Entropy, Thermostats and Chaotic Hypothesis

Giovanni Gallavotti
Fisica and I.N.F.N. Roma 1
September 4, 2006

Abstract: *The chaotic hypothesis is proposed as a basis for a general theory of nonequilibrium stationary states.*

1. *Stationary states and thermostats.*

The problem is to develop methods to establish relations between time averages of a few observables associated with a system of particles subject to work-performing external forces and to thermostat-forces that keep the energy from building up, so that it can be considered in a stationary state.

The stationary state will correspond to a probability distribution on phase space $\mathcal{F}$ so that

$$\frac{1}{\tau}\int_0^\tau F(S_t x)\,dt \xrightarrow[\tau\to\infty]{} \int_{\mathcal{F}} F(y)\,\mu(dy) \tag{1.1}$$

for all x but a set of zero volume: $x \to S_t x$ is the solution flow defined by a differential equation on $\mathcal{F}$:

$$\dot x = \mathbf{f}_{\mathbf{E}}(x) \tag{1.2}$$

where $\mathbf{f}_{\mathbf{E}}$ contains internal forces, external forces depending on a few parameters $\mathbf{E} = (E_1, \ldots, E_n)$, and thermostats forces. In general the divergence

$$\sigma(x) = -\sum_j \partial_{x_j} f_{\mathbf{E},j}(x) \tag{1.3}$$

is not zero, except in absence of external forces $\mathbf{E}$ and of thermostat forces (*i.e.* in the equilibrium case).

A fairly realistic example is the following:

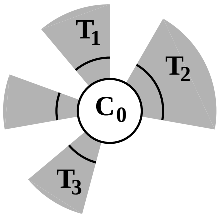

Fig.1 "Thermostats", or reservoirs, occupy finite regions outside C_0, *e.g.* sectors $C'_i \subset R^3$, $i = 1, 2\ldots$, marked T_i located beyond "buffers" $\mathcal{C}_a$: the buffers (representing a the *walls* separating the system from the thermostats) simply have their boundaries marked. The reservoir particles are constrained to have a *total* kinetic energy K_i constant, by suitable forces $\boldsymbol{\vartheta}_i$, so that their "temperatures" T_i, see (1.5), are well defined, [1]. Buffers and reservoirs have *arbitrary sizes.*

The system contains N_0 particles in a configuration $\mathbf{X}_0$ contained in $\mathcal{C}_0$ and N_i, N'_i particles in configurations that will be denoted $\mathbf{X}_i, \mathbf{X}'_i$ contained in the buffer regions $\mathcal{C}_i$, henceforth called *wall.* and in the thermostat regions $\mathcal{C}'_i$, $i = 1, \ldots, n$, respectively. The equations of motion are, for $i = 0$ and $i > 0$ respectively,

$$\begin{aligned}
\ddot{\mathbf{X}}_0 &= -\partial_{\mathbf{X}_0}(U_0(\mathbf{X}_0) + \textstyle\sum_{i>0} W_{0i}(\mathbf{X}_0, \mathbf{X}_i)) + \mathbf{E}(\mathbf{X}_0)\\
\ddot{\mathbf{X}}_i &= -\partial_{\mathbf{X}_i}(U_i(\mathbf{X}_i) + W_{0i}(\mathbf{X}_0, \mathbf{X}_i) + W_{i,i'}(\mathbf{X}_i, \mathbf{X}_{i'}))\\
\ddot{\mathbf{X}}'_i &= -\partial_{\mathbf{X}'_i}(U'_i(\mathbf{X}'_i) + W_{i,i'}(\mathbf{X}_i, \mathbf{X}'_i)) - \alpha_i\,\dot{\mathbf{X}}'_i
\end{aligned} \tag{1.4}$$

where U_i, U'_i are the interaction energies for the particles in $\mathcal{C}_i$, $i = 0, 1, \ldots, n$ and in $\mathcal{C}'_i$, $i = 1, \ldots, n$; $\mathbf{E}(\mathbf{X}_0)$ is the external force working on the system in $\mathcal{C}_0$ and $-\alpha_i \dot{\mathbf{X}}'_i$ is the *thermostat force*: which is the force prescribed by *Gauss' principle of least effort*, see Appendix A9.4 in [2], to impose the contraints ($k_B \equiv$ Boltzmann's constant)

$$\frac{1}{2}\dot{\mathbf{X}}_i'^2 = \frac{d}{2} N'_i k_B T_i, \qquad i = 1, \ldots, n \tag{1.5}$$

which gives, after a simple application of the principle,

$$\alpha_i = \frac{L_i - \dot U'_i}{N'_i\, k_B T_i} \tag{1.6}$$

where L_i is the work done per unit time by the particles $\mathbf{X}_i \in \mathcal{C}_i$ on those in $\mathbf{X}'_i \in \mathcal{C}'_i$, *i.e.* on the thermostats.

