Entropy Production in Non-Linear, Thermally Driven Hamiltonian Systems

Jean-Pierre Eckmann[1,2], *Claude-Alain Pillet*[3,4], *Luc Rey-Bellet*[1,*]
[1]Section de Mathématiques, Université de Genève, CH-1211 Genève 4, Switzerland
[2]Dept. de Physique Théorique, Université de Genève, CH-1211 Genève 4, Switzerland
[3]PHYMAT, Université de Toulon, F-83957 La Garde Cedex, France
[4]CPT-CNRS Luminy, F-13288 Marseille Cedex 09, France

Abstract. We consider a finite chain of non-linear oscillators coupled at its ends to two infinite heat baths which are at different temperatures. Using our earlier results about the existence of a stationary state, we show rigorously that for arbitrary temperature differences and arbitrary couplings, such a system has a unique stationary state. (This extends our earlier results for small temperature differences.) In all these cases, any initial state will converge (at an unknown rate) to the stationary state. We show that this stationary state continually produces entropy. The rate of entropy production is strictly negative when the temperatures are unequal and is proportional to the mean energy flux through the system.

1. Introduction

In a recent paper, [EPR], we have studied the existence of a stationary regime in a non-linear non-equilibrium setup. The model considered was that of a chain of n non-linear oscillators coupled at each of its two ends to heat baths which are infinite systems at two different temperatures. Under suitable conditions which we sketch below, it has been shown that for sufficiently small temperature differences between the baths, the complete system has a *unique* stationary state, and that every initial state converges to it. Of course, this stationary state is not an equilibrium state but a steady state in which supposedly heat flows (on average) from the hot bath to the cold one. The aim of this paper is to show first that this result extends to arbitrary temperature differences and that the heat flux across the chain is positive.

It should be noted that this is *not* a perturbative result. To prove the existence and uniqueness of the steady state for non-linear, boundary driven problems with arbitrary temperature difference is a difficult problem. See [GLP] and [GKI] for similar results for a gas of particles in a box with thermostatting boundary conditions. For other boundary driven models, see [FGS] (a 1-dimensional hard-core gas).

[*] Current address: Dept. of Mathematics and Physics, Rutgers University, Hill Center, Rutgers University, Piscataway NJ 08903, USA

Our model, whose study was started in [EPR], combines several desirable features, while still allowing for a rather complete set of rigorous results. The main features are the property of being fully Hamiltonian (as those studied in [FGS] and [SL]), with non-linear interactions, and a realistic implementation of the retro-action of the chain on the heat baths. In particular, the system is self-regulating and we do not need any Gaussian thermostats [PH, EM, H, CELS, GC1, GC2, G].

The reader should note that in our model the energy of the chain fluctuates wildly in time and there is no external dissipation term which prevents the energy of the chain from diverging to infinity. The baths can exchange energy with the chain. Also, since the potentials are not monotone, several stationary non-equilibrium states could possibly exist, each corresponding for example to one of the extrema of the potential. We show here that on the contrary, there is exactly one stationary state, no matter how large the temperature difference of the baths is.

Once the uniqueness of the steady state is established, we show that, away from equilibrium, there is a stationary, strictly positive heat flow through the chain and the (thermodynamic) entropy production is strictly negative. We also discuss briefly (heuristically) a suitable version of the Gallavotti-Cohen fluctuation relation [ECM, GC1, GC2, G, K, LS] for the entropy production in the context of our model.

2. Set-up and Notations

To make this paper accessible without the necessity of referring back to [EPR], we introduce again the model. It deals with an anharmonic chain driven at its ends by two heat baths.

The chain consists of n particles moving in $\mathbf{R}^d$, with n arbitrary but finite, and its dynamics is described by the following Hamiltonian:

$$H_{\mathrm{S}}(q_1,\dots,q_n,p_1,\dots,p_n) \;=\; \sum_{j=1}^{n}\frac{p_j^2}{2}+V(q_1,\dots,q_n)\;,$$

where the potential V is of the form:[1]

$$V(q) \;=\; \sum_{j=1}^{n}U_j^{(1)}(q_j)+\sum_{i=1}^{n-1}U_i^{(2)}(q_i-q_{i+1})\;.$$

We make the following assumptions on the potential V:

H1) Behavior at infinity: We assume that V is of the form

$$V(q) \;=\; \tfrac{1}{2}\big(q-a,\mathbf{Q}(q-a)\big)+F(q)\;,$$

where $\mathbf{Q}$ is a positive definite $(dn\times dn)$ matrix, a is a vector, and $\partial_{q_i^{(\nu)}}F\in\mathcal{F}$ for $i=1,\dots,n$ and $\nu=1,\dots,d$. Here, $\mathcal{F}$ denotes the space of those $\mathcal{C}^\infty$ functions F on $\mathbf{R}^{dn}$ for which $\partial^\alpha F(q)$ is bounded uniformly in $q\in\mathbf{R}^{dn}$, for all multi-indices α.

[1] The two-body potential is slightly more restrictive than in [EPR], since we only take functions of the coordinate differences.

H2) Coupling: Each of the $(d \times d)$ matrices

$$\mathcal{M}_i(x) \equiv \mathrm{D}^2 U_i^{(2)}(x) , \quad i = 1, \ldots, n-1 ,$$

of second derivatives, is either uniformly positive or negative definite for $x \in \mathbf{R}^d$.

Each heat bath is modeled by an infinite dimensional linear Hamiltonian system, which is a scalar field whose dynamics is governed by a d-dimensional wave equation:

$$H_{\mathrm{B}}(\phi_i, \pi_i) = \frac{1}{2} \int \mathrm{d}x \left(|\nabla \phi_i|^2 + |\pi_i|^2 \right) . \tag{2.1}$$

We will denote the heat baths by the subscripts L and R respectively. We couple the L heat bath to the first particle of the chain and the R heat bath to the n-th particle of the chain. We choose a coupling which is linear both in the field variables and in the particle variables. The total Hamiltonian of the system is then given by

$$\begin{aligned} H(q, p, \phi_{\mathrm{L}}, \pi_{\mathrm{L}}, \phi_{\mathrm{R}}, \pi_{\mathrm{R}}) = & \sum_{i \in \{\mathrm{L},\mathrm{R}\}} H_{\mathrm{B}}(\phi_i, \pi_i) + H_{\mathrm{S}}(q, p) \\ & + q_1 \cdot \int \mathrm{d}x \nabla \phi_{\mathrm{L}}(x) \rho_{\mathrm{L}}(x) + q_n \cdot \int \mathrm{d}x \nabla \phi_{\mathrm{R}}(x) \rho_{\mathrm{R}}(x) . \end{aligned} \tag{2.2}$$

We consider the heat baths at positive temperatures T_{L} and T_{R} respectively, *i.e.*, we will assume that the initial conditions of the heat baths are distributed according to the Gaussian measure with mean zero and covariance $(\cdot, \cdot)_i T_i$, where $(\cdot, \cdot)_i$ is the scalar product defined by the quadratic form (2.1), $i \in \{\mathrm{L}, \mathrm{R}\}$.

The following reduction to (essentially only) the variables of the small system is explained in detail in [EPR]: We integrate out the variables of the baths and project the dynamics on the variables of the chain. This leads to integro-differential stochastic equations. Under suitable assumptions on the coupling functions ρ_{L}, ρ_{R}, they can be expressed as Markovian equations upon introducing auxiliary variables $r_{i,m}$ with $i \in \{\mathrm{L}, \mathrm{R}\}$, and $m = 1, \ldots, M$. At the end, one obtains (see [EPR] for details) the following system of stochastic differential equations:

$$\begin{aligned} \mathrm{d}q_j(t) &= p_j(t)\mathrm{d}t , \qquad j = 1, \ldots, n , \\ \mathrm{d}p_1(t) &= -\nabla_{q_1} V(q(t))\mathrm{d}t + \sum_{m=1}^{M} r_{\mathrm{L},m}(t)\mathrm{d}t , \\ \mathrm{d}p_j(t) &= -\nabla_{q_j} V(q(t))\mathrm{d}t , \qquad j = 2, \ldots, n-1 , \\ \mathrm{d}p_n(t) &= -\nabla_{q_n} V(q(t))\mathrm{d}t + \sum_{m=1}^{M} r_{\mathrm{R},m}(t)\mathrm{d}t , \\ \mathrm{d}r_{\mathrm{L},m}(t) &= -\gamma_{\mathrm{L},m} r_{\mathrm{L},m}(t)\mathrm{d}t + \lambda_{\mathrm{L},m}^2 \gamma_{\mathrm{L},m} q_1(t)\mathrm{d}t - \lambda_{\mathrm{L},m} \sqrt{2\gamma_{\mathrm{L},m} T_{\mathrm{L}}}\, \mathrm{d}w_{\mathrm{L},m}(t) , \\ \mathrm{d}r_{\mathrm{R},m}(t) &= -\gamma_{\mathrm{R},m} r_{\mathrm{R},m}(t)\mathrm{d}t + \lambda_{\mathrm{R},m}^2 \gamma_{\mathrm{R},m} q_n(t)\mathrm{d}t - \lambda_{\mathrm{R},m} \sqrt{2\gamma_{\mathrm{R},m} T_{\mathrm{R}}}\, \mathrm{d}w_{\mathrm{R},m}(t) , \\ & \qquad m = 1, \ldots, M , \end{aligned} \tag{2.3}$$

which defines a Markov diffusion process on $\mathbf{R}^{2d(n+M)}$. Each $w_{\mathrm{L},m}$ and $w_{\mathrm{R},m}$ is a standard d-dimensional Brownian motion.

