On the definition of entropy production, via examples

Christian Maes[*], **Frank Redig**[†], **Annelies Van Moffaert**[‡]
Instituut voor Theoretische Fysica,
K.U.Leuven, B-3001 Leuven, Belgium.

June 4, 1999

Abstract

We present a definition of entropy production rate for classes of deterministic and stochastic dynamics. The point of departure is a Gibbsian representation of the steady state pathspace measure for which 'the density' is determined with respect to the time-reversed process. The Gibbs formalism is used as a unifying algorithm capable of incorporating basic properties of entropy production in nonequilibrium systems. Our definition is motivated by recent work on the Gallavotti-Cohen (local) fluctuation theorem and it is illustrated via a number of examples.

1 Introduction.

One of the more obscure concepts in nonequilibrium statistical physics is that of entropy and its production. While most people adhere to the standard textbook-formulation in the case of close to equilibrium processes, opinions start seriously deviating concerning the appropriate entropy-concepts when confronted with far from equilibrium steady states.
The question is of course in the first place not a mathematical one, but rather conceptual. The field of nonequilibrium thermodynamics is still under construction and while various mathematically precise definitions make a lot of sense *a priori*, it needs to be seen how these definitions relate to the phenomena. Yet, in the tradition of mathematical physics, [12], a mathematical treatment of various examples or models hopefully clarifies the situation and may enable an algorithmic and constructive setup, available and testable also in cases when the correct physical intuition is not immediately available.

[*]Onderzoeksleider FWO, Flanders. Email: Christian.Maes@fys.kuleuven.ac.be
[†]Post-doctoraal Onderzoeker FWO, Flanders. Email: Frank.Redig@fys.kuleuven.ac.be
[‡]Aspirant FWO, Flanders. Email: Annelies.VanMoffaert@fys.kuleuven.ac.be

In the last years, a lot of interest was generated in understanding 'entropy production' (or, entropy creation rate) in far from equilibrium steady states. In the context of certain dynamical systems, thermostatted systems in particular, entropy production appears as synonymous with 'phase space contraction rate.' In fact under certain assumptions, one proves that the change of Shannon entropy in the steady state exactly equals (minus) that contraction rate, [32, 33, 35]. The Gallavotti-Cohen theorem states a symmetry in the probability of fluctuations of this entropy creation rate, and while proven only for a limited class of systems ([16, 17]), has been observed in computer simulations in a variety of models, see e.g. the motivating experiment [10]. In all examples, the physical interpretation as entropy creation rate seems to be confirmed.

In [30] another approach to the definition of entropy production appeared. There one considers the pathspace measure in the steady state and compares it (locally) with the pathspace measure of the time-reversed process. If such a comparison can be made, the relative density can be defined (as corresponding Radon-Nikodym derivative). This density can be written as an exponential of a 'relative energy' for an interaction governing the pathspace measure with respect to the time-reversal transformation. This 'relative energy' is extensive and only contains the space-time interactions that break the time-reversal. Its density is what we call the entropy production. Under various conditions, by the very nature of its 'Gibbsian' definition, it satisfies a fluctuation theorem. It is interesting to see that it coincides with the definition given in the Gallavotti-Cohen setup when indeed the dynamics satisfies the conditions of their theorem. But it also gives a unifying description for more general dynamics if, at least for the purpose of computing macroscopic properties, the space-time trajectories are distributed via the appropriate (space-time) Gibbs measure. In particular, via this one and the same algorithm the entropy action functionals for every example of stochastic dynamics appearing in [27, 24] can be computed.

The goal of this paper is to continue and to extend the analysis of [30] mostly via specific illustrative examples. The main result is that it works: both theoretical considerations as well as exact results for examples of far from equilibrium steady state dynamics confirm our definition and the Gibbsian picture which is behind it. The material we work with here is mostly taken from [27] for the examples of stochastic dynamics, and for deterministic dynamics we mainly restrict us to the Gallavotti-Cohen setup.

The plan of the paper: We start in the next section with our Gibbsian setup and with an abstract definition of entropy production. It is then specified to the context of interacting particle systems, both stochastic and deterministic. For stochastic dynamics we consider spinflip, particle exchange and diffusion processes. There is also an example of a molecular motor. The discussion on deterministic dynamics is more descriptive as we limit ourselves mostly to explaining the relation of the Gallavotti-Cohen work with the Gibbsian setup. The last section is devoted to a brief discussion of the transient (not steady state) regime and the mathematics behind the so-called nonequilibrium work relations.

By the nature of the presented material, the reader will be confronted at the same time with elements of the Gibbs formalism, interacting particle systems, stochastic calculus, dynamical systems and thermodynamics. We have tried to make the sections more or less self-contained and we have added many references to refer to for background information.

2 Definition and main properties of entropy production

2.1 Pathspace measure: Gibbsian setup

We start with an informal description of the main algorithm used in the identification of the entropy production.
Suppose a system composed of many locally interacting components in a nonequilibrium steady state. The pathspace measure gives the microscopic distribution of the trajectories of the system compatible with the macroscopic information that is available. We take as hypothesis that for the purpose of computing the macroscopic properties of the system, this pathspace measure in fact defines a Gibbs state on the space-time configurations. For the moment, we forget how this Gibbsian description is obtained but we will come back to this in a remark below.
For the sake of simplicity let us in fact consider the case where we are dealing with a Gibbs state for a lattice spin system, see [18, 9, 36] for background. The configurations σ are then elements of $\Omega = G^{\mathbb{Z}^{d+1}}$ where G, the single spin space, is finite and $\sigma = (\sigma(n,i), n \in \mathbb{Z}, i \in \mathbb{Z}^d)$ is a space-time trajectory for $n =$ discrete time and i a spatial coordinate on the $d-$dimensional lattice $\mathbb{Z}^d$. Physically, it is better to replace the infinite space-time lattice $\mathbb{Z}^{d+1}$ by a huge space-time box W, possibly with appropriately chosen boundary conditions but what follows is easily adapted to that case. Furthermore, we have a family of macroscopic variables, formally,

$$H^\alpha = \sum_A U_A^\alpha, \alpha = 0, \ldots, m \tag{2.1}$$

each additive in 'space-time potentials' U_A^α for which we assume that U_A^α is a real-valued function only depending on the configuration in the finite set A, summable according to

$$\sum_{A \ni x} \sup_\sigma |U_A(\sigma)| < +\infty, x \in \mathbb{Z}^{d+1}. \tag{2.2}$$

Again, this condition is not strictly needed but it is the simplest choice for making sense of the differences

$$H^\alpha(\sigma) - H^\alpha(\eta) = \sum_{A \cap \Lambda \neq \emptyset} [U_A^\alpha(\sigma) - U_A^\alpha(\eta)] \tag{2.3}$$

when $\sigma, \eta \in \Omega$ coincide outside a finite set $\Lambda \subset \mathbb{Z}^{d+1}$. Moreover, thanks to the additive structure, we immediately get that (2.3) equals

$$\sum_{x \in \Lambda} \sum_{A \ni x, A \subset \Lambda} \frac{1}{|A|} [U_A^\alpha(\sigma) - U_A^\alpha(\eta)] \tag{2.4}$$

up to boundary terms $O(|\partial\Lambda|)$ when Λ is a regular, say cubic-like, region. In the same way, we will be demanding translation invariance even though this

is negotiable for some of what follows. In fact, it is good to split translations in spatial and time-like translations. We denote the translation over a lattice vector $x \in \mathbb{Z}^{d+1}$ by τ_x and the translation is time-like if $x = (n, 0)$ for some $n \in \mathbb{Z}$, space-like if $x = (0, i)$ for some $i \in \mathbb{Z}^d$. Since we are considering a system in its steady state, the condition that the potential (and the pathspace measure) is time-like translation invariant is a natural requirement, but the space-like invariance is often broken. All the same, we wish to continue with the simplest setup and global translation invariance is part of this: $U_A(\tau_x\sigma) = U_{A+x}(\sigma), where \tau_x\sigma(y) = \sigma(x + y)$.
Gibbs states corresponding to the ensemble defined via (2.1) are probability measures μ on Ω whose conditional distributions to find the trajectory σ in an arbitrary finite set $V \subset \mathbb{Z}^{d+1}$ when the configuration σ is given outside V is

$$\mu[\sigma \text{ on } V|\sigma \text{ on } V^c] = \frac{1}{Z_V(\sigma)} \exp[-\sum_\alpha E_\alpha \sum_{A\cap V\neq\emptyset} U_A^\alpha(\sigma)]. \tag{2.5}$$

Here, $Z_V(\sigma)$ is the partition function depending on the values $\sigma(x), x \in V^c$ and the E_α are conjugate variables. As a reference measure we have taken the product of counting measures on G.
Suppose now that we have a transformation θ on G for which $\theta^2 = 1$ and which leaves invariant the counting measure on G. We fix a sequence of increasing space-time cubes $\Lambda_{N,L} = \{x = (n, i) \in \mathbb{Z}^{d+1}, |n| \leq N, |i| \leq L\}$. We can then define the transformations $\pi_{\Lambda_{N,L}} = \pi_{N,L}$ on Ω for which

$$(\pi_{N,L}\sigma)(n, i) = \sigma^{N,L}(n, i) = \theta(\sigma(-n, i)) \tag{2.6}$$

when $(n, i) \in \Lambda_{N,L}$ and $(\pi_{N,L}\sigma)(x) = \sigma(x), x \in \Lambda^c_{N,L}$. The time-reversal transformation π is then defined on local observables (functions) as

$$f\pi(\sigma) = f \circ \pi(\sigma) = f(\pi_{N,L}\sigma) \tag{2.7}$$

where for the local f with dependence set $\mathcal{D}_f$ (i.e. the smallest set A so that $f(\sigma) = f(\sigma(x), x \in A)$) $\Lambda_{N,L}$ is the smallest box containing $\mathcal{D}_f$.
We assume that $U_A^0(\pi_{N,L}\sigma) = U_A^0(\sigma)$ whenever $A \subset \Lambda_{N,L}$. In other words, H^0 in (2.1) is put as a reference 'action' or 'Hamiltonian' which is invariant under the time-reversal transformation π. One should think of the Gibbs measure $\nu \sim \exp[-E_0 H^0]$ as the unperturbed equilibrium steady state. The amplitudes E_α should correspond to gradients of intensive variables each conjugate to time-reversal breaking macro-variables $H^\alpha, \alpha = 1, \ldots, m$ and they determine for each of the m considered mechanisms the amount of driving away from equilibrium.
Associated to the transformation $\pi_{N,L}$ there is the Gibbs measure $\pi_{N,L}\mu$ defined via

$$\pi_{N,L}\mu(f) = \mu(f \circ \pi_{N,L}) \tag{2.8}$$

for an observable f on Ω. In the present context, most important for us is that the measure $\pi_{N,L}\mu$ has a density with respect to the original pathspace measure μ:

$$\frac{d(\pi_{N,L}\mu)}{d\mu} = e^{-R_{N,L}}. \tag{2.9}$$

This is automatically so for Gibbs measures but the full glory of Gibbsian states is not needed as we only require in (2.9) the existence of a density corresponding to the transformations $\pi_{N,L}$. In a more formal way (which does make sense however for a finite space-time lattice $W = \Lambda_{N,L}$), what we are doing is comparing the plausibility or weight $\mu[\sigma]$ in the pathspace measure μ of a trajectory σ of the system with the weight $\mu[\sigma^{N,L}]$ where $\sigma^{N,L}(n,i) = (\theta\sigma)(-n,i)$. The condition (2.9) then amounts to asking that $\mu[\sigma^{N,L}] = 0$ whenever $\mu[\sigma] = 0$ with well-defined ratio

$$\frac{\mu[\sigma^{N,L}]}{\mu[\sigma]} = e^{-R_{N,L}}. \tag{2.10}$$

At the risk of generating confusion, we will say that the pathspace measure μ (or, the dynamics generating it via its steady state statistics) is *microscopically reversible* if there exists a transformation θ for which this property (2.9) or (2.10) is satisfied. (This should not be confused with the condition of detailed balance that we will meet later and which will make $R_{N,L}$ a boundary term, i.e. of order $|\partial\Lambda_{N,L}|$.)
Continuing with our Gibbsian setup three properties of $R_{N,L}$ are immediate.

