The Fluctuation Theorem as a Gibbs Property

Christian Maes*, Instituut voor Theoretische Fysica,
K.U.Leuven, B-3001 Leuven, Belgium.

December 18, 1998

Dedicated to the memory of Edwin T. Jaynes.

Abstract

Common ground to recent studies exploiting relations between dynamical systems and non-equilibrium statistical mechanics is, so we argue, the standard Gibbs formalism applied on the level of space-time histories. The assumptions (chaoticity principle) underlying the Gallavotti-Cohen fluctuation theorem make it possible, using symbolic dynamics, to employ the theory of one-dimensional lattice spin systems. The Kurchan and Lebowitz-Spohn analysis of this fluctuation theorem for stochastic dynamics can be restated on the level of the space-time measure which is a Gibbs measure for an interaction determined by the transition probabilities. In this note we understand the fluctuation theorem as a Gibbs property as it follows from the very definition of Gibbs state. We give a local version of the fluctuation theorem in the Gibbsian context and we derive from this a version also for some class of spatially extended stochastic dynamics.

Keywords: fluctuation theorem, large deviations, nonequilibrium, Gibbs states.

1 Context and main observations.

1.1 Scope.

The fluctuation theorem of Gallavotti and Cohen, see [10, 11, 26], asserts that for a class of dynamical systems the fluctuations in time of the phase space contraction rate obey a general law. We refer to the cited literature for additional details and precision and we only sketch here the main ingredients.

One considers a reversible smooth dynamical system $\xi \to \phi(\xi), \xi \in \Omega$. The phase space Ω is in some sense bounded carrying only a finite number of degrees of freedom (a compact and connected manifold). The transformation ϕ

*Onderzoeksleider FWO, Flanders. Email: christian.maes@fys.kuleuven.ac.be

is a diffeomorphism of Ω. The resulting (discrete) time evolution is obtained by iteration and the reversibility means that there is a diffeomorphism θ on Ω with $\theta^2 = 1$ and $\theta \circ \phi \circ \theta = \phi^{-1}$. Consider now minus the logarithm of the Jacobian determinant J which arises from the change of variables implied by the dynamics. We write $\dot{S} \equiv -\log J$. One is interested in the fluctuations of

$$w_N(\xi) \equiv \frac{1}{\rho(\dot{S})N} \sum_{-N/2}^{N/2} \dot{S}(\phi^n(\xi)), \tag{1.1}$$

for large time N. Here, ρ is the stationary probability measure (SRB measure) of the dynamics with expectations

$$\rho(f) = \lim_N \frac{1}{N} \sum_0^N f(\phi^n \xi) \tag{1.2}$$

corresponding to time-averages for almost every randomly chosen initial point $\xi \in \Omega$. This random choice refers to an absolutely continuous measure with respect to the Riemann volume element $d\xi$ on Ω (and is thought of as describing the microcanonical ensemble for Ω the energy surface). $\dot{S}(\xi)$ is the phase space contraction rate (which is identified with the entropy production rate) and one assumes (and sometimes proves) dissipativity:

$$\rho(\dot{S}) > 0. \tag{1.3}$$

It is assumed that the dynamical system satisfies some technical (ergodic) condition: it is a transitive Anosov system. This ensures that the system allows a Markov partition (and the representation via some symbolic dynamics) and the existence of the SRB measure ρ in (1.2). This technical assumption is not taken physically very serious but instead it is supposed to guide us towards general results which are true in a broader context[1]. That is what is affirmed in the so called chaotic hypothesis: "A reversible many particle system in a stationary state can be regarded as a transitive Anosov system for the purpose of computing the macroscopic properties," see also e.g. [10, 11, 7, 6, 25, 9]. The fluctuation theorem then states that $w_N(\xi)$ has a distribution $\rho_N(w)$ with respect to the stationary state ρ such that

$$\lim_N \frac{1}{N\rho(\dot{S})w} \ln \frac{\rho_N(w)}{\rho_N(-w)} = 1 \tag{1.4}$$

always. In other words, the distribution of entropy production over long time intervals satisfies some general symmetry property.
This theorem originated from numerical evidence, e.g. in [5], and it has various interesting consequences. For example, in [7], it was interpreted as extending the Green-Kubo formulas to arbitrary forcing fields for a class of non-equilibrium dynamics.

In [16], Kurchan pointed out that this fluctuation theorem also holds for certain diffusion processes. This is the context of finite systems undergoing

[1] The situation resembles here to some extent that for the ergodic hypothesis. Ergodicity is likely to be false in quite a number of realistic situations and in any event it is irrelevant. Nevertheless assuming ergodicity can lead to correct consequences.

Langevin dynamics. This was extended by Lebowitz and Spohn in [19] to quite general Markov processes. There was however no general scheme for identifying the quantity (being some analogue of (1.1)) for which the fluctuation theorem holds. Yet, from applying the fluctuation theorem in this context to simple models of stochastic dynamics, relations appeared between the entropy production and the action functional satisfying the theorem.

In this note we understand the fluctuation theorem within the Gibbs formalism. Since this formalism is often considered as giving a mathematical structure to the theory of equilibrium statistical mechanics and in order to avoid misunderstanding, we insist from the beginning that we wish to see this Gibbs formalism applied here to nonequilibrium conditions. The right way of looking at it, is to consider space-time histories drawn from a Gibbs measure. In other words, our analysis is not to be regarded as an investigation of fluctuations in an equilibrium system or as the restriction of the fluctuation theorem to equilibrium conditions. On the contrary, the observations we make can be seen as underlying and (at least in some sense) extending both the Gallavotti-Cohen and the Kurchan and Lebowitz-Spohn fluctuation theorems. Underlying because the technical conditions of the Gallavotti-Cohen work reduce to a large extent the fluctuation theorem to a statement about one-dimensional Gibbs measures. That is not very different in the Lebowitz-Spohn work where the strong chaoticity is replaced by stochasticity and the Perron-Frobenius theorem is applied to the dynamical generator as it is usually done for the transfer matrix in one-dimensional Gibbs states. The fact that something more general and typical of Gibbs states is at work here was already announced in Section 3 of [1] where Example 1.2 below was applied to the one-dimensional Ising model in an external field. Our work systematizes this remark. But for Gibbs states, the fluctuation theorem does not rely on having one dimension or on a high temperature condition. Once this is perceived, one is tempted to conclude that chaoticity assumptions, while important guides, cannot really be necessary for a fluctuation theorem or its consequences to hold. Perhaps it is more natural to assume immediately that for the purpose of computing macroscopic properties, a many particle system in a steady state should be regarded as a Gibbs system for the space-time histories. And we know that the reason for Gibbs distributions has little to do with the detailed properties of the system's dynamics but instead is based on statistical principles [2]. There is finally a second, practically speaking, more important extension of the earlier results. In the analysis below, we present a local version of the fluctuation theorem. A mechanism for the validity of a local version was already discussed in the recent [8]. This is crucial because it is only a local fluctuation theorem that leads to observable consequences and we will see that this is quite natural in a Gibbsian setup.

1.2 Disclaimers.

Our analysis below is limited in various ways including:
1. Time (and space) is discrete: a regular lattice plays the role of space-time.

[2] We have in mind the maximum entropy principle and the foundations of statistical mechanics in the theory of large deviations, see e.g. [17, 14]. This must be contrasted with the approach from the theory of dynamical systems (as summarized for example in [26]). Notice that Markov partitions do not correspond to a statistical procedure; they fully encode the dynamics.

