ANNALES DE L'I. H. P., SECTION A

GIOVANNI GALLAVOTTI

Fluctuation patterns and conditional reversibility in nonequilibrium systems

Annales de l'I. H. P., section A, tome 70, nº 4 (1999), p. 429-443
<http://www.numdam.org/item?id=AIHPA_1999__70_4_429_0>

Ann. Inst. Henri Poincaré,

Vol. 70, n° 4, 1999, p. 429-443. *Physique théorique*

Fluctuation patterns and conditional reversibility in nonequilibrium systems*

by

Giovanni GALLAVOTTI
I.H.E.S., 91440 Bures s/Yvette, France
e-mail: giovanni@ipparco.roma1.infn.it

ABSTRACT. – Fluctuations of observables as functions of time, or "*fluctuation patterns*", are studied in a chaotic microscopically reversible system that has irreversibly reached a nonequilibrium stationary state. Supposing that during a certain, long enough, time interval the average entropy creation rate has a value s and that during another time interval of the same length it has value $-s$ then we show that the relative probabilities of fluctuation patterns in the first time interval are the same as those of the reversed patterns in the second time interval. The system is "conditionally reversible" or irreversibility in a reversible system is "driven" by the entropy creation: while a very rare fluctuation happens to change the sign of the entropy creation rate it also happens that the time reversed fluctuations of all other observables acquire the same relative probability of the corresponding fluctuations in presence of normal entropy creation. A mathematical proof is sketched. © Elsevier, Paris

RÉSUMÉ. – On étudie les fluctuations des observables en tant que fonction du temps, ou « formes des fluctuations » dans un système chaotique à dynamique microscopique réversible ayant atteint de façon irréversible un état stationnaire hors d'équilibre. Si l'on suppose que pendant un laps de temps assez long le taux moyen de création d'entropie soit « s » et pendant un autre égale intervalle de temps, il soit « s », alors on montre que les probabilités relatives d'observer des formes des fluctuations dans le premier

* This work is part of the research program of the European Network on: "Stability and Universality in Classical Mechanics", # ERBCHRXCT940460.

intervalle de temps sont les mêmes que d'observer dans le second intervalle les formes « inverses », c'est-à-dire les formes « correspondantes » sous inversion du temps. On peut dire que le système montre une « réversibilité conditionnée » ou que l'irréversibilité est « pilotée » par la création d'entropie. Si une fluctuation très rare parvient à changer le signe du taux de création d'entropie alors il arrive aussi que les fluctuations « inverses » de toutes observables deviennent relativement aussi probables que les « correspondantes » en présence d'un taux normal (c'est-à-dire moyen) de création d'entropie. On esquisse une preuve mathématique.

1. ENTROPY GENERATION

We consider a reversible mechanical system governed by a smooth equation:

$$\dot{x} = f(x, \underline{G}) \tag{1.1}$$

depending on several parameters $\underline{G} = (G_1, \ldots, G_n)$ measuring the strength of the forces acting on the system and causing the evolution $x \to S_t x$ of the phase space point x representing the system state in the *phase space* $\underline{F}$ which can be, quite generally, a smooth manifold.

We suppose that the system is "thermostatted" so that motions take place on bounded smooth invariant surfaces $H(x; \underline{G}) = E$, which are level surfaces of some "level function" H. Hence we shall identify, to simplify the notations, the manifold $\underline{F}$ with this level surface.

We suppose that the flow S_t generated by (1.1) is *reversible*, *i.e.* there is an isometric smooth map I, "time reversal", of phase space "anticommuting with time" and such that $I^2 = 1$:

$$S_t I = I S_{-t} \tag{1.2}$$

i.e. $f(Ix) = -(\partial I)(x) \cdot f(x)$. If $-\sigma(x; \underline{G})$ is the rate of change of the volume element of $\underline{F}$ near x and under the flow S_t (it would be equal to $\sum_\alpha \partial_\alpha f_\alpha(x; \underline{G})$ if the space $\underline{F}$ was a euclidean space) we shall call σ the *entropy generation rate* and we suppose that it has the properties:

$$\sigma(Ix; \underline{G}) = -\sigma(x; \underline{G}), \qquad \sigma(x; \underline{0}) = 0 \tag{1.3}$$

so that at zero forcing the evolution is volume preserving (a property usually true because the non forced system is very often Hamiltonian).