Other thermostat models could be considered: however their particular structure should not influence the statistical properties of the particles in $\mathcal{C}_0$. In particular I think that replacing the container $\mathcal{C}'_i$ with an *infinite* container in which particles are initially in a state that is an equilibrium Gibbs state at temperature T_i should lead to the same results: this is a conjecture whose proof seems quite far at the moment.

In the following we shall suppose that the interaction potentials are due to bounded smooth pair interaction potentials plus, possibly hard core elastic potentials and we regard the equations (1.4) as first order equations on the phase space coordinates $x \equiv \{\dot{\mathbf{X}}_i, \mathbf{X}_i\}_{i=0}^n$. As such the equations do not conserve volume of phase space: in fact the divergence of the equations in this space is $-\sigma(x)$ with

$$\begin{aligned}
\sigma(x) =& \sum_{i>0} \frac{L_i}{k_B T_i}\frac{dN'_i - 1}{dN'_i} - \sum_{i>0}\frac{\dot U'_i}{k_B T_i}\frac{dN'_i - 1}{dN'_i} =\\
=& \sum_{i>0}\frac{L_i}{k_B T_i}\frac{dN'_i - 1}{dN'_i} + \dot\Phi
\end{aligned} \tag{1.7}$$

where $\Phi \overset{def}{=} -\sum_{i>0} \frac{U'_i}{k_B T_i}\frac{dN'_i - 1}{dN'_i}$, as it can be checked by direct computation.

Since $L_i = -\dot X'_i \cdot \partial_{X'_i} W_{i,i'} \equiv +\dot X_i \cdot \partial_{X_i} W_{i,i'} - \dot W_{i,i'}$ and the expression (1.7) is the sum over $i > 0$ of $-\frac{d}{dt}\Big(\frac{1}{2}\dot X_i^2 +$

$U_i\Big) - \dot X_i \partial_{X_i} W_{i,0}$ which has the form $\dot\Psi_i + Q_i$ where Q_i is the work per unit time done by the forces due to particles in $\mathcal{C}_0$ on the particles in $\mathcal{C}_i$: we identify therefore Q_i with the *heat* generated per unit time by the forces acting on $\mathcal{C}_0$ and transfered first to the walls $\mathcal{C}_i$ and, subsequently, to the thermostats in $\mathcal{C}'_i$.

Thus setting $\varepsilon(x) \overset{def}{=} \sum_{i>0} \frac{Q_i}{k_B T_i}$ it is (for notational simplicity, and keeping in mind that N'_i should be thought as large, we shall neglect $O({N'_i}^{-1})$)

$$\sigma(x) = \varepsilon(x) + \dot R \tag{1.8}$$

where $R(x) = -\sum_i \frac{W_{i,i'} + U'_i + U_i + \frac12 \dot X_i{}^2}{k_B T_i}$.

Remark: (1) In this model, as well as in a large number of others, one has therefore the natural interpretation of $\sigma(x)$ as the *entropy creation* per unit time: this is because for large time the average of the l.h.s., $\sigma(x)$, over a time interval and the corresponding average of $\varepsilon(x) = \sum_{i>0} \frac{Q_i}{k_B T_i}$ become equal at large time because they differ by $\frac{1}{\tau}(R(S_t x) - R(x))$, at least if R is bounded, as it is convenient to suppose for simplicity. This is a strong assumption but it will not be discussed here: it has to do with the problem of thermostats "efficiency" and its violation may lead to interesting consequences, see [1, 3].
(2) It should be noted that the walls $\mathcal{C}_i$ could be missing and the particles in $\mathcal{C}_0$ be directly in contact with the thermostats: in this case there will be no $W_{i,i'}$ but instead there will be potentials $W_{0,i'}$: the analysis would be entirely analogous with $\frac{Q_i}{k_B T_i}$ replaced by $\frac{Q'_{i'}}{k_B T_i}$ with $Q'_{i'}$ being the work per unit time done by the particles in $\mathcal{C}_0$ on the thermostat particles in $\mathcal{C}'_i$ and $R = -\sum_i \frac{W_{0,i'} + U'_i}{k_B T_i}$. In this case if the interaction potentials are bounded the R will be also bounded without any extra assumption.
(3) The L_i in Eq.(1.7) is the work ceded by the walls to thermostats: therefore it can be interpreted as the heat Q'_i ceded by the particles in $\mathcal{C}_i$ to the thermostat in $\mathcal{C}'_i$: hence the alternatice representation $\sigma(x) = \varepsilon'(x) + \dot\Phi$, (1.7), is possible with $\varepsilon'(x) = \sum_{i>0} \frac{Q'_i}{k_B T_i}$. Also in this case the remainder Φ is bounded if the interaction potentials are bounded and the discussion that follows applies to both $\varepsilon(x)$ and $\varepsilon'(x)$, which are thus equivalent for the purpose of fluctuation analysis.