We believe that going to continuous time is a technical step (which is not expected to be very difficult) and that this is irrelevant for the purpose of the paper.
2. No hard-core conditions: we take a smooth potential and all transition probabilities are bounded away from zero. In particular, this seems problematic when dealing with dynamics subject to certain conservation laws. Again, we do not think that this is essential because the Gibbs properties we use also hold for hard-core interactions. Extra care and conditions would be needed for writing down certain formulae but we believe they do not modify the main result.
3. Discrete spins: we deal with regular lattice spin systems. While some compactness of phase space is nice to have around, our results depend solely on having a large deviation principle for Gibbs states. The extent to which such a principle holds decides on the possible extensions of our results.
4. No phase transitions: while the fluctuation theorem holds quite generally, its contents can be empty when the large deviations happen on another scale than linear in time (and spatial volume). In other words, the corresponding rate function could fail to be strictly convex in which case instabilities or phase transitions are present. These 'violations' of the fluctuation theorem can of course not happen when the spatial volume is finite (for a sufficiently chaotic dynamics or for a non-degenerated stochastic dynamics) or, for infinite systems, when we are in the 'high temperature' regime. Such scenario's are of course well-documented for Gibbs states.
5. Steady states and time-homogeneity: we do not consider here the (physically very relevant) problem of forces or potentials depending on time nor do we investigate here the long time behavior of the system started in anything other than in a stationary state. In these cases, we must refer to the study of Gibbs states on half-spaces with particular boundary conditions but the main points must remain intact.
We hope to include in a future publication the extensions mentioned above. In particular, all examples that appear in [19] can be systematically obtained using the one and same algorithm that will be explained below. We will briefly illustrate such a result (for a continuous time dynamics with a conservation law) at the end (Section 3.3).

1.3 Notation and definitions.

We restrict ourselves here to lattice spin systems. For lattice we take the regular $d+1$-dimensional set $\mathbb{Z}^{d+1}, d \geq 0$. The reason for taking $d+1$ is that the extra dimension refers to the time-axis. The points of the lattice are denoted by $x, y, \ldots$ with $x = (i,n), n \in \mathbb{Z}, i \in \mathbb{Z}^d$. We can read the time by the mapping $t(x) = n$ if $x = (i,n)$. The distance between two points $x = (i,n), y = (j,m) \in \mathbb{Z}^{d+1}$ is $|x-y| \equiv \max\{|n-m|, |i-j|\}$ with $|i-j| \equiv \max\{|i_1 - j_1|, \ldots, |i_d - j_d|\}$ for the two sites $i = (i_1, \ldots, i_d), j = (j_1, \ldots, j_d) \in \mathbb{Z}^d$. The set of finite and non-empty subsets of $\mathbb{Z}^{d+1}$ is denoted by $\mathcal{S}$. For general elements of $\mathcal{S}$ we write $\Lambda, A, \ldots$; they correspond to (finite) space-time regions. $\Lambda^c = \mathbb{Z}^{d+1} \setminus \Lambda$ is the complement of Λ; $|\Lambda|$ is the cardinality of Λ.

A space-time configuration of our lattice spin system is denoted by $\sigma, \eta, \xi, \ldots$. This is a mapping $\sigma : \mathbb{Z}^{d+1} \to S$ with values $\sigma(x) \in S$ in the single site state space S which is taken finite. Ising spins have $S = \{+1, -1\}$. The set of all configurations is $\Omega_{d+1} = S^{\mathbb{Z}^{d+1}}$. By $\sigma_E, E \subset \mathbb{Z}^{d+1}$ we mean, depending on the situation, both the restriction of σ to E as well as a configuration on E, i.e. an element of S^E. The configuration $\sigma_\Lambda \eta_{\Lambda^c}$ is equal to σ on Λ and is equal to η on Λ^c.

Ω_{d+1} is equipped with the product topology and is a compact space. If we denote by $\mathcal{F}_o$ the set of all subsets of S, then $\mathcal{F}_\Lambda = \mathcal{F}_o^\Lambda$ is the Borel sigma-algebra generated by the $(\sigma(x), x \in \Lambda)$. We write $\mathcal{F} \equiv \mathcal{F}_{\mathbb{Z}^{d+1}}$; $(\Omega, \mathcal{F})$ is the measurable space of space-time configurations.

Local functions on Ω_{d+1} are real-valued functions f which are $\mathcal{F}_\Lambda$-measurable for some $\Lambda \in \mathcal{S}$. The (finite) dependence set of such a local f is denoted by D_f. A continuous function is every function on Ω_{d+1} which is the uniform limit of local functions. The uniform norm is denoted by $\|f\| \equiv \sup_\sigma |f(\sigma)|$.

Finally, configurations and functions on Ω_{d+1} can be translated over $\tau_x, x \in \mathbb{Z}^{d+1} : \tau_x \sigma(y) = \sigma(y+x), \tau_x f(\sigma) = f(\tau_x \sigma)$. If clear from the context, we also write f_x for $\tau_x f$.

We consider families of local one-to-one (invertible) transformations π_Λ on S^Λ where Λ will vary in some large enough subset of $\mathcal{S}$ (which will be specified later on). As maps on Ω_{d+1} they have the properties that

1. $$\pi_\Lambda(\sigma)(x) = \sigma(x), x \in \Lambda^c; \tag{1.5}$$

2. $$\pi_\Lambda \circ \pi_\Lambda = 1. \tag{1.6}$$

3. $$\pi_\Lambda \circ \tau_x = \pi_{\Lambda+x} \tag{1.7}$$

4. $$\pi_\Lambda(\sigma)(x) = \pi_{\Lambda'}(\sigma)(x) \tag{1.8}$$

   for all $x \in \Lambda \subset \Lambda'$.

For every function f on Ω_{d+1}, we write $\pi_\Lambda f(\sigma) \equiv f(\pi_\Lambda(\sigma))$. The (product over Λ of the) counting measure on S^Λ is invariant under π_Λ. Notice that the function $\Delta_{\pi_\Lambda} f \equiv \pi_\Lambda f - f$ satisfies $\pi_\Lambda \Delta_{\pi_\Lambda} f = -\Delta_{\pi_\Lambda} f$. We give two interesting examples of such a transformation.

Example 1.1 *Take $\Lambda = \Lambda_{N,L}$ a rectangular shaped region centered at the origin with time-extension $2N+1$ and spatial volume $(2L+1)^d$. The transformation $\pi_\Lambda(\sigma)(j,m) = \sigma(j,-m), |j| \le L, |m| \le N$ time-reverses the space-time configuration in the window $\Lambda_{N,L}$.*

Example 1.2 *Take the Ising-case $S = \{+1,-1\}$ and $\pi_\Lambda(\sigma)(x) = -\sigma(x), x \in \Lambda$ corresponding to a spin-flip in $\Lambda \in \mathcal{S}$.*

Probability measures on $(\Omega_{d+1}, \mathcal{F})$ are denoted by $\mu, \nu, \ldots$. The corresponding random field is written as $X = (X(x), x \in \mathbb{Z}^{d+1})$. The expectation of a function f is written as $\mu(f) \equiv \int f(\sigma)\mu(d\sigma)$. As *a priori* measure we take the uniform product measure $d\sigma$ with normalized counting measure as marginals, for which $\int f(\sigma) d\sigma \equiv 1/|S|^{|\Lambda|} \sum_{\sigma_\Lambda} f(\sigma_\Lambda)$ when f is $\mathcal{F}_\Lambda$-measurable.