Our analysis concerns idealized systems of the above type that are also transitive *Anosov flows* on each energy surface.

We recall that a Anosov flow S_t is a flow, without equilibrium points, solving a smooth differential equation $\dot{x} = f(x)$ on a smooth compact manifold; the flow is such that at every point y one can define a *stable tangent plane* T^s_y, and an *unstable tangent plane* T^u_y with:

1) the tangent planes T^u_y and T^s_y are linearly independent and vary continuously with y and T^s_y, T^u_y togheter with the vector $f(y)$ span the full tangent plane at y.

2) they are *covariant*: $\partial S_t T^\alpha_y = T^\alpha_{S_t y}$, $\alpha = u, s$

3) there exist constants $C,\ l > 0$ such that if $\xi \in T^s_y$ then $|\partial S_t \xi| \le Ce^{-\ lt}|\xi|$ for all $t \ge 0$, and if $\xi \in T^u_y$ then $|\partial S_{-t}\xi| \le Ce^{-\ lt}|\xi|$ for all $t \ge 0$.

A Anosov flow is *mixing* or *transitive* if given any two open sets A, B there is a $t_{A,B}$ such that $S_t A \cap B \neq \emptyset$ for $|t| > t_{A,B}$.

It follows, see [1], that there exists a unique probability distribution μ on $\underline{F}$ such that almost all initial data x (choosing them with a probability distribution proportional to the volume) generate motions that "admit a statistics μ", *i.e.*:

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T F(S_t x)dt = \int \mu(dy)F(y) \stackrel{def}{=} \langle F\rangle_+ \tag{1.4}$$

and μ is called the *SRB distribution*. The average of F with respect to μ will be denoted $\langle F\rangle_+$.

We shall further restrict our attention to transitive Anosov systems that are reversible, in the above sense, for all values of the forcing parameters $\underline{G}$ of interest and *dissipative* at $\underline{G} \neq \underline{0}$. This means that the systems we consider are such that:

$$\langle\sigma\rangle_+ > 0 \qquad \text{for } \underline{G} \neq \underline{0} \tag{1.5}$$

Under the above assumptions one can define, for $\langle\sigma\rangle_+ > 0$, the "dimensionless average entropy creation rate" p by setting:

$$p = \frac{1}{\langle\sigma\rangle_+ \tau}\int_{-\tau/2}^{\tau/2} \sigma(S_t;\, \underline{G})dt \tag{1.6}$$

Then the probability distribution of the variable p with respect to the SRB distribution μ can be written for large τ as $\pi_\tau(p)dp = e^{-\tau\zeta_\tau(p)}dp$, see [2], and the function $\zeta(p) = \lim_{\tau\to\infty} \zeta_\tau(p)$ verifies, if $\sigma_+ \equiv \langle\sigma\rangle_+ > 0$ and $|p| \le p^*$ for a suitable $p^* \ge 1$, the property:

$$\frac{\zeta(-p) - \zeta(p)}{p\langle\sigma\rangle_+} = 1 \tag{1.7}$$

which is called a *fluctuation theorem*, and is part of a class of theorems proved in [3] for discrete time systems, and in [4], for continuous time systems. This theorem can be considerably extended, as discussed in [5] and the extension can be shown to imply, in the limit $\underline{G} \to \underline{0}$ (when also $\sigma_+ \to 0$) relations that can be identified in various cases with Green–Kubo's formulae and Onsager's reciprocal relations, see also [6].