Remark 2.1. In Eqs.(2.3) the variables $r_{\mathrm{L},m}$, $r_{\mathrm{R},m}$ describe both the (random) forces exerted by the heat bath on the chain and the (dissipative) forces due to the retroaction of the heat baths on the chain. The fact that the variables $r_{\mathrm{L},m}$, $r_{\mathrm{R},m}$ obey Markovian stochastic differential equations is a consequence of our choice of coupling functions ρ_{L}, ρ_{R}. In fact, these functions are chosen in such a way that the random forces exerted by the heat baths are not white noises but have covariances which are (sums of) exponentials. Together with the fluctuation theorem relating these random forces with the dissipative forces, one obtains Markovian differential equations on the phase space consisting of the physical variables p, q, augmented by the auxiliary variables $r_{\mathrm{L},m}$, $r_{\mathrm{R},m}$.

Remark 2.2. If the temperatures of the two heat baths are the same, *i.e.*, if $T_{\mathrm{L}} = T_{\mathrm{R}}$, the stationary state of the Markov process which solves (2.3) can be written explicitly. It is given by the generalized Gibbs state

$$\mu(\mathrm{d}r, \mathrm{d}q, \mathrm{d}p) \;=\; \mu_{T_{\mathrm{L}},T_{\mathrm{L}}}(\mathrm{d}r, \mathrm{d}q, \mathrm{d}p) \;=\; Z^{-1} e^{-G^{(0)}(r,q,p)/T_{\mathrm{L}}} \mathrm{d}r\, \mathrm{d}q\, \mathrm{d}p\,, \tag{2.4}$$

where the "energy" $G^{(0)}$ is given by

$$G^{(0)}(r,q,p) \;=\; H_{\mathrm{S}}(q,p) + \sum_{m=1}^{M} \Big(\frac{r_{\mathrm{L},m}^2}{2\lambda_{\mathrm{L},m}^2} + \frac{r_{\mathrm{R},m}^2}{2\lambda_{\mathrm{R},m}^2} - q_1 \cdot r_{\mathrm{L},m} - q_n \cdot r_{\mathrm{R},m} \Big)\,. \tag{2.5}$$

The marginal of this measure on the physical phase space is given by

$$\nu(\mathrm{d}q, \mathrm{d}p) \;=\; \int \mu(\mathrm{d}r, \mathrm{d}q, \mathrm{d}p) \;=\; \frac{1}{Z'} e^{-H_{\mathrm{eff}}/T_{\mathrm{L}}} \mathrm{d}q\, \mathrm{d}p\,,$$

where the effective Hamiltonian H_{eff} is given by

$$H_{\mathrm{eff}}(q,p) \;=\; H_{\mathrm{S}}(q,p) - \tfrac{1}{2} q_1^2 \sum_{m=1}^{M} \lambda_{\mathrm{L},m}^2 - \tfrac{1}{2} q_n^2 \sum_{m=1}^{M} \lambda_{\mathrm{R},m}^2 \;\equiv\; \tfrac{1}{2} p^2 + V_{\mathrm{eff}}(q)\,. \tag{2.6}$$

It can be seen from (2.6) that the coupling between the chain and the heat baths induces a renormalization of the potential $V(q)$. In particular, because of Condition **H1**, if the coupling constants λ_{im} are too large, $V_{\mathrm{eff}}(q)$ is not confining any more, and the measure μ is not a probability measure, but only σ-finite. In the sequel we require the following:
H3) The coupling constants λ_{im}, $i \in \{\mathrm{L}, \mathrm{R}\}$, $m \in \{1, \ldots, M\}$ are such that

$$\lim_{|q| \to \infty} V_{\mathrm{eff}}(q) \;=\; \infty\,.$$

3. Uniqueness of the Stationary State

In [EPR] we proved, under the conditions **H1–H3**, the existence of an invariant measure for any temperature T_{L}, T_{R}, but the uniqueness was shown only for small temperature differences. In this section, we extend the uniqueness result to *arbitrary* temperature differences.

The uniqueness will follow from a dynamical argument: we will show that the Markov process is transitive. This is done using a (well-known) relationship between stochastic differential equations and control theory (see *e.g.*, [Ku] and references therein).

We explain the method for a general stochastic differential equation of the form

$$\mathrm{d}x(t) \,=\, b(x(t))\mathrm{d}t + \sigma dw(t) \;, \tag{3.1}$$

where $x \in \mathbf{R}^k$, $b(x)$ is a $\mathcal{C}^\infty$ vector field, $w(t)$ is a standard ℓ-dimensional Brownian motion, and σ is a $k \times \ell$ matrix. We assume that the vector field $b(x)$ is such that (3.1) has a unique solution for all $t > 0$. One then replaces $dw(t)$ in (3.1) by $u(t)\mathrm{d}t$. The function $u(t) = (u_1(t), \ldots, u_\ell(t))$ is called a control. One obtains the system of ordinary differential equations

$$\dot{x} \,=\, b(x(t)) + \sigma u(t) \;. \tag{3.2}$$

The correspondence between the two systems is established by the following result of Stroock and Varadhan [SV]. We fix an arbitrary time $\tau > 0$. Let $\mathcal{U}$ denote the set of piecewise constant functions $u : [0, \tau] \to \mathbf{R}^\ell$. Let $\mathcal{W}$ be the set of all continuous functions φ from $[0, \tau]$ to $\mathbf{R}^k$ equipped with the uniform topology and let $\mathcal{W}_x = \{\varphi \in \mathcal{W} \;:\; \varphi(0) = x\}$. We denote ξ_x the diffusion process defined by (3.1) with initial condition $\xi_x(0) = x$. Then the path ξ_x belongs almost surely to $\mathcal{W}_x$. The support of this diffusion process ξ_x on $[0, \tau]$ is, by definition, the smallest closed subset $\mathcal{S}_x$ of $\mathcal{W}_x$ such that

$$\mathbf{P}[\xi_x \in \mathcal{S}_x] \,=\, 1 \;, \tag{3.3}$$

where $\mathbf{P}$ is the probability induced by the Brownian motion w. We denote by $\varphi_x^u : [0, \tau] \to \mathbf{R}^k$ the solution of the differential equations (3.2) with control u and initial condition x. We next consider the notion of accessibility. Let x and y be two points in $\mathbf{R}^k$. The point y is called *accessible* from x at time τ ($\tau > 0$) if there is a control u such that $\varphi_x^u(\tau) \,=\, y$. The set of all points which are accessible from x at time τ is denoted $Y_\tau(x)$.

Theorem 3.1. *[SV]. One has*

$$\mathcal{S}_x \,=\, \overline{\{\varphi_x^u \;:\; u \in \mathcal{U}\,\}} \;, \tag{3.4}$$

for all $x \in \mathbf{R}^k$, and

$$\operatorname{supp} P(\tau, x, \cdot) \,=\, \overline{Y_\tau(x)} \;,$$

for all $x \in \mathbf{R}^k$ and $\tau > 0$, where $P(\tau, x, \mathrm{d}y)$ denotes the transition probability of the process ξ_x.

Remark. The first statement is explicit in [SV] and the second is a straightforward consequence of the first.

The main technical result of this section is the following

Theorem 3.2. *The control system associated with the stochastic differential equation (2.3) is strongly completely controllable, i.e.,* $\overline{Y_\tau(x)} = \mathbf{R}^{2d(n+M)}$*, for all* $x = (q, p, r_{\rm L}, r_{\rm R})$ *and all* $\tau > 0$.

Remark 3.3. We will combine this result with Theorem 3.1 and hypoellipticity to show that the invariant measure has a smooth, strictly positive density.

Remark 3.4. It should be noted that strong complete controllability (SCC) can not be deduced from Hörmander's hypoellipticity condition alone. (See *e.g.*, [IK] for examples of hypoelliptic diffusions with two invariant measures). Therefore, Theorem 3.2 contains additional information. Various sufficient conditions for SCC have been expressed in terms of differential geometry in [Ku], but these are *not* applicable to Eqs.(2.3).

Proof of Theorem 3.2. We will show the strong complete controllability of the control problem associated with (2.3) by an explicit approach using the requirement of effective coupling (condition **H2**) of the chain.