We will be dealing with Gibbs states μ in what follows; μ is a Gibbs measure with respect to the Hamiltonian H at inverse temperature β (and always with respect to the counting measure as *a priori* measure) when for every $\Lambda \in \mathcal{S}$ and for each pair of configurations $\sigma_\Lambda, \eta_\Lambda \in S^\Lambda$

$$\frac{\mu[X(x) = \sigma(x), x \in \Lambda | X(x) = \xi(x), x \in \Lambda^c]}{\mu[X(x) = \eta(x), x \in \Lambda | X(x) = \xi(x), x \in \Lambda^c]} = \exp[-\beta(H(\sigma_\Lambda \xi_{\Lambda^c}) - H(\eta_\Lambda \xi_{\Lambda^c}))] \tag{1.9}$$

for $\mu-$ almost every $\xi \in \Omega_{d+1}$. The Hamiltonian $H = \sum_A U_A$ is formally written as a sum of (interaction) potentials $U_A(\sigma) = U_A(\sigma_A)$ with well-defined relative energies $H(\sigma) - H(\eta)$ for $\{x \in \mathbb{Z}^{d+1} : \sigma(x) \neq \eta(x)\} \in \mathcal{S}$ if $\sum_{A \ni x} \|U_A\| < \infty, x \in Z^{d+1}$. Other weaker conditions than uniform absolute summability of the potential are possible. The essential Gibbs property is (1.9) which identifies the existence of a well-defined relative energy governing the relative weights of configurations that locally differ. (1.9) is the infinite volume version of the equivalent statement for finite volume Gibbs states

$$\mu_\Lambda(\sigma_\Lambda | \eta_{\Lambda^c}) = \frac{1}{Z_\Lambda^\beta(\eta)} \exp[-\beta \sum_{A \cap \Lambda \neq \emptyset} U_A(\sigma_\Lambda \eta_{\Lambda^c})], \tag{1.10}$$

with $Z_\Lambda^\beta(\eta)$ the normalizing factor (partition function with η boundary conditions).

Traditionally, Gibbs measures give the distribution of the microscopic degrees of freedom for a macroscopic system in thermodynamic equilibrium. The choice of the ensemble is determined by the experimental situation and is fixed by the choice of the relevant macro-variables. There is however no *a priori* reason to exclude nonequilibrium situations from the Gibbs formalism if considered as a procedure of statistical inference. Then the information concerning the nonequilibrium state (like obtained from measuring the currents) is incorporated in the ensemble. Moreover, as we will use in Section 3, one can in many cases explicitly construct the Gibbs states governing the space-time distribution as the path-space measure for the dynamics. The fact that these examples concern stochastic dynamics should not be regarded as a return to the strongly chaotic regime but rather as the proper way to deal with incomplete knowledge about the microscopic configuration of a system composed of a huge number of locally interacting components.

1.4 Main observation.

We start with the simplest observation. The rest will follow as immediate generalizations (with perhaps a slightly more complicated notation).

Look at (1.9). This Gibbs property implies that the image measure of μ under a transformation that affects only the spins in Λ is absolutely continuous with respect to μ with the Boltzmann-Gibbs factor as Radon-Nikodym derivative. Putting it simpler, it is an immediate consequence of the Gibbs property that for all continuous functions f

$$\mu(\pi_\Lambda f) = \mu(f W_\Lambda) \tag{1.11}$$

with $W_\Lambda \equiv \exp[-\beta \sum_{A\cap\Lambda\neq\emptyset}(\pi_\Lambda U_A - U_A)]$. But now the road is straight: take $f = W_\Lambda^{\lambda-1}$ in (1.11) and compute

$$\mu(W_\Lambda^\lambda) = \mu(W_\Lambda^{\lambda-1} W_\Lambda). \tag{1.12}$$

From (1.11) this is equal to

$$\mu(\pi_\Lambda W_\Lambda^{\lambda-1}) = \mu(W_\Lambda^{1-\lambda}) \tag{1.13}$$

where the last equality follows from $\pi_\Lambda(W_\Lambda^{\lambda-1}) = W_\Lambda^{1-\lambda}$. Thus, it is immediate that Gibbs measures satisfy

$$\mu(e^{-\lambda\beta R_\Lambda}) = \mu(e^{-(1-\lambda)\beta R_\Lambda}), \lambda \in \mathbb{R} \tag{1.14}$$

with relative energy $R_\Lambda \equiv \pi_\Lambda H - H$ corresponding to the transformation π_Λ:

$$R_\Lambda = \sum_{A\cap\Lambda\neq\emptyset} [\pi_\Lambda U_A - U_A]. \tag{1.15}$$

We now imagine the above for a sequence of volumes Λ growing to $\mathbb{Z}^{d+1}$ in a sufficiently regular manner (e.g. increasing cubes). Suppose now furthermore that μ is a Gibbs measure for a translation-invariant interaction potential and that

$$R_\Lambda(\sigma) = \sum_{x\in\Lambda} \tau_x J(\sigma) + h_\Lambda(\sigma) \tag{1.16}$$

with J a bounded continuous function and $||h_\Lambda||/|\Lambda| \to 0$ as Λ becomes infinite. This will be made explicit later on. Then, the following limit exists:

$$p(\lambda J|\mu) \equiv -\lim_\Lambda \frac{1}{|\Lambda|} \ln \mu[\exp(-\beta\lambda \sum_{x\in\Lambda} J_x)] \tag{1.17}$$

with $J_x \equiv \tau_x J$, and, from (1.14), it satisfies

$$p(\lambda J|\mu) = p((1-\lambda)J|\mu). \tag{1.18}$$

As a consequence, its Legendre transform

$$i_J(w|\mu) \equiv \sup_\lambda [p(\lambda J|\mu) - \lambda w] \tag{1.19}$$

satisfies

$$i_J(w|\mu) - i_J(-w|\mu) = -w. \tag{1.20}$$

It is not necessary (but it is possible) to employ the whole machinery of the theory of large deviations for Gibbs states to understand what this means: the probability law $P_\Lambda(w)$ for the random variable $\sum_{x\in\Lambda} J_x(X)/|\Lambda|$ as induced from the random field $(X(x), x \in \mathbb{Z}^{d+1})$ with distribution μ, behaves (for large Λ) as

$$P_\Lambda(w) \sim e^{-i_J(w|\mu)|\Lambda|}. \tag{1.21}$$

and the rate function $i_J(w|\mu)$ satisfies (1.20). Comparing this with (1.4), we see we have obtained exactly the same structure as in the Gallavotti-Cohen fluctuation theorem with practically no effort.

1.5 Plan.

We first present the fluctuation theorem in a Gibbsian context without too much reference to an underlying dynamics through which, possibly, the Gibbs states are obtained as space-time measures. Yet, to avoid misunderstanding, we repeat that we think of these Gibbs measures here as describing the steady states or symbolic dynamics for some spatially-extended non-equilibrium dynamics. They are to be thought of as distributions for the space-time histories. Via standard thermodynamic relations, we give the relation between the action functional satisfying the large deviation principle (fluctuation theorem) and the relative entropy between the forward and the backward evolution. In particular, in quadratic approximation, the Green-Kubo formula appears. Time enters explicitly in Section 3 where via the example of probabilistic cellular automata the general philosophy is illustrated.

2 Fluctuation theorem for Gibbs states.

In the present setup, we have no *a priori* reason to prefer one lattice direction over another and we fix the family of increasing cubes Λ_n of side length $n \in \mathbb{N}_0$ centered around the origin in which we are going to apply the transformations $\pi_{\Lambda_n} \equiv \pi_n$ having the properties described in Section 1.3. For every $A \in \mathcal{S}$ we write A_n for the smallest cube Λ_n (with $n = n(A)$) for which $A \subset \Lambda_n$.