The connection with applications of the above results is made via the assumption that concrete chaotic dynamical systems can be considered, "for the purpose of studying the properties of interest", as transitive Anosov flows. This assumption was called in [3] the chaotic hypothesis and makes the above results of immediate relevance for many experimental and theoretical applications to dissipative reversible systems.

The hypothesis is a reinterpretation of a principle stated by Ruelle, [7]. Applications to microscopically irreversible systems (like systems with microscopic friction or systems like the Navier Stokes equations) have also been proposed, see [8].

In this paper we shall not deal with the applications (see [9] and [10] for numerical simulations applications) but, rather, with conceptual problems of interpretation of the fluctuation theorem and of its extensions studied below.

2. FLUCTUATION PATTERNS

The fluctuation theorem can be interpreted, see above and [5], as extending to non equilibrium and to large fluctuations of time averages, gaussian fluctuations theory and Onsager reciprocity. Hence it is natural to inquire whether there are more direct and physical interpretations of the theorem (hence of the meaning of the chaotic hypothesis) when the external forcing is really different from the value 0 (that is always assumed in Onsager's theory). A result in this direction is the *conditional reversibility theorem*, discussed below.

Consider an observable F which, for simplicity, has a well defined time reversal parity: $F(Ix) = \varepsilon_F F(x)$, with $\varepsilon_F = \pm 1$. Let $F_+ = 0$ be its time average (*i.e.* its SRB average) and let $t \to \varphi(t)$ be a smooth function vanishing for $|t|$ large enough. We look at the probability, relative to the SRB distribution (*i.e.* in the "natural stationary state") that $F(S_t x)$ is $\varphi(t)$ for $t \in [-\frac{\tau}{2}, \frac{\tau}{2}]$: we say that F "follows the fluctuation pattern" φ in the time interval $t \in [-\frac{\tau}{2}, \frac{\tau}{2}]$.

No assumption on the fluctuation size, nor on the size of the forces keeping the system out of equilibrium, will be made. Besides the chaotic hypothesis we assume, however, that the evolution is time reversible *also* out of equilibrium and that the phase space contraction rate σ_+ is not zero (the results hold no matter how small σ_+ is and they make sense even if $\sigma_+ = 0$, but they become trivial).

We denote $\zeta(p,\varphi)$ the *large deviation function* for observing in the time interval $[-\frac{\tau}{2}, \frac{\tau}{2}]$ an average phase space contraction $\sigma_\tau \stackrel{def}{=} \frac{1}{\tau}\int_{-\tau/2}^{\tau/2} \sigma(S_t x)dt = p\sigma_+$ and at the same time a fluctuation pattern $F(S_t x) = \varphi(t)$. This means that the probability that the *dimensionless average entropy creation rate* p is in a open set Δ and F is in a neighborhood[1] $U_{\underline{\psi},\varepsilon}$ of φ is given by:

$$\sup_{p\in\Delta,\varphi\in U_{\underline{\psi},\varepsilon}} e^{-\tau\zeta_\tau(p,\varphi)} \tag{2.1}$$

to leading order as $\tau \to \infty$ (*i.e.* the logarithm of the mentioned probability divided by τ converges as $\tau \to \infty$ to $\sup_{p\in\Delta,\varphi\in U_{\underline{\psi},\varepsilon}} -\zeta(p,\varphi)$).

Given a reversible, dissipative, transitive Anosov flow the fluctuation pattern $t \to \varphi(t)$ and the time reversed pattern $t \to \varepsilon_F\varphi(-t)$ are then related by the following:

Conditional reversibility theorem: If F is an observable with defined time reversal parity $\varepsilon_F = \pm 1$ and if τ is large the fluctuation pattern $\varphi(t)$ and its time reversal $I\varphi(t) \equiv \varepsilon_F\varphi(-t)$ will be followed with equal likelyhood if the first is conditioned to an entropy creation rate p and the second to the opposite $-p$. This holds because:

$$\frac{\zeta(-p,I\varphi)-\zeta(p,\varphi)}{p\sigma_+} = 1 \qquad \text{for } |p| \le p^* \tag{2.2}$$

with ζ introduced above and a suitable $p^* \ge 1$.

