Entropy Production for Interacting Particle Systems

Christian MAES[1,4], Frank REDIG[2,4], and Michel VERSCHUERE[3,4]

(1) *Onderzoeksleider FWO*
(2) *Post-doctoraal onderzoeker FWO*
(3) *Aspirant FWO*
(4) *Instituut voor Theoretische Fysica, K.U.Leuven*
Celestijnenlaan 200D, B-3001 Leuven, Belgium

Abstract: The present article provides a mathematical intoduction to the concept of entropy production in stochastic dynamics. We discuss the mean rate of entropy production for a class of interacting particle systems.

Keywords: stochastic interacting particle systems, entropy production.

1 Introduction

We give the expression of the mean entropy production rate for a general class of interacting particle systems. We take the point of view that entropy production measures the deviation of microscopic time reversibility. The dynamics specifies a space-time interaction and the steady state expectation of the corresponding relative (space-time) Hamiltonian under time-reversal defines the entropy production. Its mean depends not only on the stationary measure and on the dynamics but also on the type of time-reversal. It coincides with the physical entropy production in cases where the dynamics has a natural physical interpretation.
The paper is a continuation of our studies in [1, 2, 3] and extends previous work by [6, 7, 8] to spatially extended systems in a steady state. In the next section, we give a mathematical introduction to the concept of entropy production for stochastic dynamics. The main construction can already be illustrated for the simplest Markov chains. This is extended in the third section where we derive the general expression of the mean entropy production for interacting particle systems.

2 Continuous time Markov chains

2.1 Phase space and path space

Let K be a finite set on which we have N (non-trivial) transformations $T_\alpha : K \to K, \quad \alpha = 1, \dots, N$ and an involution $\pi : K \to K, \quad \pi \circ \pi = \pi^2 = id$. We assume

that for each $\alpha = 1, \ldots, N \quad T_\alpha^{-1} = T_\beta$ for some $\beta = 1, \ldots, N$ and also $\pi T_\alpha \pi = T_\gamma$. for some $\gamma = 1, \ldots, N$. Assume also that for all $a, b \in K$, there are $n, \alpha_1, \alpha_2, \ldots, \alpha_n$ so that $T_{\alpha_1} T_{\alpha_2} \ldots T_{\alpha_n} a = b$.
In the next section, the phase space K will take the form $K = S^\Lambda$ with S a finite set (the single site state space) and $\Lambda = \Lambda_n = [-n, n]^d \cap \mathbb{Z}^d$ is a cube of size $(2n+1)^d$ in the d-dimensional regular lattice. The transformations T_α will generate a local and space-time homogeneous Markov process on S^Λ as $\Lambda \uparrow \mathbb{Z}^d$; the involution π will be of the form π_Λ generated by an involution i on S via $(\pi_\Lambda \eta)(x) = i(\eta(x)),\ x \in \Lambda, \quad \eta = (\eta(x), x \in \Lambda) \in S^\Lambda$. The precise choice of π or i depends (physically speaking) on the types of variables but this will be of no concern to us here.

The path space $D([-T, T], K)$ with starting time $-T < 0$ is the set of all trajectories or paths $\omega : [-T, T] \to K$ that have left limits and are right continuous. We denote by $\omega(t^-)$ the left limit at time t, possibly different from $\omega(t) = \omega(t^+)$. We write $\Theta\omega$ for the time-reversed path, that is $(\Theta\omega)(t) = \omega(-t)$; if the phase space time reversal π is included, we write $\Theta_\pi \omega$ for $(\Theta_\pi \omega)(t) = \pi(\omega(-t))$. Note however that in general $\Theta\omega$ or $\Theta_\pi\omega$ would not be contained in $D([-T, T], K)$ (not being right continuous) but we keep using the same notation for the obvious modification of these paths into elements of the path space. This space $D([-T, T], K)$ is equipped with the natural σ-algebra's $\mathcal{F}_t \equiv \sigma(\omega(s),\ s \in [-T, t])$.
For the Markov chains that follow next, we will have probability measures $\mathbb{P}_a$ on $D([-T, T], K)$ that concentrate on the trajectories ω for which $\omega(0) = a$ and $\omega(t^-) \neq \omega(t)$ only if $\omega(t) = T_\alpha(\omega(t^-))$ for some $\alpha = 1, \ldots, N$ and there are a finite number of those jumps. We will call such trajectories ω realizations.

2.2 Dynamics

Via probability measures on $D([-T, T], K)$ we may hope to describe certain aspects of a dynamics on K. It is easy to construct such probability measures explicitly as path space measures for a Markovian (stochastic) dynamics. For a Markov process on K, we associate rates $c(T_\alpha, \cdot) \equiv c_\alpha : K \to [0, \infty)$ to each of the $T_\alpha,\ \alpha = 1, \ldots, N$ and we consider the generator

$$Lf(a) = \sum_{\alpha=1}^{N} c_\alpha(a)[f(T_\alpha a) - f(a)] \tag{2.1}$$

We write $\lambda(a) \equiv \sum_{\alpha=1}^{N} c_\alpha(a)$ for the total rate of exiting from state $a \in K$. For convenience we put $c_\alpha(a) = 0$ when $T_\alpha a = a$ and suppose further that $c_\alpha(a) > 0$ when $T_\alpha a \neq a$.
As a reference process we consider the Markov process with rates $c^o_\alpha(a) = 1,\ a \in K$ and generator

$$L^o f(a) = \sum_{\alpha=1}^{N} [f(T_\alpha a) - f(a)] \tag{2.2}$$

so that the corresponding $\lambda^o(a) = N$.