2. *The hypothesis.*

Chaotic Hypothesis: *Motions developing on the attracting set of a chaotic system can be regarded as a transitive hyperbolic system.*

A general result is that transitive hyperbolic systems have the property (1.1), with μ a uniquely determined probability distribution on phase space, [4].

Of course a flow can be studied via a *Poincaré map* S defined by a *timing event* Σ. The latter is defined by a surface in phase space which is crossed by all trajectories infinitely many times (typically Σ is the union of a few connected surface elements $\Sigma = \cup_i \Sigma_i$, but in general it is *not* connected: *i.e.* it is a finite collection of connected pieces). The timing event occurs when a trajectory crosses Σ at a point x and time t_0: and S maps it into the next timing event Sx occurring, at some time t_1, on the trajectory $t \to S_t x$: hence $x' = Sx \overset{def}{=} S_{t_1 - t_0} x \in \Sigma$.

For model (1.4) there is a direct relation between $\sigma(x)$, $x \in \Sigma$, and the Jacobian determinant $\det \partial_x S(x)$; setting $R(t_1) \equiv R(S_{t_1 - t_0} x)$, $R(t_0) \equiv R(x)$, it is

$$\begin{aligned} -\log|\det \partial_x S(x)| &= \int_{t_0}^{t_1} \sigma(S_t x)\, dt = \\ &= \int_{t_0}^{t_1} \varepsilon(S_t x)\, dt + R(t_1) - R(t_0) = \\ &= \sum_{i>0} \frac{\int_{t_0}^{t_1} Q_i\, dt}{k_B T_i} + R(t_1) - R(t_0) \end{aligned} \tag{2.1}$$

The theory of evolutions described by flows or described by maps are therefore very closely related as the above remarks show, at least for what concerns the analysis of the entropy creation rate and its fluctuations.

The second viewpoint should be taken whenever $\sigma(x)$ has singularities: which can happen if the interaction potentials are unbounded (*e.g.* of Lennard-Jones type) or if the thermostats sizes tend to infinity, see [5].

3. Dimensionless entropy and fluctuation theorem.

Interesting properties to study are related to the fluctuations of *entropy creation* averages. Restricting the analysis to the model (1.4), define the *entropy creation* rate to be

$$\varepsilon_+ = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau \sigma(S_t x)\, dt = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau \varepsilon(S_t x)\, dt \tag{3.1}$$

by the remark at the end of Sec.1.

Assuming that the system is *dissipative*, which by definition will mean $\varepsilon_+ > 0$, consider the random variable

$$p \overset{def}{=} \frac{1}{\tau} \int_0^\tau \frac{\varepsilon(S_t x)}{\varepsilon_+}\, dt \tag{3.2}$$

that will be called the dimensionless phase space contraction and considered with the distribution inherited from the SRB-distribution μ of the system.

A general property of random variables of the form $a = \frac{1}{\tau} \int_0^\tau F(S_t x)\, dt$, which are time averages over a time τ of a smooth observable F, is that, if motions are transitive and hyperbolic, the SRB-probability distribution μ that a is in a closed interval Δ has the form

$$P_\mu(a \in \Delta) = \exp(\tau \max_{a\in\Delta} \zeta_F(a) + O(1)) \tag{3.3}$$

for $\Delta \subset (a_-, a_+)$, where $a_\pm$ are two suitable values within which the function $\zeta_F(a)$ is defined, analytic and convex; the *fluctuation interval* $[a_-, a_+]$ contains the μ–average value of F and if $\Delta \cap [a_-, a_+] = \emptyset$ the probability $P_\mu(a \in \Delta)$ tends to 0 as $\tau \to \infty$ faster than exponentially. For this reason the function $\zeta_F(a)$ can be naturally defined also for $a \notin [a_-, a_+]$ by giving it the value $-\infty$, [4, 6–8]. Finally $O(1)$ means a quantity which is bounded as $\tau \to \infty$ at Δ fixed.