2.1 Symmetry breaking potential.

In what follows we simply set $\Omega = \Omega_{d+1}$. A potential U is a real-valued function on $\mathcal{S} \times \Omega$ such that $U_A \in \mathcal{F}_A$ (i.e. only depending on the spins inside A) for each $A \in \mathcal{S}$ (put $U_\emptyset \equiv 0$). It describes the interaction between the spins in the region A. We consider a family of $m+1$ interaction potentials $(U_A^\alpha)_A, \alpha = 0, \dots, m$. We assume translation-invariance, meaning that

$$U_A^\alpha(\eta) = U_{A+x}^\alpha(\tau_x \eta), \tag{2.22}$$

for all $A \in \mathcal{S}, x \in \mathbb{Z}^{d+1}, \eta \in \Omega$. As usual we also take it that the total interaction of a finite region with the rest of the lattice is finite, i.e. we assume that the potential is uniformly absolutely summable:

$$\sum_{A \ni 0} \|U_A^\alpha\| < \infty. \tag{2.23}$$

(This assumption of uniformity is not strictly needed but it avoids irrelevant technicalities. Similarly, hard core interactions are also not excluded but extra care and assumptions would be needed.) Given the family of transformations π_n, we define the relative energies

$$R_n^\alpha \equiv \sum_{A \cap \Lambda_n \neq \emptyset} (\pi_n U_A^\alpha - U_A^\alpha), \alpha = 0, \dots, m. \tag{2.24}$$

We make a difference between the potential U^0 and the $U^\alpha, \alpha = 1, \dots, m$ from their behavior under the π_n. We assume that U^0 is invariant under the π_n in the sense that $\pi_n U^0_A = U^0_A$ whenever $n \geq n(A)$ implying that

$$\lim_n \frac{\|R^0_n\|}{|\Lambda_n|} = 0. \tag{2.25}$$

The reason for taking $m > 1$ is to allow for and to distinguish between possibly different mechanisms for breaking the symmetry of the reference interaction U^0.

We define the current associated to the symmetry breaking interaction $U^\alpha, \alpha = 1, \ldots, m$ to be

$$J^\alpha_x \equiv \lim_n \sum_{x \in A \subset \Lambda_n} \frac{1}{2|A|}(\pi_{n(A)} U^\alpha_A - U^\alpha_A). \tag{2.26}$$

J^α_x is a continuous function on Ω and, from (2.22), $J^\alpha_x(\eta) = J^\alpha_0(\tau_x \eta)$. The term 'current' is suggestive for interpreting (2.26) as the real current at the space-time point x associated to some driving of a reference steady state ν thereby breaking the time-reversal symmetry in the case of Example 1.1, see next section. We take ν to be a Gibbs state with respect to the interaction U^0, i.e. with formal Hamiltonian

$$H^0 \equiv \sum_A U^0_A, \tag{2.27}$$

(see (1.9)) for which the $\pi-$symmetry is unbroken:

$$\nu \circ \pi_n = \nu. \tag{2.28}$$

As a consequence, the currents (2.26) vanish identically in that state:

$$\nu(J^\alpha_x) = 0, \alpha = 1, \ldots, m. \tag{2.29}$$

The perturbed or driven state is denoted by μ. It is a translation-invariant Gibbs state at inverse temperature β with respect to the formal Hamiltonian

$$H \equiv H^0 + \sum_{\alpha=1}^m E^\alpha H^\alpha, \tag{2.30}$$

where the H^α are built (as in (2.27)) from the interaction potentials U^α and where the E^α are real numbers parameterizing the strength of a symmetry breaking or driving force. As before, in the definition of Gibbs states, we always take the normalized counting measure as *a priori* measure, see (1.9).

2.2 Fluctuation theorem.

Theorem 2.1 *Suppose that μ is a translation-invariant Gibbs state for the translation-invariant potential $(U_A = U^0_A + \sum_{\alpha=1}^m E^\alpha U^\alpha_A)_A$ as in the preceding subsection. The limit*

$$p(\lambda, E) \equiv -\lim_n \frac{1}{|\Lambda_n|} \ln \mu[e^{-\beta \sum_{x \in \Lambda_n} \sum_{\alpha=1}^m \lambda_\alpha J^\alpha_x}] \tag{2.31}$$

exists and satisfies

$$p(\lambda, E) = p(2E - \lambda, E) \tag{2.32}$$

for every $\lambda \equiv (\lambda_1, \ldots, \lambda_m)$ and $E \equiv (E^1, \ldots, E^m) \in \mathbb{R}^m$.

Proof: The existence of the limit is a standard result of the Gibbs formalism, see e.g. [12, 4, 27]. As announced via (1.11) the main observation leading to (2.32) is that

$$\mu(\pi_n f) = \mu(\exp[-\beta \sum_{A\cap\Lambda_n\neq\emptyset} (\pi_n U_A - U_A)] f), \tag{2.33}$$

simply because μ is a Gibbs state for the potential (U_A) at inverse temperature β. Therefore, taking numbers $h_\alpha, \alpha = 1,\dots,m$ and $f = \exp[\beta\sum_{\alpha=1}^m (1-h_\alpha)E^\alpha R_n^\alpha]$ in that formula,

$$\begin{aligned}
&\mu(\exp[-\beta R_n^0 - \beta\sum_{\alpha=1}^m h_\alpha E^\alpha R_n^\alpha]) = \\
&\mu(\exp[-\beta R_n^0 - \beta\sum_{\alpha=1}^m E^\alpha R_n^\alpha]\exp[\beta\sum_{\alpha=1}^m (1-h_\alpha)E^\alpha R_n^\alpha]) = \\
&\mu(\exp[-\beta\sum_{\alpha=1}^m (1-h_\alpha)E^\alpha R_n^\alpha]).
\end{aligned} \tag{2.34}$$

Now,

$$R_n^\alpha = 2\sum_{x\in\Lambda_n} J_x^\alpha - I_1 + I_2 \tag{2.35}$$

where both

$$I_1 \equiv \sum_{x\in\Lambda_n}\sum_{A\ni x, A\cap\Lambda_n^c\neq\emptyset} \frac{1}{|A|}(\pi_{n(A)}U_A^\alpha - U_A^\alpha) \tag{2.36}$$

and

$$I_2 \equiv \sum_{A\cap\Lambda_n, A\cap\Lambda_n^c\neq\emptyset} (\pi_n U_A^\alpha - U_A^\alpha) \tag{2.37}$$

are small of order $o(|\Lambda_n|)$ because of (2.23): $||I_i||/|\Lambda_n| \to 0$ as n goes to infinity, $i = 1, 2$. Upon inserting (2.35) into (2.34) and taking $h_\alpha = \lambda_\alpha/2E^\alpha$ (for $E^\alpha \neq 0$), we get

$$\frac{1}{|\Lambda_n|}\Big|\ln\frac{\mu(\exp[-\beta\sum_{x\in\Lambda_n}\sum_{\alpha=1}^m \lambda_\alpha J_x^\alpha])}{\mu(\exp[-\beta\sum_{x\in\Lambda_n}\sum_{\alpha=1}^m (2E^\alpha-\lambda_\alpha)J_x^\alpha])}\Big| \tag{2.38}$$

going to zero as $n\uparrow\infty$. This is exactly what was needed. ■

Remark 1: Gibbs states satisfy a large deviation principle, see e.g. [17] and [4] for additional references. As a result, (2.32) implies (1.21)-(1.20). We do not add a more precise formulation here.

Remark 2: Related to this, as is clear from the proof, the essential property is that the functionals $\{\log\frac{d(\mu\circ\pi_\Lambda)}{d\mu} : \Lambda\in\mathcal{S}\}$ satisfy a large deviation principle under μ. We speak about the (somewhat more restricted) Gibbs property because, in all cases we have in mind, the large deviations arise from Gibbsianness of the random field.

