

No Current Without Heat

Christian Maes,¹ Frank Redig,² and Michel Verschuere³

Received March 22, 2001; accepted October 2, 2001

We show for a large class of interacting particle systems that whenever the stationary measure is not reversible for the dynamics, then the mean entropy production in the steady state is strictly positive. This extends to the thermodynamic limit the equivalence between microscopic reversibility and zero mean entropy production: time-reversal invariance cannot be spontaneously broken.

KEY WORDS: Stochastic interacting particle systems; entropy production; (generalized) detailed balance.

1. INTRODUCTION

Reversibility and entropy are words with many meanings even within the context of nonequilibrium statistical mechanics. One class of models that has often been considered for learning about nonequilibrium behavior is that of interacting particle systems. These are stochastic dynamics for spatially extended systems in which particles locally interact. They are mostly toy-models remaining far from realistic in their microscopic details. Yet, it is believed that for some good purposes, the details do not matter so much and one should be concerned more with the symmetries, possible conservation laws, locality of the interaction etc. to hope to understand something about real nature.

This paper is about the relation between time-reversal invariance and the positivity of entropy production. We do this in the context of interacting particle systems following the work in refs. 1–5. The physics background

¹ Instituut voor Theoretische Fysica, K.U. Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium; e-mail: Christian.Maes@fys.kuleuven.ac.be

² T.U. Eindhoven. On leave from Instituut voor Theoretische Fysica, K.U. Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium; e-mail: f.h.j.redig@tue.nl

³ Instituut voor Theoretische Fysica, K.U. Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium; e-mail: Michel.Verschuere@fys.kuleuven.ac.be

will be discussed in Section 3. The main question is to understand why there cannot exist a “superconducting” interacting particle system in the sense of the title of this paper to be specified below. To understand the mathematical problem, let us look first at a finite Markov chain. Suppose that K is a finite set on which we have an involution $\pi: K \rightarrow K$, $\pi^2 = id$, called time-reversal. Often, the most natural choice for interacting particle systems is $\pi = id$ because we think of the state space as consisting of occupation variables or of classical spins but our mathematical set-up will be more general.

Let $(X_t, t \in [-T, T])$ be a stationary Markov process (steady state) on K with law \mathbb{P}_ρ . The subscript refers to the unique stationary probability measure ρ on K ; we assume that $\rho(a) > 0$, $a \in K$. The rate to go from a to b is denoted by $k(a, b)$, $a, b \in K$ and we assume that $k(\pi b, \pi a) = 0$ iff $k(a, b) = 0$ (dynamic reversibility). The generator is

$$Lf(a) = \sum_b k(a, b)[f(b) - f(a)]. \quad (1.1)$$

The time-reversed process of (X_t) is the stationary Markov process $(Y_t, t \in [-T, T])$ on K with $Y_t \equiv \pi X_{-t}$ having transition rates

$$\tilde{k}(a, b) \equiv k(\pi b, \pi a) \frac{\rho(\pi b)}{\rho(\pi a)}. \quad (1.2)$$

We denote its law by $\tilde{\mathbb{P}}_{\rho\pi}$ ($\rho\pi$ is stationary for (Y_t)). Of course, it easily happens that $\rho = \rho\pi$ and yet, $\mathbb{P}_\rho \neq \tilde{\mathbb{P}}_{\rho\pi}$. The corresponding generator for the time-reversed process is $\tilde{L} = \pi L^* \pi$ where the $*$ refers to the adjoint with respect to the stationary measure ρ .

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

$$\rho(a) k(a, b) = k(\pi b, \pi a) \rho(b), \quad a, b \in K \quad (1.3)$$

which is generalized (or extended) detailed balance (microscopic reversibility). For the generators, we then have $\tilde{L} = L$. Observe that (1.3) by itself implies that $\rho(a) = \rho\pi(a) \rho(b) / \rho\pi(b)$ whenever $k(a, b) \neq 0$. Applying this successively with $b_1, \dots, b_n \in K$ for which $k(a, b_1), k(b_1, b_2), \dots, k(b_n, \pi a) \neq 0$, we find that $\rho(a) = \rho\pi(a)$. On the other hand, π -reversibility implies that $\rho = \rho\pi$ is stationary.

The entropy production is the random variable obtained from taking the relative action on pathspace with respect to time-reversal, see ref. 6

for a recent review. Let ρ_{-T} be a probability measure on K which we use to sample the initial data at time $-T$ for the stochastic time-evolution generated by L . The law of this process is denoted by $\mathbb{P}_{\rho_{-T}}$. Suppose now that for this process the state at time T is described by the probability measure ρ_T . We could as well start our process (at time $-T$) from $\rho_T\pi$ and then obtain the process $\mathbb{P}_{\rho_T\pi}$. For a particular realization $\omega = (\omega(t), t \in [-T, T])$ of this process we let $\Theta_\pi\omega \equiv (\pi\omega(-t), t \in [-T, T])$ be its time-reversal. The entropy production $R_\pi(L, \rho_{-T}, T)$ is a function of the realization over the time-interval $[-T, T]$ and is then obtained as

$$R_\pi(L, \rho_{-T}, T)(\omega) = \log \frac{d\mathbb{P}_{\rho_{-T}}}{d\mathbb{P}_{\rho_T\pi}\Theta_\pi}(\omega). \tag{1.4}$$

Here we are only interested in its steady state expectation value, that is the mean entropy production rate, which in fact can be written as

$$\text{MEP}_\pi(L, \rho) = \lim_{T \uparrow \infty} \frac{1}{2T} \mathbb{E}_\rho \left[\log \frac{d\mathbb{P}_\rho}{d\tilde{\mathbb{P}}_{\rho\pi}} \right] \tag{1.5}$$

where \mathbb{E}_ρ denotes expectation with respect to \mathbb{P}_ρ . The notation $\text{MEP}_\pi(L, \rho)$ reminds us that this number depends on the transformation π , the dynamics (generated via L) and the stationary measure ρ . The mean entropy production thus measures the degree to which \mathbb{P}_ρ can be distinguished from $\tilde{\mathbb{P}}_{\rho\pi}$. The main property of the mean entropy production is then:

Proposition 1. Consider the stationary process (X_t) above with $\rho = \rho\pi$. Then, $\text{MEP}_\pi(L, \rho) = \text{MEP}_\pi(\tilde{L}, \rho) \geq 0$ with equality if and only if the process (X_t) is π -reversible.