The function $\zeta_F(a)$ is called the *large deviations rate* for the fluctuations of the observable F.

If the motions are also *reversible*, *i.e.* if there is an isometry I of phase space such that $IS_t = S_{-t}I$ or $IS = S^{-1}I$, in the case of time evolution maps, any observable F which is odd under time reversal, *i.e.* $F(Ix) = -F(x)$ will have a fluctuation interval $[-a^*, a^*]$ symmetric around the origin (and containing the SRB–average $\overline{a}$ of F).

In the case of the model (1.4) time reversibility corresponds to the velocity inversion and the evolution is reversible in the just defined sense. The fluctuation interval of $\sigma(x)/\varepsilon_+$ and of $\varepsilon(x)/\varepsilon_+$ is therefore symmetric around the origin and $p^* \geq 1$ because the averages of the two observables are 1 by definition, see (3.1),(3.2).

A general theorem that holds for transitive, hyperbolic motions is the following

Fluctuation theorem: *Given a hyperbolic, transitive and reversible system assume that the SRB average σ_+ of the phase space contraction $\sigma(x)$, i.e. that the divergence of the equations of motion (1.3), is $\sigma_+ > 0$. Consider the dimensionless phase space contraction $\sigma(x)/\sigma_+$: this is an observable which has a large deviations rate $\zeta(p)$ defined in a symmetric interval $(-p^*, p^*)$ and satisfying there*

$$\zeta(-p) = \zeta(p) - p\sigma_+ \tag{3.4}$$

Remarks: (i) The (3.4) can be regarded as valid for all p's if we follow the mentioned convention of defining $\zeta(p) = -\infty$ for $p \notin [-p^*, p^*]$.

(ii) By the chaotic hypothesis, abridged CH, it follows that a relation like (3.4) should hold for the SRB distribution of the dimensionless phase space contraction of any reversible chaotic motion with a dense attractor or, more generally, for dimensionless phase space contraction of the motions restricted to the attracting set, if a time reversal symmetry holds on the motions restricted to the attracting set, [9, 10]. Of course this is not a theorem (mainly because hyperbolicity is a hypothesis) but it should nevertheless apply to many interesting cases.

(iii) In particular it should apply to the model (1.4): actually in this case it has already been remarked that the observable $\sigma(x)/\sigma_+$ and the *dimensionless entropy creation rate* $\varepsilon(x)/\varepsilon_+$ have the *same large deviations function*; hence (3.4) should hold for the rate function of $p = \frac{1}{\tau}\int_0^\tau \sum_{a>0} \frac{Q_a}{k_B T_a\, \varepsilon_+}\, dt$:

$$\zeta(-p) = \zeta(p) - p\varepsilon_+, \qquad p \in (-p^*, p^*) \tag{3.5}$$

(iv) The latter remark is interesting because the quantity $\varepsilon(x) \overset{def}{=} \sum_{a>0} \frac{Q_a}{k_B T_a}$ has a physical meaning and can be measured in experiments like the one described in Fig.1 or in experiments for which there is not an obvious equation of motion (*i.e.* no obvious model).

(v) Therefore in applications the relation (3.5) is expected to hold quite generally and, in the general cases, it is called *fluctuation relation*, abridged FR, to distinguish it from the Fluctuation Theorem.

(vi) Furthermore the quantity $\varepsilon(x)$ is a *local* quantity as it depends only on the microscopic configurations of the system $\mathcal{C}_0$ and of the walls $\mathcal{C}_i$ in the immediate vicinity of their separating boundary. In particular the relation (3.5) does not depend on what happens in the bulk of the walls $\mathcal{C}_i$ or on the size of the thermostats $\mathcal{C}'_i$: hence the latter can be taken to infinity. One can also imagine that (3.5) remains valid in the case of infinite thermostats whose particles are initially distributed so that their emprical distribution is asymptotically a Gibbs state at temperature T_a.

(vii) The last few comments suggest quite a few tests of the chaotic hypothesis and of the corresponding fluctuation relation in various cases, see for instance [11]. Therefore the fluctuation relation, first suggested by the simulation in [12], where it has been discovered in an experiment motivated to test ideas emerging from the SRB theory, and subsequently proved as a theorem for Anosov systems in [13, 14], gave rise to the chaotic hypothesis and at the moment experiments are being designed to test its predictions.