We reconsider the stochastic differential equation (2.3). Following the procedure described above we replace the Brownian motions $w_{im}(t)$ by controls $u_{im}(t)$ in (2.3), and rewrite the system thus obtained as a system of second (and first) order equations. This leads to:

$$\begin{aligned}
\dot r_{{\rm L},m} &= -\gamma_{{\rm L},m} r_{{\rm L},m} + \lambda_{{\rm L},m}^2 \gamma_{{\rm L},m} q_1 + u_{{\rm L},m}\ , \quad m = 1, \dots, M\ , && (3.5)\\
\ddot q_1 &= -\nabla_{q_1} V(q) + \sum_{m=1}^{M} r_{{\rm L},m}\ , && (3.6)\\
\ddot q_j &= -\nabla_{q_j} V(q)\ , \qquad j = 2, \dots, n-1\ , && (3.7)\\
\ddot q_n &= -\nabla_{q_n} V(q) + \sum_{m=1}^{M} r_{{\rm R},m}\ , && (3.8)\\
\dot r_{{\rm R},m} &= -\gamma_{{\rm R},m} r_{{\rm R},m} + \lambda_{{\rm R},m}^2 \gamma_{{\rm R},m} q_n + u_{{\rm R},m}\ , \quad m = 1, \dots, M\ . && (3.9)
\end{aligned}$$

Here, we have absorbed the constants in front of the Brownian motion in (2.3) into the controls $u_{im}(t)$. We will only consider controls u of class $\mathcal{C}^\infty$. Any such function can be uniformly approximated, on any compact interval, by a piecewise constant function. Since a simple Gronwall estimate of Eq.(3.2) shows that

$$\sup_{t \in [0,\tau]} |\varphi_x^u(t) - \varphi_x^v(t)| \le C(\tau) \sup_{t \in [0,\tau]} |u(t) - v(t)|$$

holds with a constant $C(\tau)$ depending on the model, but not on u and v, we conclude that

$$\left\{\varphi_x^u(\tau) : u \in \mathcal{C}^\infty(\mathbf{R})\right\} \subset \overline{Y_\tau(x)}\ .$$

The proof of Theorem 3.2 will now be done in two parts:

Part 1: Boundary control of the chain. We start by considering the auxiliary problem of controlling a chain of $n+2$ oscillators by the motion of the two ends of the chain.

The differential equation reads

$$\ddot{q}_j = f_j(q_{j-1}, q_j, q_{j+1}) , \quad j = 1, \dots, n , \tag{3.10}$$

where $q \equiv (q_1, \dots, q_n)$ is the dynamical variable, whereas $q_0 \equiv u_L$ and $q_{n+1} \equiv u_R$ are the control variables. The smooth functions f_j are given by

$$f_j(x, y, z) \equiv -(\nabla U_j^{(1)})(y) - (\nabla U_j^{(2)})(y - z) + (\nabla U_{j-1}^{(2)})(x - y) .$$

Here, we define $U_0^{(2)}(x) = U_n^{(2)}(x) = x^2/2$.

Note that by Condition $\mathbf{H2}$, the functions $\nabla U_j^{(2)}$ are diffeomorphisms. It follows that the equation $w = f_j(x, y, z)$ can be solved for z : There exist smooth functions g_j such that $w = f_j\big(x, y, g_j(x, y, w)\big)$ for all $x, y, w \in \mathbf{R}^d$. Consequently the differential equation (3.10) is equivalent to the equation

$$q_{j+1} = g_j(q_{j-1}, q_j, \ddot{q}_j) , \quad j = 1, \dots, n . \tag{3.11}$$

Obviously, for given q_0 and q_1, this equation can be solved inductively, and in a unique way. To express this solution, let us introduce some notation. For a smooth function φ and an integer α, we shall denote the collection of the first α derivatives of φ by $\varphi^{[\alpha]} \equiv (\varphi, \dot{\varphi}, \dots, \varphi^{(\alpha)})$. We also set $p_j \equiv \dot{q}_j$ for $j = 1, \dots, n$. For $q_0 = \xi$ and $q_1 = \eta$, the inductive solution of Eq.(3.11) reads

$$\begin{array}{ccccc}
u_L & = & \xi & \equiv & G_0(\xi^{[0]}) , \\
q_1 & = & \eta & \equiv & G_1(\eta^{[0]}) , \\
p_1 & = & \dot{q}_1 & \equiv & G_2(\eta^{[1]}) , \\
q_2 & = & g_1(q_0, q_1, \dot{p}_1) & \equiv & G_3(\xi^{[0]}, \eta^{[2]}) , \\
p_2 & = & \dot{q}_2 & \equiv & G_4(\xi^{[1]}, \eta^{[3]}) , \\
q_3 & = & g_2(q_1, q_2, \dot{p}_2) & \equiv & G_5(\xi^{[2]}, \eta^{[4]}) , \\
 & \vdots & & \vdots & \\
u_R & = & g_n(q_{n-1}, q_n, \dot{p}_n) & \equiv & G_{2n+1}(\xi^{[2n-2]}, \eta^{[2n]}) .
\end{array}$$

We can organize the $2n+2$ maps G_J into a map $G\colon \mathbf{R}^{4nd} \to \mathbf{R}^{4nd}$ in the following way: Denote by (a, b) a point of $\mathbf{R}^{4nd}$, with $a \equiv (a_0, \dots, a_{2n-2}) \in \mathbf{R}^{(2n-1)d}$ and $b \equiv (b_0, \dots, b_{2n}) \in \mathbf{R}^{(2n+1)d}$. With $a^{[\alpha]} \equiv (a_0, \dots, a_\alpha)$ and $b^{[\alpha]} \equiv (b_0, \dots, b_\alpha)$, define $G(a, b) \equiv \big(a, \hat{G}(a, b)\big)$ where

$$\hat{G}(a, b) \equiv \big(G_1(b^{[0]}), G_2(b^{[1]}), G_3(a^{[0]}, b^{[2]}), \dots, G_{2n+1}(a^{[2n-2]}, b^{[2n]})\big) .$$

We have proved that, if $(u_L, q_1, \dots, q_n, u_R)$ is a solution of Eq.(3.10) on the time interval $I \subset \mathbf{R}$, then

$$(u_L^{[2n-2]}, q_1, \dot{q}_1, \dots, q_n, \dot{q}_n, u_R) = G(u_L^{[2n-2]}, q_1^{[2n]}) \tag{3.12}$$

holds on I. A simple consequence of this fact is that G is a bijection. Indeed, repeated differentiation of Eq.(3.10) gives

$$\begin{array}{ccccc} q_1^{(2)} & = & f_1(u_L, q_1, q_2) & = & F_2(u_L^{[0]}, q_1, q_2) \, , \\ q_1^{(3)} & = & \partial_t q_1^{(2)} & = & F_3(u_L^{[1]}, q_1, \dot{q}_1, q_2, \dot{q}_2) \, , \\ & \vdots & & \vdots & \\ q_1^{(2\alpha)} & = & \partial_t q_1^{(2\alpha-1)} & = & F_{2\alpha}(u_L^{[2\alpha-2]}, q_1, \dot{q}_1, \dots, q_{\alpha+1}) \, , \\ q_1^{(2\alpha+1)} & = & \partial_t q_1^{(2\alpha)} & = & F_{2\alpha+1}(u_L^{[2\alpha-1]}, q_1, \dot{q}_1, \dots, q_{\alpha+1}, \dot{q}_{\alpha+1}) \, , \\ & \vdots & & \vdots & \\ q_1^{(2n)} & = & \partial_t q_1^{(2n-1)} & = & F_{2n}(u_L^{[2n-2]}, q_1, \dot{q}_1, \dots, u_R) \, , \end{array}$$

and thus we find another functional relation $q_1^{[2n]} = \hat{F}(u_L^{[2n-2]}, q_1, \dot{q}_1, \dots, q_n, \dot{q}_n, u_R)$ for its solutions. It immediately follows that the map $F\colon (a,b) \mapsto \big(a, \hat{F}(a,b)\big)$ satisfies

$$\begin{aligned} F \circ G(u_L^{[2n-2]}, q_1^{[2n]}) &= (u_L^{[2n-2]}, q_1^{[2n]}) \, , \\ G \circ F(u_L^{[2n-2]}, q_1, \dot{q}_1, \dots, u_R) &= (u_L^{[2n-2]}, q_1, \dot{q}_1, \dots, u_R) \, , \end{aligned}$$

on every solution of Eq.(3.10). Since $u_L^{[2n-2]}(0)$ and either $q_1^{[2n]}(0)$ or $\big(q(0), \dot{q}(0), u_L(0)\big)$ can be prescribed arbitrarily, we conclude that $F = G^{-1}$.

To solve our control problem it suffices to remark that the set of solutions (u_L, q, u_R) satisfying $\big(u_L^{[2n-2]}(t_0), q_1(t_0), \dot{q}_1(t_0), \dots, u_R(t_0)\big) = (a,b)$ is identical with the set of solutions satisfying $\big(u_L^{[2n-2]}(t_0), q_1^{[2n]}(t_0)\big) = F(a,b)$. Since for $\tau > 0$ and arbitrary $(a,b), (a',b') \in \mathbf{R}^{4nd}$ one can find functions u_L and q_1 for which

$$\begin{aligned} \big(u_L^{[2n-2]}(0), q_1^{[2n]}(0)\big) &= F(a,b) \, , \\ \big(u_L^{[2n-2]}(\tau), q_1^{[2n]}(\tau)\big) &= F(a',b') \, , \end{aligned}$$

we see that the system (3.10) is strongly controllable.