This says that for finite state space Markov chains there can be no current without heat, meaning that detailed balance is equivalent with zero mean entropy production. The problem we address here is whether the same remains true in the thermodynamic limit, that is for spatially extended interacting particle systems. In this case we really should be speaking about the mean entropy production *density*, i.e., per unit volume, but we will not use this extension. Note that in this case and from now on we will not and we cannot assume in general that $\rho = \rho\pi$ even if both are stationary.

We discuss the general physics set-up and further interpretations in Section 3, after stating our mathematical results in Section 2. We start however with three examples illustrating some aspects.

1.1. Examples

Example A. We consider particles hopping on the one-dimensional lattice with a preferred direction that is itself subject to independent flips. The state space is $\{-1, +1\} \times \{0, 1\}^{\mathbb{Z}}$ and the process is determined by choosing a constant rate $c(E, \eta) = 1$ for changes from a configuration (E, η) to $(-E, \eta)$ and taking rates

$$c(x, E, \eta) = e^E \eta_x (1 - \eta_{x+1}) + e^{-E} \eta_{x+1} (1 - \eta_x)$$

for changes to $(E, \eta^{x, x+1})$ where $(\eta^{x, x+1})_y = \eta_y$ if $x \neq y \neq x+1$, and $(\eta^{x, x+1})_y = \eta_x$ when $y = x+1$ and $= \eta_{x+1}$ when $y = x$. The resulting Markov process has generator

$$\begin{aligned} Lf(E, \eta) = & \sum_x [e^E \eta_x (1 - \eta_{x+1}) + e^{-E} \eta_{x+1} (1 - \eta_x)] [f(E, \eta^{x, x+1}) - f(E, \eta)] \\ & + f(-E, \eta) - f(E, \eta). \end{aligned} \quad (1.6)$$

For invariant measure ρ we take

$$\rho(E, d\eta) \equiv \frac{1}{2} (\delta_{E, +1} + \delta_{E, -1}) \times \nu_u(d\eta)$$

where ν_u is the Bernoulli measure with specified density $u \in (0, 1)$. For time-reversal we take $\pi(E, \eta) = (-E, \eta)$ so that $\rho = \rho\pi$.

It is easy to see that the process satisfies generalized detailed balance, like (1.3), in the sense that $\rho(E, d\eta) = \rho(-E, d\eta) = \rho(E, d\eta^{x, x+1})$ and both

$$c(E, \eta) = c(-E, \eta) \quad \text{and} \quad c(x, E, \eta) = c(x, -E, \eta^{x, x+1}).$$

The last identity depends of course crucially on the fact that π is not the identity and reverses left and right as preferred direction. At the same time, as can be computed explicitly, the mean entropy production is zero. The same remains true for π a particle-hole transformation, $(\pi\eta)_x = 1 - \eta_x$, leaving the field E unchanged. Then, $\rho \neq \rho\pi$ for $u \neq 1/2$ but still generalized detailed balance holds. Finally if, instead, we were to take $\pi = \text{identity}$ as time-reversal, then we break the detailed balance condition and we obtain a strictly positive mean entropy production.

Example B. We take the simplest example of a spinflip dynamics for which the one-dimensional Ising model is stationary but not reversible (for $\pi = \text{id}$). Exactly the same can be done in two dimensions, see ref. 7. Spinflips are transformations $U_x: \sigma \rightarrow U_x(\sigma) = \sigma^x$, $x \in \mathbb{Z}$, $\sigma \in \{+1, -1\}^{\mathbb{Z}}$ for σ^x equal to σ except at the site x .

Consider the one-dimensional spinflip dynamics with the following asymmetric rates:

$$c(x, \sigma) = \exp(-2\beta\sigma_x\sigma_{x+1}). \quad (1.7)$$

The invariant measure ρ is the one-dimensional Ising model at inverse temperature β . The process starting from ρ is not time-reversal invariant and the entropy production is equal to $\text{MEP}(L, \rho) = 4\beta \tanh \beta$ (that is with time-reversal $\pi = \text{identity}$). On the other hand, this time-reversed process is easy to find; it is a spinflip process with generator

$$L^* f(\sigma) = \sum_x e^{-2\beta\sigma_x\sigma_{x-1}} [f(\sigma^x) - f(\sigma)].$$

Let us now take for time-reversal π the reflection: $(\pi\sigma)_x = \sigma_{-x}$ which leaves ρ invariant. Since

$$(\pi\sigma)^x = \pi(\sigma^{-x})$$

$L^* = \pi L \pi$ and we have in fact generalized detailed balance (1.3):

$$\frac{c(x, \sigma)}{c(-x, (\pi\sigma)^{-x})} = \frac{d\rho \circ U_x}{d\rho}(\sigma) = e^{-2\beta\sigma_x(\sigma_{x-1} + \sigma_{x+1})}.$$

The denominator in the left hand side is the rate in the original process by which $\pi U_x \sigma = \pi(\sigma^x) = (\pi\sigma)^{-x}$ is changed to $\pi\sigma$. As a result, $\text{MEP}_\pi(L, \rho) = \text{MEP}_\pi(L^*, \rho) = 0$.