In other words while it is very difficult, in the considered systems, to see an "anomalous" average entropy creation rate during a time τ (*e.g.* $p = -1$),

---

[1] By "neighborhood" $U_{\underline{\psi},\varepsilon}$ we mean that $\int_{-\tau/2}^{\tau/2}\psi(t)F(S_t x)dt$ is approximated within given $\varepsilon > 0$ by $\int_{-\tau/2}^{\tau/2}\psi(t)\varphi(t)dt$ for ψ in the finite collection $\underline{\psi} = (\psi_1, \ldots, \psi_m)$ of test functions.

it is also true that "that is the hardest thing to see". Once we see it *all the observables will behave strangely* and the relative probabilities of time reversed patterns will become as likely as those of the corresponding direct patterns under "normal" average entropy creation regime.

"A waterfall will go up, as likely as we see it going down, in a world in which for some reason the entropy creation rate has changed sign during a long enough time." We can also say that the motion on an attractor is reversible, even in presence of dissipation, once the dissipation is fixed.

The proof of the theorem is very simple and in fact it is a repetition of the fluctuation theorem proof. To be complete we sketch here the proof of the corresponding result for discrete time systems, *i.e.* for systems whose evolution is a map S of a smooth compact manifold $\mathcal{C}$, "phase space", and S is a reversible map (*i.e.* $IS = S^{-1}I$ for a volume preserving diffeomorphism I of phase space such that $I^2 = 1$) and a dissipative, transitive Anosov map (see below).

The latter systems are simpler to study than the corresponding Anosov flows because Anosov maps do not have a trivial Lyapunov exponent (the vanishing one associated with the phase space flow direction); but the techniques to extend the analysis to continuous time systems will be the same as those developed in [4] for the similar problem of proving the fluctuation theorem for Anosov flows. We dedicate the following section to the formulation of the conditional reversibility theorem for systems with discrete time evolution.

3. THE CASE OF DISCRETE TIME EVOLUTION

A description of systems which evolve chaotically in discrete time is possible, closely following the one dedicated in the previous sections to continuous time systems.

Consider a dynamical system described by a smooth map S acting on a smooth compact manifold $\mathcal{C}$ and depending on a few parameters $\underline{G}$ so that for $\underline{G} = \underline{0}$ the system is volume preserving. Suppose also that the system is time reversible, *i.e.* there is an isometric diffeomorphism I of $\mathcal{C}$ which anticommutes with S: $IS = S^{-1}I$; furthermore suppose that it is "very chaotic", *i.e.* $(\mathcal{C}, S)$ is a transitive Anosov map.

We recall that an Anosov map on a smooth compact manifold $\mathcal{C}$ is a map such that at every point $y \in \mathcal{C}$ one can define a *stable tangent plane* T^s_y, and an *unstable tangent plane* T^u_y with:

1) the tangent planes T_y^u and T_y^s are independent and vary continuously with y always spanning the full tangent plane at y.

2) they are covariant $\partial S\, T_y^\alpha = T_{Sy}^\alpha$, $\alpha = u, s$

3) there exist constants $C,\ l > 0$ such that if $\xi \in T_y^s$ then $|\partial S^n \xi| \le Ce^{-\ ln}|\xi|$ for all $n \ge 0$, and if $\xi \in T_y^u$ then $|\partial S^n \xi| \le Ce^{-\ ln}|\xi|$ for all $n \ge 0$.

A Anosov map is *mixing* or *transitive* if given any two open sets A, B there is a $n_{A,B}$ such that $S^n A \cap B \neq \emptyset$ for $|n| > n_{A,B}$.

If $(\mathcal{C}, S)$ is a transitive Anosov system the time averages of smooth observables on trajectories starting at almost all points, with respect to the volume, do exist and can be computed as integrals over phase space with respect to a probability distribution μ, which is unique and is called the SRB distribution, [11].