Part 2: Completion of the proof of Theorem 3.2. We reduce the problem of Eqs.(3.5)–(3.9) to the case dealt with in Part 1, by introducing the auxiliary variables q_0 and q_{n+1}. Recalling the definition $U_0^{(2)}(x) = U_n^{(2)}(x) = x^2/2$, we can rewrite the control problem associated with our stochastic differential equation as

$$\begin{aligned} \ddot{q}_j &= f_j(q_{j-1}, q_j, q_{j+1}) \, , \quad j = 1, \dots, n \, , \\ \sum_{m=1}^{M} r_{L,m} &= q_1 - q_0 \, , \\ \sum_{m=1}^{M} r_{R,m} &= q_{n+1} - q_n \, , \\ u_{L,m} &= \dot{r}_{L,m} + \gamma_{L,m} r_{L,m} - \lambda_{L,m}^2 \gamma_{L,m} q_1 \, , \quad m = 1, \dots, M \, , \\ u_{R,m} &= \dot{r}_{R,m} + \gamma_{R,m} r_{R,m} - \lambda_{R,m}^2 \gamma_{R,m} q_n \, , \quad m = 1, \dots, M \, , \end{aligned} \tag{3.13}$$

with the boundary conditions

$$\begin{aligned}\big(r_L(0), q_1(0), \dot q_1(0), \ldots, q_n(0), \dot q_n(0), r_R(0)\big) &= x\,,\\ \big(r_L(\tau), q_1(\tau), \dot q_1(\tau), \ldots, q_n(\tau), \dot q_n(\tau), r_R(\tau)\big) &= y\,.\end{aligned}$$

The equation $\sum_{m=1}^{M} r_{L,m} = q_1 - q_0$ serves to compensate the term $q_1 - q_0$ produced when differentiating f_1 in (3.13). Given the boundary data for r_L, q_1, q_n and r_R we obtain boundary values for q_0 and q_{n+1}. We can thus control the first equation by our previous result. This gives us $q_0, \ldots, q_{n+1}$. Selecting arbitrary functions $r_{L,2}, \ldots, r_{L,M}$ and $r_{R,2}, \ldots, r_{R,M}$ satisfying the corresponding boundary data, we define

$$\begin{aligned}r_{L,1} &\equiv q_1 - q_0 - \sum_{m=2}^{M} r_{L,m}\,,\\ r_{R,1} &\equiv q_{n+1} - q_n - \sum_{m=2}^{M} r_{R,m}\,.\end{aligned}$$

These two functions will also satisfy the boundary conditions. Finally we use the last two sets of equations to determine the control variables $u_{L,m}$ and $u_{R,m}$. This concludes the proof of Theorem 3.2.

Remark 3.5. It is obvious from the proof of Theorem 3.2 that this theorem is valid under much weaker conditions than those given in **H1**. It is enough to require that the stochastic differential equation (2.3) has a unique solution for all $t > 0$. In particular we do not need to restrict ourselves to potentials which are "quadratic at infinity" as required in the proof of the existence of the invariant measure.

The main result of this section is:

Theorem 3.6. *If Conditions* **H1–H3** *are satisfied, the Markov process which solves (2.3) has a unique invariant measure $\mu = \mu_T$. The measure μ has a $\mathcal{C}^\infty$ density $\rho(r, q, p)$. This density is an exponentially decaying, strictly positive function of r, q, and p. The invariant measure is ergodic and mixing.*

Remark 3.7. In fact, combining this result with information from [EPR], one can show that

$$\rho(r, q, p) = f(r, q, p) \exp\big(-G^{(0)}(r, q, p)/T^*\big)\,,$$

where $G^{(0)}$ was defined in (2.5), and $T^* = \max(T_\mathrm{L}, T_\mathrm{R})$. The function f is in the Schwartz space $\mathcal{S}$ when $T_\mathrm{L} \neq T_\mathrm{R}$, (and is a constant otherwise).

Proof. The proof is a combination of Theorem 3.1 and Theorem 3.2 with results in [EPR]. The existence of the invariant measure μ is proven in [EPR, Theorem 2.1]. Furthermore, using hypoellipticity, we showed that the density ρ of μ is $\mathcal{C}^\infty$. Also, the transition probabilities $P(t, x, \mathrm{d}y)$ have a smooth density, p, defined by $P(t, x, \mathrm{d}y) = p(t, x, y)\mathrm{d}y$ with $p(t, x, y) \in \mathcal{C}^\infty((0, \infty), \mathbf{R}^{2d(n+M)}, \mathbf{R}^{2d(n+M)})$.

We next show that the support of μ is all of the extended phase space $X \equiv \mathbf{R}^{2d(n+M)}$. In Theorem 3.2, we have seen that (2.3) is strongly completely controllable. By Theorem 3.1, we conclude that the support of $P(\tau, x, \cdot)$ is the whole phase space for every $x \in X$ and all $\tau > 0$. Therefore, we have, for all $t > 0$, all x, and all open sets Y, the inequality

$$P(t, x, Y) \;>\; 0\;.$$

Since $\mu(Y) = \int \mu(\mathrm{d}x) P(t, x, Y)$ (because μ is invariant), we conclude that supp $\mu = X$ and thus the density ρ is Lebesgue almost everywhere positive.

We next show that $\rho(x) > 0$, for all x. by assuming the contrary and deriving a contradiction. Assume that there is a y for which $\rho(y) = 0$. By the invariance of the measure we have, for any $t > 0$,

$$0 \;=\; \rho(y) \;=\; \int \mathrm{d}x\, \rho(x)\, p(t, x, y)\;.$$

This implies $p(t, x, y) \;=\; 0$ for Lebesgue almost all x. Since the transition kernel p is smooth, we conclude that the function $p(t, \cdot, y)$ is identically zero every $t > 0$. On the other hand, since p is the kernel of a strongly continuous semigroup, we also have $p(t, x, y) \to \delta(x - y)$ as $t \to 0$. This is a contradiction, and we have shown $\rho(y) > 0$, for all $y \in X$.

We next show uniqueness. We have just shown that every invariant measure must have a smooth, strictly positive density. Since every ergodic component is mutually singular to any other, the invariant measure is unique (and ergodic). The property of mixing of the invariant measure has been deduced from uniqueness in the proof of [EPR, Theorem 3.9] . This concludes the proof of Theorem 3.6.

Remark 3.8. We proved in [EPR, Lemma 3.7] that the density $\rho = \rho_T$ is a real analytic function of $\zeta = (T_{\mathrm{L}} - T_{\mathrm{R}})/(T_{\mathrm{L}} + T_{\mathrm{R}})$. In particular, this yields the standard perturbative results near equilibrium ($\zeta = 0$).

4. Time-Reversal, Energy Flux, and Entropy Production

In this section, we ask questions which are intimately related to the Hamiltonian nature of our model. After introducing appropriate notation, we introduce time reversal, and draw some consequences. In particular, we are able to show that the system exhibits non-zero mean energy flux as soon as $T_{\mathrm{L}} \neq T_{\mathrm{R}}$, and we relate the flux to the entropy production.

4.1. Notation

It will be useful to streamline the notation. It is convenient to introduce first $\rho_{\mathrm{L},m} = (\lambda_{\mathrm{L},m}\gamma_{\mathrm{L},m}^{1/2})^{-1} r_{\mathrm{L},m}$ and similarly for the $r_{\mathrm{R},m}$. Then the equations of motion are

$$\begin{aligned}
\mathrm{d}q_j(t) &= p_j(t)\mathrm{d}t\ , \qquad\qquad j = 1,\dots,n\ ,\\
\mathrm{d}p_1(t) &= -\nabla_{q_1} V(q(t))\mathrm{d}t + \sum_{m=1}^{M} \lambda_{\mathrm{L},m}\gamma_{\mathrm{L},m}^{1/2}\rho_{\mathrm{L},m}(t)\mathrm{d}t\ ,\\
\mathrm{d}p_j(t) &= -\nabla_{q_j} V(q(t))\mathrm{d}t\ , \qquad j = 2,\dots,n-1\ ,\\
\mathrm{d}p_n(t) &= -\nabla_{q_n} V(q(t))\mathrm{d}t + \sum_{m=1}^{M} \lambda_{\mathrm{R},m}\gamma_{\mathrm{R},m}^{1/2}\rho_{\mathrm{R},m}(t)\mathrm{d}t\ ,\\
\mathrm{d}\rho_{\mathrm{L},m}(t) &= -\gamma_{\mathrm{L},m}\rho_{\mathrm{L},m}(t)\mathrm{d}t + \lambda_{\mathrm{L},m}\gamma_{\mathrm{L},m}^{1/2} q_1(t)\mathrm{d}t - \sqrt{2}T_{\mathrm{L}}^{1/2}\,\mathrm{d}w_{\mathrm{L},m}(t)\ ,\\
\mathrm{d}\rho_{\mathrm{R},m}(t) &= -\gamma_{\mathrm{R},m}\rho_{\mathrm{R},m}(t)\mathrm{d}t + \lambda_{\mathrm{R},m}\gamma_{\mathrm{R},m}^{1/2} q_n(t)\mathrm{d}t - \sqrt{2}T_{\mathrm{R}}^{1/2}\,\mathrm{d}w_{\mathrm{R},m}(t)\ ,\\
& \qquad m = 1,\dots,M\ .
\end{aligned} \tag{4.1}$$

We can write this system in vector notation: We write the Hamiltonian of the chain (the small system) as

$$H_{\mathrm{S}}(q,p) = \frac{p^2}{2} + V(q)\ ,$$

with $q, p \in \mathbf{R}^{nd}$. The two reservoirs, L and R, are described by the variables $\rho = (\rho_{\mathrm{L}}, \rho_{\mathrm{R}}) \in \mathbf{R}^{Md} \oplus \mathbf{R}^{Md}$. The "energy" of the complete system, *i.e.*, chain and reservoirs, is then given by $G^{(0)}(r,q,p) = G^{(1)}(\rho,q,p)$, where now

$$G^{(1)}(\rho,q,p) = H_{\mathrm{S}}(q,p) + \tfrac{1}{2}\rho\cdot\Gamma\rho - q\cdot\Lambda\Gamma^{1/2}\rho\ .$$