Example C. Instead of driving the system in the bulk and breaking detailed balance via some external fields that act on each component of the system, we may also consider boundary driven processes. For this we need to start with finite volume. The simplest interesting case is that of a symmetric exclusion process on a lattice interval that is driven by independent birth and death processes at its boundaries corresponding to different chemical potentials. Take $A_n = \{-n, -n+1, \dots, n-1, n\}$ and $\eta \in \{0, 1\}^{A_n}$ a particle configuration evolving with generator

$$G_n f(\eta) = \sum_{x=-n}^{n-1} [f(\eta^{x, x+1}) - f(\eta)] + \lambda [e^{h_1 \eta^{-n}} (f(\eta^{-n}) - f(\eta)) + e^{h_2 \eta_n} (f(\eta^n) - f(\eta))]. \quad (1.8)$$

The first term corresponds to symmetric hopping with exclusion; the two last terms are giving birth and death to particles at the ends of the interval

with parameters h_1, h_2 . One can think here of particle reservoirs, to the left of the system with density $1/(1+e^{h_1})$ and to the right with density $1/(1+e^{h_2})$. For $\lambda = 0$ the system is uncoupled from the reservoirs and it has all uniform product measures as reversible measures with vanishing mean entropy production. For $\lambda \neq 0$, $h_1 \neq h_2$ this detailed balance is lost and we have positive mean entropy production. Yet, it remains of order unity, uniformly in the size n meaning that the mean entropy production density vanishes in the thermodynamic limit. This is an instance of a more general fact for interacting particle systems that will also be treated in the next section: you cannot by driving the system at its boundaries break the time-reversal invariance in the limiting infinite volume process, see Proposition 2 below.

We do not know whether there exists a time-reversal π for which (1.8) would give rise to generalized detailed balance.

In this paper we show more generally how breaking of detailed balance is strictly equivalent with non-zero mean entropy production. There is no way to get a current and at the same time to have no dissipation (zero mean entropy production).

In the next section we describe our class of models and we state our main result. In Section 3, we discuss this result and we give some more background information concerning entropy production, reversibility and time-reversal. Section 4 is devoted to the proofs.

2. MODELS AND MAIN RESULT

2.1. Dynamics

This subsection describes the assumptions and introduces the necessary notation.

The configuration space is $\Omega \equiv S^{\mathbb{Z}^d}$ where S is a finite set and \mathbb{Z}^d is the regular d -dimensional lattice. Let π be an involution on Ω . A special but important case is when $\pi = \text{identity}$. We assume here that π commutes with lattice translations τ_x , $x \in \mathbb{Z}^d$.

Let $V_0 \subset \mathbb{Z}^d$ be a finite cube containing the origin and write \mathcal{P}_0 for any specific non-empty set of transformations U_0 on Ω satisfying, for every $U_0 \in \mathcal{P}_0$, and for every $\sigma \in \Omega$:

- (i) $(U_0\sigma)(y) = \sigma(y)$, for $y \in V_0^c$,
- (ii) $U_0^{-1} \in \mathcal{P}_0$,
- (iii) $\pi\mathcal{P}_0\pi = \mathcal{P}_0$,
- (iv) If $U_0 \neq U'_0$ and $U_0\sigma \neq \sigma$, $U'_0\sigma \neq \sigma$ then $U_0\sigma \neq U'_0\sigma$ (for convenience only.)

We consider the translations $V_x \equiv \{y+x: y \in V_0\}$ and $U_x \equiv \tau_x U_0 \tau_{-x}$ to generate a dynamics via local translation invariant rates $c(U_x, \sigma)$ for the transition $\sigma \rightarrow U_x \sigma$. We assume:

- (v) Positivity: $c(U_0, \sigma) = 0$ when $U_0 \sigma = \sigma$ and if not, $c(U_0, \sigma) > 0$,
- (vi) Finite range: there is a finite $\bar{A} \subset \mathbb{Z}^d$ such that for all $\sigma, \eta \in \Omega$, and $U_0 \in \mathcal{P}_0$: $c(U_0, \sigma) = c(U_0, \sigma_{\bar{A}} \eta_{\bar{A}^c})$,
- (vii) Translation invariance: for all $x \in \mathbb{Z}^d$, $U_x \in \mathcal{P}_x$, $\sigma \in \Omega$: $c(U_x, \sigma) = c(U_0, \tau_{-x} \sigma)$.

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

$$Lf(\sigma) \equiv \sum_{x \in \mathbb{Z}^d} \sum_{U_x \in \mathcal{P}_x} c(U_x, \sigma)[f(U_x \sigma) - f(\sigma)]. \tag{2.9}$$

That is, σ is changed to η at rate $c(U_x, \sigma)$ if $\eta = U_x \sigma$. We will always write ρ for a translation invariant stationary measure for this dynamics. It can be different from $\rho\pi$ but we assume that also $\rho\pi$ is stationary. Finally, ρ and $\rho\pi$ give positive weight to all cylinders and writing $\rho^U = U\rho$, we always assume that $d\rho^{U_0}/d\rho(\sigma) \geq c > 0$, which, even in the present rather general set-up, can be expected quite generally.

For $V_0 = \{0\}$ and $S = \{+1, -1\}$, the choice $U_x \sigma = \sigma^x$ corresponds to a spinflip process. Taking $V_0 = \{0, e_1, e_2, \dots, e_d\}$ with e_α the lattice unit vectors, we can make a spin exchange process or hopping dynamics. We refer to ref. 8 for further details on constructing the infinite volume process.

