilde{\mu}}] = \sum_{\omega \in \hat{\Gamma}^{n+1}} \boldsymbol{P}_{\tilde{\mu}}(\omega) R^n_{\tilde{\mu}}(\omega) = S(\tilde{\mu}|\tilde{\rho}) - S(\tilde{\mu}_n|\tilde{\rho}) \geq 0 \tag{6.6}$$

the difference of relative entropies with respect to the stationary (reversible) probability measure $\tilde{\rho}(M) \equiv |M|/|\Gamma|$; the relative entropy is defined from $S(\tilde{\nu}|\tilde{\rho}) \equiv \sum_M \tilde{\nu}(M) \ln \tilde{\nu}(M)/\tilde{\rho}(M)$. The identities (6.6) and (4.6) are consistent since the Gibbs entropy can be written in terms of this relative entropy as $S_G(\tilde{\nu}) = \ln |\Gamma| - S(\tilde{\nu}|\tilde{\rho})$.

6.2. **Open systems.** We refer here to Section 5. In the same spirit as above, we get the Markov approximation for open systems by following the procedure of Section 5. We now get Markov processes with transition probabilities

$$q(M, M') = \boldsymbol{P}_{M \otimes \hat{\mu}^1}(M, M')$$

where we have understood $\hat{\mu}^0 = \delta(M - \cdot)$. We will again suppose for the enviroment that $\hat{\mu}^1 \pi = \hat{\mu}^1$. These transition probabilities then satisfy, from (5.11),

$$\frac{q(M, M')}{q(\pi M', \pi M)} = \exp \Delta S_B(M, M') \tag{6.7}$$

The measures $\tilde{\mu}$ of above now correspond to the distribution of the internal degrees of freedom (the open system). The important change is that detailed balance may be violated from the action of reservoirs maintained at different but fixed temperatures or chemical potentials. We can for example substitute (5.9) in (6.7) to retain only a local detailed balance condition, that is

$$\frac{q(M, M')}{q(\pi M', \pi M)} = \exp[\hat{S}^0_B(M') - \hat{S}^0_B(M) + \sum_k \beta_k \Delta E^k(M, M')]$$

Depending on the transition $M \to M'$, in particular, where this transition of the state of the system is localized, various terms in the exponential can become zero or non-zero, see also (6.14) below.

While the formal structure of the Markov approximation for open systems runs exactly similar to what we did for closed systems, cf. (6.1), we remark that its validity now requires more than what was mentioned following (6.3). In fact, a competing requirement enters if we wish to maintain assumption A2 of the previous section. Assumption A2 will be more reliable in so far as the δ (i.e., the time steps in the trajectory of reduced states) is smaller while the mixing condition on the level of reduced states that justifies the Markov approximation requires large enough δ. Again, as mentioned following assumption A2, this motivates using different time scales for the evolution of the reduced states in system and enviroment.

We now have a Markov chain (X_k) on $\hat{\Omega}^0$ with transition probabilities $q(M, M')$, and

$$\text{Prob}_{\tilde{\mu}}[X_n = M_n, \ldots, X_0 = M_0] = \tilde{\mu}(M_0) q(M_0, M_1) \ldots q(M_{n-1}, M_n)$$

is the probability of a trajectory $\bar{\omega} = (M_0, M_1, \ldots, M_n)$ when the Markov chain was started from the probability measure $\tilde{\mu}$, we have instead of (4.2):

$$\frac{\text{Prob}_{\tilde{\mu}}(X_n = M_n, \ldots, X_0 = M_0)}{\text{Prob}_{\tilde{\nu}\pi}(X_n = \pi M_0, \ldots, X_0 = \pi M_n)} = \frac{\tilde{\mu}(M_0)}{\tilde{\nu}(M_n)} \exp[\sum_{k=0}^{n-1} \ln \frac{q(M_k, M_{k+1})}{q(\pi M_{k+1}, \pi M_k)}] \quad (6.8)$$

As motivated in Section 5, its logarithm will continue to interest us as variable entropy production.

From (6.8), we see that the variable entropy production is now given by:

$$\sum_k \ln \frac{q(M_k, M_{k+1})}{q(\pi M_{k+1}, \pi M_k)} \quad (6.9)$$

Furthermore, for open systems, the relation (6.6) gets replaced with

$$\boldsymbol{E}_{\tilde{\mu}}[R^n_{\tilde{\mu}}] = \quad (6.10)$$
$$S(\tilde{\mu}_n) - S(\tilde{\mu}) + \sum_{k=0}^{n-1} \sum_{M,M'} \tilde{\mu}_k(M) q(M, M') \ln \frac{q(M, M')}{q(\pi M', \pi M)} \geq 0$$

and there is in general no way to write this as a change in relative entropies $S(\tilde{\mu}|\tilde{\rho}) - S(\tilde{\mu}_n|\tilde{\rho})$. In other words, in general, there is no role for the time-derivative of the relative Shannon entropy as total entropy production. When $\tilde{\mu} = \tilde{\mu}_n$ is stationary for the Markov chain, the right-hand side of the equality in (6.10) gives us the mean entropy production rate as

$$\sum_{M,M' \in \hat{\Gamma}} \tilde{\mu}(M) q(M, M') \ln \frac{q(M, M')}{q(\pi M', \pi M)} \quad (6.11)$$

which (up to the inclusion of the time-reversal involution π) is the standard expression for an effective Markovian dynamics modeling a nonequilibrium steady state, see e.g. [21, 6]. Note that if $\tilde{\mu}$ is stationary under updating with transition probabilities $q(M, M')$, then $\tilde{\mu}\pi$ is stationary under updating with the transition probabilities $\Theta q(M, M') \equiv q(\pi M', \pi M)\tilde{\mu}(\pi M')/\tilde{\mu}(\pi M)$ for the time-reversed process. It is then easy to see that the mean entropy production rate is positive and equal for both stationary processes. Or, the mean entropy production is time-reversal invariant. This again is ultimately a consequence of the dynamic reversibility of the microscopic dynamics and it yields interesting by-products (like Onsager reciprocities) as discussed in [15].
For the pathwise expression of the entropy production rate, we look back at (6.9). The entropy production per time-step is

$$\sigma_B^n(\omega) \equiv \frac{1}{n}\sum_{k=0}^{n-1} \ln \frac{q(\omega_k, \omega_{k+1})}{q(\pi\omega_{k+1}, \pi\omega_k)} \tag{6.12}$$