Here, $\Gamma = \Gamma_{\mathrm{L}} \oplus \Gamma_{\mathrm{R}}$, where Γ_i is the diagonal $(M \times M)$ matrix $\mathrm{diag}(\gamma_{i,1},\dots,\gamma_{i,M})$, with $i \in \{\mathrm{L},\mathrm{R}\}$. Note that by assumption, the γ's are all strictly positive. We also define Λ as the $(2Md \times nd)$ matrix given by

$$q\cdot\Lambda\rho = q_1\cdot\Lambda_{\mathrm{L}}\rho_{\mathrm{L}} + q_n\cdot\Lambda_{\mathrm{R}}\rho_{\mathrm{R}} = q_1\sum_{m=1}^{M}\lambda_{\mathrm{L},m}\rho_{\mathrm{L},m} + q_n\sum_{m=1}^{M}\lambda_{\mathrm{R},m}\rho_{\mathrm{R},m}\ .$$

With these notations, the equations of motion can be written as:

$$\begin{aligned}
\mathrm{d}q &= \nabla_p G^{(1)}\,\mathrm{d}t = p\,\mathrm{d}t\ ,\\
\mathrm{d}p &= -\nabla_q G^{(1)}\,\mathrm{d}t = -\big(\nabla_q V(q) - \Lambda\Gamma^{1/2}\rho\big)\mathrm{d}t\ ,\\
\mathrm{d}\rho &= -\nabla_\rho G^{(1)}\,\mathrm{d}t - (2T^{1/2})\mathrm{d}w = -\big(\Gamma\rho - \Gamma^{1/2}\Lambda^{\mathrm{T}} q\big)\mathrm{d}t - (2T^{1/2})\mathrm{d}w\ .
\end{aligned}$$

Here, $w = w_{\mathrm{L}} \oplus w_{\mathrm{R}} = (w_{\mathrm{L},1}, \dots, w_{\mathrm{L},M}, w_{\mathrm{R},1}, \dots, w_{\mathrm{R},M})$ is a $2Md$-dimensional standard Brownian motion, and T is the $(2M \times 2M)$ diagonal temperature matrix

$$T = \mathrm{diag}(T_{\mathrm{L}}, \dots, T_{\mathrm{L}}, T_{\mathrm{R}}, \dots, T_{\mathrm{R}}) .$$

It is useful to introduce the (final!) change of variables $s = \rho - F^{\mathrm{T}}q$, where $F = \Lambda\Gamma^{-1/2}$. In terms of these variables, one can introduce the effective potential

$$V_{\mathrm{eff}}(q) = V(q) - \tfrac{1}{2} q \cdot \Lambda\Lambda^{\mathrm{T}} q , \tag{4.2}$$

and the "energy" is now $G(s, q, p) = G^{(1)}(\rho, q, p)$ with

$$G(s, q, p) = \tfrac{1}{2}p^2 + V_{\mathrm{eff}} + \tfrac{1}{2}s \cdot \Gamma s . \tag{4.3}$$

Finally, with the adjoint change in the derivatives $\nabla_q \to \nabla_q - F\nabla_s$, the equations of motion become

$$\begin{aligned}
\mathrm{d}q &= \nabla_p G \mathrm{d}t = p\,\mathrm{d}t , \\
\mathrm{d}p &= -(\nabla_q - F\nabla_s) G \mathrm{d}t = -\big(\nabla_q V_{\mathrm{eff}}(q) - F\Gamma s\big)\mathrm{d}t , \\
\mathrm{d}s &= -(\nabla_s + F^{\mathrm{T}}\nabla_p) G\,\mathrm{d}t - (2T^{1/2})\mathrm{d}w = -\big(\Gamma s + F^{\mathrm{T}}p\big)\mathrm{d}t - (2T^{1/2})\mathrm{d}w .
\end{aligned} \tag{4.4}$$

Notation. In the sequel, we shall write G_p for $\nabla_p G$ and G_q for $\nabla_q G$ (these are vectors with nd components), and G_s for $\nabla_s G$ (this is a vector with $2Md$ components).

The generator L of the diffusion process takes, in the variables s, q, p, the form

$$L = \nabla_s \cdot T\nabla_s - G_s \cdot \nabla_s + \big(G_p \cdot \nabla_q - G_q \cdot \nabla_p\big) + \big((FG_s) \cdot \nabla_p - G_p \cdot F\nabla_s\big) . \tag{4.5}$$

If f is a function on the phase space X, we let

$$S^t f(x) = \big(e^{Lt} f\big)(x) = \int f\big(\xi_x(t)\big) d\mathbf{P}(w) .$$

The associated Fokker-Planck operator L^{T} is the adjoint of L in the space $\mathrm{L}^2(\mathbf{R}^{d(2M+2n)}, \mathrm{d}x)$, *i.e.*,

$$L^{\mathrm{T}} = \nabla_s \cdot T\nabla_s + \nabla_s \cdot G_s - \big(G_p \cdot \nabla_q - G_q \cdot \nabla_p\big) - \big((FG_s) \cdot \nabla_p - G_p \cdot F\nabla_s\big) . \tag{4.6}$$

Remark 4.1. The density ρ of the invariant measure is the (unique) normalized solution of the equations

$$L^{\mathrm{T}}\rho = 0 .$$

4.2. The Entropy Production σ

We now establish the relation between the energy flux and the entropy production. Since we are dealing with a Hamiltonian setup, the energy flux is naturally defined by the time derivative of the mean evolution S^t of the effective energy, $H_{\text{eff}}(q,p) = p^2/2 + V_{\text{eff}}(q)$. Differentiating, we get from the equations of motion

$$\partial_t S^t H_{\text{eff}} \;=\; S^t L H_{\text{eff}} \;,$$

$$L H_{\text{eff}} \;=\; p \cdot (-\nabla_q V_{\text{eff}} + F\Gamma s) + \nabla_q V_{\text{eff}} \cdot p \;=\; p \cdot F\Gamma s \;.$$

We define the total flux by $\Phi = p \cdot F\Gamma s$, and inspection of the definition of F and Γ leads to the identification of the flux at the left and right ends of the chain:

$$\Phi \;=\; \Phi_{\text{L}} + \Phi_{\text{R}} \;,$$

with

$$\begin{aligned} \Phi_{\text{L}} \;&=\; p_1 \cdot \Lambda_{\text{L}} \Gamma_{\text{L}}^{1/2} s_{\text{L}} \;, \\ \Phi_{\text{R}} \;&=\; p_n \cdot \Lambda_{\text{R}} \Gamma_{\text{R}}^{1/2} s_{\text{R}} \;. \end{aligned}$$

Note that $\Lambda_{\text{L}} \Gamma_{\text{L}}^{1/2} s_{\text{L}}$ is the net force exerted by the left bath on the chain. Therefore, $\Phi_{\text{L}} = p_1 \cdot \Lambda_{\text{L}} \Gamma_{\text{L}}^{1/2} \rho_{\text{L}} - L\, q_1 \cdot \Lambda_{\text{L}}^2 q_1/2$ is, up to a time-derivative which vanishes in the stationary state, the total power dissipated by the left bath. A similar interpretation holds for Φ_{R}. Furthermore, observe that

$$\langle \Phi \rangle_\mu \;=\; 0 \;, \tag{4.7}$$

where, generally,

$$\langle f \rangle_\mu \;\equiv\; \int \mu(\mathrm{d}x) f(x) \;.$$

The Equation (4.7) holds because $\Phi = L H_{\text{eff}}$ and $L^{\mathrm{T}} \mu = 0$.

We next proceed to define the entropy production in the setting of our model. Since we have been able to identify the energy flux on the ends of the chain, we *define* the (thermodynamic) entropy production σ by

$$\sigma \;=\; \frac{\Phi_{\text{L}}}{T_{\text{L}}} + \frac{\Phi_{\text{R}}}{T_{\text{R}}} \;=\; p \cdot F T^{-1} \Gamma s \;. \tag{4.8}$$

We refer to [CL] and references therein for a detailed discussion of the various types of entropy production in non-equilibrium stationary states. In Subsection 4.4, we will explain, in the context of our model, the relationship between the entropy production σ and the Gibbs entropy.

4.3. Time-Reversal, Generalized Detailed Balance Condition, and Negativity of the Entropy Production

Definition. We define the "time-reversal" map J by $\big(Jf\big)(s,q,p) = f(s,q,-p)$. This map is the projection onto the space of the s,q,p of the time-reversal of the Hamiltonian flow (on the full phase space of chain plus baths) defined by the original problem (2.2).