Proposition 2.1 *Given the above hypothesis for a translation invariant Gibbs measure μ,*

- *(Extensivity) There exists a continuous function $\dot{s}$ (constructed in (2.15) below) so that (with $\dot{s}_x = \dot{s} \circ \tau_x$)*

$$R_{N,L} = \sum_{x\in\Lambda_{N,L}} \dot{s}_x + O(NL^{d-1} + L^d) \tag{2.11}$$

  with order $|O|/N \leq cL^{d-1}$ if $L \leq N$.

- *(Symmetry) Let $P_{N,L}$ denote the law of $R_{N,L}$ as induced by μ and let $\tilde{P}_{N,L}$ denote the law of $-R_{N,L}$. Then,*

$$\frac{d\tilde{P}_{N,L}}{dP_{N,L}}(y) = e^{-y}. \tag{2.12}$$

  In particular, for each complex number $z \in \mathbb{C}$,

$$\mu(e^{-zR_{N,L}}) = \mu(e^{-(1-z)R_{N,L}}). \tag{2.13}$$

- *(Positive expectation)*

$$\mu(R_{N,L}) \geq 0 \tag{2.14}$$

  with equality only if $R_{N,L}$ is zero $\mu-$almost surely.

Proof: The extensivity is a standard result of the Gibbs formalism. With the definition

$$\dot{s} = \sum_{A\ni 0} \frac{1}{|A|} \sum_{\alpha=1}^{m} E_\alpha[U_A^\alpha \pi - U_A^\alpha] \tag{2.15}$$

it is readily checked, using (2.3) (see also [30]), that

$$R_{N,L} = \sum_{A\cap\Lambda_{N,L}\neq\emptyset} \sum_\alpha E_\alpha[U_A^\alpha \pi_{N,L} - U_A^\alpha] \tag{2.16}$$

can be written as a sum over the $x \in \Lambda_{N,L}$ of translations of $\dot{s}$ (as in (2.11)) up to boundary terms of order $\partial\Lambda_{N,L}$.

The symmetry is a consequence of (2.9) and $R_{N,L}\pi_{N,L} = -R_{N,L}$. Furthermore, there is the trivial identity

$$\mu(e^{-(1-z)R_{N,L}}) = \pi_{N,L}\mu(e^{zR_{N,L}}) = \mu(e^{-zR_{N,L}}). \tag{2.17}$$

Finally, since of course $\mu(e^{-R_{N,L}}) = 1$, the positivity of the expectation value follows from Jensen's inequality. (Remark: in fact, all moments of $R_{N,L}$ have non-negative expectation as is easy to find from the symmetry (2.13) in its characteristic function.) ■

We define the **entropy production** (creation rate) as the expectation value of $\dot{s}$:

$$\mu(\dot{s}) = \sum_{\alpha=1}^{m} E_\alpha \mu(J^\alpha) \tag{2.18}$$

where the currents J^α are

$$J^\alpha = \sum_{A \ni 0} \frac{1}{|A|}[U_A^\alpha \pi - U_A^\alpha]. \tag{2.19}$$

The rest of the paper is devoted to explaining and motivating this definition. This will be done as follows. We first list some of the properties of our definition in the present setup. This is repeated in a somewhat more abstract setting in the next subsection which however can guide us in many different cases. We then illustrate it via examples.

(2.18) attempts to define the entropy production directly as bilinear in thermodynamic forces E_α and fluxes J^α and it inherits the properties of $R_{N,L}$ in Proposition 2.1:

Proposition 2.2 *Under the hypotheses above for a translation invariant Gibbs measure μ,*

- $$\mu(\dot{s}) \geq 0 \tag{2.20}$$
  with equality only if $\dot{s} = 0$ $\mu-$almost surely (in which case $\mu\pi$ and μ must be Gibbs measures with respect to the same (physically equivalent) potential $(U_A\pi) \sim (U_A)$.

- *If μ is extremal, then*
  $$\mu(\dot{s}) = \lim_{N,L} \frac{1}{|\Lambda_{N,L}|} R_{N,L} \tag{2.21}$$
  $\mu-$almost surely.

- *Let $Q_{N,L}$ denote the law of $\sum_{x\in\Lambda_{N,L}} \dot{s}_x/|\Lambda_{N,L}|$ and let $\tilde{Q}_{N,L}$ denote the law of $-\sum_{x\in\Lambda_{N,L}} \dot{s}_x/|\Lambda_{N,L}|$. Then,*
  $$\lim_{N,L} \frac{1}{|\Lambda_{N,L}|} \ln \frac{d\tilde{Q}_{N,L}}{dQ_{N,L}}(y) = -y. \tag{2.22}$$
  In particular, the limit
  $$p(\lambda, E) = -\lim_{N,L} \frac{1}{|\Lambda_{N,L}|} \ln \mu(e^{-\sum_{\alpha,x\in\Lambda_{N,L}} \lambda_\alpha J_x^\alpha}) \tag{2.23}$$

exists and equals $p(2E-\lambda, E)$ *for all* $\Lambda = (\lambda_1, \ldots, \lambda_m), E = (E_1, \ldots, E_m) \in \mathbb{R}^m$.

Proof: (2.21) is a consequence of the law of large numbers using that μ is a translation invariant extremal Gibbs state. The rest of the proof was already given in [30].■

The symmetry of (2.23) for the generating function of $\dot{s}$ gives rise to relations between correlation functions when $p(\lambda, E)$ is differentiable with respect to λ and E. Most important is the following version of the Green-Kubo formula (also already proven in [30])

Corollary 2.1

$$L_{\alpha\gamma} = \frac{\partial}{\partial E_\alpha}\mu(\frac{\partial}{\partial E_\gamma}\dot{s})(E=0) = \sum_x \nu(J^\gamma J_x^\alpha). \tag{2.24}$$

Obviously, (2.24) is (Onsager-) symmetric in interchanging α with γ.

Remark: As far as we see there are two main approaches connecting the Gibbs formalism with nonequilibrium statistical mechanics. The first one is easiest to grasp for the 'mechanically inclined.' It is obtained by the explicit construction of the pathspace measure of various dynamics as a Gibbs measure. In fact (and fortunately), the full identification of the interaction is not needed for our purposes as we are only interested in a specific 'relative energy' (as in (2.9-2.10)). The second approach is more familiar to the 'statistically inclined.' It involves identifying Gibbs measures as solutions of the maximum entropy principle, given space-time information on macroscopic variables such as steady state currents. It does not build the pathspace measure from the microscopic dynamics but instead tries to predict macroscopic behavior from a Gibbs measure on the space-time trajectories which is statistically compatible with certain experimentally available data. We refer to the method of [22] and to specific examples (in the quantum domain) in e.g. [1]. In this paper, we deal exclusively with the first approach. It should however be realized that, while the first approach works equally well for systems containing only few degrees of freedom, the stone-wall character of the laws of irreversible thermodynamics can only be expected for systems containing a large number of degrees of freedom.

2.2 Abstract definition

Here we start from an abstract probability space $(\Omega, \mathcal{F}, \mu)$. The set Ω has to be thought of as the set of discrete or continuous time paths of a particle system. Next we consider an index set $\mathcal{S}, \leq$ equipped with the partial order $\leq$, and an increasing family of sigma-fields $\{\mathcal{F}_s, s \in \mathcal{S}\}, \mathcal{F}_s \subset \mathcal{F}, \forall s \in \mathcal{S}$, with $\mathcal{F} = \sigma(\cup_{s\in\mathcal{S}}\mathcal{F}_s)$. Typically $\mathcal{S}$ has to be thought of as some set of increasing space-time windows, e.g.

$$\mathcal{S} = \{[0,T], T \geq 0\}$$

or

$$\mathcal{S} = \{V \times [0,T], T \geq 0, V \subset \mathbb{Z}^d(\text{ finite })\}$$

ordered by inclusion. Consider a set of transformations $\{\pi_s, s \in \mathcal{S}\}$, indexed by $\mathcal{S}$ such that the following four conditions are satisfied:

- (i) $\pi_s : \Omega \to \Omega$ is $\mathcal{F}_s$ measurable for all $s \in \mathcal{S}$, i.e., $\pi_s^{-1}(A) \in \mathcal{F}_s$ for all $A \in \mathcal{F}_s$.
- (ii) $\forall s \in \mathcal{S}, \forall \omega \in \Omega, \pi_s \circ \pi_s(\omega) = \omega$
- (iii) $\forall s \in \mathcal{S}, \pi_s \mu$ is absolutely continuous with respect to μ.
- (iv)$\forall t > s$, $\forall$ $\mathcal{F}_s$ measurable f: $\mathbb{E}(f \circ \pi_t | \mathcal{F}_s) = f \circ \pi_s$

The transformation π_s has to be thought of as the time-reversal transformation in the window indexed by $s \in \mathcal{S}$.
By condition (iii) above, we can write the Radon-Nikodym derivative

$$\frac{d(\pi_s \mu)}{d\mu} = e^{-R_s} \tag{2.25}$$

and we immediately obtain, by the same proof, the analogue of Proposition 2.1:

Proposition 2.3 *For all complex numbers $z \in \mathbb{C}$,*

$$\mu(e^{-zR_s}) = \mu(e^{-(1-z)R_s}) \tag{2.26}$$

and $\mu(R_s) \geq 0$ with equality iff $R_s = 0$ $\mu-$ almost surely.

In the examples that follow we will always identify the function R_s by means of a Girsanov formula. The procedure is always to refer to a time-reversal invariant process ν and to compute the ratio

$$\frac{d\mu}{d\nu}(\pi_s(\omega)) / \frac{d\mu}{d\nu}(\omega) = e^{-R_s(\omega)}. \tag{2.27}$$

Since we have not insisted on a natural notion of spatial translations in our abstract setup we cannot expect to obtain an analogue to (2.11) which for statistical mechanics is of prime importance as it is related to the extensiveness of the entropy production. Note however that from condition (iv)

$$M_t := \mathbb{E}[\frac{d(\pi_t \mu)}{d\mu} | \mathcal{F}_t] \tag{2.28}$$

is a martingale and hence $\mathbb{E}(-\log M_t)$ is a non-decreasing function of time t ($-\log M_t$ is a non-negative submartingale). Most of the time, there will be a natural choice for an increasing function $\alpha : \mathcal{S} \to \mathbb{R}_+$ being a suitable normalization in the sense that $\sup_{s \in \mathcal{S}} \alpha(s) = +\infty$ and the limit (free energy functional)

$$p(\lambda) = -\lim_s \frac{1}{\alpha(s)} \ln \mu(e^{-\lambda R_s}) \tag{2.29}$$

exists (and is non-trivial) for $\lambda \in \mathbb{R}$. By Proposition 2.3 we have automatically $p(\lambda) = p(1 - \lambda)$, see (2.23).
We then call

$$\mu(\dot{s}) = \lim_s \frac{1}{\alpha(s)} \mu(R_s) \geq 0 \tag{2.30}$$

the entropy production corresponding to $(\Omega, \mathcal{F}, \mu)$ and transformations π_s. This is the analogue of (2.18-2.21) in the general context.