(viii) The theorem will be referred as FT. It is often written in the form, see (3.3),(3.4),

$$\lim_{\tau\to\infty} \frac{1}{\tau} \log \frac{P_\mu(p \in \Delta)}{P_\mu(p \in -\Delta)} = \sigma_+ \max_{p\in\Delta} p \tag{3.6}$$

for $\Delta \subset (0, p^*)$ or in the more suggestive, although slightly imprecise, form:

$$\lim_{\tau\to\infty} \frac{1}{\tau} \log \frac{P_\mu(p)}{P_\mu(-p)} = p\,\sigma_+ \tag{3.7}$$

which can be regarded valid for $p \in (-p^*, p^*)$.

(ix) It is natural to think that the special way in which the thermostats are implemented is not important as long as the notion of temperature of the thermostats is clearly understood. For instance an alternative thermostat could be a stochastic one with particles bouncing off the walls with a Maxwellian velocity distribution at temperature depending on the wall hit. In this context the experiment in [15] appears to give an interesting confirmation.

4. *Extending Onsager-Machlup's fluctuations theory*

A remarkable theory on nonequilibrium fluctuations has been started by Onsager and Machlup, [16, 17], and concerns fluctuations near equilibrium and, in fact, it only deals with properties of derivatives with respect to the external forces parameters $\mathbf{E}$ *evaluated* at $\mathbf{E}=\mathbf{0}$.

The object of the analysis are *fluctuation patterns*: the question is which is the probability that the successive values of $F(S_tx)$ follow, for $t\in[-\tau,\tau]$, a preassigned sequence of values, that I call *pattern* $\varphi(t)$, [18].

In a reversible hyperbolic and transitive system consider n observables $F_1,\ldots,F_n$ which have a well defined parity under time reversal $F_j(Ix)=\pm F_j(x)$. Given n functions $\varphi_j(t)$, $j=1,\ldots,n$, defined for $t\in[-\frac{\tau}{2},\frac{\tau}{2}]$ the question is: which is the probability that $F_j(S_tx)\sim\varphi_j(t)$ for $t\in[-\frac{\tau}{2},\frac{\tau}{2}]$? the following *FPT theorem* gives an answer:

Fluctuation Patterns Theorem: *Under the assumptions of the fluctuation theorem given F_j,φ_j, and given $\varepsilon>0$ and an interval $\Delta\subset(-p^*,p^*)$ the joint probability with respect to the SRB distribution*

$$\begin{aligned}&\frac{P_\mu(|F_j(S_tx)-\varphi_j(t)|_{j=1,\ldots,n}<\varepsilon,p\in\Delta)}{P_\mu(|F_j(S_tx)\mp\varphi_j(-t)|_{j=1,\ldots,n}<\varepsilon,-p\in\Delta)}=\\&=\exp(\tau\max_{p\in\Delta}p\,\sigma_++O(1))\end{aligned}\tag{4.1}$$

where the sign choice $\mp$ is opposite to the parity of F_j and $p\overset{def}{=}\frac{1}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}}\frac{\sigma(S_tx)}{\sigma_+}\,dt$.

Remarks: (i) The FPT theorem means that "all that has to be done to change the time arrow is to change the sign of the entropy production", *i.e.* the *time reversed processes occur with equal likelyhood as the direct processes if conditioned to the opposite entropy creation.* This is made clearer by rewriting the above equation in terms of probabilities *conditioned on a preassigned value of* p; in fact up to $e^{O(1)}$ it becomes, [2], for $|p|<p^*$:

$$\frac{P_\mu(|F_j(S_tx)-\varphi_j(t)|_{j=1,\ldots,n}<\varepsilon,\,|p)}{P_\mu(|F_j(S_tx)\mp\varphi_j(-t)|_{j=1,\ldots,n}<\varepsilon,\,|-p)}=1\tag{4.2}$$

(ii) An immediate consequence is that defining f_i the averages $f_i\overset{def}{=}\frac{1}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}}F_j(S_tx)$ then the SRB probability that $f_1,\ldots,f_n$ occur in presence of an entropy creation rate p is related to the occurrence of $\mp f_1,\ldots,\mp f_n$ in presence of the opposite entropy creation rate: in a slightly imprecise form, see remark (viii) in Sec.3 and (3.7), this means that

$$\lim_{\tau\to\infty}\frac{1}{\tau}\log\frac{P_\mu(f_1,\ldots,f_n,p)}{P_\mu(\mp f_1,\ldots,\mp f_n,-p)}=p\,\sigma_+.\tag{4.3}$$