Note again that $\sigma_B^n(\Theta\omega) = -\sigma_B^n(\omega)$ and that, when $\tilde{\mu}$ is stationary,

$$R_{\tilde{\mu}}^n(\omega)/n = \frac{1}{n}\ln\frac{\tilde{\mu}(\omega_0)}{\tilde{\mu}(\omega_n)} + \sigma_B^n(\omega)$$

This leads again as in (5.13) and in (5.15) almost directly to a Gallavotti-Cohen symmetry, [7, 16].

For a continuous time Markov chain (X_t) on a finite set with transition rates $k(M, M')$, similarly, the entropy production rate in the distribution μ is given by

$$\sigma(\mu) \equiv \frac{1}{2}\sum_{M,M'} [k(M, M')\mu(M) - k(M', M)\mu(M')] \ln \frac{k(M, M')\mu(M)}{k(M', M)\mu(M')} \tag{6.13}$$

(We have set $\pi =$ identity for simplicity.) Let us take

$$k(M, M') = k_0(M, M')e^{\varepsilon\psi(M,M')/2}$$

where k_0 is the rate for a detailed balance evolution with unique reversible measure μ_0. We assume that there is a unique stationary measure μ_ε with ε measuring the distance from equilibrium:

$$\frac{k(M, M')}{k(M', M)} = \frac{\mu_0(M')}{\mu_0(M)} \exp[\varepsilon\psi^{\mathrm{as}}(M, M')] \tag{6.14}$$

Here $\psi^{\mathrm{as}}(M, M') = -\psi^{\mathrm{as}}(M', M) = [\psi(M, M') - \psi(M', M)]/2$ originates in some driving. We did not indicate it but the $k(M, M')$ and therefore the functional σ in (6.13) depend now on ε. One can then check that $\sigma(\mu)$ is minimal for a probability measure $\mu^\star$ which coincides with μ_ε to first order in ε (minimum entropy production principle). A special case of this calculation can be found in [6]. We give the general statement and argument in Appendix C.

We next apply the above scheme for a Markov approximation for closed systems to a diffusion process that appeared in the Onsager-Machlup paper, [19].

7. Application: Gaussian fluctuations

As we have argued before, the entropy production appears as the source term of time-reversal breaking in the logarithm of the probability for a preassigned succession of thermodynamic states. Such calculations were already done to study the fluctuations in irreversible processes in the work of Onsager and Machlup in 1953, [19]. Our previous section is some extension of this, as we will now indicate.

We only redo the very simplest case of [19], their Section 4 for a single thermodynamic variable α obeying the equation (in their notation)

$$R\dot{\alpha} + s\alpha = \epsilon \tag{7.1}$$

We do not explain here the origin of this equation except for mentioning that R relates the thermodynamic force to the flux $\dot{\alpha}$ (assumption of linearity). The constant s finds its origin in an expansion of the thermodynamic entropy function $S_T(\alpha) = S_T(0) - s\alpha^2/2$ around equilibrium. For every $\alpha, S_T(\alpha)$ is the equilibrium entropy when the system is constrained to this macroscopic value and can be identified with the Boltzmann entropy $\hat{S}_B(\alpha)$ (up to the thermodynamic limit) which is also defined outside equilibrium. From the expansion of the entropy, the thermodynamic force $dS_T/d\alpha$ depends linearly on the variable α (Gaussian fluctuations). The right-hand side of (7.1) is purely random (white noise) with variance $2R$. In this way the oscillator process $d\alpha = -s/R\ \alpha dt + \sqrt{2/R}\, dW_t$ with W_t a standard Wiener process, is obtained for the variable α.

The work in [19] is then to calculate the probability of 'any path.' These are the trajectories we had before. With the current methods of stochastic calculus, this is not so difficult.

We proceed with (7.1). Using the path-integral formalism we can write the 'probability' of any path $\omega = (\alpha(\tau), \tau \in [0,t])$ with respect to the flat path space 'measure' $d\omega = [d\alpha(\tau)]$:

$$\mathrm{Prob}_{\tilde{\mu}}(\omega) \simeq \tilde{\mu}(\alpha(0))\, e^{-\mathcal{A}(\omega)}$$

for some initial distribution $\tilde{\mu}$ and with action functional

$$\mathcal{A}(\omega) \equiv \frac{1}{4}\int_0^t R(\dot{\alpha}(\tau) + \gamma\alpha(\tau))^2 d\tau \tag{7.2}$$

for $\gamma \equiv s/R$. There is no problem to make mathematical sense of this; for example the cross-product

$$\int_0^t \alpha\dot{\alpha}d\tau = \int_0^t \alpha \circ d\alpha$$

is really a Stratonovich integral and the exponent of the square $\dot{\alpha}^2$ can be combined with the flat path space measure to define the Brownian reference measure. More to the point here is that the integrand in the action functional $\mathcal{A}$ can be rewritten as

$$R\dot{\alpha}^2(\tau) + \frac{s^2}{R}\alpha^2(\tau) + \frac{d}{d\tau}(s\alpha^2(\tau))$$

The last term is minus twice the variable entropy production rate $\dot{S}_T(\alpha)$. It is the only term in the integrand that is odd under time-reversal. So if we take the ratio as in (4.4) but here with $\pi =$ identity, we get, rigorously,

$$\ln \frac{d\boldsymbol{P}_{\tilde{\mu}}}{d\boldsymbol{P}_{\tilde{\mu}_t}\Theta}(\omega) = [S_T(\alpha(t)) - S_T(\alpha(0))] + [-\ln\tilde{\mu}_t(\alpha(t)) + \ln\tilde{\mu}(\alpha(0))] \tag{7.3}$$

so that, just as in (4.4), indeed the change in thermodynamic entropy is obtained from the source term in the action functional that breaks the time-reversal invariance.