For reversible transitive Anosov maps a fluctuation theorem analogous to the one for flows can be formulated as follows. Let $J(x) = \partial S(x)$ be the jacobian matrix of the transformation S.

The quantity $\overline{\sigma}(x) = -\log \Lambda(x)$ will be called the *entropy production* per timing event so that $e^{-\overline{\sigma}(x)}$ is the *phase space volume contraction* per event: *it will be called entropy creation rate per event.* This is the analogue of the divergence of the equations of motion in the continuous time systems considered in §1,2.

The non negativity of the time average, *i.e.* of the SRB average, of $\overline{\sigma}(x)$ is in fact a theorem that could be called the *H–theorem* of reversible non equilibrium statistical mechanics, [12]. It can be also shown, *c.f.r* [12], that the average of $\overline{\sigma}(x)$ with respect to the SRB distribution can vanish only if the latter has a density with respect to the volume.

Therefore we shall call *dissipative* systems for which the time average of $\overline{\sigma}(x)$ is positive, [3].

We call $\overline{\sigma}_\tau(x)$ the partial average of $\overline{\sigma}(x)$ over the part of trajectory centered at x (in time): $S^{-\tau/2}x, \ldots,\ S^{\tau/2-1}x$. Then we can define the *dimensionless entropy production* $p = p(x)$ via:

$$\overline{\sigma}_\tau(x) = \frac{1}{\tau} \sum_{j=-\tau/2}^{\tau/2-1} \overline{\sigma}(S^j x) \stackrel{def}{=} \langle \overline{\sigma} \rangle_+ p \tag{3.1}$$

where $\langle \overline{\sigma} \rangle_+$ is the infinite time average $\int_{\mathcal{C}} \overline{\sigma}(y)\mu(dy)$, if μ is the "forward statistics" of the volume measure (*i.e.* is the SRB distribution), and τ is any integer.

Let F be an observable with time reversal parity $\varepsilon_F = \pm 1$ and let $n \longrightarrow \varphi(n)$ be a function vanishing for $|n|$ large enough, say $|n| > n_0$. If Δ, U_j, $|j| < N \le n_0$, are open intervals with U_j centered at $\varphi(j)$ and τ is given, the probability that $p(x) \in \Delta, F(S^j x) \in U_j$ is given by:

$$\sup_{p\in\Delta,\varphi'(j)\in U_j} e^{-\tau\zeta(p,\varphi')} \tag{3.2}$$

to leading order as $\tau \longrightarrow \infty$ for some $\zeta(p, \varphi)$. And the following theorem holds:

Conditional reversibility theorem (discrete time case): *Given F, its fluctuation patterns φ, considered in a time interval $[t - \frac{1}{2}\tau, t + \frac{1}{2}\tau]$ during which the entropy creation rate has average p, have the same relative probability as their time reversed patterns $I\varphi(n) = \varepsilon_F\varphi(-n)$ in a time interval $[t' - \frac{1}{2}\tau, t' + \frac{1}{2}\tau]$ during which the entropy creation rate has opposite average $-p$.*

Analytically this is again expressed by property (2.2) of the large deviation function $\zeta(p, \varphi)$. The proof of this result is a repetition of the proof of the fluctuation theorem of [3] and we sketch it in the following section for the case of reversible dissipative transitive Anosov maps.

4. "CONDITIONAL REVERSIBILITY" AND "FLUCTUATION" THEOREMS

We need a few well known, but sometimes quite non trivial, geometrical results about Anosov maps.

The stable planes form an "integrable" family in the sense that there are locally smooth manifolds having everywhere a stable plane as tangent plane, and likewise the unstable manifolds are integrable. This defines the notion of stable and unstable manifold through a point $x \in \mathcal{C}$: they will be denoted W_x^s and W_x^u, see [13]. Globally such manifolds wrap around in the phase space $\mathcal{C}$ and, in *transitive* systems, the stable and unstable manifolds of each point are dense in $\mathcal{C}$, see [11], [13].