2.3 Relative Hamiltonian

The processes generated via (2.1) and (2.2) give both rise to a probability measure (path space measure) on $D([-T,T],K)$ upon fixing the initial state $\omega(-T) = a$. Let $\mathbb{P}_a$ denote the law of this first process generated via (2.1) and let $\mathbb{P}_a^0$ stand for the reference process. We then have that $\mathbb{P}_a$ has a density with respect to $\mathbb{P}_a^0$ given by

$$\frac{d\mathbb{P}_a}{d\mathbb{P}_a^0} = \exp[-\mathcal{H}_{a,T}]\, e^{2NT} \tag{2.3}$$

where the Hamiltonian in the exponent equals

$$\mathcal{H}_{a,T}(\omega) \equiv \int_{-T}^{T} \lambda(\omega(s))ds - \sum_{s\in\mathcal{S}_T(\omega)} \log\, k(\omega(s^-),\omega(s)) \tag{2.4}$$

where

$$k(a,b) \equiv \sum_{\alpha:T_\alpha a=b} c_\alpha(a), \lambda(a) = \sum_b k(a,b)$$

and with $\mathcal{S}_T(\omega) \equiv \{s \in [-T,T] : T_\alpha(\omega(s^-)) = \omega(s) \neq \omega(s^-) \text{ for some } \alpha = 1,\dots,N\}$ the jump times, for all realizations ω starting at $\omega(-T) = a$. More generally, we define the Hamiltonian $\mathcal{H}_T$ evaluated in a realization ω to be

$$\mathcal{H}_T(\omega) \equiv \int_{-T}^{T} \lambda(\omega(s))ds - \sum_{s\in\mathcal{S}_T(\omega)} \log\, k(\omega(s^-),\omega(s)) \tag{2.5}$$

We put $\mathcal{H}_T(\omega) = +\infty$ for other paths $\omega \in D([-T,T],K)$ that are not realizations of the process at hand.
Clearly, ω is a realization iff $\Theta_\pi\omega$ is a realization because every jump $\omega \to T_\alpha\omega$ at time s in ω corresponds uniquely with a jump $\pi T_\alpha\omega \to \pi\omega = \pi T_\alpha^{-1}\pi(\pi T_\alpha\omega)$ at time $-s$ in the path $\Theta_\pi\omega$. It is thus possible to define the relative Hamiltonian $R_{T,\pi}$ for such realizations ω as

$$\begin{aligned} R_{T,\pi}(\omega) &\equiv \mathcal{H}_T(\Theta_\pi\omega) - \mathcal{H}_T(\omega) \\ &= \int_{-T}^{T} [\lambda\circ\pi(\omega(s)) - \lambda(\omega(s))]ds + \sum_{s\in\mathcal{S}_T(\omega)} \log\frac{k(\omega(s^-),\omega(s))}{k(\pi\omega(s),\pi\omega(s^-))} \end{aligned} \tag{2.6}$$

2.4 Entropy production

The entropy produced in a realization ω over the time-interval $[-T,T]$ is defined as $R_{T,\pi}(\omega)$, the relative Hamiltonian with respect to time-reversal as written down in

(2.6). Note that $R_{T,\pi}(\Theta_\pi \omega) = -R_{T,\pi}(\omega)$.
In many interesting physical situations, the choice of the involution π is exactly so that k is $\pi-$invariant:

$$k(a,b) = k(\pi a, \pi b),\;\; a,b \in K \tag{2.7}$$

implying $\lambda = \lambda \circ \pi$ and reducing (2.6) to the form

$$R_T(\omega) = \sum_{s\in \mathcal{S}_T(\omega)} \log \frac{k(\omega(s^-),\omega(s))}{k(\omega(s),\omega(s^-))} \tag{2.8}$$

In the form (2.8) it is clear that if $\rho(a)k(a,b) = k(b,a)\rho(b)$ for some positive ρ, then $R_T(\omega)$ is telescoping and becomes equal to $\log \rho(\omega(T)) - \log \rho(\omega(-T))$ which is no longer of order T. This will come back in section 3.3.

2.5 Stationary Measure

Suppose that ρ is a stationary probability measure for the Markov chain above. This means that $\rho(a) > 0,\;\; a\in K,\;\; \sum_a \rho(a) = 1$ and

$$\sum_{b\in K} \rho(b)\, k(b,a) = \sum_{b\in K} k(a,b)\rho(a) = \lambda(a)\rho(a) \tag{2.9}$$

Note that if we assume (2.7), then $\rho = \rho\pi = \rho\circ\pi$. This is because then $\lambda(a) = \lambda(\pi a)$ and, from (2.9),

$$\begin{aligned}
\sum_{b\in K} \rho\pi(b) k(b,a) &= \sum_b \rho(b) k(\pi b, a) \\
&= \sum_b \rho(b) k(b, \pi a) \\
&= \lambda(a)\rho\pi(a)
\end{aligned} \tag{2.10}$$

which implies that also $\rho\pi$ is stationary. Our assumptions then imply that necessarily, under (2.9) and (2.7), $\rho = \rho\pi$. This needs to be reconsidered in the case of (infinite) interacting particle systems where different stationary measures can coexist. We will however in all cases just assume that both ρ and $\rho\pi$ are stationary, i.e. both (2.9) and

$$\sum_b \rho\pi(b) k(b,a) = \lambda(a)\rho\pi(a) \tag{2.11}$$

are verified. At any rate, this has still little to do with 'microscopic reversibility' or with 'the condition of detailed balance' which will be discussed next.