When we deal with "fluctuations of entropy production", then we mean fluctuations of the random variable $R_s/\alpha(s)$. In the identification of R_s, we always want to obtain R_s "up to boundary terms". More precisely, given a suitable normalization α, we define $R_s \doteq R'_s$ iff for all $\lambda > 0$

$$\limsup_s \frac{1}{\alpha(s)} \log \int \exp\left[(\lambda|R_s - R'_s|)\right] d\mu = 0. \tag{2.31}$$

This of course implies that R_s and R'_s give rise to the same free energy functional. We also denote $\frac{d(\pi_s\mu)}{d\mu} \doteq 1$ iff $p(\lambda) = 0$ for all $\lambda \geq 0$. When $R_s \doteq 0$ we say that R_s is "negligible". In this case, we say that the dynamics is time-reversal invariant.

3 Stochastic dynamics

In this section we discuss a variety of examples. In each particular case, from the Gibbs representation of the pathspace measure of the stochastic process (as in (2.9-2.10), we can identify the entropy production by comparison with the time-reversed process.

3.1 Spinflip dynamics

Spinflip processes (SFP) are continuous time Feller processes taking values in the space $X = \{+1,-1\}^{\mathbb{Z}^d}$ of Ising spin configurations. The spin at site $i \in \mathbb{Z}^d$ is flipped according to

$$\text{Prob}[\sigma_{t+dt}(i) = -\xi(i)|\sigma_t = \xi] = c(i,\xi)dt, \tag{3.1}$$

where the spin flip rate $c(i,\xi)$ is assumed strictly positive, translation invariant and local. The main difference with the automata considered in [30] is the sequential updating whereas the "discreteness" of time is replaced by the event times of a Poisson process. The generator of a SFP is defined on local functions via

$$Lf(\xi) = \sum_{i\in\mathbb{Z}^d} c(i,\xi)\left(f(\xi^i) - f(\xi)\right), \tag{3.2}$$

where ξ^i denotes the configuration obtained from ξ by flipping the spin at i, i.e. $\xi^i(j) = -\xi(i)$ if $j = i$ and $\xi^i(j) = \xi(j)$ for $j \neq i$. See [28] for the existence of the process with generator L and for more details on SFP. A probability measure ρ on X is called invariant or stationary if for all local f, $\int d\rho Lf = 0$. If we start the process from ρ then we obtain of course a stationary process. We denote by μ its pathspace measure. The spinflip rates are satisfying the condition of detailed balance if

$$\frac{c(i,\xi)}{c(i,\xi^i)} = \exp[-(H(\xi^i) - H(\xi))], \tag{3.3}$$

for some Hamiltonian H on X (see [28] for more details). If the spinflip rates satisfy the condition of detailed balance, then the Gibbs measures with respect to the Hamiltonian H are "reversible", which means that the process started at one of them is time-reversal invariant. In order to apply the general formalism of Section 2, we identify:

$$\Omega := \Omega_{T_0,\Lambda_{L_0}} := D([-T_0,T_0], \{-1,+1\}^{\Lambda_{L_0}}), \tag{3.4}$$

the set of cadlag trajectories of spin configurations in the finite volume $\Lambda_{L_0} := [-L_0, L_0]^d \cap \mathbb{Z}^d$ in the finite time interval $[-T_0, T_0]$. Elements of Ω are denoted by ω, σ. The time-reversal transformation is defined in windows $[-T, T] \times \Lambda_L$ for $T \leq T_0$ and $L \leq L_0$:

$$\begin{aligned} (\pi_{T,L}\sigma)_t(i) &= \sigma_{-t}(i) \text{ if } t \in [-T, T], i \in \Lambda_L \\ &= \sigma_t(i) \text{ otherwise.} \end{aligned} \tag{3.5}$$

Notice that this transformation does not preserve the right-continuity of the paths at the jumping times but we can modify the jumps of $\pi_{T,L}(\sigma)$ making it again an element of Ω. With a slight abuse of notation we write $\pi_{T,L}(\sigma)$ for this cadlag modification of (3.5).

Lemma 3.1

$$\frac{d\pi_{T,L}\mu}{d\mu} = \exp(-R_{T,L}), \tag{3.6}$$

with

$$R_{T,L}(\sigma) = \sum_{i \in \Lambda_L} \int_{-T}^{T} \log \frac{c(i, \sigma_s)}{c(i, \sigma_s^i)} dN_s^i + G_{T,L}(\sigma), \tag{3.7}$$

where

$$N_s^i = \frac{1}{2} \sum_{t \in [-T, s]} |\sigma_{t^+}(i) - \sigma_{t^-}(i)|$$

is the basic jump process, counting the number of flips at site i in the time interval $[-T, s]$ and $G_{T,L}$ is negligible in the sense (2.31).

Proof: This is a direct application of the Girsanov formula. We introduce the time-reversal invariant reference process of independent spinflips, i.e. the process with generator

$$L_0 f(\xi) = \sum_{i \in \Lambda_{L_0}} \left[f(\xi^i) - f(\xi) \right] \tag{3.8}$$

and denote by μ_ξ^0 its pathspace measure starting from $\xi \in \{-1, +1\}^{\Lambda_{L_0}}$, i.e. $\sigma(t = 0, .) \equiv \sigma_{t=0} \equiv \sigma_0 = \xi$. We also denote by μ the pathspace measure of the process with generator L starting from its stationary measure ρ, and by μ_0 the same for the process with generator L_0, starting from the stationary measure ρ_0. ¿From the Girsanov formula for point processes (see [29] p314) we obtain

$$\frac{d\mu_\xi}{d\mu_\xi^0} = \exp\left\{ \sum_i \int_{-T}^{T} \log c(i, \sigma_s) dN_s^i - \left[\int_{-T}^{T} c(i, \sigma_s) ds - 2T \right] \right\}, \tag{3.9}$$

Using time-reversal invariance of the reference process we obtain:

$$\begin{aligned} R_{L,T} &= -\log \frac{d(\pi_{T,L}\mu)}{d\mu} \\ &\doteq -\log \frac{d(\pi_{T,L}\mu)}{d\mu_0} + \log \frac{d\mu}{d\mu_0} && (3.10) \\ &\doteq \sum_{i \in \Lambda} \int_0^T \log \frac{c(i, \sigma_s)}{c(i, \sigma_s^i)} dN_s^i && (3.11) \end{aligned}$$

In the first step we put '$\doteq$' because the process μ_0 satisfies $\frac{d\pi_{T,L}\mu_0}{d\mu_0} \doteq 1$, in the last step because we omitted the term coming from the initial measures (ρ for the process, and ρ_0 for the reference process). ■

For the entropy production we obtain from Lemma 3.1

$$\mu(\dot{s}) = \lim_{T,\Lambda} \frac{1}{2T|\Lambda|} \mu \left(\sum_{i \in \Lambda} \int_{-T}^{T} \log \frac{c(i, \sigma_s)}{c(i, \sigma_s^i)} dN_s^i \right). \tag{3.12}$$

Use that $N_s^i - \int_{-T}^{s} c(i, \sigma_t) dt$ is a martingale. By stationarity and translation invariance of ρ, we obtain from (3.12)

$$\begin{aligned} \mu(\dot{s}) &= \lim_{T,\Lambda} \frac{1}{2T|\Lambda|} \mu \left(\sum_{i \in \Lambda} \int_{-T}^{T} \log \frac{c(i, \sigma_s)}{c(i, \sigma_s^i)} dN_s^i \right) \\ &= \int \rho(d\xi) c(0, \xi) \log \frac{c(0, \xi)}{c(0, \xi^0)}. \end{aligned} \tag{3.13}$$

If μ is time-reversal invariant, then (3.13) is zero: this is the case when ρ is a Gibbs measure with respect to the Hamiltonian of (3.3), i.e., when

$$\int \rho(d\sigma) c(x, \sigma) [f(\sigma^x) - f(\sigma)] = 0, \tag{3.14}$$

for all f local and for all $x \in \mathbb{Z}^d$.

Example (cf. [25]): Consider the one-dimensional spinflip dynamics with the following asymmetric rates:

$$c(x, \sigma) = \exp(-\beta \sigma(x) \sigma(x+1)). \tag{3.15}$$

The invariant measure ρ is the one-dimensional Ising model at inverse temperature $\beta/2$. The process starting from ρ is not time-reversal invariant. For the entropy production we find, after an easy calculation:

$$\mu(\dot{s}) = 2\beta \tanh(\frac{\beta}{2}) > 0. \tag{3.16}$$

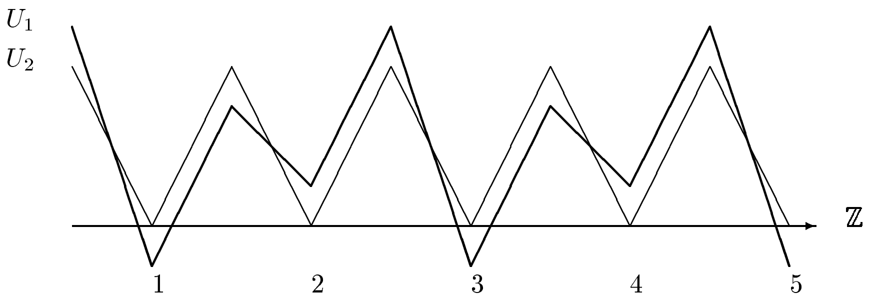
It would be interesting to understand how this could have been guessed from a physical interpretation of the model.

3.2 Molecular Motor

We illustrate our method for a simple model of a molecular motor. This example was suggested to us by H. Wagner. Since, at the time of writing this paper, we received the preprint [21] which deals with a similar model, we omit giving context and motivation and we refer to [21, 23, 31] for details. Here, after defining the simplest variant of the model, we proceed at once with our algorithm.

We consider a point particle that can jump on the one-dimensional lattice ($\mathbb{Z}$) under the influence of a periodic potential U. This potential U can take two different shapes U_1 and U_2 (see figure) and it switches between them at a rate Γ. The state of the particle can be characterized by a couple (x, α) where x

describes the position of the particle and $\alpha \in \{1,2\}$ specifies in which potential the particle moves. The potential U_2 is symmetric and periodic with period 1 while U_1 has minima of two different heights bounded by asymmetric energy barriers. It is also periodic but with period 2.



To be concrete, let us put for $n \in \mathbb{Z}$, $U_2(n) = 0$, $U_1(2n) = \epsilon = -U_1(2n+1)$. The asymmetry of the barriers implies that in U_1 the rates for jumping to the right are different from those for jumping to the left. The transitions occur at the rates:

$$\begin{aligned} & r\left((2n+1,1) \to (2n+2,1)\right) = \gamma_1 \\ & r\left((2n+1,1) \to (2n,1)\right) = \delta_1 \\ & r\left((2n,1) \to (2n+1,1)\right) = \gamma_2 \\ & r\left((2n,1) \to (2n-1,1)\right) = \delta_2 \\ & r\left((x,1) \to (x,2)\right) = \Gamma \\ & r\left((x,2) \to (x,1)\right) = \Gamma \\ & r\left((x,2) \to (x \pm 1,2)\right) = \gamma. \end{aligned} \tag{3.17}$$

In the figure above we have taken $\delta_1 < \gamma_1 < \gamma_2 < \delta_2$. It is the combination of the energy difference and the height of the energy barrier between adjacent minima that determines the rates in a concrete model.