Onsager and Machlup use the expression (7.2) for the action functional to derive a variational principle that extends the so called Rayleigh Principle of Least Dissipation. The idea is to take t very small and to seek the $\alpha(t)$ which will maximize the probability $\text{Prob}[\alpha(t)|\alpha(0)]$. Or, what is the most probable value of the flux $\dot{\alpha}$ when you start from $\alpha(0)$? This then determines the most probable path. This means that we should maximize

$$-\mathcal{A} = [-\frac{1}{4}R\dot{\alpha}^2 - \frac{s^2}{4R}\alpha^2 + \frac{\dot{S}_T(\alpha)}{2}]t$$

over all possible $\dot{\alpha}$, or that we should take

$$\dot{S}_T - \Phi(\dot{\alpha}) = \max.$$

where $\Phi(\dot{\alpha}) \equiv R\dot{\alpha}^2/2$ is the so called dissipation function. In other words, the maximum (over the flux) of the difference of the entropy production rate and the dissipation function determines the most probable path given an initial value for the thermodynamic variable. We mention this here not only because it is a central topic in the Onsager-Machlup paper but because this Rayleigh principle is often confused with the minimum entropy production principle that we had at the end of Section 6. In fact, the Rayleigh principle is more like a maximum entropy production principle (similar to the Gibbs variational principle in equilibrium statistical mechanics) enabling the search for the typical histories. Of course, its solution is just (7.1) for $\epsilon = 0$, i.e., the deterministic evolution for the thermodynamic variable, cf. [11]. The minimum entropy production principle on the other hand, attempts to characterize the stationary states as those where the entropy production rate is minimal. Both principles have serious limitations.

8. Phase space contraction

A more recent attempt to model nonequilibrium phenomena that was largely motivated by concerns of simulation and numerical work, involves so called thermostated dynamics, see [3, 5]. These are again as in the previous section, effective models but now using a deterministic dynamics. First, non-Hamiltonian external forces are added to the original Hamiltonian equations of motion to keep the system outside equilibrium. Since then, energy is no longer conserved and the system would escape the compact surface of constant energy, one adds 'thermostat forces', maintaining the energy fixed. There are other possible choices but they do not matter here. The resulting dynamics no longer preserves the phase space volume. We will keep the same notation as in Section 3 to denote the discretized dynamics; f is still an invertible transformation on Γ satisfying dynamic reversibility $\pi f \pi = f^{-1}$ but now the Liouville measure is not left invariant. It is important to remember that Γ does no longer represent the phase space of the total system (subsystem plus reservoirs); it is the phase space of the subsystem while the action of the environment is effectively incorporated in f. This environment has two functions at once: it drives the subsystem in a nonequilibrium state and it consists of a reservoir in which all dissipated heat can leak.

In the same context it has been repeatedly argued that the phase space contraction plays the role of entropy production, see e.g. [20, 7]. For thermostated dynamics, there are indeed good reasons to identify the two and various examples, mostly applied in numerical work, have illustrated this. Yet, from a more fundamental point of view, this needs an argument. To start, there is the simple observation that entropy can change in closed Hamiltonian systems while there is no phase space contraction. Moreover, even when used for open systems in the steady state regime, entropy production as commonly understood in irreversible thermodynamics is more than a purely dynamical concept. It is also a statistical object connecting the microscopic complexity with macroscopic behavior. That was also the reason to introduce the reduced states and the partitions $\hat{\Gamma}, \hat{\Omega}$. It is therefore interesting to see how and when phase space contraction relates to the concept of entropy production that we have introduced before.

Since the set-up is here somewhat different from that of Section 3, we denote here the state space by $\mathcal{M}$ instead of by Γ. It need not be the set of microstates (as in thermostated dynamics); it may be the set of possible values for some hydrodynamic variables, more like our set $\hat{\Gamma}$. We think of $\mathcal{M}$ as a bounded closed and smooth region of $\mathbb{R}^d$. Still, the dynamics f is assumed dynamically reversible (which would fail for irreversible hydrodynamics).

Suppose we have probability densities μ and ν on $\mathcal{M}$. We replay (4.2) or (5.4) but now on the space $\mathcal{M}$. For every function Φ_n on $\mathcal{M}^{n+1}$, let Φ_n^* be given as $\Phi_n^*(x_0, x_1, \dots, x_n) \equiv \Phi_n(\pi x_n, \pi x_{n-1}, \dots, \pi x_0)$. We find that

$$\begin{aligned}\int \Phi_n^*(y, fy, \dots, f^n y)\nu\pi(y)dy &= \int \Phi_n(\pi f^n y, \dots, \pi y)\nu\pi(y)dy \\ &= \int \Phi_n(f^{-n} y, \dots, y)\nu(y)dy\end{aligned}$$

using dynamic reversibility. Now change variables $y = f^n x$,

$$\int \Phi_n^*(x, fx, \dots, f^n x)\nu\pi(x)dx = \int \Phi_n(x, \dots, f^n x)\frac{\nu(f^n x)}{\mu(x)}|\frac{df^n}{dx}(x)|\mu(x)dx$$

or

$$\langle \Phi_n^* \rangle_{\nu\pi} = \langle \Phi_n r_n \rangle_\mu, \; r_n(x) = \frac{\nu(f^n x)}{\mu(x)}|\frac{df^n}{dx}(x)|$$

This should again be compared with (4.2) and with (5.4). In particular, we see that the phase space contraction, or more precisely, minus the logarithm of the Jacobian determinant $-\ln|df^n/dx|$, replaces the total entropy production (of the total system) we had before:

$$S_B(f^n x) - S_B(x) \longrightarrow -\ln|df^n/dx| \tag{8.1}$$

This requires the dynamical reversibility; without it, even this purely formal identification is not justified.