2.6 Time-reversed distribution

Let $(X_t, t \in [-T,T])$ be the stationary process (steady state) starting from the stationary measure ρ. Its law on $D([-T,T],K)$ is denoted by $\mathbb{P}_\rho$. Assume that $\rho = \rho\pi$. The time-reversed distribution of (X_t) is by definition the stationary process $(Y_t, t \in [-T,T])$ with

$$Y_t \equiv \pi\, X_{-t}, \quad t \in [-T,T] \tag{2.12}$$

We denote its law on the same set of trajectories by $\tilde{\mathbb{P}}_\rho$ where we indicated that ρ is stationary for (Y_t). It is easy to see that (Y_t) is also a Markov process on K concentrating on the same realizations as the original (X_t) (after the obvious modifications making the trajectories right-continuous). The probabilities of such realizations ω will however in general be different for $\mathbb{P}_\rho$ and $\tilde{\mathbb{P}}_\rho$ as they may have different transition rates. A little computation immediately gives that the rates of (Y_t) are given by

$$\tilde{k}(a,b) \equiv k(\pi b, \pi a)\frac{\rho(b)}{\rho(a)} \tag{2.13}$$

with $\sum_b \tilde{k}(a,b) = \lambda(\pi a)$ (use (2.11), $\rho\pi$ is invariant). Of course, if $\rho = \rho\pi$, $\tilde{\tilde{k}} = k$. The corresponding generator for the time-reversed process is

$$\begin{aligned}
\tilde{L}f(a) &= \sum_b \tilde{k}(a,b)[f(b) - f(a)] \\
&= \sum_b k(\pi b, \pi a)\frac{\rho(b)}{\rho(a)}[f(b) - f(a)]
\end{aligned}$$

which expressed via the transformations T_α reads

$$\begin{aligned}
\tilde{L}f(a) &= \sum_\alpha c_\alpha(T_\alpha^{-1}\pi a)\frac{\rho(\pi T_\alpha^{-1}\pi a)}{\rho(a)}[f(\pi T_\alpha^{-1}\pi a) - f(a)] \\
&= \sum_\alpha c(\pi T_\alpha^{-1}\pi, \pi T_\alpha a)\frac{\rho(T_\alpha a)}{\rho(a)}[f(T_\alpha a) - f(a)]
\end{aligned} \tag{2.14}$$

Under the assumptions (2.7) this becomes

$$\tilde{L}f(a) = \sum_\alpha c_\alpha(T_\alpha^{-1}a)\frac{\rho(T_\alpha^{-1}a)}{\rho(a)}[f(T_\alpha^{-1}a) - f(a)] \tag{2.15}$$

We can also verify from the invariance of $\rho\pi$, (2.11), that

$$\tilde{L}^* = \pi L \pi$$

where the $*$ refers to the adjoint with respect to the stationary measure ρ:

$$\sum_a \rho(a) f(a) \tilde{L} g(a) = \sum_a \rho(a) \pi L \pi f(a) g(a) \tag{2.16}$$

with $\pi L \pi f(a) \equiv L(\pi f)(\pi a) = \sum_b k(\pi a, \pi b) f(b) - \lambda(\pi a) f(a)$.

We say that the process $(X_t)_{-T}^T$ is **π-reversible** if $\mathbb{P}_\rho = \tilde{\mathbb{P}}_\rho$. This implies that the stationary measure ρ satisfies

$$\rho(a) k(a,b) = k(\pi b, \pi a) \rho(b) \tag{2.17}$$

For the symmetry (2.7), (2.17) reduces to the usual detailed balance condition $\rho(a)k(a,b) = k(b,a)\rho(b)$. Note that if $k(a, \pi a) \neq 0,\ a \in K$, then (2.17) by itself implies that $\rho = \rho\pi$. Moreover, the relation (2.17) combined with the stationarity of $\rho\pi$, (2.11), implies that $\lambda = \lambda \circ \pi$ because summing it over b gives

$$\rho(a)\lambda(a) = \sum_b k(b, \pi a) \rho\pi(b) = \rho\pi(\pi a) \lambda(\pi a)$$

As a result, (2.17) reduces the entropy production $R_{T,\pi}(\omega)$ in a realization ω to the (temporal) boundary term

$$\log\, \rho(\omega(T)) - \log\, \rho(\omega(-T)) \tag{2.18}$$

Finally, $\lambda = \lambda \circ \pi$ combined with (2.17) directly gives

$$L^* = \pi L \pi$$

which, in turn, implies the stationarity of ρ and of $\rho\pi$ and the relation (2.17).
We have in the equivalences discussed above been careful to mention when we use the stationarity of $\rho\pi$ (distinguishing it from the invariance of ρ) eventhough under our assumptions $\rho = \rho\pi$ when both are stationary. We have done this as a preparation to interacting particle systems because it is very well possible there that ρ is different from $\rho\pi$ and both are stationary.

2.7 Mean entropy production

We define the **Mean Entropy Production** in the stationary process (X_t) as the limit

$$\mathrm{MEP}_\pi(k, \rho) = \lim_{T\uparrow\infty} \frac{1}{2T} \mathbb{E}_\rho(R_{T,\pi}) \tag{2.19}$$

where we take the relative Hamiltonian from (2.6). We will first give a number of equivalent definitions to (2.19) (showing existence of the limit and including a more explicit version) and then state the main property.