We suppose that

$$\frac{\gamma_2}{\delta_1} = \frac{\delta_2}{\gamma_1} = \mathrm{e}^{2\epsilon/T_s} \tag{3.18}$$

for thermal transitions at the temperature T_s in the potential U_1. Notice that from this condition of detailed balance both potentials (separately) give rise to a time-reversal invariant dynamics. However switching between the two potentials at rate $\Gamma > 0$ destroys time-reversal invariance and can produce a netto current and a non-zero entropy production. The physical reason is that the jumping of the particle from site to site is a thermal process occurring at temperature T_s while the switching between the two potentials is also a thermal process but occurring at a different temperature $T_r = +\infty$. Through the system there is a net passage of heat from the reservoir at temperature T_r to the system at temperature T_s. As we will now see, our algorithm nicely captures this.
We distinguish the following four different states:

$$(x,h) \quad \leftrightarrow \quad \text{‘state’}$$

$$\begin{array}{rc} (2n+1,1) & 1 \\ (2n,1) & 2 \\ (2n+1,2) & 3 \\ (2n,2) & 4 \end{array} \tag{3.19}$$

Denote by p_i the probability to find the particle in state i. Then

$$\begin{aligned} \frac{dp_1}{dt} &= (\gamma_2+\delta_2)p_2 - (\gamma_1+\delta_1)p_1 + \Gamma(p_3-p_1), \\ \frac{dp_2}{dt} &= (\gamma_1+\delta_1)p_1 - (\gamma_2+\delta_2)p_2 + \Gamma(p_4-p_2), \\ \frac{dp_3}{dt} &= 2\gamma(p_4-p_3) + \Gamma(p_1-p_3), \\ \frac{dp_4}{dt} &= 2\gamma(p_3-p_4) + \Gamma(p_2-p_4). \end{aligned} \tag{3.20}$$

¿From (3.20) we easily obtain the following probabilities in the stationary measure: $p_i = 1/4 + (-1)^{i+1}a_i$, where $a_1 = a_2 = a$ and $a_3 = a_4 = b$, with

$$a = \frac{1}{4}\frac{(\gamma_2+\delta_2-\gamma_1-\delta_1)(4\gamma+\Gamma)}{(4\gamma+\Gamma)\Sigma+4\gamma\Gamma}, \tag{3.21}$$

$$b = \frac{1}{4}\frac{(\gamma_2+\delta_2-\gamma_1-\delta_1)\Gamma}{(4\gamma+\Gamma)\Sigma+4\gamma\Gamma}, \tag{3.22}$$

and $\Sigma = \gamma_1+\delta_1+\gamma_2+\delta_2$.

Let us now introduce the particle current in the stationary state:

$$\begin{aligned} J^+ &= p_1\gamma_1 - p_2\delta_2 + (p_3-p_4)\gamma & (3.23) \\ &= p_2\gamma_2 - p_1\delta_1 + (p_4-p_3)\gamma. & (3.24) \end{aligned}$$

Expression (3.23) is the net current to the right at odd and (3.24) at even lattice sites. Since there are no sources or sinks both expressions must be equal (as they are). Using the explicit expression for the stationary state we obtain

$$J^+ = \frac{1}{2}\frac{\gamma\Gamma(\gamma_1+\gamma_2-\delta_1-\delta_2)}{(4\gamma+\Gamma)\Sigma+4\gamma\Gamma}. \tag{3.25}$$

Apart from this particle current we can introduce yet another current. It is the heat current measuring the amount of energy that is displaced through our system.

$$\begin{aligned} J_Q &= \Gamma\ (p_4-p_2)\ \epsilon + \Gamma\ (p_1-p_3)\ \epsilon & (3.26) \\ &= \Gamma\ T_s\ (p_4-p_2)\log\frac{\delta_2}{\gamma_1}, & (3.27) \end{aligned}$$

where we used (3.18) and stationarity. J_Q is the rate at which energy or heat flows from the reservoir to our system at temperature T_s.

Let us now turn to the setup of Section 2.2. We put $\Omega = D([-T,T],\{1,2,3,4\})$ the set of paths on the states (3.19), $\mathcal{S} = \{[-T,T]:\ T \geq 0\}$ the set of increasing time windows,

$$\pi_T(\omega)(t) = \omega(-t) \tag{3.28}$$

and $\alpha([-T,T]) = 2T$. Let μ be the space-time extension of the stationary state $\rho = (p_1, p_2, p_3, p_4)$. Then

$$\begin{aligned} R_T(\omega) &= -\log\left[\frac{d(\pi_T\mu)}{d\mu}(\omega)\right] \\ &= \int_{-T}^{T} \log\frac{r(\omega_{t^-}, \omega_{t^+})}{r(\omega_{t^+}, \omega_{t^-})} dN_t, \end{aligned} \tag{3.29}$$

where N_t denotes a mean one Poisson process.

Therefore the entropy production corresponding to $(\Omega, \mathcal{F}, \mu)$ is

$$\begin{aligned} \mu(\dot{s}) = \lim_T \frac{1}{2T}\mu(R_T) = \\ p_1\left[\gamma_1 \log\frac{\gamma_1}{\delta_2} + \delta_1 \log\frac{\delta_1}{\gamma_2}\right] + p_2\left[\gamma_2 \log\frac{\gamma_2}{\delta_1} + \delta_2 \log\frac{\delta_2}{\gamma_1}\right]. \end{aligned} \tag{3.30}$$

Using $\gamma_1\gamma_2 = \delta_1\delta_2$ together with the explicit expression for the stationary state we can rewrite this as

$$\mu(\dot{s}) = \Gamma\ [p_4 - p_2]\ \log\frac{\delta_2}{\gamma_1} = J_Q \cdot \frac{1}{T_s}. \tag{3.31}$$

Again we recognize the form 'current $\cdot$ field', where the role of the field is played by an inverse temperature difference. Since we have taken Γ constant independent of the particle position, the effective temperature of the second reservoir is $T_r = \infty$ and thus $\mu(\dot{s}) = J_Q/T_s = J_Q/T_s - J_Q/T_r$. The particle current J^+ does not contribute to the entropy production because we did not include an external load, see [21]. Remark also that $\mu(\dot{s}) \geq 0$ with equality only if $J_Q = 0$. This happens when $\gamma\ \Gamma = 0$ or $\gamma_1 = \delta_2$, $\delta_1 = \gamma_2$ ($\epsilon = 0$) or $\gamma_1 = \delta_1 = \gamma_2 = \delta_2$ ($T_s = \infty$).
This shows that from our algorithm, the exact expression for the entropy production is obtained. Of course, for a thermodynamic interpretation, one really should be imagining lots of particles.

3.3 Particle exchange dynamics

For particle exchange processes (PEP), the configuration space is $X = \{0,1\}^{\mathbb{Z}^d}$, where $\xi(i) = 1, 0$ is interpreted as the presence respectively absence of a particle at lattice site i. A PEP is a Feller process on X with generator given by

$$Lf(\xi) = \sum_{|i-j|=1} c(i,j,\xi)\left[f(\xi^{ij}) - f(\xi)\right] \tag{3.32}$$

where $|i-j| = \sum_{\alpha=1}^{d} |i_\alpha - j_\alpha| = 1$ and

$$\xi^{ij}(k) = \begin{cases} \xi(j) & k = i \\ \xi(i) & k = j \\ \xi(k) & \text{otherwise.} \end{cases} \tag{3.33}$$

The exchange rates $c(i,j,\xi)$ are strictly positive, local and translation invariant. As for spinflip processes, we put

$$\Omega = \Omega_{L_0,T_0} = D([-T_0,T_0], \{0,1\}^{\Lambda_{L_0}}), \tag{3.34}$$

$$\begin{aligned}\pi_{T,L}(\omega)_t(i) &= \omega_{-t}(i) \; t\in[-T,T], i\in\Lambda_L \\ &= \omega_t(i) \;\; \text{otherwise} \;, \end{aligned} \tag{3.35}$$

where $T \le T_0$, $L \le L_0$. We then have the following analogue of Lemma 3.1:

Lemma 3.2

$$\frac{d\pi_{T,L}\mu}{d\mu} = \exp\left(-R_{T,L}\right), \tag{3.36}$$

with

$$R_{T,L}(\omega) = \sum_{i,j\in\Lambda_L, |i-j|=1} \int_{-T}^{T} \log \frac{c(i,j,\omega_s)}{c(i,j,\omega_s^{ij})} dN_s^{ij} + G_{T,L}(\omega), \tag{3.37}$$

where

$$N_s^{ij} = \sum_{t\in[-T,s]} \frac{1}{2}|\omega_{t^+}(i) - \omega_{t^+}(j) - \omega_{t^-}(i) + \omega_{t^-}(j)|, \tag{3.38}$$

and $G_{T,L}$ *is negligible in the sense (2.31).*

Proof: The lemma follows again from the Girsanov formula. The basic jump processes are now indexed by nearest neighbour bonds rather than by sites. As time-reversal invariant reference process we choose the simple symmetric exclusion process with generator

$$L_0 f(\xi) = \sum_{|i-j|=1} \left[f(\xi^{ij}) - f(\xi)\right]. \tag{3.39}$$

Again, μ_ξ and μ_ξ^0 denote the pathspace measures starting from a configuration $\xi \in X$ under the original and reference dynamics respectively, and μ and μ_0 denote the pathspace measures starting from a stationary measure ρ (resp. ρ_0). We obtain

$$\frac{d\mu_\xi}{d\mu_\xi^0} = \exp\left\{ \sum_{|i-j|=1} \int_{-T}^{T} \log c(i,j,\omega_s) dN_s^{ij} - \left[\int_{-T}^{T} c(i,j,\omega_s) ds - 2T \right] \right\}, \tag{3.40}$$

and

$$\frac{d(\pi_{T,L}\mu)}{d\mu} \doteq \exp\left\{ \sum_{|i-j|=1, \{i,j\}\subset\Lambda_L} \int_{-T}^{T} -\log \frac{c(i,j,\omega_s)}{c(i,j,\omega_s^{ij})} dN_s^{ij} \right\}. \tag{3.41}$$

■

The entropy production is

$$\begin{aligned}\mu(\dot{s}) &= \lim_{T,\Lambda} \frac{1}{2T|\Lambda|} \int d\mu \left(\sum_{|i-j|=1, \{i,j\}\subset\Lambda} \int_{-T}^{T} \log \frac{c(i,j,\sigma_s)}{c(i,j,\sigma_s^{ij})} dN_s^{ij} \right) \\ &= \lim_{\Lambda} \frac{1}{|\Lambda|} \sum_{|i-j|=1, \{i,j\}\subset\Lambda} \int \rho(d\xi) c(i,j,\xi) \log \frac{c(i,j,\xi)}{c(i,j,\xi^{ij})}. \end{aligned} \tag{3.42}$$

If ρ is translation invariant then

$$\mu(\dot{s}) = \frac{1}{2} \sum_{|i|=1} \int \rho(d\xi) c(0,i,\xi) \log \frac{c(0,i,\xi)}{c(0,i,\xi^{0i})}. \tag{3.43}$$

Examples: Asymmetric exclusion processes. We consider the case of bulk driven diffusive lattice gases as was done in [27], see also [39]. For the asymmetric exclusion process in dimension $d = 1$,

$$c(i,i+1,\xi) = \xi(i)(1-\xi(i+1))\frac{\mathrm{e}^{E/2}}{2} + \xi(i+1)(1-\xi(i))\frac{\mathrm{e}^{-E/2}}{2}. \tag{3.44}$$

the Bernoulli measure ρ, with particle density q is a stationary (non-reversible) measure. Denoting by μ its space-time extension, we find for the entropy production

$$\begin{aligned} \mu(\dot{s}) &= \int c(0,1,\xi) \log \frac{c(0,1,\xi)}{c(0,1,\xi^{01})} \rho(d\xi) \\ &= E\ q\ (1-q)\ \sinh(\frac{E}{2}). \end{aligned} \tag{3.45}$$

The particle current is defined by

$$j_{T,L} = \frac{1}{2TL^d} \sum_{\{i,j\} \subset \Lambda_L, |i-j|=1} \int_{-T}^{T} [\omega_s(i) - \omega_s(j)] dN_s^{ij} \tag{3.46}$$

Notice the relation

$$\frac{\partial R_{T,L}}{\partial E} \doteq 2TL^d j_{T,L}. \tag{3.47}$$

In the limit $T, L \uparrow \infty$, the expectation of the particle current is precisely

$$j(E) = \int \rho(d\xi) c(0,1,\xi)(\xi(0) - \xi(1)) = q\ (1-q)\ \sinh(\frac{E}{2}). \tag{3.48}$$

This is also the quantity appearing as the current in the hydrodynamic equation which in this case is the Burgers' equation for a space-time density profile $q_t(r), r \in \mathbb{R}$:

$$\frac{\partial q}{\partial t} = \sinh(\frac{E}{2}) \frac{\partial}{\partial r}(q(1-q)). \tag{3.49}$$

Hence the entropy production can be written in the form

$$\mu(\dot{s}) = E.j(E), \tag{3.50}$$

which corresponds to the dissipated power in a resistor according to Ohm's law. As soon as the process is not time-reversal invariant (i.e. $E \neq 0$), the entropy production (3.45) is strictly positive.