Looking further to compare with Proposition 4.1 and (5.8), we can take

$\nu(x) = \mu_n(x)$, the time-evolved density. Then,

$$\nu(f^n x) = \mu(x)|\frac{df^{-n}}{dx}(f^n x)|, \quad r_n(x) = 1$$

so that the formal analogue of the right-hand side of (4.4) and (5.8) now becomes

$$-\ln \mu_n(f^n x) + \ln \mu(x) - \ln |\frac{df^n}{dx}(x)| = 0 \tag{8.2}$$

But if we believe in our algorithm for computing the mean entropy production as in (4.6) for closed systems and as in (5.17) for open systems, the expectation of (8.2) with respect to μ should give us the mean entropy production; it remains of course zero:

$$-\int dx \mu_n(x) \ln \mu_n(x) + \int dx \mu(x) \ln \mu(x) - \int dx \mu(x) \ln \frac{df^n}{dx}(x) = 0 \tag{8.3}$$

In other words, we find that the mean entropy production is zero. Heuristically, this is quite natural by the very philosophy of the thermostated dynamics; the change of entropy in the subsystem is exactly canceled by the change of entropy in the environment. That is: the difference in Shannon entropies is given by the expected phase space contraction. This is known since at least [1].

It is true that the above and in particular (8.2) concerns the transient regime and that the above calculation cannot be repeated for the stationary measure as it may be singular. Yet, this property may be considered as an artifact of the infinitely fine resolution in $\mathcal{M}$ and we can remove it by taking a finite partition $\hat{\mathcal{M}}$ of $\mathcal{M}$. We need a generalization of (4.2) for dynamics that do not preserve the phase space volume, $\rho f^{-1} \neq \rho$, with $\rho(dx) = dx$ the flat measure on $\mathcal{M}$. Using the notation of Section 4, we now get

$$\frac{\text{Prob}_{\hat{\mu}\times\rho}(x_n \in M_n, \ldots, x_0 \in M_0)}{\text{Prob}_{\hat{\mu}_n \pi\times\rho}(x_n \in \pi M_0, \ldots, x_0 \in \pi M_n)} = \frac{\hat{\mu}(M_0)}{\hat{\mu}_n(M_n)} \frac{\rho(M_n)}{\rho(M_0)} \frac{\rho(\cap_{j=0}^n f^{-j} M_j)}{\rho f^n(\cap_{j=0}^n f^{-j} M_j)} \tag{8.4}$$

by using again the dynamic reversibility of the map f. In the stationary regime, the formal analogue of the entropy production rate equals

$$\lim_n \frac{1}{n} \ln \frac{\text{Prob}_{\hat{\mu}\times\rho}(x_n \in M_n, \ldots, x_0 \in M_0)}{\text{Prob}_{\hat{\mu}_n \pi\times\rho}(x_n \in \pi M_0, \ldots, x_0 \in \pi M_n)} = \lim_n \frac{1}{n} \ln \frac{\rho(\cap_{j=0}^n f^{-j} M_j)}{\rho f^n(\cap_{j=0}^n f^{-j} M_j)} \tag{8.5}$$

Note that while this is true for every finite partition $\hat{\mathcal{M}}$, it fails for the finest partition where $\hat{\mathcal{M}}$ would coincide with the original phase space $\mathcal{M}$. The above formula may be further elaborated, assuming that the partition $\hat{\mathcal{M}}$ is generating for f. (This would not be true for the partition $\hat{\Gamma}$ corresponding to the physical coarse-graining induced by a set of thermodynamic variables). Let $x \in \mathcal{M}$ be fixed and choose $M_j = M(f^j x)$. Using the notation $M_x^{(n)} = \cap_{j=0}^n f^{-j} M_j$, we have $M_x^{(n+1)} \subset M_x^{(n)}$ and $\cap_n M_x^{(n)} = \{x\}$. Suppose now that the following limits are equal:

$$\lim_n \frac{1}{n} \sum_{k=0}^n \ln \frac{\rho(f^k M_x^{(n)})}{\rho f(f^k M_x^{(n)})} = \lim_n \frac{1}{n} \sum_{k=0}^n \ln \frac{d\rho}{d(\rho f)}(f^k x)$$

Clearly, $(d\rho f / d\rho)(x)$ is a general form of the phase space contraction (the Jacobian determinant of f). The right-hand side takes its ergodic average. If we sample $x \in \mathcal{M}$ from the flat measure ρ, we could suppose that these ergodic averages converge to the expected phase space contraction for some distribution μ on $\mathcal{M}$. That would for example be guaranteed under some chaoticity assumptions for the dynamics f; in particular if the dynamical system allows a SRB state μ, [20]. We can then combine the previous two relations and find that, for ρ-almost every $x \in \mathcal{M}$, the mean entropy production rate gets the form

$$\lim_n \frac{1}{n} \ln \frac{\mathrm{Prob}_{\hat{\mu} \times \rho}(x_n \in M(f^n x), \dots, x_0 \in M(x))}{\mathrm{Prob}_{\hat{\mu}_n \pi \times \rho \pi}(x_n \in \pi M(x), \dots, x_0 \in \pi M(f^n x))} = \mathbb{E}_\mu \Big(\ln \frac{d\rho}{d(\rho f)} \Big) \tag{8.6}$$

This is exactly the mean entropy production rate one works with in thermostated dynamics, see e.g. [20]. Comparing it with (4.6) and (5.14)-(5.17), it does indeed replace the mean entropy production as computed from the algorithms in Sections 4 and 5.

Acknowledgment

We thank S. Goldstein for insisting on clarifying the connections with the Boltzmann entropy. We thank J. Bricmont for reading a first draft of the paper. We thank both for useful discussions.