Theorem 2.1 *For the stationary process (X_t) with $\rho = \rho\pi$, the mean entropy production is given by the following equivalent expressions*

1.

$$MEP_\pi(k,\rho) = \sum_{a,b} \rho(a)k(a,b)\log\frac{k(a,b)}{k(\pi b,\pi a)} \tag{2.20}$$

2.

$$MEP_\pi(k,\rho) = \lim_{T\uparrow\infty}\frac{1}{2T}\mathbb{E}_\rho[\log\frac{d\mathbb{P}_\rho}{d\tilde{\mathbb{P}}_\rho}] \tag{2.21}$$

3.

$$MEP_\pi(k,\rho) = \lim_{\delta\downarrow 0}\frac{1}{\delta}\sum_{a,b}\mathbb{P}_\rho[X_0=a,X_\delta=b]\log\frac{\mathbb{P}_\rho[X_0=a,X_\delta=b]}{\tilde{\mathbb{P}}_\rho[Y_0=a,Y_\delta=b]} \tag{2.22}$$

4.

$$MEP_\pi(k,\rho) = \frac{1}{2}\sum_{a,b}\rho(a)[k(a,b)-\tilde{k}(a,b)]\log\frac{k(a,b)}{\tilde{k}(a,b)} \tag{2.23}$$

Proof: From (2.3)-(2.6),

$$\frac{d\tilde{\mathbb{P}}_\rho}{d\mathbb{P}_\rho} = \exp(-R_{T,\pi})\frac{\rho(\pi\omega(-T))}{\rho(\omega(T))}$$

By this observation (2.21) is clearly equivalent to (2.19). The more explicit (2.20) for $\mathrm{MEP}_\pi(k,\rho)$ can be derived using (2.6), (2.13) and the stationarity of ρ. It is important to observe that

$$\sum_{a,b}\rho(a)k(a,b)\log\frac{\rho(b)}{\rho(a)} = 0$$

vanishes identically, again by stationarity. To prove (2.22) it suffices to note that

$$\begin{aligned}
\mathbb{P}_\rho(X_0=a,X_\delta=b) &= \rho(a)\delta_{a,b}(1-\delta\lambda(a)) + \delta\rho(a)k(a,b) + O(\delta^2)\\
\tilde{\mathbb{P}}_\rho(Y_0=a,Y_\delta=b) &= \rho(a)\delta_{a,b}(1-\delta\lambda(\pi a)) + \delta\rho(b)k(\pi b,\pi a) + O(\delta^2)
\end{aligned}$$

with $\delta_{a,b}$ the Kronecker delta.
For (2.23), we use that

$$\sum_{a,b}\rho(a)\tilde{k}(a,b)\log\frac{k(a,b)}{\tilde{k}(a,b)} = \sum_{a,b}\rho(\pi a)k(a,b)\log\frac{k(\pi b,\pi a)\rho\pi(b)}{k(a,b)\rho\pi(a)}$$

and the rest is a simple computation. ■

The main property of the mean entropy production can be summarized as follows:

Theorem 2.2 *Consider the stationary process* (X_t) *above with* $\rho = \rho\pi$*.* $MEP_\pi(k,\rho) = MEP_\pi(\tilde{k},\rho) \geq 0$ *with equality if and only if the process* (X_t) *is* π*-reversible.*

Proof: From (2.23) above, the mean entropy production is clearly non-negative and is zero iff $k = \tilde{k}$. Moreover, interchanging k and $\tilde{k}$ leaves the mean entropy production invariant. ■

2.8 Physical entropy production

It is important to check whether the previously introduced definition(s) of (mean) entropy production have a physical relevance. After all, entropy production in a nonequilibrium steady state has a meaning in physics, be it that it mostly appears in a somewhat phenomenological set-up in the context of linear response theory, see e.g. [12]. There are various answers to this problem of identification. In various concrete examples where every physicist would agree what to call (physical) entropy production, our definitions above indeed give rise to the correct expression, see the verifications in [2, 13]. A second answer deals with general theoretical consequences of the above definitions. In [3, 2] we have been able to connect our definition with the work of Gallavotti-Cohen, e.g. [10, 11], and we have shown how our approach quite naturally gives rise to a fluctuation theorem and resulting Green-Kubo type relations. We will also not discuss these points here any further (except for the open questions in Section 4). Further motivations have already appeared in [8] and seem to go back so far as Kirchoff's current law, [9]. One point is that under the natural assumption (2.7), the mean entropy production can be directly expressed as the bilinear form

$$MEP(k,\rho) = \frac{1}{2}\sum J_{ab}(\rho)A_{ab}(\rho) \tag{2.24}$$

where, assuming (2.7),

$$J_{ab}(\rho) \equiv k(a,b)\rho(a) - k(b,a)\rho(b)$$

can be identified with the thermodynamic flux between the states a and b, and is linear in ρ, while

$$A_{ab}(\rho) \equiv \log\frac{\rho(a)k(a,b)}{\rho(b)k(b,a)}$$

is the thermodynamic force by which the system is driven from equilibrium. This is most transparant in case we have the form $k(a,b)/k(b,a) = \exp -[V(b) - V(a)] + E\Phi_{ab}$

(which is called local detailed balance in [13]) where E is some amplitude (external field) and $\Phi_{ab} = -\Phi_{ba}$ is related to the microscopic current involving states a and b. Similar discussions have already appeared in [6, 7, 8, 13], e.g. about the relation with the time-derivative of the relative entropy of a time-evolved measure with respect to the stationary state.
Most important to us however seem the thermodynamic analogues of the identities (2.21) and (2.22). It appears that the mean entropy production is the relative entropy between the forward and the backward steady state. That these states are different is a consequence of maintaining a current in the system. The system is then able to perform work and the entropy production measures the rate at which entropy is produced in the reservoirs.