2.2. Mean Entropy Production

Put $A_n = [-n, n]^d \cap \mathbb{Z}^d$ for large n and define $A_n^\#$ as the maximal subset of A_n , such that for all $x \in A_n^\#$ and $U_x \in \mathcal{P}_x$, $c(U_x, \sigma)$ depends only on coordinates inside A_n , and $V_x \subset A$. Consider now the Markov chain on S^{A_n} with generator

$$L_n f(\sigma) \equiv \sum_{x \in A_n^\#} \sum_{U_x \in \mathcal{P}_x} c(U_x, \sigma)[f(U_x \sigma) - f(\sigma)] \tag{2.10}$$

and started from a probability measure ρ_{-T} on S^{A_n} at time $-T$. The measure at time T is denoted by ρ_T . Via a Girsanov formula this dynamics gives rise to a Hamiltonian (or action functional) on space-time trajectories

ω (as in refs. 2 and 6), with corresponding relative energy with respect to time-reversal given by the entropy production (1.4) and here equal to

$$R_\pi(L_n, \rho_{-T}, T)(\omega) = \ln \rho_{-T}(\omega(-T)) - \ln \rho_T(\omega(T)) + \Delta S_e(\omega) \quad (2.11)$$

with

$$\begin{aligned} \Delta S_e(\omega) = & \sum_{x \in A_n^*} \sum_{U_x \in \mathcal{D}_x} \int_{-T}^T \log \frac{c(U_x, \omega(s^-))}{c(\pi U_x^{-1} \pi, \pi U_x \omega(s^-))} dN_s^{U_x}(\omega) \\ & + \int_{-T}^T [c(U_x, \pi \omega(s)) - c(U_x, \omega(s))] ds \end{aligned} \quad (2.12)$$

where $N_t^{U_x}(\omega) \equiv \sum_{-T \leq s \leq t} I(\omega(s) = U_x(\omega(s^-)) \neq \omega(s^-))$ is the number of times the transformation U_x appeared in the realization ω up to time $t \in [-T, T]$. The expression (2.12) must be interpreted as the variable entropy produced in the reservoirs (environment) when the microscopic system configuration moves from $\omega(-T)$ to $\omega(T)$: To get the total variable entropy production (2.11) one should add to (2.12) the corresponding change in the system's entropy, that are the first two terms in (2.11). However, when taking steady state averages, this part vanishes (the entropy of the stationary system does not change on average). We can therefore define the mean entropy production for the interacting particle system as

$$\text{MEP}_\pi(L, \rho) \equiv \lim_n \lim_{T \uparrow \infty} \frac{1}{2 |A_n| T} \mathbb{E}_\rho^{n, T}(\Delta S_e) \quad (2.13)$$

$\mathbb{E}_\rho^{n, T}$ denotes the expectation with respect to the path space measure, in the stationary distribution ρ , restricted to trajectories within S^{A_n} . In other words, the mean entropy production is the expectation of the time-reversal breaking part in the space-time action functional governing the dynamics. We refer to ref. 2 for a mathematical discussion on the existence of the limit (2.13) and for a proof of its non-negativity. We refer to refs. 3, 4, and 6 and Section 3 for further background.

2.3. Results

The main question is to see whether for a dynamics where the time-reversal symmetry is explicitly broken (in the sense that there is no detailed balance), there still can be zero mean entropy production (dissipationless steady state). Our main result says that this is impossible.

Main Theorem. Under the conditions above, $\text{MEP}_\pi(L, \rho) = \text{MEP}_\pi(L, \rho\pi) = 0$ implies that the dynamics satisfies (generalized) detailed balance in the sense that for all U_0

$$c(\pi U_0^{-1}\pi, \pi U_0\sigma) \frac{d\rho^{U_0}}{d\rho}(\sigma) = c(U_0, \sigma) \quad \rho\text{-a.s.} \quad (2.14)$$

Note that (2.14) is really the analogue of (1.3). Observe also here that (2.14) implies that the densities $d\rho^{U_0}/d\rho$ are invariant under replacing ρ by $\rho\pi$. This follows from rewriting (2.14) from right to left with $\sigma \rightarrow \pi U_0\sigma$ and $U_0 \rightarrow \pi U_0^{-1}\pi$:

$$\begin{aligned} c(\pi U_0^{-1}\pi, \pi U_0\sigma) &= c(\pi\pi U_0\pi, \pi\pi U_0^{-1}\pi\pi U_0\sigma) \frac{d\rho^{\pi U_0^{-1}\pi}}{d\rho}(\pi U_0\sigma) \\ &= c(U_0, \sigma) \frac{d\rho\pi}{d\rho\pi^{U_0}}(\sigma) \end{aligned}$$

and comparing it with the original (2.14).

We call \mathcal{P}_0 complete if every local transformation $h: \Omega \rightarrow \Omega$ can be written as a composition of U_x : i.e., if $h = U_{x_1} \cdots U_{x_n}$ for some $x_1, \dots, x_n \in \mathbb{Z}^d$.

Corollary 1. If $\text{MEP}_\pi(L, \rho) = \text{MEP}_\pi(L, \rho\pi) = 0$ and if \mathcal{P}_0 is complete and π is continuous, then ρ is a reversible Gibbs measure for the dynamics defined above.

In ref. 2 the converse to these results was already shown: Suppose that the rates satisfy

$$c(U_x, \sigma) = c(\pi U_x^{-1}\pi, \pi U_x\sigma) \exp(-H(U_x\sigma) + H(\sigma)). \quad (2.15)$$

This is again the analogue of (1.3). The energy difference in (2.15) should be interpreted in terms of an absolutely convergent sum of potentials:

$$H(\sigma_A \eta_{A^c}) - H(\xi_A \eta_{A^c}) = \sum_{A \cap A^c \neq \emptyset} (V(A, \sigma_A \eta_{A^c}) - V(A, \xi_A \eta_{A^c})), \quad (2.16)$$

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

$$\sum_{A \ni 0} \max_{\sigma \in S^A} |V(A, \sigma)| < +\infty. \quad (2.17)$$

Then,

$$\text{MEP}_\pi(L, \rho) = \text{MEP}_\pi(L, \rho\pi) = 0$$

When we combine the above we obtain a final

Corollary 2. Under the conditions of Corollary 1, if there is one translation invariant stationary measure ρ for which $\rho = \rho\pi$ and $\text{MEP}_\pi(L, \rho) = 0$, then also $\text{MEP}_\pi(L, \nu) = 0$ for all translation invariant stationary measures ν and they are all Gibbsian for the same potential.