Appendix A. Kac ring model

The scheme of Section 4 can also be applied to every model dynamics sharing the property of dynamical reversibility with Hamiltonian dynamics. To illustrate this and in order to specify some quantities that have appeared above, we briefly discuss the so called Kac ring model. We refer to the original [12] for the context and to [2] for more

discussion.
The microscopic state space is $\Omega = \{-1,+1\}^N$. Its elements are denoted by $x = (\eta; v) = (\eta_1, \ldots, \eta_N; v)$ and the kinematic time-reversal is $\pi x = (\eta_1, \ldots, \eta_N; -v)$. The microscopic dynamics f_ε depends on parameters $\varepsilon_i = \pm 1, i = 1, \ldots, N$, and is defined as

$$f_\epsilon(\eta_1, \ldots, \eta_N; +1) \equiv (\varepsilon_N \eta_N, \varepsilon_1 \eta_1, \ldots, \varepsilon_{N-1} \eta_{N-1}; +1)$$
$$f_\epsilon(\eta_1, \ldots, \eta_N; -1) \equiv (\varepsilon_1 \eta_2, \varepsilon_2 \eta_3, \ldots, \varepsilon_N \eta_1; -1)$$

so that $f_\varepsilon = \pi f_\varepsilon^{-1} \pi$ (dynamic reversibility). The only information about the parameters is that $\sum_i \varepsilon_i = mN$ for some fixed m.
Since the "velocity" v is conserved, we can as well study the dynamics on $\Gamma = \{-1,+1\}^N$ (fixing $v = +1$) and to each microstate η we associate the macroscopic variable $\alpha(\eta) \equiv \sum_i \eta_i / N$. This introduces the partition $\hat{\Gamma}$ containing $N+1$ elements. For example, the set $M(\eta) \in \hat{\Gamma}$ contains all the $\sigma \in \Gamma$ for which $\alpha(\eta) = \alpha(\sigma)$ and $|M(\eta)| = C_N((\alpha(\eta)+1)N/2)$ the binomial factor. Trajectories can therefore be identified with a sequence of macroscopic values α_j. We are interested in the case of finite (but possibly long) trajectories while taking N extremely large. In the simplest approximation, this means that we let $N \to +\infty$. It can be shown that for the overwhelming majority of the ε_i, the macroscopic value α_n after n time steps behaves as $\alpha_n = m^n \alpha_0$ with α_0 the initial macro-value. The limiting evolution on the level of macrostates is therefore deterministic but not time-reversal invariant. Equilibrium corresponds to $\alpha = 0$. The entropy production rate (per degree of freedom) when the system has macro-value α is $(1+\alpha)\ln\sqrt{1+\alpha} + (1-\alpha)\ln\sqrt{1-\alpha} - (1+m\alpha)\ln\sqrt{1+m\alpha} - (1-m\alpha)\ln\sqrt{1-m\alpha} = (1-m^2)\alpha^2/2$ up to second order in α .

Various examples of applying exactly the algorithm of Section 6 to compute the entropy production and to study its fluctuations have appeared before, see [16, 17, 18]. We just add here how the Markov approximation for the Kac ring model looks like.
The transition probability can be read from (6.1): for finite N, $p(\alpha, \alpha')$ is the probability that the macroscopic value equals α' after one time step, when the process was started from a randomly chosen η with macroscopic value $\sum_i \eta_i / N = \alpha$:

$$p(\alpha, \alpha') = \frac{1}{C_N((\alpha+1)N/2)} |\{\eta \in \Gamma : \sum_i \eta_i = \alpha N \text{ and } \sum_i \varepsilon_i \eta_i = \alpha' N\}|$$

Depending on the parameters ε_i, this will often be zero, certainly when N is large and α' is far from equal to $m\alpha$. On the other hand, when $\alpha' = m\alpha \pm \sqrt{1-m^2}/\sqrt{N}$, the transition will be possible but damped as $\exp[-N(\alpha' - m\alpha)^2/2(1-m^2)]$. It is therefore interesting to study the evolution on the level of the rescaled variables $\sqrt{N}\alpha$; these are the fluctuations. This takes us back to Section 7. In equation (7.1), we should take $R = 1/(1-m)$ and $s = 1$. The solution of the Rayleigh principle is of course here found from maximizing the transition probability $p(\alpha,\alpha')$ and this happens when $\alpha' = m\alpha$. As always with this principle, see [11], this does not teach us anything new; it only gives a variational characterization of the hydrodynamic evolution.

Appendix B. Hamiltonian dynamics of composed systems

In order to clear up the content of assumptions A1 and A2 of Section 5, we demonstrate here how it naturally emerges in the framework of Hamiltonian dynamics. We again have in mind a composed system consisting of a *system* thermally coupled to an environment, the latter having the form of a few subsystems (reservoirs).
Let T be a finite set whose elements label the individual particles of the total system. That means that we are really considering a solid (and not a fluid). To every particle $i \in T$, there is associated a position and momentum variable $x_i \equiv (q_i, p_i)$. Given a configuration x, we put $x_\Lambda \equiv (q_\Lambda, p_\Lambda)$ for the coordinates of particles belonging to the system $\Lambda \subset T$. We thus decompose the set of particles T by splitting the total system into a *system* and m reservoirs, $T = \Lambda \cup V^1 \cup \ldots \cup V^m$. We assume that the Hamiltonian of the total system may be written in the form $H(x) = H^0(x_\Lambda) + \sum_{k=1}^m H^k(q_{\bar{V}^k}, p_{V^k})$ where $\bar{V}^k \equiv V^k \cup \partial_k\Lambda$ and $\partial_k\Lambda \subset \Lambda$ is the set of all particles of the system coupled to the $k-$th reservoir. Moreover, we need the assumption that the reservoirs are *mutually separated* in the sense $\partial_k\Lambda \cap \partial_\ell\Lambda = \emptyset$ whenever $k \neq \ell$. To be specific, consider the following form of the Hamiltonian:

$$H^0(x_\Lambda) = \sum_{i\in\Lambda}\big[\frac{p_i^2}{2m_i} + U_i(q_i)\big] + \sum_{(ij)\subset\Lambda}\Phi(q_i - q_j) \tag{B.1}$$

$$H^k(q_{\bar{V}^k}, p_{V^k}) = \sum_{i\in V^k}\big[\frac{p_i^2}{2m_i} + U_i(q_i)\big] + \sum_{(ij)\subset V^k}\Phi(q_i - q_j) + \sum_{\substack{i\in V^k\\ j\in\partial_k\Lambda}}\Phi(q_i - q_j) \tag{B.2}$$