It is not necessary to have a non-vanishing particle current in order to have strictly positive entropy production. Consider e.g. the one-dimensional asymmetric exclusion process with generator

$$Lf(\xi) = \sum_{i,j \in \mathbb{Z}} p(j-i)\xi(i)(1-\xi(j)) \left[f(\eta^{ij}) - f(\eta)\right] \tag{3.51}$$

where $p : \mathbb{Z} \to \mathbb{R}^+$ satisfies $\sum_i p(i) = 1$, $\sum_i ip(i) = 0$ but $p(i) \neq p(-i)$. Then we find, starting from the Bernoulli measure ρ with density q:

$$\mu(\dot{s}) = q(1-q) \sum_i (p(i) - p(-i)) \log \frac{p(i)}{p(-i)} \tag{3.52}$$

which is clearly strictly positive, whereas the hydrodynamical equation of this process is a linear diffusion equation (i.e., we do not have macroscopic drift).

Remark: The asymmetric exclusion process has stationary measures which are non-translation invariant but reversible (i.e., the corresponding process is time-reversal invariant). When we start the process from such a measure, the expectation of $R_{T,L}$ in (3.37) is zero but the random variable is not negligible in the sense of (2.31). This shows that the reversible measure has to be translation invariant in order to conclude negligibility of $R_{T,L}$.

3.4 Diffusion processes

In this section we consider, following [27], three examples of diffusion processes. A fluctuation theorem for diffusion processes was first discussed in [24]. Via the Girsanov formula we find a Gibbsian representation for the path space measure and we can apply our formalism to obtain the entropy production. We will use a similar notation as in [27].

Example : Diffusion with drift

Consider the stochastic differential equation ([29])

$$dx_t = [(c(x_t) - \frac{1}{2}a.\nabla U(x_t)) + \frac{1}{2}\nabla.a(x_t)]dt + \sqrt{a(x_t)}dW_t, \tag{3.53}$$

The covariance matrix $a(x)$ is strictly positive, sufficiently smooth and bounded, and $U \in \mathcal{C}^2$ satisfies $\int_{\mathbb{R}^d} \exp(-U(x))dx < \infty$. Furthermore we suppose that the drift c is sufficiently confining so that there exists a unique stationary probability measure which is absolutely continuous with respect to Lebesgue measure. To put ourselves in the context of Section 2.2 we identify

$$\begin{aligned}
&\Omega := \Omega_T := \mathcal{C}([-T,T], \mathbb{R}^d) \\
&\mathcal{S} := \{[-T,T] : T \geq 0\} \\
&\pi_T(\omega)(s) := \omega(-s) \\
&\alpha([-T,T]) := 2T.
\end{aligned} \tag{3.54}$$

The transformation π_T is time-reversal in the time window $[-T,T]$. As before we denote by ρ the stationary measure of the process x_t and μ the corresponding pathspace measure.

Lemma 3.3

$$\frac{d\pi_T\mu}{d\mu} = \exp(-R_T), \tag{3.55}$$

with

$$R_T(\omega) = 2\int_{-T}^{T} a^{-1}(\omega_s)c(\omega_s) \circ d\omega_s + G_T, \tag{3.56}$$

where $\circ$ stands for Stratonovich integral and G_T is negligible in the sense of (2.31).

Proof: If $c \equiv 0$ in (3.53), then the corresponding process $\{y_t : t \in [-T,T]\}$ satisfying

$$dy_t = [-\frac{1}{2}a.\nabla U(y_t) + \frac{1}{2}\nabla.a(y_t)]dt + \sqrt{a(y_t)}dW_t, \tag{3.57}$$

is time-reversal invariant when started (at $t = -T$) from the stationary measure

$$\rho_0(dx) := \frac{\exp[-U(x)]dx}{\int_{\mathbb{R}^d} \exp[-U(x)]dx}. \tag{3.58}$$

¿From the Girsanov formula ([29]) we obtain the following expression for the Radon-Nikodym derivative of the pathspace measure of the process $\{x_t : t \in [-T,T]\}$ with respect to the process $\{y_t : t \in [-T,T]\}$, starting at the same point $u = x_{t=-T} = y_{t=-T}$:

$$\frac{d\mu_u}{d\mu_u^0}(\omega) = \exp\left\{\int_{-T}^{T} a^{-1}(\omega_s)c(\omega_s)d\omega_s - \frac{1}{2}\int_{-T}^{T} a^{-1}(\omega_s)\right.$$
$$\left.\left[(c - \frac{1}{2}a.\nabla U + \frac{1}{2}\nabla.a)^2(\omega_s) - (-\frac{1}{2}a.\nabla U + \frac{1}{2}\nabla.a)^2(\omega_s)\right] ds\right\}. \tag{3.59}$$

In this expression the integrals are Ito-integrals, i.e.,

$$\int_{-T}^{T} f(\omega_s)d\omega_s = \lim_{n\uparrow\infty, \Delta s\downarrow 0} \sum_{j=1}^{n} f(\omega_{s_{j-1}})(\omega_{s_j} - \omega_{s_{j-1}}), \tag{3.60}$$

where $-T = s_0 < s_1 < \ldots < s_n = T$, $\Delta s_j = s_j - s_{j-1}$, is a subdivision of the interval $[-T,T]$, such that $\max_j \Delta s_j \downarrow 0$ for $n \uparrow \infty$. The limit in (3.60) has to be understood in L^2-sense, and it is important to evaluate f in the points $\omega_{s_{j-1}}$ (the left endpoint) in order to obtain the Ito-integral. If we evaluate in the midpoints we obtain the Stratonovich integral, i.e., for continuous f:

$$\begin{aligned}\int_{-T}^{T} f(\omega_s)\circ d\omega_s &:= \lim_{n\uparrow\infty, \Delta s\downarrow 0} \sum_{j=1}^{n} f(\frac{\omega_{s_{j-1}} + \omega_{s_j}}{2})(\omega_{s_j} - \omega_{s_{j-1}}) \\ &= \lim_{n\uparrow\infty, \Delta s\downarrow 0} \frac{1}{2}\sum_{j=1}^{n} [f(\omega_{s_{j-1}}) + f(\omega_{s_j})](\omega_{s_j} - \omega_{s_{j-1}})\end{aligned} \tag{3.61}$$

where $\circ$ stands for Stratonovich. The procedure to obtain $d(\pi_T\mu)/d\mu_0$ is now straightforward: we plug in the time reversed path $\pi_T(\omega)$ in the Girsanov formula, and use the reversibility of $\{y_t : 0 \le t \le T\}$ to obtain

$$\frac{d(\pi_T\mu)}{d\mu}(\omega) = \frac{\frac{d\mu}{d\mu_0}(\pi_T(\omega))}{\frac{d\mu}{d\mu_0}(\omega)} \tag{3.62}$$

Next use the following:

Lemma 3.4 *1.* $[\int_{-T}^{T} f(\omega_s)d\omega_s]\circ\pi_T = -2\int_{-T}^{T} f(\omega_s)\circ d\omega_s + \int_{-T}^{T} f(\omega_s)d\omega_s$.

2. $[\int_{-T}^{T} f(\omega_s)ds]\circ\pi_T = \int_{-T}^{T} f(\omega_s)ds$.

Proof: 1. Use expression (3.60) for the Ito integral:

$$\begin{aligned}
\left(\int_{-T}^{T} f(\omega_s)d\omega_s\right)\circ\pi_T &= \left[\lim_{n\uparrow\infty,\Delta s\downarrow 0}\sum_{j=1}^{n} f(\omega_{s_{j-1}})(\omega_{s_j}-\omega_{s_{j-1}})\right]\circ\pi_T \\
&= -\lim_{n\uparrow\infty,\Delta s\downarrow 0}\sum_{j=1}^{n} f(\omega_{s_j})(\omega_{s_j}-\omega_{s_{j-1}}) \\
&= -2\int_{-T}^{T} f(\omega_s)\circ d\omega_s + \int_{-T}^{T} f(\omega_s)d\omega_s. \qquad (3.63)
\end{aligned}$$

where we used (3.61).
2. The second equality holds trivialy. ■
¿From (3.59), (3.62), and this lemma we obtain finally

$$\frac{d(\pi_T\mu)}{d\mu} \doteq \exp\left[-2\int_{-T}^{T} a^{-1}(\omega_s)c(\omega_s)\circ d\omega_s\right], \qquad (3.64)$$

where we put $\doteq$ because we start the process $\{x_t : -T \le t \le T\}$ from its true stationary measure and not from the reversible measure of the process $\{y_t : -T \le t \le T\}$, hence we have to include an extra factor of order one (as $T \uparrow \infty$, since these two measures are absolutely continuous) in the expression (3.64). This finishes the proof of Lemma 3.3. ■
The entropy production corresponding to $(\Omega, \mathcal{F}, \mu)$ is

$$\begin{aligned}
\mu(\dot{s}) &= -\lim_{T}\frac{1}{2T}\mathbb{E}_\rho\left[\log\frac{d(\pi_T\mu_u)}{d\mu_u}\right] = \lim_{T}\frac{1}{T}\mathbb{E}_\rho\left[\int_{-T}^{T} a^{-1}(\omega_s)c(\omega_s)\circ d\omega_s\right] \\
&= 4\int \rho(dx)[(a^{-1}(x)c(x)).c(x) + \frac{1}{2}\nabla.c(x)]. \qquad (3.65)
\end{aligned}$$

When $c = -1/2a.\nabla\tilde{U}$ we have

$$-2\int_{-T}^{T} a^{-1}(x_s)c(x_s)\circ dx_s = \tilde{U}(x_T) - \tilde{U}(x_{-T}), \qquad (3.66)$$

which is a boundary term. Hence, in the case of a conservative driving force we get $\mu(\dot{s}) = 0$ as it should since then the process $\{x_t : \ -T \le t \le T\}$ is time-reversal invariant.

Example: Langevin Equation with temperature gradient:
Consider the following stochastic differential equation

$$\left\{\begin{array}{rcl} dx_t &=& v_t dt \\ dv_t &=& \left[-\frac{\nabla U}{m}(x_t) - \gamma(x_t)v_t\right]dt + \sqrt{\frac{2\gamma(x_t)}{m\beta(x_t)}}dW_t, \end{array}\right. \qquad (3.67)$$

which describes a particle subject to friction, a conservative force and a random force. We take $U \in \mathcal{C}^2$ and $\gamma, \beta \in \mathcal{C}^2$ positive, bounded and bounded away from zero. This guarantees the existence of a unique stationary probability measure absolutely continuous with respect to Lebesgue measure. To apply the formalism of Section 2.2, we identify $\Omega = \Omega_T := \mathcal{C}([-T,T],\mathbb{R}^d)$ the velocity paths, $\mathcal{S} := \{[-T,T] : T \ge 0\}$, $\alpha([-T,T]) := 2T$ and $\pi_T\omega(s) = -\omega(-s)$. The minus sign in the transformation π_T comes from the fact that $\omega(t)$ is interpreted as the velocity at time t.