The covariance of the stable and unstable planes implies that if $J(x) = \partial S(x)$ is the jacobian matrix of S we can regard its action mapping the tangent plane T_x onto T_{Sx} as "split" linearly into an action on the stable plane and one on the unstable plane: *i.e.* $J(x)$ restricted to the stable plane becomes a linear map $J^s(x)$ mapping T_x^s to T_{Sx}^s. Likewise one can define the map $J^u(x)$, [13].

We call $\Lambda_u(x), \Lambda_s(x)$ the determinants of the jacobians $J^u(x), J^s(x)$: their product differs from the determinant $\Lambda(x)$ of $\partial S(x)$ by the ratio of the sine of the angle $a(x)$ between the planes T^s_x, T^u_x and the sine of the angle $a(Sx)$ between T^s_{Sx}, T^u_{Sx}: hence $\Lambda(x) = \frac{\sin a(Sx)}{\sin a(x)}\Lambda_s(x)\Lambda_u(x)$. We set also:

$$\Lambda_{u,\tau}(x) = \prod_{j=-\tau/2}^{\tau/2-1} \Lambda_u(S^j x), \quad \Lambda_{s,\tau}(x) = \prod_{j=-\tau/2}^{\tau/2-1} \Lambda_s(S^j x),$$
$$\Lambda_\tau(x) = \prod_{j=-\tau/2}^{\tau/2-1} \Lambda(S^j x) \tag{4.1}$$

Time reversal symmetry implies that $W^s_{Ix} = IW^u_x, W^u_{Ix} = IW^s_x$ and:

$$\Lambda_\tau(x) = \Lambda_\tau(Ix)^{-1}, \quad \Lambda_{s,\tau}(Ix) = \Lambda_{u,\tau}(x)^{-1}, \quad \Lambda_{u,\tau}(Ix) = \Lambda_{s,\tau}(x)^{-1}$$
$$\sin a(x) = \sin a(Ix) \tag{4.2}$$

Given the above geometric–kinematical notions the SRB distribution μ can be represented by assigning suitable weights to small phase space cells. This is very similar to the representation of the Maxwell–Boltzmann distributions of equilibrium states in terms of suitable weights given to phase space cells of equal Liouville volume.

The phase space cells can be made consistently as small as we please and, by taking them small enough, one can achieve an arbitrary precision in the description of the SRB distribution μ, in the same way as we can approximate the Liouville volume by taking the phase space cells small.

The key to the construction is a Markov partition: this is a partition $\mathcal{E} = (E_1, \ldots, E_{\mathcal{N}})$ of the phase space $\mathcal{C}$ into $\mathcal{N}$ cells which are *covariant* with respect to the time evolution, see [13], in a sense that we do not attempt to specify here, and with respect to time reversal in the sense that $IE_j = E_{j'}$ for some j', see [14].

Given a Markov partition $\mathcal{E}$ we can "refine" it "consistently" as much as we wish by considering the partition $\mathcal{E}_T = \vee_{-T}^{T} S^{-j}\,\mathcal{E}$ whose cells are obtained by "intersecting" the cells of $\mathcal{E}$ and of its S–iterates; the cells of $\mathcal{E}_T$ become exponentially small with $T \to \infty$ as a consequence of the hypebolicity. In each $E_j \in \mathcal{E}_T$ one can select a point x_j ("center": quite arbitrary and not uniquely defined, see [3], [14]) so that Ix_j is the point selected in IE_j. Then we evaluate the expansion rate $\Lambda_{u,2T}(x_j)$ of S^{2T} as a map of the unstable manifold of $S^{-T}x_j$ to that of $S^T x_j$.