Notation. To obtain simple formulas for the entropy production σ we write the strictly positive density ρ of the invariant measure μ as

$$\rho \,=\, Je^{-R}e^{-\varphi}\ , \tag{4.9}$$

where we have introduced the quantity

$$R \,=\, R(s) \,=\, \tfrac{1}{2}s\cdot\Gamma T^{-1}s\ . \tag{4.10}$$

Let L^* denote the adjoint of L in the space $\mathcal{H}_\mu \,=\, \mathrm{L}^2(X,\mathrm{d}\mu)$ associated with the invariant measure μ with density ρ. In terms of the adjoint L^{T} on $\mathrm{L}^2(X,\mathrm{d}s\,\mathrm{d}q\,\mathrm{d}p)$, we have

$$L^* \,=\, \rho^{-1}L^{\mathrm{T}}\rho\ . \tag{4.11}$$

Let $L_\lambda \,=\, L+\lambda\sigma$, where $\lambda\in\mathbf{R}$. (This definition is suggested by the paper [K], see below.) We have the following important symmetry property:

Theorem 4.2. *One has the operator identity*

$$Je^{-J\varphi}(L_\lambda)^*e^{J\varphi}J \,=\, L_{1-\lambda}\ . \tag{4.12}$$

In particular, one has

$$Je^{-J\varphi}L^*e^{J\varphi}J-L \,=\, \sigma\ . \tag{4.13}$$

Remark 4.3. This relation may be viewed as a generalization to non-equilibrium of the detailed balance condition (at equilibrium, one has $JL^*J-L=0$).

Remark 4.4. Recently, a lot of interest has been generated in the wake of papers by Gallavotti and Cohen, [GC1, GC2, G, and references therein], in which intriguing relations for the fluctuations of the entropy production have been found. These papers dealt first with numerical experiments by [ECM], which were then abstracted to the general context of dynamical systems. In further work, these ideas have been successfully applied to thermostatted systems modeling non-equilibrium problems. In the papers [K] and [LS] these ideas have been further extended to non-equilibrium models described by stochastic dynamics. In the context of our model, the setup is as follows: One considers the observable

$$W(t) \,=\, \int_0^t \mathrm{d}\tau\,\sigma\big(\xi_x(\tau)\big)\ .$$

By ergodicity, $\lim_{t\to\infty} t^{-1}W(t) = \langle\sigma\rangle_\mu$, for μ-almost all x. We are interested in the rate function $\hat{e}$ for the large deviations of $W(t)/(t\langle\sigma\rangle_\mu)$, and want to argue (heuristically) that it satisfies

$$\hat{e}(w) - \hat{e}(-w) = -w\langle\sigma\rangle_\mu \,. \tag{4.14}$$

In particular this means that at equal temperatures, when $\langle\sigma\rangle_\mu = 0$, the fluctuations are symmetric around the mean 0, while at unequal temperatures, the odd part is linear in w and proportional to the mean entropy production. This is the celebrated Gallavotti-Cohen fluctuation theorem.

The rate function $\hat{e}$ is characterized by the relation

$$\inf_{w\in I} \hat{e}(w) = -\lim_{t\to\infty} \frac{1}{t} \log \mathbf{P}\left(\frac{W(t)}{t\langle\sigma\rangle_\mu} \in I\right) .$$

Under suitable conditions it can be expressed as the Legendre transform of the function

$$e(\lambda) \equiv -\lim_{t\to\infty} t^{-1} \log\left\langle e^{-\lambda W(t)}\right\rangle_\mu .$$

Formally, $-e(\lambda)$ can be represented as the maximal eigenvalue of L_λ. Observing now the relation (4.12), one sees immediately that

$$e(\lambda) = e(1-\lambda) \,. \tag{4.15}$$

This in turn implies (4.14). A rigorous derivation of the program outlined above lacks several technical ingredients, in particular, more spectral information about L_λ seems to be necessary.

The relation (4.12) has a number of other consequences which we enumerate now, before going to the proof of Theorem 4.2. It allows to prove that the entropy production is negative in our model and it yields an interesting symmetry relation (see Theorem 4.6).

Proposition 4.5. *One has the following identities (between functions):*

$$L\varphi = \sigma + |T^{1/2}\nabla_s \varphi|^2 \,, \tag{4.16}$$

$$L^* J\varphi = -\sigma - |T^{1/2}\nabla_s J\varphi|^2 \,. \tag{4.17}$$

Here, $|f|^2 \equiv f\cdot f$.

Theorem 4.6. *In the stationary state μ the entropy production satisfies the identity:*

$$\langle\sigma\rangle_\mu = -\langle|T^{1/2}\nabla_s\varphi|^2\rangle_\mu = -\langle|T^{1/2}\nabla_s J\varphi|^2\rangle_\mu \le 0 \,. \tag{4.18}$$

Remark 4.7. In Subsection 4.5, we will show that the heat flux is non-zero for unequal temperatures by showing that the entropy production in the stationary state satisfies:

$$\langle\sigma\rangle_\mu = 0 \,,$$

if and only if $T_{\mathrm{L}} = T_{\mathrm{R}}$.

The remainder of this subsection is devoted to the proofs of Theorem 4.2, Proposition 4.5, and Theorem 4.6.

Proof of Theorem 4.2. We show the identity

$$e^{R} J L^{\mathrm{T}} J e^{-R} \;=\; L + \sigma\ . \tag{4.19}$$

Starting with the relation

$$e^{R} \nabla_s e^{-R} \;=\; \nabla_s - (\nabla_s R) \;=\; \nabla_s - T^{-1} G_s\ ,$$

we get, using the definition of L^{T},

$$\begin{aligned} e^{R} J L^{\mathrm{T}} J e^{-R} \;=\; & \nabla_s \cdot T \nabla_s - G_s \cdot \nabla_s \\ & + \big(G_p \cdot \nabla_q - G_q \cdot \nabla_p\big) + \big(F\Gamma s \cdot \nabla_p - p \cdot F \nabla_s\big) \\ & + G_p \cdot F T^{-1} G_s\ . \end{aligned}$$

Note that the sum of all the terms except the last equals L, while the last equals

$$G_p \cdot F T^{-1} G_s \;=\; p \cdot F T^{-1} \Gamma s \;=\; \frac{p_1 F_{\mathrm{L}} \Gamma_{\mathrm{L}} s_{\mathrm{L}}}{T_{\mathrm{L}}} + \frac{p_n F_{\mathrm{R}} \Gamma_{\mathrm{R}} s_{\mathrm{R}}}{T_{\mathrm{R}}} \;=\; \sigma\ .$$

We have thus shown (4.19). Combining (4.19) with the expressions (4.11) and (4.9) for L^* and ρ we obtain the identity:

$$L + \sigma \;=\; e^{R} J L^{\mathrm{T}} J e^{-R} \;=\; e^{R} J e^{-R} e^{-J\varphi} L^* e^{J\varphi} e^{R} J e^{-R} \;=\; J e^{-J\varphi} L^* e^{J\varphi} J\ , \tag{4.20}$$

which is (4.12) for $\lambda = 0$, *i.e.*, Eq.(4.13). Observing now that $J \sigma J = -\sigma$, we obtain

$$J e^{-J\varphi} L^*_{\lambda} e^{J\varphi} J \;=\; L_{1-\lambda}\ ,$$

and thus conclude the proof of Theorem 4.2.

Proof of Proposition 4.5. From (4.12) we obtain the identity

$$J L^* J \;=\; e^{\varphi} (L + \sigma) e^{-\varphi}\ . \tag{4.21}$$

A straightforward computation shows that, for any smooth function f, we have the following operator identity

$$e^{f} L e^{-f} \;=\; L - 2(\nabla_s f) \cdot T \nabla_s - (Lf) + |T^{1/2} (\nabla_s f)|^2\ . \tag{4.22}$$

Applying (4.22) with $f = \varphi$ we obtain from Eq.(4.21) the operator identity

$$J L^* J \;=\; L - 2(\nabla_s \varphi) \cdot T \nabla_s - (L\varphi) + |T^{1/2} (\nabla_s \varphi)|^2 + \sigma\ . \tag{4.23}$$

Since $L^* = \rho^{-1} L^{\mathrm{T}} \rho$ and $L^{\mathrm{T}}\rho = 0$ we have $L^* 1 = 0$. Applying the operator identity (4.23) to the function 1 and noting that $J1 = 1$ we get

$$0 \;=\; JL^*J1 \;=\; -L\varphi + |T^{1/2}(\nabla_s\varphi)|^2 + \sigma \;, \tag{4.24}$$

and this is the identity (4.16). With this, (4.23) simplifies to

$$JL^*J \;=\; L - 2(\nabla_s\varphi)\cdot T\nabla_s \;. \tag{4.25}$$

Applying the operator identity (4.25) to the function φ, and using (4.24) we get

$$\begin{aligned} JL^*J\varphi \;&=\; L\varphi - 2|T^{1/2}(\nabla_s\varphi)|^2 \\ &=\; \sigma - |T^{1/2}(\nabla_s\varphi)|^2 \;, \end{aligned}$$

or, equivalently,

$$L^*J\varphi \;=\; -\sigma - |T^{1/2}(\nabla_s J\varphi)|^2 \;,$$

which proves (4.17). With this we have concluded the proof of Proposition 4.5.