(iii) In particular if F_j are odd under time reversal and p can be expressed as an (obviously odd) function of $f_1,\ldots,f_n$: $p=\pi(f_1,\ldots,f_n)$ the (4.3) can be written, [18],

$$\lim_{\tau\to\infty}\frac{1}{\tau}\log\frac{P_\mu(f_1,\ldots,f_n)}{P_\mu(-f_1,\ldots,-f_n)}=\pi(f_1,\ldots,f_n)\,\sigma_+\tag{4.4}$$

for $\pi(f_1,\ldots,f_n)\in(-p^*,p^*)$: a particular case of this relation is relevant for Kraichnan's theory of turbulence, [19].
(iv) An interesting application, [20, 21], of(4.3) with $j_j(x)=\partial_{E_j}\sigma(x)$ is that, setting $J_j=\mu(j_j)\equiv\langle j_i\rangle_\mu$, it is

$$L_{jk}=\partial_{E_k}J_j|_{\mathbf{E}=\mathbf{0}}=L_{kj}\tag{4.5}$$

Since in several interesting cases J_j have the interpretation of "thermodynamic currents" (*i.e.* currents divided by k_BT if T is the temperature) generated by the "thermodynamic forces" E_j the (4.5) have the interpretation of *Onsager reciprocal relations.* In fact also the expressions

$$L_{jk}=\frac{1}{2}\int_{-\infty}^{\infty}\mu(\sigma_k(S_tx)\sigma_j(x))_{\mathbf{E}=\mathbf{0}}\,dt\tag{4.6}$$

follow from FPT and have the interpretation of *Green–Kubo formulae.* The above relations have been derived under the extra simplification that $\sigma\equiv0$ for $\mathbf{E}=\mathbf{0}$ which is satisfied in several cases, see comment following Eq.(3.5) in [21]. However what is really necessary is that $\langle\sigma\rangle_{\mathbf{E}=\mathbf{0}}=\mathbf{0}$, which is an even weaker assumption because the analysis in [20] is, *verbatim*, unchanged if instead of $\sigma=0$ for $\mathbf{E}=\mathbf{0}$ one has $\langle\sigma\rangle_{\mathbf{E}=\mathbf{0}}=\mathbf{0}$.
(v) The assumption of reversibility at $\mathbf{E}\neq\mathbf{0}$, which is necessary for the FPT, is not really necessary to derive (4.6) (hence (4.5)) as shown in [22] where such relations are derived under the only assumption that for just $\mathbf{E}=\mathbf{0}$ the motions is reversible.
(vi) A further application of FPT is its relation with the theory of intermittency, see [23, 24].
(vii) The above analysis and the arbitrariness of the walls $\mathcal{C}_i$ hints that even if the thermostating mechanism is quite different, for instance it is generated by viscous forces $-\nu_i\dot{\mathbf{X}}_i$ hence not reversible, nevertheless the quantity $\varepsilon(x)$ will satisfy a FR.
(viii) In any event it appears that the total phase space divergence $\sigma(x)$ is not directly physically relevant and in fact *it is not physically meaninglful.* Since it differs from the physically measurable entropy production $\varepsilon(x)$ by a total derivative it can only be used to infer properties of the latter, as done in the FR: of course a FR will hold also for $s(x)$ in the reversible cases. However given the possibly very large (arbitrarily large) size of the contributions to $\sigma(x)$ due to the total derivative $\dot{R}(x)$ to (2.1), or (1.8), the time scale for the large fluctuations of $p'=\frac{1}{\tau}\int_0^\tau\frac{\sigma(S_tx)}{\varepsilon_+}$ easily becomes unobservably large while

the time scale for the fluctuations of $\varepsilon(x)$ remains independent on the size of the walls $\mathcal{C}_i$ and of the thermostats $\mathcal{C}'_i$.