A caveat in the above main result is to understand better the relation between $\text{MEP}_\pi(L, \rho)$ and $\text{MEP}_\pi(L, \rho\pi)$. To this we can only add that $\text{MEP}_\pi(\pi L\pi, \rho) = \text{MEP}_\pi(L, \rho\pi)$, as can be verified from a direct computation starting with (4.24). The simplest illustration of all this was already obtained in ref. 5 for a spinflip process. Example B, (1.7), deals with a spinflip process but there the time-reversal π does not commute with translations. As will be seen from the proof, that is indeed not essential as long as the dynamics and the stationary measure are translation invariant. Of course, one should then be extra careful with condition (iii) but also this can be modified accordingly. It will also be clear that more general lattice structures and configuration spaces can be employed (e.g., already in Example A).

Finally, for completeness we come back to the situation of Example C in Section 1.1. For this we must leave the translation invariant infinite volume context and ask whether boundary driven interacting particle systems can give rise to non-vanishing mean entropy production density in the thermodynamic limit. The question can be formalized as follows. We consider a process on S^{A_n} with generator G_n generalizing (1.8)

$$G_n f(\sigma) \equiv L_n f(\sigma) + \sum_{\substack{A \subset A_n \setminus A_n^\# \\ \text{diam } A \leq r}} \sum_{\eta \in S^A} k_A^{(n)}(\sigma, \eta) [f(\sigma^{A, \eta}) - f(\sigma)]$$

where $\sigma^{A, \eta} \equiv \sigma_{A^c} \eta_A$ equals σ outside the set A which has a diameter (maximal lattice distance within) less than a given constant r .

Here the generator L_n is given by (2.10) but with rates verifying condition (2.15) for a finite range potential, and rates $k_A^{(n)}(\sigma, \eta)$ as in (1.8) inducing configurational changes at the boundary of A_n . We further assume that the $k_A^{(n)}(\sigma, \eta)$ are uniformly bounded from below and from above. In other words, we have a bulk dynamics generated by L_n with rates satisfying (generalized) detailed balance, and at the boundary the configuration can change quite arbitrarily (but in a local and bounded way). We

suppose that ρ_n is the unique stationary measure of this dynamics and for simplicity we only treat the case $\pi = id$. We are interested in the mean entropy production $\text{MEP}(G_n, \rho_n)$ defined in (1.5) (with $\pi = id$).

Proposition 2. There is a constant K so that $\text{MEP}_\pi(G_n, \rho_n) \leq Kn^{d-1}$.

The proofs of the above results are postponed to Section 4.

3. DISCUSSION

We briefly discuss some concepts that are important for our result.

3.1. Time-Reversal

By this we usually mean a transformation on phase space Ω which, for a many-particle system, is defined particle-wise or, for spatially extended systems, is sufficiently local. Physically speaking, its precise nature follows from kinematical considerations on the dynamical variables. In classical mechanics, it reverses the momenta of all the particles but in the presence of say an electromagnetic potential, considered part of the system, one can add an extra transformation reversing also the magnetic field and thus making the Lorentz force time-reversal invariant. In our case, we have a configuration space $\Omega = S^{\mathbb{Z}^d}$ with \mathbb{Z}^d the d -dimensional lattice and S a finite set. Time-reversal is an involution π on Ω , $\pi^2 = id$. Time-reversal extends to a transformation θ_π on path-space, as introduced for (1.4), by reversing the trajectories. That is, if we have a trajectory $(\omega_t, t \in [-T, T])$ then the time-reversed trajectory $\theta_\pi(\omega)$ is given by $(\theta_\pi(\omega))_t \equiv \pi\omega_{-t}$.

3.2. Reversibility

Dynamic reversibility is a property of the dynamics itself under time-reversal. It says that if one trajectory ω of the system is possible, so is its time-reversed $\theta_\pi(\omega)$. For a deterministic system where $\omega_t = \phi(t)\omega_0$ with $\phi(t)$ an invertible flow on phase space, it says that $\phi(t)^{-1} = \pi\phi(t)\pi$, that is a symmetry that anticommutes with the time evolution. For a stochastic dynamics this is implied by assuming that if a transition $\sigma \rightarrow U\sigma$ is possible (positive transition rate), then also the same is true for its time-reversal $\pi U\sigma \rightarrow \pi\sigma$.

Microscopic reversibility is a consequence of dynamic reversibility in case of an equilibrium dynamics. For our purposes here we do not make a distinction with the condition of detailed balance. When the dynamics is driven away from equilibrium, the resulting stochastic model will not satisfy

detailed balance. Usually this produces a current in the system (but that need not be true in general, see an example in ref. 4). On the other hand, a net current signifies the breaking of the detailed balance condition. In general we like to distinguish between two classes of finite volume dynamics where microscopic reversibility is explicitly broken. These are boundary driven versus bulk driven dynamics depending on the extensivity of the perturbation from an equilibrium dynamics. In the bulk driven case, one usually verifies so called *local detailed balance*, i.e., (2.15) is changed into

$$c(U_x, \sigma) = c(\pi U_x^{-1} \pi, \pi U_x \sigma) \exp(-H(U_x \sigma) + H(\sigma)) e^{E\Phi(U_x \sigma, \sigma)}$$

where E is some amplitude of an external field and Φ is antisymmetric, $\Phi(\pi\eta, \pi\sigma) = -\Phi(\sigma, \eta)$, see e.g., ref. 1. Note also that then, necessarily, the relative energies $H(U_x \sigma) - H(\sigma)$ are invariant under exchanging H with $H\pi$.

In boundary driven systems, the process becomes non-translation invariant and the rates remain of the form (2.15) in the bulk (that is for x well inside the considered finite volume) while more or less arbitrary on the boundary. This was the case for Example C in Section 1.1 and was formalized for Proposition 2. Note that there is in fact an example of a boundary driven system where uniformly in the size of the system a bulk current can be maintained. This is the nonequilibrium harmonic crystal treated in refs. 9 and 10 where the heat flux is proportional to the boundary temperature difference rather than to the temperature gradient (infinite heat conductivity in the thermodynamic limit). Such “superconductors” do not exist in the context of interacting particle systems as discussed in the present paper.