Remark 3: The theorem above provides a **local** version of the fluctuation theorem since the measure μ lives on a much larger (in fact, infinite) volume than the size of the observation window Λ_n. The relations (1.14) and (2.34) are identities exactly verified for the finite volumes Λ_n. This is similar to the local fluctuation theorem of [8]. Notice also that the limit $p(\lambda, E)$ exists and remains unchanged if instead of taking the sequence of cubes Λ_n we take volumes Λ growing to $\mathbb{Z}^{d+1}$ in the van Hove sense, see e.g. [12, 4, 27]. This will be exploited in the next section (Theorem 3.1) to separate time from the spatial volume.

Remark 4: The fluctuation theorem is formulated here (and elsewhere) on a volume-scale, anticipating large deviations which are exponentially small in the volume, see (1.21). This is certainly the typical behavior at high temperatures. However, the same reasoning of the proof above remains equally valid for other — less disordered — regimes where the large deviations may happen on another scale. As an example, suppose that

$$a(\lambda, E) \equiv -\lim_n \frac{1}{n^d} \ln \mu[e^{-\frac{\beta}{n}\sum_{x\in\Lambda_n}\sum_{\alpha=1}^m \lambda_\alpha J_x^\alpha}]. \tag{2.39}$$

Then, remembering that $\Lambda_n \sim n^{d+1}$, it also satisfies

$$a(\lambda, E) = a(2E - \lambda, E). \tag{2.40}$$

Such a scaling is applied in the study of large deviations in the phase coexistence regime where the probability of a droplet of the wrong phase is only exponentially small in the surface of that droplet.

2.3 Thermodynamic relation.

As mentioned in the introduction, the original context of the fluctuation theorem concerned the large deviations in the entropy production rate of a dynamical system. Since we have not specified any dynamics here, we must postpone a related discussion to the next section. Yet, we can compare with the thermodynamic potentials.

To start define the energy function

$$\Phi_0(U) \equiv \sum_{A\ni 0} \frac{U_A}{|A|} \tag{2.41}$$

and its translations $\Phi_x(U)(\eta) = \Phi_0(U)(\tau_x \eta)$. We define the free energy density for the interaction U as

$$P(U) \equiv \lim_n \frac{1}{|\Lambda_n|} \ln \sum_{\sigma\in\Omega_{\Lambda_n}} \exp[-\beta \sum_{A\subset\Lambda_n} U_A(\sigma)]). \tag{2.42}$$

This coincides with

$$P(U) = \lim_n \frac{1}{|\Lambda_n|} \ln Z^\beta_{\Lambda_n}(\eta) \tag{2.43}$$

of (1.10) for all boundary conditions η.

Finally, the entropy density of a translation-invariant probability measure μ is

$$s(\mu) \equiv -\lim_n \frac{1}{|\Lambda_n|} \sum_{\sigma\in\Omega_{\Lambda_n}} \mu_n[\sigma] \ln \mu_n[\sigma] \geq 0 \tag{2.44}$$

where $\mu_n[\sigma]$ is the probability for the measure μ to find the configuration σ in the box Λ_n (and $0 \ln 0 = 0$). The relative entropy density between two translation-invariant probability measures μ and ρ (with $\rho_n(\sigma) = 0$ implying $\mu_n(\sigma) = 0$)

is

$$s(\mu|\rho) \equiv \lim_n \frac{1}{|\Lambda_n|} \sum_{\sigma \in \Omega_{\Lambda_n}} \mu_n[\sigma] \ln \frac{\mu_n[\sigma]}{\rho_n[\sigma]} \geq 0. \tag{2.45}$$

If μ is a translation-invariant Gibbs measure (at inverse temperature β) for the interaction U, then

$$P(U) = s(\mu) - \beta\mu(\Phi_0(U)). \tag{2.46}$$

For a given interaction (U_A) we also like to have around the free energy functional $F(U,\rho)$ defined for translation-invariant probability measures ρ by

$$F(U,\rho) \equiv s(\rho) - \beta\rho(\Phi_0(U)). \tag{2.47}$$

We have, besides $F(U,\mu) = P(U)$ for the Gibbs measures μ with respect to U, see (2.46), that

$$P(U) > F(U,\rho) \tag{2.48}$$

for all translation-invariant probability measures ρ which are not Gibbs measures for U at inverse temperature β (Gibbs' variational principle).

The $\pi-$ transformed interaction potential πU is defined via

$$\pi U_A = \pi_{n(A)} U_A \tag{2.49}$$

and the $\pi-$ transformed measure $\pi\mu$ is obtained by its expectations for all local functions f:

$$\pi\mu(f) = \mu \circ \pi_n(f) \tag{2.50}$$

for $n = n(f)$ so that $D_f \subset \Lambda_n$. Clearly, $P(\pi U) = P(U)$ by the assumed $\pi-$invariance of the counting measure, see (1.3). (This also follows from observing that $-P(\pi U) + P(U) = p(2E,E) = p(0,E) = 0$ by (2.32) and (2.31).) For the same reason, $s(\pi\mu) = s(\mu)$ and if μ is a Gibbs measure for U, then $\pi\mu$ is a Gibbs measure for πU (and vice versa). (To avoid trivialities, it is understood that the interaction πU is not physically equivalent with U as long as some $E^\alpha \neq 0$.)

We next show that the averaged current (whose fluctuations are investigated in Theorem 2.1) is always (strictly) positive as it equals a relative entropy density. To link it also to a free energy production we must require that the free energy $P(U + t(\pi U - U))$ is differentiable with respect to t at $t = 0$. For this (see e.g. [12]), it suffices e.g. that

$$\sum_{A \ni 0} |A| \|U_A^0\|, \sum_{\alpha=1}^m E^\alpha \sum_{A \ni 0} |A| \|U_A^\alpha\| < 1. \tag{2.51}$$

Proposition 2.1 *For the Gibbs measure $\mu \neq \pi\mu$ of Theorem 2.1,*

$$s(\pi\mu|\mu) = s(\mu|\pi\mu) = 2\beta \sum_{\alpha=1}^m E^\alpha \mu(J_0^\alpha) > 0 \tag{2.52}$$

and, under the assumption (2.51), is also given via

$$\sum_{\alpha=1}^m E^\alpha \mu(J_0^\alpha) = -\frac{1}{2\beta} \frac{\partial}{\partial t} P(U + t(\pi U - U))(t = 0). \tag{2.53}$$

Proof: The positivity follows from the variational principle:

$$s(\mu) - \beta\mu(\Phi_0(U)) = P(U) > F(U, \pi\mu) = s(\mu) - \beta\pi\mu(\Phi_0(U)). \tag{2.54}$$

As is well known the relative entropy (2.45) can be rewritten as a difference of free energies: $s(\pi\mu|\mu) = P(U) - F(U, \pi\mu)$. We can now use that

$$\beta\mu(\Phi_0(\pi U) - \Phi_0(U)) = -\frac{\partial}{\partial t}P(U + t(\pi U - U))(t = 0) \tag{2.55}$$

is exactly equal to $2\beta\sum_{\alpha=1}^{m} E^\alpha \mu(J_0^\alpha)$, as required.■

Remark 1: The positivity of (2.52) should be compared with (1.3). The positivity of the entropy production is discussed in [23, 24]. The positivity of (2.52) just follows here from the Gibbs' variational principle: with $\dot{s}_n \equiv \sum E^\alpha \sum_{x\in\Lambda_n} J_x^\alpha$, for $\mu-$ almost every $\sigma, \dot{s}_n(\sigma)/|\Lambda_n| \to s(\mu|\pi\mu)/2\beta > 0$ where the almost sure convergence assumes that μ is a phase. That $s(\mu|\pi\mu)$ has something to do with entropy production will become clear in the next section when a dynamics and the time-reversal operation is considered.
Remark 2: Thinking about $s(\mu|\pi\mu)$ as entropy production, (2.52) gives the usual bilinear expression in terms of thermodynamic fluxes and forces. Remember that the dependence of $p(\lambda, E)$ on E in (2.31) comes from the state μ. The E^α correspond to field strengths or amplitudes producing energy- or particle flow. Of course, on the formal level above, the distinction must remain arbitrary and one can of course include the E^α in the potentials U_A^α.