Proof of Theorem 4.6. Theorem 4.6 is a simple consequence of Proposition 4.5. From (4.16), using the invariance of the measure μ, we get

$$\begin{aligned} \langle\sigma\rangle_\mu \;&=\; \langle L\varphi\rangle_\mu - \langle|T^{1/2}(\nabla_s\varphi)|^2\rangle_\mu \\ &=\; -\langle|T^{1/2}(\nabla_s\varphi)|^2\rangle_\mu \;, \end{aligned}$$

which yields the first equality in (4.18). The second inequality is obtained in the same way using (4.17). We have

$$\begin{aligned} \langle\sigma\rangle_\mu \;&=\; -\langle L^*J\varphi\rangle_\mu - \langle|T^{1/2}(\nabla_s J\varphi)|^2\rangle_\mu \\ &=\; -\langle|T^{1/2}(\nabla_s J\varphi)|^2\rangle_\mu \;, \end{aligned} \tag{4.26}$$

where the last equality in (4.26) follows from the identity

$$\langle L^*J\varphi\rangle_\mu \;=\; \int \mathrm{d}x(L^{\mathrm{T}}\rho J\varphi) \;=\; \int \mathrm{d}x\rho J\varphi(L1) \;=\; 0 \,.$$

It is obvious from (4.26) that the entropy production in the stationary state is a non-positive quantity and this concludes the proof of Theorem 4.6.

Other observables for the entropy production. The analysis done for the entropy production σ can be repeated for other observables, (see also [LS] for a similar generalization). A family of such observables can be obtained by replacing the conjugation operator $e^{J\varphi}J$ of (4.12) by any conjugation operator of the form $e^f e^{J\varphi}J$, where $f = f(q,p)$ satisfies $Jf = f$.[1]

[1] The operators $e^f e^{J\varphi}J$ are all formally selfadjoint on $\mathcal{H}_\mu$.

Interesting examples are obtained when one considers the energy flux between position j and the $j+1$ on the chain, $j = 1, \ldots, n-1$. To this end we write the Hamiltonian H_{eff} as follows:

$$H_{\text{eff}}(q,p) \;=\; \sum_{i=1}^{n} H_i(q,p) \;,$$

where

$$\begin{aligned}
H_1(q,p) \;&=\; \frac{p_1^2}{2} + U_1^{(1)}(q_1) - \frac{1}{2} q_1^2 \sum_{m=1}^{M} \lambda_{\text{L},m}^2 + \frac{1}{2} U_1^{(2)}(q_1 - q_2) \;, \\
H_i(q,p) \;&=\; \frac{p_i^2}{2} + U_i^{(1)}(q_i) + \frac{1}{2} U_{i-1}^{(2)}(q_{i-1} - q_i) + \frac{1}{2} U_i^{(2)}(q_i - q_{i+1}) \;, \quad i = 2, \ldots, n-1 \;, \\
H_n(q,p) \;&=\; \frac{p_n^2}{2} + U_n^{(1)}(q_n) - \frac{1}{2} \sum_{m=1}^{M} \lambda_{\text{R},m}^2 \, q_n^2 + \frac{1}{2} U_{n-1}^{(2)}(q_{n-1} - q_n) \;.
\end{aligned}$$

For any $j = 1, \ldots, n-1$, we choose $f = -S_j$, where

$$S_j(q,p) \;=\; \frac{1}{T_{\text{L}}} \sum_{i=1}^{j} H_i(q,p) + \frac{1}{T_{\text{R}}} \sum_{i=j+1}^{n} H_i(q,p) \;.$$

We write now the invariant density ρ as

$$\rho \;=\; J e^{-R} e^{-S_j} e^{-\psi_j} \;,$$

i.e., $\psi_j = \varphi - S_j$. Variants of computations done above show that we have the operator identity, similar to (4.20):

$$e^{S_j} e^{R} J L^{\text{T}} J e^{-R} e^{-S_j} \;=\; L + \sigma_j \;,$$

where σ_j is given by the relation

$$\sigma_j \;=\; \sigma - L S_j \;, \tag{4.27}$$

since S_j does not depend on the variable s. Our choice of S_j has been made in such a way that

$$\sigma_j \;=\; \Big(\frac{1}{T_{\text{L}}} - \frac{1}{T_{\text{R}}}\Big) \, (p_j - p_{j+1}) \cdot \nabla U^{(2)}(q_j - q_{j+1}) \;,$$

i.e., σ_j is the energy flux between position j and $j+1$ on the chain multiplied by the difference between the inverse temperatures of the heat baths. Using next that $J S_j = S_j$ one derives easily a relation corresponding to (4.12), namely,

$$J e^{-J\psi_j} (L + \lambda \sigma_j)^* e^{J\psi_j} J \;=\; L + (1-\lambda)\sigma_j \;. \tag{4.28}$$

We have thus found $n-1$ "entropy productions" σ_j, which, under the assumptions made for σ, satisfy a fluctuation theorem. Note that these entropy productions are all different observables, but, because of Eq.(4.27), the expectations of σ and σ_j in the stationary state μ coincide.

4.4. Relation with the Gibbs Entropy

We give now a second proof of the negativity of entropy production in our model using the Gibbs entropy.

Let ν_0 be a probability measure in the variables $x = (s, q, p)$ and let ν_t denote the time evolution of ν_0 given by

$$\nu_t(\mathrm{d}x) = \int \nu_0(\mathrm{d}x')P(t, x', x) .$$

Because of the properties of the transition probabilities $P(t, x', x)$ proven in Sect. 3, we see that ν_t is a probability measure (for any $t > 0$) with a smooth positive density denoted f_t in the sequel. The time evolution of f is then given by the equation

$$\partial_t f_t = L^{\mathrm{T}} f_t .$$

We define the Gibbs entropy as

$$S(f) = -\int \mathrm{d}x\, f(x) \log f(x) ,$$

and we compute next the entropy change in time. We get:

$$\begin{aligned}\partial_t S(f_t) &= -\int \mathrm{d}x\, (\partial_t f_t)(1 + \log f_t)\\ &= -\bigl(L^{\mathrm{T}} f_t, (1 + \log f_t)\bigr)\\ &= -\bigl(f_t, L \log f_t\bigr)\\ &= -\bigl(f_t, f_t^{-1} L f_t\bigr) + \bigl(f_t, |T^{1/2}\nabla_s \log f_t|^2\bigr) \equiv X_1 .\end{aligned}$$

The last term is the (additional) contribution from the second order derivative (in s) appearing in L when it acts on $\log f$. We can transform X_1 further by writing it as

$$X_1 = -(L^{\mathrm{T}}1, f_t) + \bigl(f_t, |T^{1/2}\nabla_s \log f_t|^2\bigr) . \tag{4.29}$$

Since $L^{\mathrm{T}}1 = \mathrm{Tr}\,\Gamma$ the first term in (4.29) is equal to $-\mathrm{Tr}\,\Gamma$. We use the definition (4.10) of R and the analog of (4.9) to define φ_t:

$$Je^{-R}e^{-\varphi_t} = f_t . \tag{4.30}$$

Since $JRJ = R$, we see that $-\log f_t = R + J\varphi_t$. Expanding the second term in (4.29) we obtain

$$\bigl(f_t, |T^{1/2}\nabla_s \log f_t|^2\bigr) = \bigl(f_t, |T^{-1/2}\Gamma s|^2\bigr) + \bigl(f_t, |T^{+1/2}\nabla_s J\varphi_t|^2\bigr) + 2(f_t, \Gamma s \cdot \nabla_s J\varphi_t) . \tag{4.31}$$

Since we have the relation

$$\nabla_s f_t = -f_t\bigl(\nabla_s R + \nabla_s J\varphi_t\bigr) ,$$

we obtain

$$f_t \nabla_s J \varphi_t \;=\; -f_t \nabla_s R - \nabla_s f_t \;.$$

Using this and integrating by parts, we rewrite the third term in (4.31) as

$$2(f_t, \Gamma s \cdot \nabla_s J \varphi_t) \;=\; -2 \int \mathrm{d}x \, \Gamma s \cdot \bigl(f_t \nabla_s R + \nabla_s f_t \bigr) \;=\; -2(f_t, |T^{-1/2} \Gamma s|^2) + 2 \mathrm{Tr}\, \Gamma \;.$$

Altogether we obtain

$$\partial_t S(f_t) \;=\; \mathrm{Tr}\, \Gamma - (f_t, |T^{-1/2} \Gamma s|^2) + \bigl(f_t, |T^{1/2} \nabla_s J \varphi_t|^2 \bigr) \;.$$

Using the identity

$$LR \;=\; \mathrm{Tr}\, \Gamma - |T^{-1/2} \Gamma s|^2 - \sigma \;,$$

we obtain finally,

$$\partial_t S(f_t) \;=\; \int \mathrm{d}x \; f_t \sigma + \int \mathrm{d}x \; f_t LR + \int \mathrm{d}x \; f_t |T^{1/2} \nabla_s J \varphi_t|^2 \;. \tag{4.32}$$

In line with the ideas of [CL], we can write this last identity in the form:

$$\partial_t S(f_t) - \langle \sigma \rangle_{\nu_t} \;=\; \bigl\langle |T^{1/2} \nabla_s J \varphi_t|^2 \bigr\rangle_{\nu_t} + \partial_t \langle R \rangle_{\nu_t} \;.$$

This shows that the (rearrangement) entropy produced in addition to the thermodynamic entropy σ is a positive quantity, up to a (time-) boundary term. Also note that if $\nu_t = \mu$, *i.e.*, if the system is in the stationary state, then we get the identity

$$0 \;=\; \langle \sigma \rangle_\mu + \langle |T^{1/2} J \nabla_s \varphi|^2 \rangle_\mu \;, \tag{4.33}$$

which we already found in Theorem 4.6.