For what follows, we consider another decomposition of the energy of the system in the form $H^0(x_\Lambda) = h^0(q_\Lambda, p_{\Lambda^0}) + \sum_{k=1}^n h^k(x_{\partial_k\Lambda})$ where

$\Lambda^0 \equiv \Lambda \setminus \cup_k \partial_k \Lambda$. We can take, for instance,

$$h^0(q_\Lambda, p_{\Lambda^0}) = \sum_{i\in\Lambda^0} \bigl[\frac{p_i^2}{2m_i} + U_i(q_i)\bigr] + \sum_{(ij)\subset\Lambda} \Phi(q_i - q_j) \tag{B.3}$$

$$h^k(x_{\partial_k\Lambda}) = \sum_{i\in\partial_k\Lambda} \bigl[\frac{p_i^2}{2m_i} + U_i(q_i)\bigr] \tag{B.4}$$

If the trajectory $\omega(\tau) \equiv (q(\tau), p(\tau))$ is a solution of the Hamiltonian equations of motion, then the time-derivative of the energy of each reservoir is in terms of Poisson brackets:

$$\begin{aligned} \frac{dH^k}{d\tau}(\omega(\tau)) &= \{H^k, H\}(\omega(\tau)) = \{H^k, H^0\}(\omega(\tau)) \\ &= \{H^k, h^k\}(\omega(\tau)) \end{aligned} \tag{B.5}$$

A similar calculation yields

$$\begin{aligned} \frac{dh^k}{d\tau}(\omega(\tau)) &= \{h^k, H\}(\omega(\tau)) \\ &= \{h^k, H^0\}(\omega(\tau)) + \{h^k, H^k\}(\omega(\tau)) \end{aligned} \tag{B.6}$$

where in the last equality we used the assumption that the reservoirs are mutually separated. Combining the above equations and integrating them over the time interval (t_0, t) one gets

$$\begin{aligned} H^k(\omega(t)) - H^k(\omega(t_0)) &= h^k(\omega_{\partial_k\Lambda}(t_0)) - h^k(\omega_{\partial_k\Lambda}(t)) \\ &\quad - \int_{t_0}^t d\tau \sum_{\substack{i\in\partial_k\Lambda \\ j\in\Lambda^0}} \frac{p_i(\tau)}{m_i} \Phi'(q_i(\tau) - q_j(\tau)) \end{aligned} \tag{B.7}$$

Notice that the right-hand side depends only on the restriction $\omega_\Lambda(t)$ of the trajectory ω. Therefore, assuming that $\omega(t)$, $\omega'(t)$ are two solutions of the equations of motion such that $\omega_\Lambda(\tau) = \omega'_\Lambda(\tau)$, $\tau \in (t_0, t)$, the heat flow into the $k-$th reservoir, $Q^k_{(t_0,t)}(\omega) \equiv H^k(\omega(t)) - H^k(\omega(t_0))$ satisfies $Q^k_{(t_0,t)}(\omega) = Q^k_{(t_0,t)}(\omega')$. Put differently, the heat current into each reservoir, being a state quantity from the point of view of the total system, is also a functional of the (complete) trajectory of the system. This motivates assumption A2. Moreover, from the above calculation, also the assumption A1 becomes plausible as we can expect that the right-hand side of (B.7) is of order $(t - t_0)|\partial_k\Lambda|$.

Remark: Note that the decomposition of the total energy into local parts is not unique due to the presence of interaction. The above claim is only true for the decomposition in which the interaction energy between the system end each reservoir is taken as part of the energy of the reservoir. However, the difference between this reservoir energy and others can only be of the order of $|\partial_k \Lambda|$ which is again sufficient. Furthermore, all possible decompositions become undistinguishable in the regime of weak coupling.

Appendix C. Minimum entropy production principle

In this appendix we examine the validity of the minimum entropy production principle in case of Markov chains breaking the detailed balance condition, as promised at the end of Section 6.. We use the same notation as there, namely we consider a continuous time Markov chain (X_t) on a finite state space with transition rates $k_\varepsilon(M, M')$. The latter are parameterized by ε measuring the distance from equilibrium. More precisely, let $k_\varepsilon(M, M') = k_0(M, M') \exp[\varepsilon\psi(M, M')/2]$ where the Markov chain with rates $k_0(M, M')$ has a unique reversible measure μ_0, i.e. ,

$$\mu_0(M)k_0(M, M') = \mu_0(M')k_0(M', M) \tag{C.1}$$

We also assume that $\mu_0(M) \neq 0$ for all M. The stationary measure μ_ε is a solution of the equation

$$\sum_M [\mu_\varepsilon(M)k_\varepsilon(M, M') - \mu_\varepsilon(M')k_\varepsilon(M', M)] = 0 \tag{C.2}$$

for all M'. Write this measure in the form $\mu_\varepsilon = \mu_0(1 + \varepsilon f + o(\varepsilon))$ with the normalization condition $\sum_M \mu_0(M)f(M) = 0$. Then a simple calculation yields the following (linearized) equation for stationarity:

$$\sum_M \mu_0(M)k_0(M, M')[f(M) - f(M') + \psi^{\mathrm{as}}(M, M')] = 0 \tag{C.3}$$

where we used $\psi^{\mathrm{as}}(M, M')$ to denote the asymmetrical part of the driving, $\psi^{\mathrm{as}}(M, M') = [\psi(M, M') - \psi(M', M)]/2$. This equation with the constraint $\sum_M \mu_0(M)f(M) = 0$ has always a solution, we will assume it is unique. Notice that, up to first order in ε, only the asymmetric part of the driving deforms the stationary measure.