2.4 Green-Kubo formula.

It has been observed in other places, [16, 19, 7], that the fluctuation theorem quite directly gives rise to various familiar formulae of linear response. We will not pursue this matter here very far except for repeating the simplest derivations.

Assuming smoothness of the free energy in the external fields, we differentiate (2.32) with respect to E^γ and $\lambda_\alpha, \alpha, \gamma = 1, \ldots, m$ at $E = \lambda = 0$:

$$\frac{\partial}{\partial E^\gamma}\frac{\partial}{\partial \lambda_\alpha}p(0,0) = -\frac{\partial}{\partial E^\gamma}\frac{\partial}{\partial \lambda_\alpha}p(0,0) - 2\frac{\partial}{\partial \lambda_\gamma}\frac{\partial}{\partial \lambda_\alpha}p(0,0). \tag{2.56}$$

On the other hand,

$$\frac{\partial}{\partial \lambda_\alpha}p(0,E) = \beta\mu(J_0^\alpha) \tag{2.57}$$

while

$$\frac{\partial}{\partial \lambda_\gamma}\frac{\partial}{\partial \lambda_\alpha}p(0,0) = -\beta^2\sum_x \nu(J_0^\alpha J_x^\gamma). \tag{2.58}$$

Conclusion,

$$\frac{\partial}{\partial E^\gamma}\mu(J_0^\alpha)(E = 0) = \beta\sum_x \nu(J_0^\alpha J_x^\gamma), \tag{2.59}$$

and the change in relative entropy $s(\mu|\pi\mu)$ (see (2.52)) from Proposition 2.1 is in quadratic approximation for small E given by

$$s(\mu|\pi\mu) = 2\beta^2\sum_{\alpha,\gamma} E^\alpha E^\gamma \sum_x \nu(J_0^\alpha J_x^\gamma). \tag{2.60}$$

Equation (2.59) is a standard Green-Kubo relation while (2.60) expresses the relative entropy density $s(\mu|\pi\mu)$ (or change in free energy) in terms of the current-current correlations (with the obvious analogues of Onsager symmetries). In conclusion, we have identified a (model-dependent) continuous function

$$\dot{S}(\sigma) \equiv \sum_{\alpha} E^{\alpha} J_0^{\alpha}(\sigma) \tag{2.61}$$

with $\pi\dot{S} = -\dot{S}, \mu(\dot{S}) > 0, \nu(\dot{S}) = 0$ and symmetric response matrix

$$\frac{\partial}{\partial E^{\gamma}}\mu(\frac{\partial}{\partial E^{\alpha}}\dot{S})(E=0) = \beta \sum_{x} \nu(J_0^{\alpha} J_x^{\gamma}). \tag{2.62}$$

Symmetries in higher order terms can be obtained by taking higher derivatives of the generating formula (2.32).
The notation $\dot{S}$ should not be read as a time-derivative (change of entropy in time). More appropriate will be to regard $\mu(\dot{S})/2$ as the limit $(S_f - S_i)/T$ as time T goes to infinity of the total change of entropy $S_f - S_i$ in a reservoir during the nonequilibrium process. The reservoir is initially in equilibrium with thermodynamic entropy S_i and after absorbing the heat dissipated by the nonequilibrium process it reaches a new equilibrium with entropy S_f. We will come back to this once time has been explicitly introduced (in Section 3).

3 Fluctuation theorem for PCA.

PCA (short for probabilistic cellular automata) are discrete time parallel updating stochastic dynamics for lattice spin systems. They are used in many contexts but we see them here as interesting examples of non-equilibrium dynamics. We refer to [13, 18] for details and examples and we restrict ourselves here to the essentials we need. We work with time-homogeneous translation-invariant nearest-neighbor PCA which are specified by giving the single-site transition probabilities

$$0 < p_{i,n}(a|\sigma) = p_i(a|\sigma(j, n-1), |j-i| \leq 1) < 1, a \in S, \sigma \in \Omega_{d+1}. \tag{3.63}$$

This defines a Markov process $(X_n)_{n=0,1\ldots}$ on Ω_d for which for all finite $V \subset \mathbb{Z}^d$,

$$\text{Prob}\,[X_n(i) = a_i, \forall i \in V | X_{n-1}] = \prod_{i \in V} p_i(a_i | X_{n-1}(j), |i-j| \leq 1), a_i \in S \tag{3.64}$$

with some given initial configuration $X_0 = \xi \in \Omega_d$. Notice that we have kept the notation σ for a general configuration on the space-time lattice. Remember that $x = (i, n) \in \mathbb{Z}^{d+1}$ stands for a space-time point with time-coordinate n at site $i \in \mathbb{Z}^d$.
The π_Λ are restricted to time-reversal transformations and the volumes Λ are to grow first in the time-direction (for a fixed spatial window).

3.1 Steady state fluctuation theorem.

If we take a translation-invariant stationary state ρ of a PCA as above, then its Markov extension defines a translation-invariant Gibbs measure μ for the (formal) Hamiltonian

$$H(\sigma) = -\sum_{i,n} \ln p_{i,n}(\sigma(i,n)|\sigma(\cdot, n-1)). \tag{3.65}$$

We refer to [13, 18] for a precise formulation. μ describes the distribution of the space-time configurations in the steady state and its restriction to any spatial layer is equal to the stationary state ρ we started from. To characterize ρ, one must study the projection of μ to a layer (see [20] for a variational characterization of such a projection).

Since μ is Gibbsian we can try applying the theory of the previous section. Most interesting is to consider a sequence of rectangular boxes $\Lambda_{L,N} \equiv \{x = (i,n) \in \mathbb{Z}^{d+1} = \mathbb{Z}^d \times \mathbb{Z} : |i| \le L, |n| \le N\}$. The idea is that we wish to keep the spatial size L much smaller than the time-extension $N >> L$. As transformation we take $\pi_{\Lambda_{L,N}} \equiv \pi_{L,N}$ corresponding to a time-reversal:

$$\pi_{L,N}\sigma(i,n) = \sigma(i,-n), |n| \le N, |i| \le L \tag{3.66}$$

and $\pi_{L,N}\sigma(i,n) = \sigma(i,n)$ whenever $(i,n) \notin \Lambda_{L,N}$.