3.3. Entropy Production

In phenomenological thermodynamics, entropy production appears in open driven systems as the product of thermodynamic fluxes and forces. The forces are gradients of intensive quantities (like chemical potential) generating the currents. The entropy production is identified from a balance equation for the time-derivative of an entropy density which is defined in systems close to equilibrium. The definition of entropy production as we use it here in statistical mechanics comes from refs. 2–5, 11–13 and we refer to the review.⁽⁶⁾ The mean entropy production appears there and in (1.4)–(1.5) as a relative entropy (density) for the process with respect to its time-reversal. That immediately invites the following thought (we are grateful to Senya Shlosman for pointing to this): In equilibrium statistical mechanics, if two translation invariant Gibbs measures have zero relative entropy density, then they must both be Gibbsian for the same interaction potential (but not

necessarily equal, e.g., because of spontaneous symmetry breaking). Apply this to the space-time measures obtained for the process \mathbb{P}_ρ and the time-reversed process $\mathbb{P}_\rho\Theta$ as introduced for (1.4). Here we take $\pi = \text{identity}$ to avoid extra complications. In some sense, both processes are Gibbs measures. Thus, if the mean entropy production is zero, then the process itself and its time-reversal have the same (space-time) action functional. Because they also have the same marginals ρ , they must in fact coincide (hence no spontaneous time-reversal breaking). Hence, zero mean entropy production implies microscopic reversibility. While convincing on a superficial level, unfortunately the details of this argument are technically cumbersome and a direct sufficiently general proof along this line has not been found.

The only more recent paper that we know of concerning time-reversal symmetry and the relation with entropy production is ref. 14. The set-up there is however quite different from ours. Time-reversal symmetry is there associated with the anticommutation of an involution with the time evolution, what we have called dynamic reversibility in the above. In our discussions here, we deal with spatially extended stochastic dynamics and the breaking of microscopic reversibility.

4. PROOFS

Lemma 1. Under the conditions of Section 2.1, for a translation invariant stationary measure ν ,

$$\sum_{U_0 \in \mathcal{F}_0} \int d\nu(\sigma) c(U_0, \sigma) \log \frac{d\nu^{U_0}}{d\nu}(\sigma) = 0. \tag{4.18}$$

Proof. Let \mathcal{F}_A be the σ -field generated by $\sigma_x, x \in A$. Denote by ν_A , respectively $\nu_A^{U_0}$ the \mathcal{F}_A -restrictions of ν and ν^{U_0} . Then we have

$$\frac{d\nu_A^{U_0}}{d\nu_A} = \mathbb{E}_\nu \left[\frac{d\nu^{U_0}}{d\nu} \middle| \mathcal{F}_A \right].$$

Since $d\nu^{U_0}/d\nu \in L^1(d\nu)$ for all U_0 , we find using the martingale convergence theorem that

$$\lim_{A \uparrow \mathbb{Z}^d} \frac{d\nu_A^{U_0}}{d\nu_A} = \frac{d\nu^{U_0}}{d\nu}, \tag{4.19}$$

in $L^1(d\nu)$. Let $\tilde{\nu}$ be the product measure on Ω having as marginals the uniform measure on S . From stationarity applied to the local function $f_A = d\nu_A/d\tilde{\nu}_A$ we find

$$\begin{aligned}
0 &= \sum_{x \in A'} \sum_{U_x \in \mathcal{P}_x} \int d\nu(\sigma) c(U_x, \sigma) \left[\log \frac{d\nu_{A'}^{U_x}}{d\tilde{\nu}_A} - \log \frac{d\nu_A}{d\tilde{\nu}_A} \right] \\
&= \sum_{x \in A'} \sum_{U_x \in \mathcal{P}_x} \int d\nu(\sigma) c(U_x, \sigma) \log \frac{d\nu_{A'}^{U_x}}{d\nu_A} \\
&= \sum_{x \in A'} \sum_{U_x \in \mathcal{P}_x} \int d\nu(\sigma) c(U_x, \sigma) \log \frac{d\nu^{U_x}}{d\nu} \\
&\quad + \sum_{x \in A'} \sum_{U_x \in \mathcal{P}_x} \int d\nu(\sigma) c(U_x, \sigma) \left[\log \frac{d\nu_{A'}^{U_x}}{d\nu_A} - \log \frac{d\nu^{U_x}}{d\nu} \right] \\
&= |A'| \sum_{U_0 \in \mathcal{P}_0} \int d\nu(\sigma) c(U_0, \sigma) \log \frac{d\nu^{U_0}}{d\nu}(\sigma) \\
&\quad + \sum_{x \in A'} \sum_{U_x \in \mathcal{P}_x} \int d\nu(\sigma) c(U_x, \sigma) F_A^{U_x}(\sigma).
\end{aligned}$$

The last equality uses translation invariance. We have used the notation $A' \equiv \{x \in \mathbb{Z}^d \mid V_x \cap A \neq \emptyset\}$ and the expression

$$F_A^{U_x}(\sigma) \equiv \left(\log \frac{d\nu_{A'}^{U_x}}{d\nu_A} - \log \frac{d\nu^{U_x}}{d\nu} \right).$$

We thus have

$$\begin{aligned}
&\left| \sum_{U_0 \in \mathcal{P}_0} \int d\nu(\sigma) c(U_0, \sigma) \log \frac{d\nu^{U_0}}{d\nu}(\sigma) \right| \\
&\leq \frac{1}{|A'|} \sum_{x \in A'} \sum_{U_x \in \mathcal{P}_x} \left| \int d\nu(\sigma) c(U_x, \sigma) F_A^{U_x}(\sigma) \right| \\
&\leq M \frac{1}{|A'|} \sum_{x \in A'} \sum_{U_0 \in \mathcal{P}_0} \int d\nu |F_{A-x}^{U_0}|, \tag{4.20}
\end{aligned}$$

by the translation invariance of ν , and M bounds the rates. Now we use the general fact that if f_n converges to f in $L^1(d\nu)$ and both f_n, f are bounded from below by some constant $c > 0$, then $\log f_n$ converges to $\log f$ in $L^1(d\nu)$. This fact implies that for any given $\varepsilon > 0$, we can choose $A \subset \mathbb{Z}^d$ such that for all $A' \supset A$:

$$\max_{U_0 \in \mathcal{P}_0} \int d\nu |F_A^{U_0}| \leq \frac{\varepsilon}{2MN}, \quad \text{with } |\mathcal{P}_0| \equiv N.$$