3 Thermodynamic limit: interacting particle systems

3.1 Dynamics

To each site $x \in \mathbb{Z}^d$, we associate a variable $\sigma(x)$ taking values in a finite set S. The configuration space is denoted by $\Omega \equiv S^{\mathbb{Z}^d}$. We have an involution π on Ω obtained from an involution i on S with $\pi\sigma(x) \equiv i(\sigma(x))$. Local versions can be obtained via $\pi_\Lambda\sigma(x) \equiv i(\sigma(x))$ for $x \in \Lambda, \pi_\Lambda\sigma(x) \equiv \sigma(x)$ for $x \in \Lambda^c \equiv \mathbb{Z}^d \setminus \Lambda$.
In a while we will be considering $\Omega-$valued Markov processes. The spatial degree of freedom makes it possible to introduce the concepts of translation invariance and of locality. Translations on the configuration space are denoted by τ_x: $\tau_x\sigma(y) = \sigma(y+x)$. A function f on configuration space is called local if there exists a finite set $\Lambda \subset \mathbb{Z}^d$ such that for $\sigma, \eta \in \Omega$ agreeing on Λ, $f(\sigma) = f(\eta)$. A quite general interacting particle system can now be defined by introducing positive, translation invariant and local rates $k(\sigma, \eta_\Lambda\sigma_{\Lambda^c})$ by which $\sigma \in \Omega$ is changed into the configuration $\eta_\Lambda\sigma_{\Lambda^c}$ which equals σ on Λ^c and equals $\eta \in \Omega$ on Λ, where Λ runs over some specified set of finite subsets of $\mathbb{Z}^d$. To be somewhat more specific we fix a finite cube $\Lambda_0 \subset \mathbb{Z}^d$, containing the origin and define $\mathcal{P}_0$ to be a set of transformations U_0 affecting only the spins inside Λ_0, i.e., for $U_0 \in \mathcal{P}_0$, and for every $\sigma \in \Omega$:

i. $U_0\sigma \in \Omega$, and $(U_0\sigma)(y) = \sigma(y)$, for $y \in \Lambda_0^c$,

ii. for every $U_0 \in \mathcal{P}_0$, $U_0^{-1} \in \mathcal{P}_0$.

Moreover, it is assumed (with some abuse of notation) that $\pi\mathcal{P}_0\pi = \mathcal{P}_0$. We further denote $\Lambda_x \equiv \{y + x : y \in \Lambda_0\}$ and $U_x \equiv \tau_x U_0 \tau_{-x}$.
The dynamics is generated via local translation invariant rates $c(U_x, \sigma)$ for the transition $\sigma \to U_x\sigma$. We assume the following:

i. Positivity: $c(U_0, \sigma) = 0$ when $U_0\sigma = \sigma$ and if not, $c(U_0, \sigma) \geq \epsilon$ for some $\epsilon > 0$,

ii. Finite range: there is a finite $\bar{\Lambda} \subset \mathbb{Z}^d$ such that for all $\sigma, \eta \in \Omega$, and $U_0 \in \mathcal{P}_0$: $c(U_0, \sigma) = c(U_0, \sigma_{\bar{\Lambda}} \eta_{\bar{\Lambda}^c})$,

iii. Translation invariance: for all $x \in \mathbb{Z}^d$, $U_x \in \mathcal{P}_x$, $\sigma \in \Omega$: $c(U_x, \sigma) = c(U_0, \tau_{-x}\sigma)$,

iv. $U_0\sigma \neq U_0'\sigma$ for all $\sigma \in \Omega, U_0 \neq U_0' \in \mathcal{P}_0$ for which $U_0\sigma \neq \sigma$ and $U_0'\sigma \neq \sigma$. (this is for convenience only.)

The generator L corresponding to the given rates is now defined on local functions f as

$$Lf(\sigma) \equiv \sum_{x \in \mathbb{Z}^d} \sum_{U_x \in \mathcal{P}_x} c(U_x, \sigma)[f(U_x\sigma) - f(\sigma)] \tag{3.25}$$

We refer to [4] for further details on constructing the infinite volume process.

3.2 Mean Entropy production

From now on we use the sets $\Lambda_n = [-n, n]^d \cap \mathbb{Z}^d$ for n large and their corresponding subsets $\Lambda_n^\sharp$, meaning the (possibly empty) maximal subset of Λ_n, such that for all $x \in \Lambda_n^\sharp$, and $U_x \in \mathcal{P}_x$, $c(U_x, \sigma)$ depends only on coordinates inside Λ, and $\Lambda_x \subset \Lambda$. Consider now the Markov chain (as in section 2) on $\Omega_{\Lambda_n} \equiv S^{\Lambda_n}$ with generator

$$L_\Lambda f(\sigma) \equiv \sum_{x \in \Lambda_n^\sharp} \sum_{U_x \in \mathcal{P}_x} c(U_x, \sigma)[f(U_x\sigma) - f(\sigma)]$$

Just as in section 2.3, we can compute the corresponding relative Hamiltonian

$$\begin{aligned} R_{T,n,\pi}(\omega) &= \sum_{x \in \Lambda_n^\sharp} \sum_{U_x \in \mathcal{P}_x} \int_{-T}^{T} \log \frac{c(U_x, \omega(s^-))}{c(\pi U_x^{-1}\pi, \pi U_x \omega(s^-))} dN_s^{U_x} \\ &+ \int_{-T}^{T} [c(U_x, \pi\omega(s)) - c(U_x, \omega(s))] ds \end{aligned} \tag{3.26}$$

where $N_t^{U_x} \equiv \sum_{-T \leq s \leq t} I\left(\omega(s) = U_x(\omega(s^-) \neq \omega(s^-)\right)$ is the number of times the transformation U_x appeared in the realization ω up to time t. (3.26) must be compared with (2.6) but we have omitted putting additional subscripts to the π or the U_x meaning that they are restricted here to work on elements $\omega(s) \in \Omega_\Lambda$.
The mean entropy production for the interacting particle system is now defined as

$$\mathrm{MEP}_\pi(\mathcal{P}_0, \rho) \equiv \lim_{n,T \uparrow \infty} \frac{1}{2|\Lambda_n|T} \mathbb{E}_\rho^{n,T}(R_{T,n,\pi}) \tag{3.27}$$