Lemma 3.5

$$\frac{d\pi_T\mu}{d\mu} = \exp(-R_T), \tag{3.68}$$

with

$$R_T = \int_{-T}^{T} \frac{1}{2}[mv_s^2 + U(x_s)] \circ d\beta(x_s) + G_T, \tag{3.69}$$

where G_T is negligible in the sense of (2.31).

Proof: We introduce the reference process

$$\begin{cases} dX_t &= V_t dt \\ dV_t &= \left[-\frac{\nabla U}{m}(X_t) - \frac{\gamma(X_t)}{\beta(X_t)}\right] dt + \sqrt{\frac{2\gamma(X_t)m}{\beta(X_t)}} dW_t. \end{cases} \tag{3.70}$$

In this process, $\{V_t,\ t \geq 0\}$ is π-invariant when started from the measure

$$\rho_0(dV) = \frac{\mathrm{e}^{-1/2mV^2} dV}{\int \mathrm{e}^{-1/2mV^2} dV}.$$

Let μ_u and μ_u^0 be the path measures in the two processes, starting at a common initial condition $u = (x_0, v_0) = (X_0, V_0) \in \mathbb{R}^d$. ¿From the Girsanov formula, we obtain

$$\begin{aligned} \log \frac{d\mu_u}{d\mu_u^0} &= \int_{-T}^{T} \frac{1}{2} m(1-\beta(X_s))V_s dV_s \\ &\quad + \frac{1}{2}\int_{-T}^{T} \nabla U(X_s)V_s(1-\beta(X_s))ds \\ &\quad + \int_{-T}^{T} \Phi(X_s, V_s)ds, \end{aligned} \tag{3.71}$$

where the last integral is π_T-invariant:

$$\left[\int_{-T}^{T} \Phi(X_s, V_s)ds\right] \circ \pi_T = \int_{-T}^{T} \Phi(X_s, V_s)ds. \tag{3.72}$$

Hence, using Lemma 3.4,

$$\begin{aligned} R_T &\doteq \log \frac{d\mu}{d\mu_0} - \log \frac{d\mu}{d\mu_0} \circ \pi_T \\ &\doteq \int_{-T}^{T} m(1-\beta(x_s))v_s \circ dv_s + \int_{-T}^{T} \nabla U(x_s)v_s(1-\beta(x_s))ds \\ &= \int_{-T}^{T} \frac{1}{2} m(1-\beta(x_s)) \circ dv_s^2 + \int_{-T}^{T} (1-\beta(x_s)) \circ dU(x_s) \\ &\doteq \int_{-T}^{T} \left[\frac{1}{2}mv_s^2 + U(x_s)\right] \circ d\beta(x_s), \end{aligned} \tag{3.73}$$

where in the last line we have put '$\doteq$' because we omitted the boundary terms of the partial integration. ■

We obtain for the entropy production:

$$\mu(\dot{s}) = \lim_T \frac{1}{2T}\mu(R_T) = \lim_T \frac{1}{2T}\mathbb{E}_\rho \left[\int_{-T}^{T} \left[\frac{1}{2}mv_s^2 + U(x_s)\right] \circ d\beta(x_s)\right] \tag{3.74}$$

where once again, ρ is a stationary measure and μ is its space-time extension.

Example 3: Langevin equation with non-conservative driving force.

We now put β constant in (3.67) and consider

$$\begin{cases} dx_t &= v_t dt \\ dv_t &= \left[\frac{1}{m}F(x_t) - \gamma(x_t)v_t\right] dt + \sqrt{\frac{2\gamma(x_t)}{m\beta}}dW_t. \end{cases} \tag{3.75}$$

where again F is supposed to be sufficiently confining so that the velocity process v_t has a unique stationary probability measure which is absolutely continuous with respect to Lebesgue measure.

Lemma 3.6 *As in (3.5),*

$$\frac{d\pi_T\mu}{d\mu} = \exp(-R_T), \tag{3.76}$$

with now

$$R_T = \int_{-T}^{T} \beta F(x_s)v_s ds + G_T, \tag{3.77}$$

where G_T is negligible in the sense of (2.31).

Proof: The reversible reference process now reads

$$\begin{cases} dX_t &= V_t dt \\ dV_t &= -\gamma(X_t)V_t dt + \sqrt{\frac{2\gamma(X_t)}{m\beta}}dW_t, \end{cases} \tag{3.78}$$

i.e., in this process $\{V_t,\ t \in [-T,T]\}$ is π-invariant when started from the measure

$$\rho_0(dV) = \frac{\mathrm{e}^{-1/2mV^2}dV}{\int \mathrm{e}^{-1/2mV^2}dV}.$$

¿From the Girsanov formula we obtain

$$\begin{aligned} \log\frac{d\mu_u}{d\mu_u^0} &= \int_{-T}^{T} \frac{\beta}{2\gamma(X_s)}F(X_s)dV_s \\ &\quad -\frac{1}{2}\int_{-T}^{T} \frac{\beta}{2m\gamma(X_s)}F^2(X_s)ds \\ &\quad +\frac{1}{2}\int_{-T}^{T} \beta F(X_s)V_s ds \end{aligned} \tag{3.79}$$

and thus

$$R_T = \log\frac{d\mu}{d\mu_0} - \log\frac{d\mu}{d\mu_0}\circ\pi_T \doteq \int_{-T}^{T}\beta F(x_s)v_s ds. \tag{3.80}$$

We used that for a $\mathcal{C}^1$ function $\phi : \mathbb{R}^d \to \mathbb{R}^d$

$$\left[\int_{-T}^{T}\phi(X_s)dV_s\right]\circ\pi_T - \left[\int_{-T}^{T}\phi(X_s)dV_s\right] = 0. \tag{3.81}$$

This can be seen from the following calculation:

$$
\begin{aligned}
&\left|\left(\int_{-T}^{T}\phi(X_s)dV_s\right)\circ\pi_T - \int_{-T}^{T}\phi(X_s)dV_s\right| \\
&= \left|\lim_{n\uparrow\infty,\Delta S\downarrow 0}\sum_{i=1}^{n}\phi(X_{s_i})(V_{s_i}-V_{s_{i-1}}) - \lim_{n\uparrow\infty,\Delta S\downarrow 0}\sum_{i=1}^{n}\phi(X_{s_{i-1}})(V_{s_i}-V_{s_{i-1}})\right| \\
&= \lim_{n\uparrow\infty,\Delta S\downarrow 0}\sum_{i=1}^{n}\left[\phi(\int_{-T}^{s_i}V_r dr) - \phi(\int_{-T}^{s_{i-1}}V_r dr)\right]\left[V_{s_i}-V_{s_{i-1}}\right] \\
&\leq \|\phi'\|_\infty \sup_{0\leq t\leq T}|V_s|\sqrt{\langle V,V\rangle_T}\lim_{n\uparrow\infty,\Delta S\downarrow 0}\sqrt{\sum_{i=1}^{n}(s_i-s_{i-1})^2} \\
&= 0 \qquad (3.82)
\end{aligned}
$$

where $\langle V,V\rangle_T$ denotes the quadratic variation process:

$$
< V,V >_T = \lim_{n\uparrow\infty,\Delta s\downarrow 0}\sum_{i=1}^{n}\left(V_{s_i}-V_{s_{i-1}}\right)^2. \qquad (3.83)
$$

■

For this example we conclude that the entropy production corresponding to $(\Omega,\mathcal{F},\mu)$ is given by

$$
\mu(\dot{s}) = \lim_T \frac{1}{T}\int d\mu\left[\int_{-T}^{T}\beta F(x_s)v_s ds\right]. \qquad (3.84)
$$

If F is conservative, i.e. $F(X_s) = -\nabla U(X_s)$, then the integral $-\int_{-T}^{T}\beta F(X_s)V_s ds = U(X_T) - U(X_{-T})$ is a boundary term, making $\mu(\dot{s}) = 0$, expressing the time reversal symmetry.

Notice that (3.84) is not dependent on γ, so that at least formally in the limit $\gamma\to 0$ ("zero noise limit"), we can use the same expression.

Remark: There is one generalization which we have not considered so far and which is physically rather important. It concerns the case of dynamics with memory. When the time-scales of a reservoir and a subsystem have not been infinitely separated, memory effects are present. Fortunately, in principle our setup can handle this case. After all, a Markovian dynamics just corresponds to a short range interaction in the time-direction for our space-time Gibbs measure. If this interaction happens to be long range (but summable in the appropriate sense), the formalism still applies unchanged. Remark indeed that in order to apply the Girsanov formula for the stochastic differential equation (3.53), we do not need that the drift is a function $c_t(x) = c(x_t)$; it suffices that c_t is an adapted process, hence x_t does not need to to be a Markov process. The only problem is to find the correct analogue of Lemma 3.4 which will now envolve anticipating stochastic integrals.

4 Deterministic dynamics

Dynamical systems: In recent years new ideas in nonequilibrium statistical mechanics have emerged from the study of smooth dynamical systems, see e.g.

[35]. In particular, the chaotic character of a time evolution plays an important role and, depending on author and context, is argued to be responsible for the positivity of entropy production and for the positivity of transport coefficients. This, at first sight, is paradoxical since the results obtained from the theory of dynamical systems apply to small systems. It is certainly not the case that nonequilibrium behavior (such as macroscopic irreversibility) is compatible with systems having only a few degrees of freedom. One suggestion to bridge this gap is contained in the Chaoticity Hypothesis of Gallavotti and Cohen, [13, 14, 16, 17]. We will come back to this below. We start with some general remarks.