Choose now $A \subset \mathbb{Z}^d$ so large that

$$\frac{|\{x \in A' : A+x \cap A^c \neq \emptyset\}|}{|A'|} \leq \frac{\varepsilon}{2MN \sup_{W, U_0} \|F_W^{U_0}\|_{L^1(dv)}}.$$

We then conclude that

$$\begin{aligned} & \left| \sum_{U_0 \in \mathcal{P}_0} \int dv(\sigma) c(U_0, \sigma) \log \frac{dv^{U_0}}{dv}(\sigma) \right| \\ & \leq \frac{1}{|A'|} \sum_{x \in A', A+x \subset A} \frac{\varepsilon}{2} + MN \frac{|\{x \in A' : A+x \cap A^c \neq \emptyset\}|}{|A'|} \sup_{W, U_0} \|F_W^{U_0}\|_{L^1(dv)} \\ & \leq \varepsilon. \quad \blacksquare \end{aligned} \tag{4.21}$$

Proof of Main Theorem. Define

$$\bar{c}(U_0, \sigma) \equiv \bar{c}(\pi, \rho; U_0, \sigma) \equiv c(\pi U_0^{-1} \pi, \pi U_0 \sigma) \frac{d\rho^{U_0}}{d\rho}(\sigma)$$

and substitute it in

$$\sum_{U_0 \in \mathcal{P}_0} \int d\rho(\sigma) [c(U_0, \sigma) - \bar{c}(U_0, \sigma)] \log \frac{c(U_0, \sigma)}{\bar{c}(U_0, \sigma)}. \tag{4.22}$$

We get four terms, (4.22) =

$$\begin{aligned} & \sum_{U_0 \in \mathcal{P}_0} \left[\int d\rho(\sigma) \log \frac{c(U_0, \sigma)}{c(\pi U_0^{-1} \pi, \pi U_0 \sigma)} + \int d\rho(\sigma) c(U_0, \sigma) \log \frac{d\rho}{d\rho^{U_0}} \right. \\ & \quad + \int d\rho^{U_0}(\sigma) c(\pi U_0^{-1} \pi, \pi U_0 \sigma) \log \frac{c(\pi U_0^{-1} \pi, \pi U_0 \sigma)}{c(U_0, \sigma)} \\ & \quad \left. + \int d\rho^{U_0}(\sigma) c(\pi U_0^{-1} \pi, \pi U_0 \sigma) \log \frac{d\rho^{U_0}}{d\rho} \right]. \end{aligned} \tag{4.23}$$

The second term is zero by Lemma 1. The fourth term is also zero because, using condition (iii), we can change $\pi U_0^{-1} \pi \rightarrow U_0$ in the sum over \mathcal{P}_0 getting it equal to

$$\sum_{U_0 \in \mathcal{P}_0} \int d\rho \pi(\sigma) c(U_0, \sigma) \log \frac{d\rho \pi}{d(\rho \pi)^{U_0}}$$

which is zero, again by Lemma 1 applied to the stationary measure $\rho\pi$. Again using (iii), we can also rewrite the third term as

$$\sum_{U_0 \in \mathcal{P}_0} \int d\rho\pi(\sigma) \log \frac{c(U_0, \sigma)}{c(\pi U_0^{-1}\pi, \pi U_0\sigma)}.$$

Therefore, what remains of (4.23) is the sum of the first and the third term so that (4.22) equals

$$\sum_{U_0 \in \mathcal{P}_0} \left[\int d\rho(\sigma) \log \frac{c(U_0, \sigma)}{c(\pi U_0^{-1}\pi, \pi U_0\sigma)} + \int d\rho\pi(\sigma) \log \frac{c(U_0, \sigma)}{c(\pi U_0^{-1}\pi, \pi U_0\sigma)} \right].$$

We now recall that the mean entropy production (2.13) equals

$$\begin{aligned} \text{MEP}_\pi(L, \rho) = \sum_{U_0 \in \mathcal{P}_0} & \left(\int d\rho(\sigma) c(U_0, \sigma) \log \frac{c(U_0, \sigma)}{c(\pi U_0^{-1}\pi, \pi U_0\sigma)} \right. \\ & \left. + \int d\rho(\sigma) [c(U_0, \pi\sigma) - c(U_0, \sigma)] \right). \end{aligned} \quad (4.24)$$

This was derived from (2.12) in ref. 2. We conclude therefore that (4.22) equals $\text{MEP}_\pi(L, \rho) + \text{MEP}_\pi(L, \rho\pi)$ which is zero by hypothesis. This implies the statement of the Theorem. ■

Proof of Corollary 1. Since the Radon–Nikodym derivative of ρ^{U_0} with respect to ρ is a local function for all U_0 and since by assumption, we can generate with the U_0 all local excitations σ' from σ , it means that ρ has a continuous version for its local conditional distributions. ■

Proof of Corollary 2. From the main result and Corollary 1 it follows that ρ is a translation invariant stationary Gibbs measure and (2.15) must be satisfied. All other translation invariant stationary measures must be Gibbsian and for the same potential, see e.g., ref. 7. From the results in ref. 2 as cited above the statement of Corollary 2, it follows that every other stationary translation invariant measure must have zero mean entropy production. ■