By $\mathbb{E}_\rho^{n,T}$, we mean the expectation with respect to the path space measure, started from the stationary state ρ, restricted to trajectories within S^{Λ_n}. The first part of the next theorem then proves the existence of the limit (3.27). A second property of (3.27) that requires a proof is its positivity. In order to complete this task, we will here apply a similar argument as used in (2.22) for the case of finite Markov chains. Denote by μ_δ the joint distribution of (X_0, X_δ), where X_0 is distributed according to ρ. Let ν_δ denote the joint probability distribution of $(\pi X_\delta, \pi X_0)$ where X_0 is distributed according to $\rho\pi$. μ_δ and ν_δ are measures on $(S \times S)^{\mathbb{Z}^d}$. Consider the relative entropy S_n of the finite volume Λ_n- restrictions of μ_δ, ν_δ.

$$\begin{aligned} S_n(\mu_\delta|\nu_\delta) &\equiv \sum_{\sigma_{\Lambda_n}, \sigma'_{\Lambda_n}} \mu_\delta(\sigma_{\Lambda_n}, \sigma'_{\Lambda_n}) \log \frac{\mu_\delta(\sigma_{\Lambda_n}, \sigma'_{\Lambda_n})}{\nu_\delta(\sigma_{\Lambda_n}, \sigma'_{\Lambda_n})} \qquad (3.28) \\ &= \sum_{\sigma_{\Lambda_n}, \sigma'_{\Lambda_n}} \mathbb{P}_\rho[X_{0,\Lambda_n} = \sigma_{\Lambda_n}, X_{\delta,\Lambda_n} = \sigma'_{\Lambda_n}] \log \frac{\mathbb{P}_\rho[X_{0,\Lambda_n} = \sigma_{\Lambda_n}, X_{\delta,\Lambda_n} = \sigma'_{\Lambda_n}]}{\mathbb{P}_{\rho\pi}[X_{0,\Lambda_n} = \pi\sigma'_{\Lambda_n}, X_{\delta,\Lambda_n} = \pi\sigma_{\Lambda_n}]} \end{aligned}$$

Theorem 3.1 *For a stationary and translation invariant measure ρ,*

i.

$$\begin{aligned} MEP_\pi(\mathcal{P}_0, \rho) &= \sum_{U_0 \in \mathcal{P}_0} \Big(\int \rho(d\sigma)\, c(U_0, \sigma) \log \frac{c(U_0, \sigma)}{c(\pi\, U_0^{-1}\pi, \pi\, U_0\sigma)} \\ &+ \int \rho(d\sigma)\, [c(U_0, \pi\sigma) - c(U_0, \sigma)] \Big) \qquad (3.29) \end{aligned}$$

ii.

$$MEP_\pi(\mathcal{P}_0, \rho) = \lim_{n\to\infty} \lim_{\delta \downarrow 0} \frac{1}{|\Lambda_n|\delta} S_n(\mu_\delta, \nu_\delta) \qquad (3.30)$$

Proof: *i.* Start by noting that $N_t^{U_x} - \int_{-T}^t c(U_x, \omega_s) ds$ is a martingale. By the stationarity of ρ, the limit $T \to \infty$ can be taken immediately

$$\begin{aligned} \mathrm{MEP}_\pi(\mathcal{P}_0, \rho) &= \lim_{n\uparrow\infty} \frac{1}{|\Lambda_n|} \sum_{x \in \Lambda_n^\sharp} \sum_{U_x \in \mathcal{P}_x} \int \rho(d\sigma) c(U_x, \sigma) \log \frac{c(U_x, \pi\sigma)}{c(\pi U_x^{-1}\pi, \pi U_x\sigma)} \\ &+ \int \rho(d\sigma)[c(U_x, \pi\sigma) - c(U_x, \sigma)] \\ &= \lim_{n\uparrow\infty} \frac{|\Lambda_n^\sharp|}{|\Lambda_n|} \sum_{U_0 \in \mathcal{P}_0} \int \rho(d\sigma) c(U_0, \sigma) \log \frac{c(U_0, \pi\sigma)}{c(\pi U_0^{-1}\pi, \pi U_0\sigma)} \\ &+ \int \rho(d\sigma)[c(U_0, \pi\sigma) - c(U_0, \sigma)] \end{aligned}$$

By translation invariance of ρ together with the fact $|\Lambda_n \setminus \Lambda_n^\sharp| = 0(n^{d-1})$, the first part of the theorem follows.
ii. We have

$$\begin{aligned}
& S_n(\mu_\delta|\nu_\delta) \\
= & \ \delta\{\sum_{\sigma_{\Lambda_n}} \sum_{x\in\tilde{\Lambda}_n} \sum_{U_x\in\mathcal{P}_x} \left[\int I(\xi_{\Lambda_n} = \sigma_{\Lambda_n}) c(U_x,\xi) \log \frac{\int I(\xi_{\Lambda_n} = \sigma_{\Lambda_n}) c(U_x,\xi)\rho(d\xi)}{\int I(\xi_{\Lambda_n} = U_x\sigma_{\Lambda_n}) c(\pi U_x^{-1}\pi, \pi U_x\xi)\rho(d\xi)}\right] \\
+ & \sum_{\sigma_{\Lambda_n}} \sum_{x\in\tilde{\Lambda_n}} \sum_{U_x\in\mathcal{P}_x} \int \rho(d\xi)(c(U_x,\pi\xi) - c(U_x,\xi)) I(\xi_{\Lambda_n} = \sigma_{\Lambda_n})\} + O(\delta^2)
\end{aligned}$$

where $\tilde{\Lambda}_n \equiv \{x \in \mathbb{Z}^d : \Lambda_x \cap \Lambda_n \neq \emptyset\}$. We rewrite