Consider a discrete time dynamics on a compact connected Riemannian manifold X defined by the map $T: X \rightarrow X$. Extra conditions will have to be added but we prefer to postpone them until they are explicitly needed. Trajectories starting from $v \in X$ are sequences $(v, Tv, T^2v, \ldots)$. For every trajectory-segment $(v_1, v_2, \ldots, v_n)$ we must have $v_k = Tv_{k-1}$.
We assume that T is invertible and that there exists an involution θ on X which leaves the Riemann metric dv on X invariant, $\theta^2 = 1$, so that $\theta \circ T^{-1} = T \circ \theta$. This is called reversibility. To see the relation with (2.9) or (2.10), it is crucial to observe that the trajectory-segment $(v_{-n}, v_{-n+1}, \ldots, v_n)$ is allowed if and only if the trajectory segment $(\theta v_n, \theta v_{n-1}, \ldots, \theta v_{-n})$ is allowed. One of the segments is then of the form $(T^{-n}v, T^{-n+1}v, \ldots, T^n v)$ and the other $(\theta T^n v, \theta T^{n-1}v, \ldots, \theta T^{-n}v)$ is allowed because $T(\theta T^k v) = \theta T^{k-1}v$. Another related consequence comes from the fact that if $T^n v = v$ for a certain n, then also $T^n w = w$ for $w = \theta v$. In other words, θ is a bijection on the set of $n-$periodic points Fix $T^n = \{v \in X, T^n v = v\}$.
Next, since we want to study steady state properties, comes considering time-invariant states. Most of the time, there are plenty of them and it is important to select the natural ones. This can of course only be decided from the (partial) information we have on the particular system we are interested in (for example via initial conditions, symmetries etc.) At any rate, steady states μ describing the statistics of trajectories are completely determined by the selection of an invariant measure ρ in the sense that the only randomness in a trajectory comes from the initial data. It is therefore somewhat artifical to use another notation for an invariant measure ρ and for its corresponding steady state (pathspace measure) μ since, for an observable f that depends on the configuration of the system at times $n_1, n_2, \ldots, n_k$, we have

$$\mu(f(v_{n_1}, \ldots, v_{n_k})) = \rho(f(v, T^{n_2 - n_1}v, \ldots, T^{n_k - n_1}v). \tag{4.1}$$

As far as we are aware, there are two main strategies to connect steady states to Gibbs measures, getting the Gibbs formalism of Section 2 at work also for deterministic dynamics. The first strategy uses the concept of Markov partition and symbolic dynamics and, for our purposes, is most useful for (if not limited to) mixing Anosov diffeomorphisms T; we refer to [3, 4, 5, 6, 37, 38] for older results and newer extensions. The second strategy uses the definition of Capocaccia in [7], and the equivalence of Gibbs states and equilibrium states for homeomorphisms satisfying expansiveness and specification, see e.g. [19].
Whatever strategy is taken, the entropy production would be defined in a similar way as for stochastic dynamics: suppose that we can find a function $\dot{S}_N(v) = \sum_{n=-N}^{N} \dot{s}(T^n v)$ with $\dot{s}(\theta(Tv)) = -\dot{s}(v)$ such that for all continuous f

on X,

$$\limsup_N \frac{1}{N} \log \frac{\int \rho(dv) f(v) e^{-\dot{S}_N(v)}}{\int \rho(dv) f(\theta v)} = 0, \tag{4.2}$$

then we call $\rho(\dot{s})$ the entropy production of the dynamics T in the state ρ. This, again, should be compared with the situation for Gibbs states $\tilde{\mu}$ (here, one-dimensional) with θ in (4.2) replacing the π of Section 2, see (2.9). The main difference is that there (local) approximators $\pi_{N,L}$ of π appear, in the sense that $\lim \mu(f \circ \pi_{N,L}) = \mu(f \circ \pi)$ for all continuous functions f and that was constantly used for stochastic dynamics. Here however, in general, there is no obvious candidate for such approximating θ_N for θ.

Equation (4.2) is not yet sufficient to derive a fluctuation theorem for $\dot{S}_N$ since we would need a function f in (4.2) depending on N. Suppose however that we have the following strengthening of (4.2):

$$\limsup_N \frac{1}{N} \log \frac{\int \rho(dv) \exp(NF(\mu_N(v)) \exp(-\dot{S}_N(v))}{\int \rho(dv) \exp(NF(\mu_N(\theta v))} = 0, \tag{4.3}$$

where $\mu_N(v) = \frac{1}{2N+1} \sum_{i=-N}^{N} \delta_{T^i v}$ is the empirical distribution and F is an arbitrary weakly continuous function on the space of probability measures on X. If (4.3) holds, then the following symmetry relation holds

$$\limsup_N \frac{1}{N} \log \frac{\int \rho(dv) \exp(-\lambda \dot{S}_N(v))}{\int \rho(dv) \exp(-(1-\lambda)\dot{S}_N(v))} = 0. \tag{4.4}$$

For Anosov systems this program can be completed.

Gallavotti-Cohen theorem: The Gallavotti-Cohen fluctuation theorem can be seen as a result about a symmetry in the fluctuations of the phase space contraction rate in the theory of smooth reversible dissipative dynamical systems, see [16, 17, 35]. It selects a class of dynamical systems (so-called mixing Anosov diffeomorphisms) where via the existence of Markov partitions and symbolic dynamics a one-to-one relation with a one-dimensional Gibbs measure $\tilde{\mu}$ with an exponentially decaying interaction can be established.
Consider minus the logarithm of the Jacobian determinant J which arises from the change of variables implied by the dynamics. We write $\dot{s} \equiv -\log J$, the phase space contraction rate, and the object of interest is

$$w_N(v) \equiv \frac{1}{2\rho(\dot{s})N} \sum_{-N}^{N} \dot{s}(T^n(v)), \tag{4.5}$$

for large time N. ρ is the SRB measure of the dynamics which arises naturally from

$$\rho(f) = \lim_N \frac{1}{N} \sum_{0}^{N} f(T^n v) \tag{4.6}$$

corresponding to time-averages for almost every randomly chosen initial point $v \in X$ with respect to the Riemann volume element dv on X. One assumes (and sometimes proves) dissipativity:

$$\rho(\dot{s}) > 0. \tag{4.7}$$

The fluctuation theorem of Gallavotti and Cohen states that $w_N(v)$ has a distribution $\rho_N(p)$ with respect to the stationary (SRB) state ρ such that

$$\lim_N \frac{1}{N\rho(\dot{s})p} \ln \frac{\rho_N(p)}{\rho_N(-p)} = 1 \tag{4.8}$$

always. In other words, the distribution of w_N for N large satisfies some general symmetry property. The reader will of course recognize the relation with (2.22).

The technical (mixing Anosov) assumption assures the uniform hyperbolicity of the dynamical system. The use of symbolic dynamics converts the study of entropy production into the framework of statistical mechanics for one-dimensional lattice spin systems with an exponentially decaying interaction. This is intrinsic in the proof of [16, 17] and it was explicitly remarked in Section 3 of [2].
The steady state μ (corresponding to the SRB-state ρ) is the $\gamma-$image of a translation invariant Gibbs measure $\tilde{\mu}$ on $\Omega := G^{\mathbb{Z}}$, corresponding to the $d = 0$ case of the previously considered interacting particle systems. The (de)coding $\gamma : \Omega \to X$ is continuous one-to-one almost everywhere and satisfies $T \circ \gamma = \gamma \circ \tau$ where $\tau = \tau_1$ is the shift on $\mathbb{Z}$. This is brought about via a finite (Markov) partition $(I_a, a \in G)$ of X from which we define $\ell(v) = a$ if $v \in I_a$. Given $\sigma = \gamma^{-1}v \in \Omega$, there are many $w \in X$ for which $\ell(T^n w) = \sigma(n), |n| \leq N$. In the same way there could be many w' for which $\ell(T^n w') = \ell(\theta \circ T^{-n} v) = \ell(T^n \circ \theta v), |n| \leq N$. What we are interested in is to take the ratio of the corresponding weights according to our pathspace measure μ (as in (2.9) or (2.10)). Following our general scheme, the logarithm of this must be related to the entropy production just as in (4.2).
The reversibility plays as follows on the symbols. First of all, again because of the Anosov character of T, we can choose the partition such that $\ell(\theta v) = \tilde{\theta}\ell(v), v \in X$ for some involution $\tilde{\theta}$ on G. Define $\pi = \gamma^{-1} \circ \theta \circ \gamma$. In the same way we can define the local transformation $\pi_N(\sigma)(n) = \tilde{\theta}(\sigma(-n)), |n| \leq N, = \sigma(n)$ for $|n| > N$ and obviously, $g \circ \pi_N \to g \circ \pi$ for continuous g on Ω. In the same way

$$\theta_N = \gamma \circ \pi_N \circ \gamma^{-1} \tag{4.9}$$

approximates θ. But now, (4.2) relies on a statement about the Gibbs measure $\tilde{\mu}$ and the transformation π_N, just as in our Section 2: we can apply the theory of large deviations for Gibbs states, see [26].

The reason why the phase space contraction appears as entropy production is that the SRB state μ (or, after transforming to symbolic sequences, $\tilde{\mu}$) is a Gibbs state with respect to the interaction $-\log J^u$ (which is Hölder continuous) where $J^u > 0$ is the expanding or unstable Jacobian. The Jacobian determinant satisfies $J(\theta \circ T) = J^{-1}$ and θ interchanges the stable with the unstable directions $(J^s(\theta T v) = J^u(v)^{-1})$ so that

$$-\log J^u(\theta T v) + \log J^u(v) = \log J(v). \tag{4.10}$$

As a consequence, we recover the expression (2.15) where the entropy production is related to the 'relative energy' after time-reversal:

$$\sum_{k=-N}^{N} \log J(T^k v) = \sum_{k=-N}^{N} [(-\log J^u(T^k \theta v)) - (-\log J^u(T^k v))] + O(1) \tag{4.11}$$

where $0(1)/N \to 0$ as $N \uparrow \infty$.

The reason why the quantity $\dot{s}$ can be identified with the change of entropy in the steady state follows from the following simple calculation. Define the (Shannon) entropy of a probability distribution $m(dv) = m(v)dv$ on X as

$$S(m) = -\int dv m(v) \log m(v). \tag{4.12}$$

If m is the density at time n, then, under the dynamics, the density at time $n+1$ is

$$m'(v) = \frac{m(T^{-1}v)}{J(T^{-1}v)} \tag{4.13}$$

and the change in entropy (gained by the system) is therefore

$$S(m') - S(m) = \int dv m(v) \log J(v). \tag{4.14}$$

Taking n to infinity, the empirical probability distribution approaches the SRB distribution ρ, as in (4.6). Therefore, the amount of entropy produced by the system per time unit is (4.7), see also [32, 33, 35].

Even though the preceding discussion was mentioning mostly technical points that are part of the theory of (Anosov) dynamical systems, this was certainly not the final goal of the authors. This is summarized via their chaoticity hypothesis: "A reversible many particle system in a stationary state can be regarded as a transitive Anosov system for the purpose of computing the macroscopic properties," see also e.g. [16, 17, 14, 13, 34, 15]. In fact, various numerical experiments have shown extremely good agreement with the symmetries predicted by Gallavotti and Cohen. e.g. in [2]. The theorem originated from numerical evidence in [10]. These computer experiments are carried out via so called thermostatted systems. These are dynamical systems where mechanical forces are replacing the action of reservoirs in keeping the energy of the system constant. More theoretically, the theorem has various interesting consequences. For example [14] has been extending Green-Kubo type formulas to arbitrary forcing fields for a class of nonequilibrium dynamics, see also [35]. We see it therefore as a major argument (and motivation) in favor of the suggested definition of entropy production.