Define the current

$$J_{i,n}(\sigma) \equiv \ln p_i(\sigma(i,n)|\sigma(\cdot,n-1)) - \ln p_i(\sigma(i,n-1)|\sigma(\cdot,n)). \tag{3.67}$$

Notice that in contrast with the previous section, we do not specify here the unperturbed state (but one can always take some homogeneous product measure) and we take $m = 1 = 2E$ for simplicity. $J_{i,n}$ is a local function and it is the space-time translate of J_0. In the same way as in (2.24), we define

$$R_{L,N}(\sigma) \equiv H(\pi_{L,N}\sigma) - H(\sigma). \tag{3.68}$$

Starting from (3.65) $R_{L,N}$ can be written out as a finite sum but most important is that

$$R_{L,N}(\sigma) = \sum_{n=-N+1}^{N-1} \sum_{|i| \le L-1} J_{i,n}(\sigma) + G_{L,N}(\sigma), \tag{3.69}$$

where

$$\|G_{L,N}\| \le c(2N+1)(2L+1)^{d-1} + c'(2L+1)^d \le c(d)NL^{d-1}, \tag{3.70}$$

with a constant $c(d)$ depending on the dimension d and on the transition probabilities (3.63). We are therefore in a position to repeat the fluctuation Theorem 2.1 in that context.

Theorem 3.1 *Take $L = L(N) \le N$ growing to infinity as $N \uparrow \infty$. The limit*

$$e(\lambda) \equiv -\lim_N \frac{1}{|\Lambda_{L,N}|} \ln \mu(\exp[-\lambda \sum_{x \in \Lambda_{L-1,N-1}} J_{i,n}]) \tag{3.71}$$

exists for all real λ and

$$e(\lambda) = e(1-\lambda). \tag{3.72}$$

Moreover, for fixed L,

$$e_{L,N}(\lambda) \equiv -\frac{1}{N} \ln \mu(\exp[-\lambda \sum_{x \in \Lambda_{L-1,N-1}} J_{i,n}]) \tag{3.73}$$

(which, generally, is of order L^d) satisfies

$$|e_{L,N}(\lambda) - e_{L,N}(1-\lambda)| \leq c(\lambda, d)L^{d-1} \tag{3.74}$$

uniformly in $N \uparrow \infty$.

Proof: The proof is a copy of the proof of Theorem 2.1. As before, we have automatically, from the Gibbs property (as in (1.14)), that

$$\mu(\exp[-\lambda R_{L,N}]) = \mu(\exp[-(1-\lambda)R_{L,N}]). \tag{3.75}$$

We now substitute (3.69) and use the estimate (3.70) to perform the limits.■

Remark 1: One may wonder about the existence of the limit $e_L(\lambda) \equiv \lim_N e_{L,N}(\lambda)$ for fixed L. This is certainly expected when the steady-state μ is a high temperature Gibbs state. In that case, the limit $\lim_L e_L(\lambda)/L^d = \lim_L e_L(1-\lambda)/L^d$ satisfies (3.72).
Remark 2: Some quite similar results were discussed already in [8]. There however the dynamics was deterministic (weakly coupled strongly chaotic maps). There again, the methods of [2, 3, 22, 15] can reduce the problem to a higher dimensional symbolic dynamics and the methods of the previous section are ready for use.

3.2 Entropy production.

The measure μ gives the probability distribution of the space-time histories in a steady-state. It is therefore natural to consider $s(\mu)$ (see (2.44)) as its specific entropy rate (i.e, entropy per unit volume and per unit time). In terms of the stationary state ρ we have (see [13]) that

$$s(\mu) = -\rho(\sum_{a\in S} p_0(a|\sigma(\cdot,-1)) \ln p_0(a|\sigma(\cdot,-1)). \tag{3.76}$$

On the other hand, the free energy density vanishes identically for PCA (because of the normalization in (3.65), see [13, 18]), so that, from (2.54),

$$P(U) = 0, P(U) - F(U,\pi\mu) = -s(\mu) + \beta\pi\mu(\Phi_0(U)). \tag{3.77}$$

Hence, still in the notation of the previous section, whenever $P(U) = 0$ (which is verified for PCA),

$$-s(\mu|\pi\mu) = s(\mu) - \beta\pi\mu(\Phi_0(U)). \tag{3.78}$$

(This formula is not correct when we replace in it μ by $\pi\mu$.) That is interesting because we found that now $s(\mu|\pi\mu) > 0$ is minus the specific entropy rate $s(\mu)$ modulo a term which is linear in μ. Writing this out in our present notation, this is nothing else than

$$-\mu(J_0) = -\rho(\sum_{a\in S} p_0(a|\sigma(\cdot,-1)) \ln p_0(a|\sigma(\cdot,-1)) - \mu(-\ln p_0(\sigma(0)|\sigma(\cdot,1)). \tag{3.79}$$

The first term to the right is the specific entropy rate (3.76) (always positive) and the second term (linear in μ) subtracts from this exactly so much that the net-result to the left vanishes in the case of time-reversal symmetry (detailed balance). Of course, as in Proposition 2.1 we have an equality between the averaged current in the steady state μ and the relative entropy $s(\mu|\pi\mu)$ (remember that we took $m = 1 = 2E!$). We can therefore conclude that indeed $\mu(J_0)$ or $s(\mu|\pi\mu)$ must be regarded as the (positive) entropy production by our dynamics. The current associated to the breaking of time-reversal symmetry gives rise to nothing else than the local (in space-time) entropy production whose fluctuations we have investigated in Theorems 2.1 and 3.1. The points made in Section 2.4 related to the Green-Kubo formula remain unaltered and we do not repeat them here.

Yet, to obtain a physically inspiring picture, we should connect the above analysis to measurable quantities. The (second part of the) second law of thermodynamics connects the thermodynamic entropy of an initial and final equilibrium state after some thermodynamically irreversible process has taken place. In an adiabatic non-quasi-stationary process the entropy can only increase: $S_f > S_i$. If we now were to rerun the process in the opposite direction, simply by (thermodynamically) inverting all the currents (by changing the sign of all gradients of the intensive variables), again the entropy would increase and by the same amount as before ($s(\mu) = s(\pi\mu)$) and we would reach a new equilibrium with entropy equal to $S_i + 2(S_f - S_i)$. While we lack at this point a more formal understanding, we believe that our entropy production exactly measures that difference: $[S_f - S_i] - [S_i - S_f] = 2(S_f - S_i) = s(\mu|\pi\mu) > 0$. More generally and depending on the physical realization of the process, these considerations must apply to the relevant thermodynamic potential and 'entropy production' must for example be replaced by 'work done' or 'free energy production.'

We will further illustrate this by an example in the following subsection but it is interesting to remark already that $s(\mu|\pi\mu)$ reproduces, via the formal analogies on the level of the variational principle (both for Gibbs and for SRB states), the entropy production in the context of the theory of dynamical systems. There we have that the entropy production is given by (1.3) with $\rho(\dot{S})$ equal to the sum of the positive Lyapunov exponents with respect to ϕ^{-1} minus the sum of positive Lyapunov exponents with respect to ϕ. If ρ is singular with respect to $d\xi$ and has no vanishing Lyapunov exponent, then $\rho(\dot{S}) > 0$, see [23]. In our case, $s(\mu|\pi\mu) = \beta\pi\mu(\Phi_0(U)) - \beta\mu(\Phi_0(U)) = P(U) - F(U, \pi\mu) > 0$.

3.3 Illustration.

We take here a closer look at the current (3.67) for Markov chains. The spatial degree of freedom $i \in \mathbb{Z}^d$ has now disappeared and we must study

$$J_n(\sigma) \equiv \ln p(\sigma(n)|\sigma(n-1)) - \ln p(\sigma(n-1)|\sigma(n)) \tag{3.80}$$

for $\sigma \in \Omega_1$ and transition probabilities

$$\text{Prob}\,[X_n = a|X_{n-1} = b] = p(a|b), a, b \in S \tag{3.81}$$

for the stationary S-valued Markov chain X_n. The steady state μ is now a homogeneous one-dimensional Gibbs measure and its single-time restriction is the stationary measure ρ on $S : \sum_b p(a|b)\rho(b) = \rho(a), a \in S$.