$$\begin{aligned}
\lim_{\delta\downarrow 0} \frac{1}{\delta} S_n(\mu_\delta|\nu_\delta) &= \sum_{\sigma_{\Lambda_n}} \sum_{x\in\Lambda_n^\sharp} \sum_{U_x\in\mathcal{P}_x} \rho(\sigma_{\Lambda_n}) c(U_x,\sigma_{\Lambda_n}) \log \frac{c(U_x,\sigma_{\Lambda_n})\rho(\sigma_{\Lambda_n})}{c(\pi U_x^{-1}\pi, \pi U_x\sigma_{\Lambda_n})\rho(U_x\sigma_{\Lambda_n})} \\
&+ \sum_{\sigma_{\Lambda_n}} \sum_{x\in\Lambda_n^\sharp} \sum_{U_x\in\mathcal{P}_x} \int \rho(d\sigma_{\Lambda_n})(c(U_x,\pi\sigma_{\Lambda_n}) - c(U_x,\sigma_{\Lambda_n})) \\
&+ o(|\Lambda_n|)
\end{aligned}$$

Comparing this expression with that of (3.29), we still have to prove that

$$\lim_{n\to\infty} \frac{1}{|\Lambda_n|} \sum_{\sigma_{\Lambda_n}} \sum_{x\in\Lambda_n^\sharp} \sum_{U_x\in\mathcal{P}_x} \rho(\sigma_{\Lambda_n}) c(U_x,\sigma_{\Lambda_n}) \log \frac{\rho(\sigma_{\Lambda_n})}{\rho(U_x\sigma_{\Lambda_n})} = 0 \tag{3.31}$$

Put $F_{\Lambda_n}(\sigma) \equiv -\log\rho(\sigma_{\Lambda_n})$; this function depends only on spins inside Λ_n. (3.31) can be rewritten as

$$\lim_{n\to\infty} \frac{1}{|\Lambda_n|} \int \rho(d\sigma) \sum_{x\in\Lambda_n^\sharp} \sum_{U_x\in\mathcal{P}_x} c(U_x,\sigma)[F_{\Lambda_n}(U_x\sigma) - F_{\Lambda_n}(\sigma)] \tag{3.32}$$

By stationarity of ρ (3.32) equals

$$\lim_{n\to\infty} \frac{1}{|\Lambda_n|} \int \rho(d\sigma_{\Lambda_n}) \sum_{x\in\Lambda_n^{\sharp\,c}} \sum_{U_x\in\mathcal{P}_x} c(U_x,\sigma_{\Lambda_n})[F_{\Lambda_n}(U_x\sigma) - F_{\Lambda_n}(\sigma)]$$

The only $x \in \Lambda_n^{\sharp\,c}$ for which $F_{\Lambda_n}(U_x\sigma) - F_{\Lambda_n}(\sigma)$ is non-zero are $x \in \Lambda_n^{\sharp\,c} \cap \tilde{\Lambda_n}$, hence the integral is of the order $|\tilde{\Lambda}_n \setminus \Lambda_n| + |\Lambda_n \setminus \Lambda_n^\sharp| = O(n^{d-1})$, which proves (3.31). ■

3.3 Detailed balance

Given our configurational time reversal π, we say that the dynamics satisfies the condition of (generalized) detailed balance with respect to the Hamiltonian H if for all $x \in \mathbb{Z}^d, \sigma \in \Omega, U_x \in \mathcal{P}_x$,

$$c(U_x, \sigma) = c(\pi U_x^{-1}\pi, \pi U_x \sigma) \exp(-H(U_x\sigma) + H(\sigma)). \tag{3.33}$$

This is the analogue of (2.17). The energy difference in (3.33) should be interpreted in terms of an absolutely convergent sum of potentials:

$$H(\sigma_\Lambda \eta_{\Lambda^c}) - H(\xi_\Lambda \eta_{\Lambda^c}) = \sum_{A \cap \Lambda \neq \emptyset} \left(V(A, \sigma_\Lambda \eta_{\Lambda^c}) - V(A, \xi_\Lambda \eta_{\Lambda^c})\right), \tag{3.34}$$

where $(V(A, \cdot) : S^A \to (-\infty, +\infty), A$ finite subsets of $\mathbb{Z}^d)$, is a translation invariant (uniformly) absolutely summable potential:

$$\sum_{A \ni 0} \max_{\sigma \in S^A} |V(A, \sigma)| < +\infty \tag{3.35}$$

The following theorem states that detailed balance makes the mean entropy production zero.

Theorem 3.2 *Suppose that the transformation rates satisfy (3.33) for the potential $V(A, \cdot)$, and that ρ and $\rho\pi$ are translation invariant stationary distributions. Then,*

$$\mathrm{MEP}_\pi(\mathcal{P}_0, \rho) = \mathrm{MEP}_\pi(\mathcal{P}_0, \rho\pi) = 0$$