5 Transient regime

So far we have been considering the (nonequilibrium) steady state. However, our setup can just as well be applied to transient regimes as e.g. decribed in [8, 20]. We briefly comment here on the mathematics behind the so-called nonequilibrium work relations and how they fit into our framework. Let us start with the mathematics in the easiest example of a (discrete time) Markov chain σ_n on a finite configuration space X. Consider a probability measure ρ_i on X, $\rho_i(\xi) > 0$, $\xi \in X$, as initial state and the corresponding pathspace measure P_N on X^{N+1} for which

$$\begin{aligned} P_N\left(f(\sigma_0,\dots,\sigma_N)\right) &= \mathbb{E}_{\rho_i}\left[f(\sigma_0,\dots,\sigma_N)\right] \\ &= \sum f(\sigma_0,\dots,\sigma_N) p(\sigma_N|\sigma_{N-1})\dots p(\sigma_1|\sigma_0)\rho_i(\sigma_0) \end{aligned} \tag{5.1}$$

with transition probabilities $p(\xi|\xi') > 0$, $\xi, \xi' \in X$. It is immediately verified that for all pairs of probability measures ρ, ρ' on X with $\rho(\xi)$, $\rho'(\xi) > 0$, $\xi \in X$, we have the relation

$$\mathbb{E}_{\rho'}\left[f \circ \pi_N\right] = \mathbb{E}_{\rho}\left[f \mathrm{e}^{-\dot{S}_N + \log \rho'(\sigma_N) - \log \rho(\sigma_0)}\right] \tag{5.2}$$

where $(\pi_N \sigma)_n = \sigma_{N-n}$ and

$$\dot{S}_N(\sigma) = \sum_{k=1}^{N} \log \frac{p(\sigma_k|\sigma_{k-1})}{p(\sigma_{k-1}|\sigma_k)} \tag{5.3}$$

corresponds to the steady state entropy production, see [30]:

$$\frac{\dot{S}_N(\sigma)}{N} \to \mu(\dot{s})$$

μ-almost surely, in the unique stationary measure. We now write

$$\rho(\xi) = \frac{\mathrm{e}^{-\beta H(\xi)}}{Z} \tag{5.4}$$

$$\rho'(\xi) = \frac{\mathrm{e}^{-\beta H'(\xi)}}{Z'} \tag{5.5}$$

and we choose

$$f(\sigma_0, \ldots, \sigma_N) = \mathrm{e}^{(1-\lambda)[\dot{S}_N(\sigma) - \log \rho'(\sigma_N) + \log \rho(\sigma_0)]} \quad (\lambda \in \mathbb{C}). \tag{5.6}$$

Substitution in (5.2) gives

$$\mathbb{E}_{\rho'}\left[\mathrm{e}^{-(1-\lambda)[\dot{S}_N(\sigma) + \beta(H(\sigma_N) - H'(\sigma_0))]}\right] = \frac{Z}{Z'} \mathbb{E}_{\rho}\left[\mathrm{e}^{-\lambda[\dot{S}_N(\sigma) + \beta(H'(\sigma_N) - H(\sigma_0))]}\right]. \tag{5.7}$$

In particular, for $\lambda = 1$, we get

$$P_N\left[\mathrm{e}^{-[\dot{S}_N(\sigma) + \beta(H'(\sigma_N) - H(\sigma_0))]}\right] = \frac{Z'}{Z} \tag{5.8}$$

where started in $\rho_i = \rho$. This is a variant of the nonequilibrium work relation appearing in [8, 20]. More generally, such relations are easy to produce in the context of stochastic dynamics (as in Section 3) by using the appropriate form of the trivial identity $\mu\left(\frac{d(\pi_{N,L}\mu)}{d\mu}\right) = 1$.

Instead of considering a stochastic dynamics on our system, we could also consider a Hamiltonian time evolution on our system plus reservoir. To be specific, let us take the configuration space of the form $X = X_s \times X_r$ with a deterministic invertible transformation $T : X \to X$. The momenta and positions of the particles in the system are collected in the (first) variable v and w will stand for the reservoir variable. T preserves the elementary volume: $d(T^{-1}(v,w))/d(v,w) = 1$, $d(v,w) = dvdw$. Suppose that the initial state ρ_i has a density

$$\rho_i(v,w) = \frac{\mathrm{e}^{-H(v)/T_i}}{Z_i} \mathrm{e}^{-S_i(w)} \tag{5.9}$$

with partition function $Z_i = \int dv \mathrm{e}^{-H(v)/T_i}$ and $\int \mathrm{e}^{-S_i} dw = 1$. We also define

$$\rho_f(v,w) = \frac{\mathrm{e}^{-H(v)/T_f}}{Z_f} \mathrm{e}^{-S_f(w)}. \tag{5.10}$$

Then,

$$\begin{aligned} &\int f\left((v,w), T(v,w)\right) \rho_i(v,w) dv dw \\ = &\int f\left(T^{-1}(v,w),(v,w)\right) \frac{\rho_i(T^{-1}(v,w))}{\rho_f(v,w)} \rho_f(v,w) dv dw \end{aligned} \tag{5.11}$$

and thus, taking

$$f\left((v,w),(v'w')\right) = \mathrm{e}^{-(1-\lambda)[H(v')/T_f - H(v)/T_i + S_f(w') - S_i(w)]}, \tag{5.12}$$

we have

$$\begin{aligned} &\rho_i \left[\mathrm{e}^{-(1-\lambda)[H(v')/T_f - H(v)/T_i + S_f(w') - S_i(w)]} \right] \\ = \; &\rho_f \left[\mathrm{e}^{-\lambda[H(v_{-1})/T_i - H(v)/T_f + S_i(w_{-1}) - S_f(w)]} \right] \frac{Z_f}{Z_i}, \end{aligned} \tag{5.13}$$

where $(v_i, w_i) = T^i(v,w)$, $i = \pm 1$. Taking $\lambda = 0$ recovers the main relation of [20]. If ρ_f would really correspond to the 'final' state (the image of ρ_i under T) then $\rho_f(S_f) - \rho_i(S_i)$ would be the change in entropy and $\log Z_f/Z_i = \exp[-(F_f/T_f - F_i/T_i)]$ would correspond to a change in Helmholtz free energy F.

Acknowledgement: We thank Prof. H. Wagner for stimulating discussions and for suggesting the molecular motor model of Section 3.2.

References

[1] Antal, T., Rácz, Z. and Sasvári, L. (1997) Nonequilibrium steady state in a quantum system: one-dimensional transverse Ising model with energy current, *Phys. Rev. Lett.* **78**, 167.

[2] Bonetto, F., Gallavotti, G. and Garrido, P. (1997) Chaotic principle: an experimental test, *Physica D* **105**, 226.

[3] Bricmont, J. and Kupiainen, A. (1997) Infinite dimensional SRB measures, *Physica D* **103**, 18–33.

[4] Bricmont, J. and Kupiainen, A. (1996) High temperature expansions and dynamical systems, *Comm. Math. Phys.* **178**, 703–732.

[5] Bowen, R. (1970) Markow partitions for Axiom A diffeomorphisms, *Amer. J. Math.* **92**, 725-747.

[6] Bowen, R. (1973) Symbolic dynamics for hyperbolic flows, *Amer. J. Math.* **95**, 429-460.

[7] Capocaccia, D. (1976) A definition of Gibbs' states for a compact set with $\mathbb{Z}^\nu$-action, *Comm. Math. Phys.* **48**, 85–88.

[8] Crooks, G.E. (1999) The Gallavotti-Cohen fluctuation theorem and the nonequilibrium work relation for free energy defferences, *preprint*, cond-mat/9901352 v2.

[9] van Enter, A.C.D., Fernández, R. and Sokal A.D. (1993) Regularity properties and pathologies of position-space renormalization transformations: scope and limitations of Gibbsian theory, *J. Stat. Phys.* **72**, 879–1167.

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

[11] G. Eyink, J.L. Lebowitz, H. Spohn, *Microscopic origin of hydrodynamic behavior: entropy production and the steady state*, in: 'Chaos, Soviet-American Perspectives in Nonlinear Science', Hg. D.K Campbell, p. 367. American Institute of Physics, 1990

[12] Gallavotti, G. (1994) Perturbation theory, *Mathematical physics towards the 21st century*, eds. R.N. Sen and A Gersten, Ben Gurion University of the Negev Press.

[13] Gallavotti, G. (1996) Chaotic hypothesis: Onsager reciprocity and fluctuation-dissipation theorem, *J. Stat. Phys.* **84**, 899-926.

[14] Gallavotti, G. (1996) Extension of Onsager's reciprocity to large fields and the chaotic hypothesis, *Phys. Rev. Lett.* **77**, 4334–4337.

[15] Gallavotti, G. (1998) Chaotic dynamics, fluctuations, nonequilibrium ensembles, *Chaos* **8**, 384–392.

[16] Gallavotti, G. and Cohen, E.G.D. (1995) Dynamical ensembles in nonequilibrium statistical mechanics, *Phys. Rev. Lett.* **74**, 2694–2697.

[17] Gallavotti, G. and Cohen, E.G.D. (1995) Dynamical ensembles in stationary states, *J. Stat. Phys.* **80**, 931–970.

[18] Georgii, H.-O. (1988) *Gibbs measures and phase transitions*, de Gruyter, Berlin · New York.

[19] Haydn, N. T. and Ruelle, D. (1992) Equivalence of Gibbs and equilibrium states for homeomorphisms satisfying expansiveness and specification, *Comm. Math. Phys.* **148**, 155-167.

[20] Jarzynski, C. (1998) Microscopic analysis of Clausius-Duhem processes, *preprint*, cond-mat/9802249.

[21] Jarzynski, C., Mazonka, O. (1999) Feynman's ratchet and pawl: an exactly solvable model, *preprint*.

[22] Jaynes, E.T. (1989) Clearing up Mysteries; the Original Goal, in: Proceedings of the 8'th International Workshop in Maximum Entropy and Bayesian Methods, Cambridge, England, August 1–5, 1988; J. Skilling, Editor; Kluwer Academic Publishers, Dordrecht, Holland. See also in *Papers on Probability, Statistics, and Statistical Physics*, D. Reidel Publishing Co., Dordrecht, Holland, R.D. Rosenkrantz, Editor. Reprints of 13 papers. See also http://bayes.wustl.edu/etj/node1.html.

[23] Jülicher, F., Ajdari, A. and Prost, J. (1997) Modeling molecular motors, *Rev. Mod. Phys.* **69**, No **4**, 1269-1281.

[24] Kurchan, J. (1998) Fluctuation theorem for stochastic dynamics, *J. Phys. A: Math. Gen.* **31**, 3719–3729.

[25] Künsch, H. (1984) Non reversible stationary measures for infinite interacting particle systems, *Z. Wahrsch. Verw. Gebiete* **66**, 407.

[26] Lanford III, O.E. (1973) Entropy and equilibrium states in classical statistical mechanics, in *Statistical Mechanics and Mathematical Problems (Batelle Seattle Rencontres 1971)*, Lecture Notes in Physics No. 20 (Springer-Verlag, Berlin), 1–113. Comets, F. (1986) Grandes déviations pour des champs de Gibbs sur $\mathbb{Z}^d$, *C.R. Acad. Sci. Paris I* **303**, 511–513. Olla, S. (1988) Large deviations for Gibbs random fields, *Prob. Th. Rel. Fields* **77**, 343–357.

[27] Lebowitz, J.L. and Spohn, H. (1999) A Gallavotti-Cohen Type Symmetry in the Large Deviation Functional for Stochastic Dynamics, to appear in J. Stat. Phys.

[28] Liggett, T.M. (1985) *Interacting Particle Systems*, Springer, Berlin, Heidelberg, New York.

[29] Lipster, R.S. and Shiryayev, A.N. (1977) *Statistics of Random Processes I & II*, springer-Verlag, New York, Heidelberg, Berlin.

[30] Maes, C. (1999) The Fluctuation Theorem as a Gibbs Property, to appear in J. Stat. Phys.

[31] Magnasco, M.O. (1994) Molecular combustion motors, *Phys. Rev. Lett.* **72**, No. **16**, 2656-2659.

[32] Ruelle, D. (1996) Positivity of entropy production in nonequilibrium statistical mechanics, *J. Stat. Phys.* **85**, 1–25.

[33] Ruelle, D. (1997) Entropy production in nonequilibrium statistical mechanics, *Comm. Math. Phys.* **189**, 365–371.

[34] Ruelle, D. (1978) Sensitive dependence on initial conditions and turbulent behavior of dynamical systems, *Annals of the New York Academy of Sciences* **356**, 408–416.

[35] Ruelle, D. (1998) Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics. Rutgers University Lecture Notes, October-November 1997-98 (unpublished).

[36] Simon, B. (1993) *The Statistical Mechanics of Lattice Gases*, Volume 1, Princeton University Press, Princeton.

[37] Sinai, Ya. G. (1968) Markov partitions and C-diffeomorphisms, *Functional Anal. Appl.* **2**, 61-82.

[38] Sinai, Ya. G. (1968) Construction of Markov partitions, *Functional Anal. Appl.* **2**, 245-253.

[39] H. Spohn, Large Scale Dynamics of Interacting Particles. Springer, Heidelberg 1991.