The steady state expectation of the current (3.80) is

$$\mu(J) = \sum_b \rho(b) \sum_a p(a|b)[\ln p(a|b) - \ln p(b|a)]. \tag{3.82}$$

Now use that the transition probabilities $q(\cdot|\cdot)$ for the reversed chain ($Y_n \equiv X_{-n})_n$ (with distribution $\pi\mu$ but with the same stationary measure ρ) are given by

$$q(a|b) \equiv \text{ Prob } [X_n = a|X_{n+1} = b] = p(b|a)\frac{\rho(a)}{\rho(b)}. \tag{3.83}$$

Since $\sum_b \rho(b)\sum_a p(a|b)\ln\rho(b) = \sum_b \rho(b)\ln\rho(b) = \sum_b \rho(b)\sum_a p(a|b)\ln\rho(a)$, we can substitute (3.83) into (3.82) ($q(a|b)$ for $p(b|a)$) with no extra cost and we obtain

$$\mu(J_0) = \rho(S(p|q)) \tag{3.84}$$

where

$$S(p|q) = \sum_a p(a|\cdot)\ln\frac{p(a|\cdot)}{q(a|\cdot)} \geq 0. \tag{3.85}$$

is the relative entropy between the forward and the backward transition probabilities. (3.85) is zero only if the Markov chain is time-reversible (in which case $\mu = \pi\mu$). Then, (3.83) for $q(a|b) = p(a|b)$ becomes the detailed balance condition. Relation (3.84) is nothing but (2.52) specified to the context of Markov chains.

A second less trivial and physically interesting illustration can be taken from a model of hopping conductivity. It is a bulk driven diffusive lattice gas where charged particles, subject to an on-site exclusion, hop on a ring in the presence of an electric field. The configuration space is $\Omega_1 = \{0,1\}^{\mathcal{T}}$ with $\xi(i) = 0$ or 1 depending on whether the site $i \in \mathcal{T}$ is empty or occupied. We take for $\mathcal{T}$ the set $\{1,\ldots,\ell\}$ with periodic boundary conditions. To each bond $(i, i+1)$ in the ring and independently of all the rest there is associated a Poisson clock (with rate 1). If the clock rings and $\xi(i) = 1, \xi(i+1) = 0$ then the particle at i jumps to $i+1$ with probability p. If on the other hand, $\xi(i) = 0, \xi(i+1) = 1$ the particle jumps to i with probability q. Therefore, the 'probability per unit time' to make the transition from ξ to $\xi^{i,i+1}$ (in which the occupations of i and $i+1$ are interchanged) is given by the exchange rate

$$c(i, i+1, \xi) = p\xi(i)(1 - \xi(i+1)) + q\xi(i+1)(1 - \xi(i)). \tag{3.86}$$

and should be thought of as a continuous time analogue of (3.63). It is natural to call $E = \ln p/q$ the electric field. This model is called the asymmetric simple exclusion process and it is also considered in [19]. Strictly speaking, it is not a PCA but a continuous time process with sequential updating. However, since it is a jump process, the change with respect to the PCA of above just amounts to randomizing the time between successive transitions.

Each uniform product measure ρ is time-invariant for this process and we consider the steady state μ starting in this invariant state. If we now consider a realization σ of the process in which at a certain time, when the configuration is $\xi \in \Omega_1$, a particle hops from site i to $i+1$, then the time-reversed trajectory shows a particle jumping from $i+1$ to i. The contribution of this event to the entropy production is therefore

$$\ln c(i, i+1, \xi) - \ln c(i, i+1, \xi^{i,i+1}) = E[\xi(i)(1-\xi(i+1)) - \xi(i+1)(1-\xi(i))]. \tag{3.87}$$

This formula is the continuous time analogue of (3.80) or (3.67) (but we do not take $E = 1/2$ here) with ξ the configuration right before the jump and $\xi^{i,i+1}$

the configuration right after the jump in the trajectory σ. Of course, this jump in σ itself happens with a rate $c(i, i+1, \xi)$. We see therefore that the derivative of (3.87) with respect to E has expectation

$$\mu(J_{i,t}) = \rho(c(i, i+1, \xi)[\xi(i)(1-\xi(i+1)) - \xi(i+1)(1-\xi(i))]) = (p-q)u(1-u) \tag{3.88}$$

for $u \equiv \rho(\xi(i))$ the density. (3.88) is indeed the current as it appears in the hydrodynamic equation, here the Burgers equation, through which a density profile evolves. The fluctuations of the particle current satisfy (2.32) or (3.72) (with $E = 1/2$), see also [19]. The entropy production (as in (3.84)-(3.85)) is

$$\frac{1}{2} s(\mu|\pi\mu) = \rho(c(i, i+1, \xi) \ln \frac{c(i, i+1, \xi)}{c(i, i+1, \xi^{i,i+1})}) = E(p-q)u(1-u) \tag{3.89}$$

which is the field times the current and is left invariant by changing E into $-E$. If, to be specific, we take $p = 1/(1+e^{-E}) = 1-q$, then, in quadratic approximation,

$$\frac{1}{2} s(\mu|\pi\mu) = u(1-u)E^2 \tag{3.90}$$

which is the dissipated heat through a conductor in an electric field E with Ohmic conductivity $u(1-u) = \mu(J_0^2)(E=0) = \rho(c(0,1,\xi)[\xi(0)(1-\xi(1)) - \xi(1)(1-\xi(0))]^2)$ given by the variance of the current. This model (together with the models discussed in [19]) illustrates that the methods exposed in the present paper are not restricted to just PCA. We have restricted us here to a somewhat informal treatment of the aspects concerning the entropy production in the model as it will be included in a future publication dealing with the local fluctuation theorem, [21].

4 Concluding remark.

It does not seem unreasonable that Gibbs' variational principle determining the conditions of equilibrium can be generalized to certain nonequilibrium conditions. In this note we have shown that describing the steady state via the standard methods of the Gibbs formalism leads directly to the fluctuation theorem. This is true close or far from equilibrium because it follows quite generally from the defining Gibbs property itself. From this 'Gibbsian' point of view, applying the local fluctuation theorem to various specific models is to add specific observable consequences to the studies of E.T. Jaynes, [14].

References

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

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

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

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

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

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

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

[8] Gallavotti, G. (1998) A local fluctuation theorem. Preprint.

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

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

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

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

[13] Goldstein, S., Kuik, R., Lebowitz, J.L. and Maes, C. (1989) From PCA's to Equilibrium Systems and Back, *Comm. Math. Phys.* **125**, 71–79.

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

[15] Miaohuang Jiang and Pesin, Y.B. (1997) Equilibrium Measures for Coupled Map Lattices: Existence, Uniqueness and Finite-Dimensional Approximations. Preprint.

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

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

[18] Lebowitz, J.L., Maes, C. and Speer, E.R. (1990) Statistical mechanics of probabilistic cellular automata, *J. Stat. Phys.* **59**, 117–170.

[19] Lebowitz, J.L. and Spohn, H. (1998) The Gallavotti-Cohen Fluctuation Theorem for Stochastic Dynamics, Rutgers University preprint.

[20] Maes, C., Redig, F. and Van Moffaert, A. (1998) The restriction of the Ising model to a layer. Preprint.

[21] Maes, C., Redig, F. and Van Moffaert, A. (1998) Work in progress.

[22] Pesin, Y.B. and Sinai, Y.G. (1991) Space-time chaos in chains of weakly interacting hyperbolic mappings, *Adv. Sov. Math.* **3**, 165–198.

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

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

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

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

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