Proof: Let us first assume that $\rho = \rho\pi$. Using translation invariance of ρ, we obtain from subsituting (3.33) in (3.29)

$$\mathrm{MEP}(\rho) = -\frac{1}{|\Lambda_\mathrm{n}|} \sum_{\mathrm{x} \in \Lambda_\mathrm{n}} \sum_{\mathrm{U_x} \in \mathcal{P}_\mathrm{x}} \int \rho(\mathrm{d}\sigma_{\Lambda_\mathrm{n}})\, \mathrm{c}(\mathrm{U_x}, \sigma_{\Lambda_\mathrm{n}}) \sum_{\mathrm{A} \cap \Lambda_\mathrm{x} \neq \emptyset} [\mathrm{V}(\mathrm{A}, \mathrm{U_x}\sigma_\Lambda) - \mathrm{V}(\mathrm{A}, \sigma_\Lambda)] \tag{3.36}$$

for every Λ_n. We split up the third sum in (3.36) according to $A \subset \Lambda_n$ or not. The first part is equal to

$$-\frac{1}{|\Lambda_n|} \sum_{x \in \Lambda_n} \sum_{U_x \in \mathcal{P}_x} \int \rho(d\sigma_{\Lambda_n})\, c(U_x, \sigma) \left[H_\Lambda(U_x\sigma) - H_\Lambda(\sigma)\right] \tag{3.37}$$

for $H_{\Lambda_n}(\sigma) \equiv \sum_{A \subset \Lambda_n} V(A, \sigma)$. Since, by the stationarity of ρ, (3.37) would be zero if summed over all x, we see that its absolute value is bounded by

$$\frac{2 \sup_\sigma c(U_0, \sigma)}{|\Lambda_n|} \sum_{x \in \Lambda_n^c} \sum_{U_x \in \mathcal{P}_x,\ \Lambda_x \cap \Lambda_n \neq \emptyset} \sum_{A \cap \Lambda_x \neq \emptyset} \sup_{\sigma_{\Lambda_n}} |V(A, \sigma)| \tag{3.38}$$

Since from (3.35), $\sum_{A \cap \Lambda_x \neq \emptyset} \sup_\sigma |V(A, \sigma)| < C|\Lambda_0|$ and $|\{U_x \in \mathcal{P}_x\}| \leq |S|^{|\Lambda_0|}!$, we can estimate the first term (3.37) by

$$\mathrm{C'}\frac{1}{|\Lambda_n|} |\{x \in \Lambda_n^c : \Lambda_n \cap \Lambda_x \neq \emptyset\}| \tag{3.39}$$

which converges to zero as $\Lambda \uparrow \mathbb{Z}^d$.
A similar analysis can be made for the second part, that is

$$-\frac{1}{|\Lambda_n|} \sum_{x\in\Lambda_n} \sum_{U_x\in\mathcal{P}_x} \int \rho(d\sigma_{\Lambda_n})\, c(U_x, \sigma_{\Lambda_n}) \sum_{A\cap\Lambda_x\neq\emptyset,\ A\cap\Lambda_n^c\neq\emptyset} [V(A, U_x\, \sigma_{\Lambda_n}) - V(A, \sigma_{\Lambda_n})]$$

This converges to zero by (3.35) and translation invariance of V. ■
When $\rho \neq \rho\pi$, we apply the previous to the measure $\rho' = 1/2(\rho + \rho\pi)$ which is clearly π-invariant. Use then that $\text{MEP}(\rho') = 0$ implies $\text{MEP}(\rho) = \text{MEP}(\rho\pi) = 0$ because $\text{MEP} \geq 0$ (by (3.30)) and because MEP is affine by definition (3.27).

4 Additional remarks and questions

The specific enterprise that was started in the present paper concerns a mathematical study of entropy production in spatially (infinitely) extended systems. We list here some of the many additional questions concerning general aspects of the theory. Of course, specific models will allow many additional (specific) questions.

- Is it correct that for the interacting particle systems that have appeared in the present paper, vanishing mean entropy production implies that the process is reversible? More specifically, can one show that

$$MEP_\pi(\mathcal{P}_0, \rho) = 0 \Rightarrow c(U_0, \sigma) = c(\pi U_0^{-1}\pi, \pi U_0 \sigma)\frac{d\rho \circ U_0}{d\rho}, U_0 \in \mathcal{P}_0 \tag{4.40}$$

  where we assume that $\rho \circ U_0$ has a density with respect to ρ. If the set $\mathcal{P}_0$ is sufficiently rich (in the sense that any configuration can be changed in any other configuration that locally differs by applications of the U_x), that would imply that zero mean entropy production can only occur if ρ is a π-reversible Gibbs measure. Additionally, does this 'no current without heat' result fail in case the dynamics or the involution π is no longer (quasi-)local? A partial result has been worked out in [3] for the case of spinflip dynamics.

- Is there a thermodynamic analogue of the circulation formulae that have been developed in [6, 7, 8] for the mean entropy production?

- Can one extend the local fluctuation theorems for the entropy production obtained in [1] for probabilistic cellular automata to the case of interacting particle systems, including those that have a conservation law?

References

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

[2] Maes C., Redig F., Van Moffaert A., *On the definition of entropy production via examples* J. Math. Phys. (March 2000).

[3] Maes C., Redig F., *Positivity of entropy production* To appear in J. Stat. Phys.

[4] Liggett T. M., *Interacting particle systems* Springer-Verlag, New York, Heidelberg, Berlin. (1985)

[5] Kipnis C., Landim C., *Scaling limits of interacting particle systems* Springer-Verlag, New York, Heidelberg, Berlin. (1999)

[6] Qian M. P., Qian M., Qian C., *Circulations of markov chains with continuous time and probability interpretation of some determinants* Sci. Sinica **27**, 470-481. (1984)

[7] Qian M. P., Qian M., *The entropy production and reversibility of Markov processes* Proceedings of the first world congress Bernoulli soc. 1988, 307-316

[8] Schnakenberg J., *Network theory of behavior of master equation systems* Rev. Mod. Phys. **48**, 4, 571-585. (1976)

[9] Kirchhoff G., Poggendorffs Ann. Phys. Chem., **72**, 495. (1847)

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

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

[12] De Groot S. R., Mazur P., *Non- equilibrium thermodynamics* North- Holland publishing company, Amsterdam (1962)

[13] Lebowitz J. L., Spohn H., *A Gallavotti-Cohen type symmetry in the large deviation functional for stochastic dynamics* J. Stat. Phys. **95**, 333-365. (1999)