4.5. Strict Positivity of the Heat Flux

We first show that the thermodynamic entropy production, as defined in (4.8), is negative in our model. As an immediate consequence we will show that, in the stationary state, energy is flowing from the hotter heat bath to the colder one.

Theorem 4.8. *The entropy production σ satisfies:*

$$\langle \sigma \rangle_\mu \;=\; 0 \,,$$

if and only if $T_{\mathrm{L}} = T_{\mathrm{R}}$.

Proof. Note that if $T_{\mathrm{L}} = T_{\mathrm{R}}$, then $\sigma = \partial_t S^t H_{\mathrm{eff}} / T_{\mathrm{L}} \big|_{t=0}$ and therefore $\langle \sigma \rangle_\mu = 0$. We will show that if $T_{\mathrm{L}} \neq T_{\mathrm{R}}$, then $\langle \sigma \rangle_\mu \neq 0$. We will proceed by assuming the converse, namely $\langle \sigma \rangle_\mu = 0$, and produce a contradiction. The assumption implies, by (4.18) that $\langle |T^{1/2}\nabla_s\varphi|^2\rangle_\mu = 0$. Since ρ is positive, this means that $\nabla_s\varphi = 0$, and therefore φ does not depend on the s variables. From (4.16) we obtain

$$0 = -L\varphi + |T^{1/2}\nabla_s\varphi|^2 + \sigma = -L\varphi + \sigma .$$

Using the definition of L and σ and the fact that φ does not depend on s, we obtain the equation

$$0 = \bigl(p\cdot\nabla_q\varphi - (\nabla_q V_{\rm eff})\cdot\nabla_p\varphi\bigr) + F\Gamma s\cdot(\nabla_p\varphi - T^{-1}p) .$$

Since φ does not depend on s we get

$$\begin{aligned} p\cdot\nabla_q\varphi - (\nabla_q V_{\rm eff})\cdot\nabla_p\varphi &= 0 , \\ \nabla_{p_1}\varphi &= T_{\rm L}^{-1}p_1 , \\ \nabla_{p_n}\varphi &= T_{\rm R}^{-1}p_n . \end{aligned} \tag{4.34}$$

We will show that this system of linear equations has no solution unless $T_{\rm L} = T_{\rm R}$. To see this we consider the system of equations

$$\begin{aligned} p\cdot\nabla_q\varphi - (\nabla_q V_{\rm eff})\cdot\nabla_p\varphi &= 0 \\ \nabla_{p_1}\varphi &= T_{\rm L}^{-1}p_1 . \end{aligned} \tag{4.35}$$

This system has a solution which is given by $H_{\rm eff}(q,p)/T_{\rm L}$. We claim that this the unique solution (up to an additive constant) of (4.35).

If this holds true, then the only solution of (4.34) is given by $H_{\rm eff}(q,p)/T_{\rm L}$ and this is incompatible with the third equation in (4.34) when $T_{\rm L} \neq T_{\rm R}$.

Since (4.35) is a linear inhomogeneous equation, it is enough to show that the only solutions of the homogeneous equation

$$\begin{aligned} p\cdot\nabla_q\varphi - (\nabla_q V_{\rm eff})\cdot\nabla_p\varphi &= 0 , \\ \nabla_{p_1}\varphi &= 0 , \end{aligned} \tag{4.36}$$

are the constant functions. Since $\nabla_{p_1}\varphi = 0$, φ does not depend on p_1, we conclude that the first equation in (4.36) reads

$$p_1\cdot\nabla_{q_1}\varphi + f_1(q_1,\dots,q_n,p_2,\dots p_n) = 0 ,$$

where f_1 does not depend on the variable p_1. Thus we see that $\nabla_{q_1}\varphi = 0$ and therefore φ does not depend on the variable q_1 either. By the first equation in (4.36) we now get

$$-\nabla_{q_1}U_1^{(2)}(q_1-q_2)\cdot\nabla_{p_2}\varphi + f_2(q_2,\dots,q_n,p_2,\dots,p_n) = 0 ,$$

where f_2 does not depend on p_1 and q_1. By condition **H2** we see that $\nabla_{p_2}\varphi = 0$ and hence f does not depend on p_2. Iterating the above procedure we find that the only solutions of (4.36) are the constant functions. This concludes the proof of Theorem 4.8.

Corollary 4.9. *The stationary state $\mu = \mu_{T_{\mathrm{L}},T_{\mathrm{R}}}$ produces a non-vanishing mean heat flux in the direction from the hotter to the colder reservoir. The mean heat flux vanishes only if $T_{\mathrm{L}} = T_{\mathrm{R}}$.*

Proof. The entropy production σ is given by

$$\sigma \;=\; \frac{\Phi_{\mathrm{L}}}{T_{\mathrm{L}}} + \frac{\Phi_{\mathrm{R}}}{T_{\mathrm{R}}}\;,$$

where Φ_{L} is the energy flow from the left heat bath to the chain and similarly for Φ_{R}. In the stationary state we have, by (4.7),

$$\langle \Phi_{\mathrm{L}} + \Phi_{\mathrm{R}} \rangle_\mu \;=\; 0\;,$$

and therefore

$$\langle \Phi_{\mathrm{L}} \rangle_\mu \;=\; -\langle \Phi_{\mathrm{R}} \rangle_\mu\;.$$

We obtain from Theorem 4.8, for $T_{\mathrm{L}} \neq T_{\mathrm{R}}$:

$$0 \;>\; \langle \sigma \rangle_\mu \;=\; \Big(\frac{1}{T_{\mathrm{L}}} - \frac{1}{T_{\mathrm{R}}}\Big)\langle \Phi_{\mathrm{L}} \rangle_\mu\;.$$

If, say, $T_{\mathrm{L}} > T_{\mathrm{R}}$, we get $\langle \Phi_{\mathrm{L}} \rangle_\mu > 0$ and thus energy flows from the hotter to the cooler heat bath.

Acknowledgments. We would like to thank G. Ben Arous, Ch. Mazza, H. Spohn, D. Stroock, A.-S. Sznitman, and L.E. Thomas for encouragements and helpful discussions, and V. Zagrebnov for useful questions and comments. This work was supported in part by the Fonds National Suisse.

References

[CELS] N.I. Chernov, G.L. Eyink, J.L. Lebowitz, and Ya. G. Sinai: Steady-state electric conduction in the periodic Lorentz gas. Commun. Math. Phys. **154**, 569–601 (1993).

[CL] N.I. Chernov and J.L. Lebowitz: Stationary nonequilibrium states in boundary driven Hamiltonian systems: Shear flow. Preprint. (1997).

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

[EM] D. Evans and G. Morriss: *Statistical Mechanics of Nonequilibrium Liquids*, New York, Academic Press (1990).

[EPR] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet: Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. Commun.Math.Phys., in print.

[FGS] J. Farmer, S. Goldstein, and E.R. Speer: Invariant states of a thermally conducting barrier. J. Stat. Phys. **34**, 263–277 (1984).

[G] G. Gallavotti: Chaotic hypothesis and universal large deviations properties. Doc. Math. J. DMV Extra Volume ICM I 65–93 (1998).

[GC1] G. Gallavotti and E.G.D. Cohen: Dynamical ensembles in nonequilibrium statistical mechanics. Phys. Rev. Lett. **74**, 2694–2697 (1995).

[GC2] G. Gallavotti and E.G.D. Cohen: Dynamical ensembles in stationary states. J. Stat. Phys, **80**, 931–970 (1995).

[GKI] S. Goldstein, C. Kipnis, and N. Ianiro: Stationary states for a mechanical system with stochastic boundary conditions.. J. Stat. Phys. **41**, 915–939 (1985).

[GLP] S. Goldstein, J.L. Lebowitz, and E. Presutti: Stationary states for a mechanical system with stochastic boundaries. In *Random Fields, (Colloquia Mathematicae Societatis János Bolyai*. Amsterdam, North-Holland **27** (1981).

[H] W.G. Hoover: *Computational Statistical Mechanics*, Elsevier (1991).

[IK] K. Ishihara and H. Kunita: A classification of the second order degenerate elliptic operators and its probabilistic characterization. Z. Wahrsch. und Verw. Geb. **39**, 235–254 (1974).

[K] J. Kurchan: Fluctuation theorem for stochastic dynamics. J. Phys. **A31**, 3719–3729 (1998).

[Ku] H. Kunita: Supports of diffusion processes and controllability problems. In *Proc. Intern. Symp. SDE Kyoto*. 163–185 (1976).

[LS] J.L. Lebowitz and H. Spohn: The Gallavotti-Cohen fluctuation theorem for stochastic dynamics. Preprint. (1998).

[PH] H.A. Posch and W.G. Hoover: Equilibrium and non equilibrium Lyapunov spectra for dense fluids and solids. Phys. Rev. **A39**, 2175–2188 (1989).

[SL] H. Spohn and J.L. Lebowitz, Stationary non-equilibrium: states of infinite harmonic systems. Commun. Math. Phys. **54**, 97 (1977).

[SV] D.W. Stroock and S.R.S. Varadhan: On the support of diffusion processes with applications to the strong maximum principle. In *Proc. 6-th Berkeley Symp. Math. Stat. Prob.*. **III**, 333–368 (1972).