(ix) Finally the FPT should not be confused with an identity which has the same form as (4.1),(4.2) with P_μ replaced by P_{μ_0} with μ_0 an *equilibrium Gibbs' distribution*, the interval $[-\tau,\tau]$, where the patterns $\varphi_j(t)$ are defined, replaced by $[0,\tau]$ and p replaced by by the *non normalized* average phase space contraction $a=\frac{1}{\tau}\int_0^\tau \sigma(S_tx)\,dt$. The (trivial) identity is, under the assumption of reversibility of the evolution,

$$\frac{P_{\mu_0}(|F_j(S_tx)-\varphi_j(t)|_{j=1,\ldots,n}<\varepsilon,\,|\,a)}{P_{\mu_0}(|F_j(S_tx)\mp\varphi_j(\tau-t)|_{j=1,\ldots,n}<\varepsilon,\,|\,-a)}=1 \tag{4.7}$$

which is obtained by the argument given in [25] for the special case $n=0$ (*i.e.* for just the fluctuations of a) and it should be compared with the very different (4.3).

5. JF, BF and fluctuation relations

An immediate consequence of FT is that

$$\langle e^{-\int_0^\tau \varepsilon(S_tx)\,dt}\rangle_{SRB}=e^{O(1)} \tag{5.1}$$

i.e. $\langle e^{-\int_0^\tau \varepsilon(S_tx)\,dt}\rangle_{SRB}$ stays bounded as $\tau\to\infty$. This is a relation that I will call *Bonetto's formula* and denote it BF, see Eq.(9.10.4) in [2]; it can be also written, somewhat imprecisely and for mnemonic purposes, [26],

$$\langle e^{-\int_0^\tau \varepsilon(S_tx)\,dt}\rangle_{SRB}\xrightarrow[\tau\to\infty]{}1 \tag{5.2}$$

which *would be exact* if the FT in the form (3.7) held for finite τ (rather than in the limit as $\tau\to\infty$).

This relation bears resemblance to *Jarzinsky's formula*, henceforth JF, which deals with a canonical Gibbs distribution (in a finite volume) corresponding to a Hamiltonian $H_0(p,q)$ and temperatute $T=(k_B\beta)^{-1}$, and with a time dependent family of Hamiltonians $H(p,q,t)$ which interpolates between H_0 and a second Hamiltonian H_1 as t grows from 0 to 1 (in suitable units) which is called *a protocol*.

Imagine to extract samples (p,q) with a canonical probability distribution $\mu_0(dpdq)=Z_0^{-1}e^{-\beta H_0(p,q)}dpdq$, with Z_0 being the canonical partition function, and let $S_{0,t}(p,q)$ be the solution of the Hamiltonian *time dependent* equations $\dot p=-\partial_qH(p,q,t),\dot q=\partial_pH(p,q,t)$ for $0\le t\le 1$. Then JF, [27, 28], gives:

Let $(p',q')\overset{def}{=}S_{0,1}(p,q)$ and let $W(p',q')\overset{def}{=}H_1(p',q')-H_0(p,q)$, then the distribution $Z_1^{-1}e^{-\beta H_1(p',q')}dp'dq'$ is exactly equal to $\frac{Z_0}{Z_1}e^{-\beta W(p',q')}\mu_0(dpdq)$. Hence

$$\langle e^{-\beta W}\rangle_{\mu_0}=\frac{Z_1}{Z_0}=e^{-\beta\Delta F(\beta)} \tag{5.3}$$

where the average is with respect to the Gibbs distribution μ_0 and ΔF is the free energy variation between the equilibrium states with Hamiltonians H_1 and H_0 respectively.

Remark: (i) The reader will recognize in this *exact identity* an instance of the Monte Carlo method. Its interest lies in the fact that it can be implemented *without actually knowing* neither H_0 nor H_1 nor the *protocol* $H(p,q,t)$. If one wants to evaluate the difference in free energy bewteen two equilibrium states at the same temperature of a system that one can construct in a laboratory then "all one has to do" is

(a) to fix a protocol, *i.e.* a procedure to transform the forces acting on the system along a well defined *fixed once and for all* path from the initial values to the final values in a fixed time interval ($t=1$ in some units), and

(b) measure the energy variation W generated by the machines implementing the protocol. This is a really measurable quantity at least in the cases in which W can be interpreted as the work done on the system, or related to it.

Then average of the exponential of $-\beta W$ with respect to a large number of repetition of the protocol. This can be useful even, and perhaps mainly, in biological experiments.

(ii) If the "protocol" conserves energy (like a Joule expansion of a gas) or if the difference $W=H_1(p',q')-H_0(p,q)$ has zero average in the equilibrium state μ_0 we get, by Jensen's inequality (*i.e.* by the convexity of the exponential function $\langle e^A\rangle\ge e^{\langle A\rangle}$), that $\Delta F\le 0$ as expected from Thermodynamics.

(iii) The measurability of W is a difficult question, to be discussed on a case by case basis. It is often possible to identify it with the "work done by the machines implementing the protocol".