We now compare this result with that of the minimum entropy production principle. Recall that the entropy production rate is the functional on measures

$$\sigma_\varepsilon(\mu) = \sum_{M,M'} \mu(M)k_\varepsilon(M, M') \ln \frac{\mu(M)k_\varepsilon(M, M')}{\mu(M')k_\varepsilon(M', M)} \tag{C.4}$$

The first observation is that it is convex. So, the constrained variational problem $\sigma_\varepsilon(\mu^\star) = \min$, $\sum_M \mu^\star(M) = 1$, is equivalent to solving the equation

$$\frac{\delta}{\delta\mu}\big[\sigma_\varepsilon - \lambda\sum_M \mu(M)\big](\mu^\star) = 0 \tag{C.5}$$

together with $\sum_M \mu^\star(M) = 1$. We again linearize this equation by writing $\mu^\star_\varepsilon = \mu_0(1+\varepsilon f^\star + o(\varepsilon))$ and after some calculation we get

$$\frac{1}{\mu_0(M)}\sum_{M'}\mu_0(M)k_0(M,M')[f^\star(M) - f^\star(M') + \psi^{\mathrm{as}}(M,M') - \lambda] = 0 \tag{C.6}$$

Observe that for $\lambda = 0$ this equation is equivalent to (C.3). Therefore, if the minimizing point $\mu^\star$ is unique, it must correspond to $f^\star \equiv f$ with f being the normalized solution of (C.3).

Note that in higher orders the minimum entropy production principle fails as a variational principle for the stationary measure. But even to linear order, outside the context of Markov processes, the principle can be questioned both for its correctness and for its usefulness, see [11].

Appendix D. Systems in local thermodynamic equilibrium

In this appendix we connect our presentation of Section 5 with the standard formulations of irreversible thermodynamics. We go about this in a rather formal way, trying to safeguard the simplicity of the explanation.

We consider the system itself to be large (yet small when compared with the reservoirs) and we split it further into (still large) subsystems around a spatial point r. We assume that these subsystems are in (local) equilibrium so that the change of entropy $\hat{S}^0_B(\bar\omega_n) - \hat{S}^0_B(\bar\omega_0)$ of the system (appearing in (5.8) or (5.9)) is a change of (maximal) Boltzmann entropy when the energy in the subsystem around r moves from the value $U(r,0)$ to $U(r,t)$. That is,

$$\hat{S}^0_B(\bar\omega_n) - \hat{S}^0_B(\bar\omega_0) = \sum_r [S_B(r,U(r,t)) - S_B(r,U(r,0))]$$

where $S_B(r,U)$ is the logarithm of the phase space volume of the subsystem around r corresponding to energy value U. We start from (5.9). It gets the form

$$\Delta S_B(\bar\omega) = \sum_r [S_B(r,U(r,t)) - S_B(r,U(r,0))] + \sum_k \beta_k \Delta E^k(\bar\omega) \tag{D.1}$$

Here the first sum runs over the subsystems of the system under consideration, while the second sum is taken over all reservoirs. We introduce the temperature of the r-subsystem as $\beta(r,\tau) = (\partial S_B/\partial U)(r, U(r,\tau))$, and then

$$\Delta S_B(\bar{\omega}) = \sum_r \int_0^t d\tau \beta(r,\tau) \frac{dU}{d\tau}(r,\tau) + \sum_k \beta_k \Delta E^k(\bar{\omega}) \tag{D.2}$$

We use $J(r,r',\tau)$ to denote the energy current at time τ from the r-subsystem to the r'-subsystem. Similarly, $J^k(r,\tau)$ stands for the energy current from the r-subsystem to the $k-$th reservoir. The conservation of energy then implies the equalities

$$\frac{dU}{d\tau}(r,\tau) + \sum_y J(r,r',\tau) + \sum_k J^k(r,\tau) = 0 \tag{D.3}$$

and

$$\Delta E^k(\omega) = \sum_r \int_0^t d\tau \; J^k(r,\tau) \tag{D.4}$$

The currents are antisymmetric: $J(r,r',\tau) = -J(r',r,\tau)$. The entropy production now becomes

$$\Delta S_B(\bar{\omega}) = \int_0^t d\tau \big[\sum_k \sum_r (\beta_k - \beta(r,\tau)) J^k(r,\tau) + \sum_{r,r'} \beta(r,\tau) J(r',r,\tau) \big] \tag{D.5}$$

The first term on the right is a surface sum. Its origin is the entropy current. We assume that every subsystem is coupled to at most one reservoir. In the continuum, if r is at the boundary of the system with the $k-$th reservoir, then in fact $\beta(r,\tau) = \beta_k$. Hence, for the first term, either $J^k(r,\tau) = 0$ or $\beta(r,\tau) = \beta_k$ which makes it vanish. When we are dealing with closed systems, then $J^k(r,\tau) = 0$ by definition. Using further the antisymmetry of the bulk currents, we obtain

$$\Delta S_B(\bar{\omega}) = \int_0^t d\tau \sum_r \nabla\beta(J)(r,\tau) \tag{D.6}$$

where we used the notation

$$\nabla\beta(J)(r,\tau) \equiv \sum_{r'} \frac{\beta(r',\tau) - \beta(r,\tau)}{2} J(r,r',\tau) \tag{D.7}$$

This is already close to the standard formulations in which the entropy production rate equals a thermodynamic force times a current. Indeed, assuming that the decomposition of the system into subsystems has a natural space structure, say as the regular $\mathbb{Z}^d$-lattice, and that the current exchanges take place only between neighboring subsystems (via their common interface), we can write $\nabla\beta(J)(r,\tau) \simeq \nabla\beta(r,\tau) \cdot \vec{J}(r,\tau)$

(the derivative taken in the discrete sense). The (total) entropy production is then $\Delta S_B(\bar\omega) = \int_0^t d\tau \sum_r \sigma(r,\tau)$ with space-time entropy production rate

$$\sigma(r,\tau) = \nabla\beta(r,\tau)\cdot\vec{J}(r,\tau)$$

as sought.