Proof of Proposition 2. From the definition (1.5) we must first compute the relative action under time reversal, that is

$$R_n \equiv \log \frac{d\mathbb{P}_{\rho_n}}{d\tilde{\mathbb{P}}_{\rho_n}}$$

This can be done via a Girsanov formula and we obtain the analogue of (2.12):

$$R_n(\omega) = \sum_{x \in A_n^\#} \sum_{U_x \in \mathcal{P}_x} \int_{-T}^T \log \frac{c(U_x, \omega(s^-))}{c(U_x^{-1}, U_x \omega(s^-))} dN_s^{U_x}(\omega) + \sum_{\substack{A \subset A_n \setminus A_n^\# \\ \text{diam}(A) \leq r}} \sum_{\sigma \in S^A} \int_{-T}^T \log \frac{k_A^{(n)}(\omega(s^-), \omega(s^-)^{A, \sigma})}{k_A^{(n)}(\omega(s^-)^{A, \sigma}, \omega(s^-))} dN_s^{A, \sigma, n}(\omega).$$

The first integral is really a sum over all the times when the trajectory makes a jump from the action of one of the U_x ; the second integral is a sum over all times when a configuration σ is replacing $\omega(s^-)$ in a set A on the boundary. In order to further clarify this formula, let us first look at trajectories where no boundary transitions take place (or, what amounts to the same, take $k \equiv 0$ for the moment). Then, we only keep the first term, that is just (2.12), in case $\pi = id$:

$$\sum_{x \in A_n^\#} \sum_{U_x \in \mathcal{P}_x} \int_{-T}^T \log \frac{c(U_x, \omega(s^-))}{c(U_x^{-1}, U_x \omega(s^-))} dN_s^{U_x}.$$

But if we insert the detailed balance condition (2.15), the above expression telescopes to

$$H(\omega(-T)) - H(\omega(T))$$

and the mean entropy production is zero by stationarity.

Turning to the general case we let $\{s_i\}_{i=1}^q$ be the set of times at which boundary transitions occur in the sets A_i , $i = 1, \dots, q$, for the trajectory ω . These are random but we fix them as $-T \leq s_1 < s_2 < \dots < s_q \leq T$. The important thing to realize now is that while the perfect telescoping of above is broken at each of these times, it can be restored by adding and subtracting. More precisely, we have

$$R_n(\omega) = H(\omega(-T)) - H(\omega(s_1^-)) + H(\omega(s_1)) - H(\omega(s_2^-)) + \dots + H(\omega(s_q)) - H(\omega(T)) + \log \frac{k_{A_1}^{(n)}(\omega(s_1^-), \omega(s_1))}{k_{A_1}^{(n)}(\omega(s_1), \omega(s_1^-))} + \log \frac{k_{A_2}^{(n)}(\omega(s_2^-), \omega(s_2))}{k_{A_2}^{(n)}(\omega(s_2), \omega(s_2^-))} + \dots + \log \frac{k_{A_q}^{(n)}(\omega(s_q^-), \omega(s_q))}{k_{A_q}^{(n)}(\omega(s_q), \omega(s_q^-))}.$$

But by the absolute convergence of the interaction potential we have

$$|H(\omega(s_i^-) - H(\omega(s_i)))| \leq rC$$

for some constant C , since $\omega(s_i^-)$ and $\omega(s_i)$ only differ in the set A_i . Therefore the telescoping of the terms involving energy differences can be restored upon inserting q terms of order unity.

As for the other terms, we have assumed uniform boundedness so that we get

$$|R_n(\omega)| \leq q \left(rC + \log \frac{M}{\epsilon} \right)$$

where M and ϵ are constant upper and lower bounds for the transition rates $k^{(n)}$. As the expectation of $q = q(\omega)$ under \mathbb{E}_{ρ_n} is proportional to $T |\partial \mathcal{A}_n|$, the proposition is proved. ■

ACKNOWLEDGMENTS

C.M. thanks Senya Shlosman for some very useful discussions.

REFERENCES

1. J. L. Lebowitz and H. Spohn, A Gallavotti–Cohen type symmetry in the large deviation functional for stochastic dynamics, *J. Stat. Phys.* **95**:333–365 (1999).
2. C. Maes, F. Redig, and M. Verschuere, Entropy production for interacting particle systems, *Markov Proc. Rel. Fields* **7**:119–134 (2001).
3. C. Maes, The fluctuation theorem as a Gibbs property, *J. Stat. Phys.* **95**:367–392 (1999).
4. C. Maes, F. Redig, and A. Van Moffaert, On the definition of entropy production via examples, *J. Math. Phys.* **41**:1528–1554 (2000).
5. C. Maes and F. Redig, Positivity of entropy production, *J. Statist. Phys.* **101**:3–16 (2000).
6. C. Maes, *Statistical Mechanics of Entropy Production: Gibbsian Hypothesis and Local Fluctuations*, preprint from cond-mat/0106464.
7. H. Künsch, Non reversible stationary measures for infinite interacting particle systems, *Z. Wahrsch. Verw. Gebiete* **66**:407 (1984).
8. T. M. Liggett, *Interacting Particle Systems* (Springer-Verlag, New York, Heidelberg, Berlin, 1985).
9. Z. Rieder, J. L. Lebowitz, and E. Lieb, Properties of a harmonic crystal in a stationary nonequilibrium state, *J. Math. Phys.* **8**:1073–1078 (1967).
10. H. Nakazawa, On the lattice thermal conduction, *Suppl. Prog. Theor. Phys.* **45**:231–262 (1970).
11. M. P. Qian, M. Qian, and C. Qian, Circulations of markov chains with continuous time and probability interpretation of some determinants, *Sci. Sinica* **27**:470–481 (1984).

12. M. P. Qian and M. Qian, *The Entropy Production and Reversibility of Markov Processes*, Proceedings of the first world congress Bernoulli soc. (1988), pp. 307–316.
13. J. Schnakenberg, Network theory of behavior of master equation systems, *Rev. Mod. Phys.* **48**(4):571–585 (1976).
14. G. Gallavotti, Breakdown and regeneration of time reversal symmetry in nonequilibrium Statistical Mechanics, *Physica D* **112**:250–257 (1998).