The two formulae (5.2) and (5.3) are however quite different:

(1) the $\int_0^\tau\sigma(S_tx)\,dt$ is an entropy creation rather than the energy variation W.

(2) the average is over the SRB distribution of a stationary state, in general out of equilibrium, rather than on a canonical equilibrium state.

(3) the BF says that $\langle e^{-\int_0^\tau \varepsilon(S_tx)\,dt}\rangle_{SRB}$ is bounded, (5.1), as $\tau\to\infty$ rather than being 1 exactly. However a careful analysis of the meaning of W would lead to concluded that also JF necessitates corrections, particularly in thermostatted systems, [28].

The JF has proved useful in various equilibrium problems (to evaluate the free energy variation when an equilibrium state with Hamiltonian H_0 is compared to one with Hamiltonian H_1); hence it has some interest to investigate whether (5.2) can have some consequences.

If a system is in a steady state and produces entropy at rate ε_+ (*e.g.* a living organism feeding on a background) the FT (3.4) and is consequence BF, (5.2), gives us informations on the the fluctuations of entropy production, *i.e.* of heat produced, and (5.2) *could be useful*, for instance, to check that all relevant heat transfers have been properly taken into account.

Acknowledgements: *Source of a talk at the Durham Symposium "Dynamical Systems and Statistical Mechanics", 3-13 July 2006. I am indebted to E.G.D. Cohen for stimulating and clarifying comments.*

[1] G. Gallavotti, Chaos **16**, 023130 (+7) (2006).
[2] G. Gallavotti, *Statistical Mechanics. A short treatise* (Springer Verlag, Berlin, 2000).
[3] P. Garrido and G. Gallavotti, cond-mat/06?????? **??**, ?? (2006).
[4] R. Bowen and D. Ruelle, Inventiones Mathematicae **29**, 181 (1975).
[5] F. Bonetto, G. Gallavotti, A. Giuliani, and F. Zamponi, Journal of Statistical Physics **123**, 39 (2006).
[6] Y. G. Sinai, Russian Mathematical Surveys **27**, 21 (1972).
[7] Y. G. Sinai, *Lectures in ergodic theory*, Lecture notes in Mathematics (Princeton University Press, Princeton, 1977).
[8] Y. G. Sinai, *Topics in ergodic theory*, vol. 44 of *Princeton Mathematical Series* (Princeton University Press, 1994).
[9] F. Bonetto, G. Gallavotti, and P. Garrido, Physica D **105**, 226 (1997).
[10] F. Bonetto and G. Gallavotti, Communications in Mathematical Physics **189**, 263 (1997).
[11] F. Bonetto, G. Gallavotti, A. Giuliani, and F. Zamponi, Journal of Statistical Mechanics (cond-mat/0601683) p. P05009 (2006).
[12] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Physical Review Letters **70**, 2401 (1993).
[13] G. Gallavotti and E. G. D. Cohen, Physical Review Letters **74**, 2694 (1995).
[14] G. Gentile, Forum Mathematicum **10**, 89 (1998).
[15] F. Bonetto, N. Chernov, and J. L. Lebowitz, Chaos **8**, 823 (1998).
[16] L. Onsager and S. Machlup, Physical Review **91**, 1505 (1953).
[17] L. Onsager and S. Machlup, Physical Review **91**, 1512 (1953).
[18] G. Gallavotti, chao-dyn/9703007, in Annales de l' Institut H. Poincaré **70**, 429 (1999).
[19] R. Chetrite, J. Y. Delannoy, and K.Gawedzki, nlin.CD/0606015 (2006).
[20] G. Gallavotti, Physical Review Letters **77**, 4334 (1996).
[21] G. Gallavotti, Journal of Statistical Physics **84**, 899 (1996).
[22] G. Gallavotti and D. Ruelle, Communications in Mathematical Physics **190**, 279 (1997).
[23] G. Gallavotti, Journal of Mathematical Physics (physics/0001071) **41**, 4061 (2000).
[24] G. Gallavotti, Markov processes and Related fields (nlin.CD/0003025) **7**, 135 (2001).
[25] E. G. D. Cohen and G. Gallavotti, Journal of Statistical Physics **96**, 1343 (1999).
[26] G. Gallavotti, Chaos **8**, 384 (1998).
[27] C. Jarzynski, Journal of Statistical Physics **98**, 77 (1999).
[28] C. Jarzynski, Physical Review Letters **78**, 2690 (1997).