References

[1] L. Andrey, *The rate of entropy change in non-Hamiltonian systems*, Phys. Lett. **11A**, 45-46 (1985).

[2] J. Bricmont, *Bayes, Boltzmann and Bohm: Probability in Physics.* In: Chance in Physics, Foundations and Perspectives. Eds. J. Bricmont, D. Dürr, M.C. Galavotti, G. Ghirardi, F. Petruccione, and N. Zanghi, (Springer-Verlag, 2002).

[3] J.R. Dorfman, An Introduction to chaos in nonequilibrium statistical mechanics. (Cambridge University Press, Cambridge, 1999).

[4] D.J. Evans, E.G.D. Cohen and G.P. Morriss, *Probability of second law violations in steady flows*, Phys. Rev. Lett. **71**, 2401-2404 (1993).

[5] S. Sarman, D.J. Evans and P.T. Cummings, *Recent developments in non-Newtonian molecular dynamics*, Physics Reports, Elsevier (Ed. M.J. Klein), **305**, 1–92 (1998).

[6] G.L. Eyink, J.L. Lebowitz and H. Spohn, *Microscopic origin of hydrodynamic behavior: Entropy production and the steady state*, Chaos (Soviet-American Perspectives on Nonlinear Science), ed. D. Campbell, 367–391 (1990).

[7] G. Gallavotti and E.G.D. Cohen, *Dynamical ensembles in nonequilibrium Statistical Mechanics*, Phys. Rev. Letters **74**, 2694–2697 (1995). *Dynamical ensembles in stationary states*, J. Stat. Phys. **80**, 931–970 (1995).

[8] S. Goldstein, *Boltzmann's Approach to Statistical Mechanics.* In: Chance in Physics, Foundations and Perspectives. Eds. J. Bricmont, D. Dürr, M.C. Galavotti, G. Ghirardi, F. Petruccione, and N. Zanghi, (Springer-Verlag, 2002).

[9] C. Jarzynski, *Nonequilibrium Equality for Free Energy Differences*, Phys. Rev. Lett. **78**, 2690–2693 (1997).

[10] E.T. Jaynes, *Gibbs vs Boltzmann Entropies*, Am. J. Phys. **33**, 391–398 (1965). *Papers on Probability, Statistics and Statistical Physics*, Ed. R. D. Rosencrantz (Reidel, Dordrecht 1983).

[11] E.T. Jaynes, *The minimum entropy production principle*, Ann Rev. Phys. Chem. **31**, 579–600 (1980). *Papers on Probability, Statistics and Statistical Physics*, Ed. R. D. Rosencrantz (Reidel, Dordrecht 1983). Note however that historically, the Rayleigh principle (Lord Rayleigh, Phil Mag. **26**, 776 (1913)), was much closer to the variational principle in mechanics.

[12] M. Kac, *Probability and Related Topics in the Physical Sciences*, (Interscience Pub., New York 1959).

[13] J.L. Lebowitz, *Reduced Description in Nonequilibrium Statistical Mechanics.* In: Proceedings of International Conference on Nonlinear Dynamics, December 1979, Annals New York Academy of Sciences **357**, 150–156 (1980).

[14] J.L. Lebowitz, *Microscopic Origins of Irreversible Macroscopic Behavior*, Physica A **263**, 516–527 (1999). Round Table on Irreversibility at STATPHYS20, Paris, July 22, 1998.

[15] M. Liu, *The Onsager Symmetry Relation and the Time Inversion Invariance of the Entropy Production*, Archive `cond-mat/9806318`.

[16] C. Maes, *Fluctuation theorem as a Gibbs property*, J. Stat. Phys. **95**, 367–392 (1999).

[17] C. Maes, F. Redig and A. Van Moffaert, *On the definition of entropy production via examples*, J. Math. Phys. **41**, 1528–1554 (2000).

[18] C. Maes, F. Redig and M. Verschuere, *From Global to Local Fluctuation Theorems*, Moscow Mathematical Journal **1**, 421–438 (2001).

[19] L. Onsager and S. Machlup, *Fluctuations and Irreversible Processes*, Phys. Rev. **91**, 1505–1512 (1953). More generally Onsager-Machlup consider in their second paper (starting on page 1512) a second-order process for extensive variables $(\alpha, \beta) \equiv (\alpha_1, \ldots, \alpha_m; \beta_1, \ldots, \beta_m)$ parametrizing the partition $\hat{\Gamma}$ that we had before. The difference between the α's and the β's arises from their different behavior under applying the time reversal π; the β's are the so called 'velocity' variables that change their sign if the time is reversed; the α's are even functions under time reversal. In all cases, they are thermodynamic variables that are sums of a large number of molecular variables.

[20] D. Ruelle, *Smooth Dynamics and New Theoretical Ideas in Nonequilibrium Statistical Mechanics*, J. Stat. Phys. **95**, 393–468 (1999).

[21] J. Schnakenberg, *Network theory of behavior of master equation systems*, Rev. Mod. Phys. **48**, 571-585 (1976).

[22] Ya. G. Sinai, Introduction to Ergodic Theory, Princeton University Press (1976).

[23] R. Zwanzig, *Memory Effects in Irreversible Thermodynamics*, Phys. Rev. **124**, 983–992 (1961). See also: S. Nakajima, Prog. Theor. Phys. **20**, 948 (1958); M. Mori, Prog. Theor. Phys. **32**, 423 (1965).