Using the elements $E_j \in \mathcal{E}_T$ as cells we can define approximations "as good as we wish" to the SRB distribution μ because for all smooth observables F defined on $\mathcal{C}$ it is:

$$\int \mu(dy)F(y) = \lim_{T\to\infty} \int m_T(dy)F(y) \stackrel{def}{=} \lim_{T\to\infty} \frac{\sum_{E_j\in\mathcal{E}_T} F(x_j)\Lambda^{-1}_{u,2T}(x_j)}{\sum_{E_j\in\mathcal{E}_T} \Lambda^{-1}_{u,2T}(x_j)} \tag{4.3}$$

where $m_T(dy)$ is implicitly defined here by the ratio in the *r.h.s.* of (4.3). This deep theorem of Sinai, [11], (and [1] in the continuous time case), see also [3], [14], is the basis of the technical part of our reversibility theorem.

If Δ_a denotes an interval $[a, a+da]$ we first evaluate the probability, with respect to $m_{\tau/2}$ of (4.3), of the event that $a(x_j) = \overline{\sigma}_\tau(x)/\langle\overline{\sigma}\rangle_+ \in \Delta_p$ and, also, that $F(S^n x) \in \Delta_{\varphi_n}$ for $|n| < \tau$, divided by the probability (with respect to the same distribution) of the time reversed event that $a(x) = \overline{\sigma}_\tau(x)/\langle\overline{\sigma}\rangle_+ \in \Delta_{-p}$ and, also, $F(S^n x) \in \Delta_{\varepsilon_F\varphi_{-n}}$ for $|n| < \tau$; *i.e.* we compare the probability of a fluctuation pattern φ in presence of average dissipation p and that of the time reversed pattern in presence of average dissipation $-p$. This is essentially:

$$\frac{\sum_{j,\, a(x_j)=p, F(S^n x_j)=\varphi(n)} \Lambda^{-1}_{u,\tau}(x_j)}{\sum_{j,\, a(x_j)=-p, F(S^n x_j)=\varepsilon_F\varphi(-n)} \Lambda^{-1}_{u,\tau}(x_j)} \tag{4.4}$$

Since $m_{\tau/2}$ in (4.3) is only an approximation to μ_+ *an error is involved in using* (4.4) as a formula for the same ratio computed by using the true SRB distribution μ instead of $m_{\tau/2}$.

It can be shown that this "first" approximation (among the two that will be made) can be estimated to affect the result only by a factor bounded above and below uniformly in τ, p, [3]. This is *not* completely straightforward: in a sense this is perhaps the main technical problem of the analysis. Further mathematical details can be found in [3] and in [4].

Remark. – There are other representations of the SRB distributions that seem more appealing than the above one based on the (not too familiar, see [14]) Markov partitions notion. The simplest is perhaps the *periodic orbits representation* in which the role of the cells is taken by the periodic orbits. However I do not know a way of making the argument that leads to (4.4) *keeping under control* the approximations and which does not rely on Markov partitions. And in fact I do not know of *any expression* of the SRB distribution that is not proved by using the very existence of Markov partitions.

We now try to establish a one to one correspondence between the addends in the numerator of (4.4) and the ones in the denominator, aiming at showing that corresponding addends have a *constant ratio* which will, therefore, be the value of the ratio in (4.4).

This is possible because of the reversibility property: it will be used in the form of its consequences given by the relations (4.2).

The ratio (4.4) can therefore be written simply as:

$$\frac{\sum_{j,\, a(x_j)=p, F(S^n x_j)=\varphi(n)} \Lambda_{u,\tau}^{-1}(x_j)}{\sum_{j,\, a(x_j)=-p, F(S^n x_j)=\varepsilon_F \varphi(-n)} \Lambda_{u,\tau}^{-1}(x_j)} \equiv \frac{\sum_{j,\, a(x_j)=p, F(S^n x_j)=\varphi(n)} \Lambda_{u,\tau}^{-1}(x_j)}{\sum_{j,\, a(x_j)=p, F(S^n x_j)=\varphi(n)} \Lambda_{s,\tau}(x_j)} \tag{4.5}$$

where $x_j \in E_j$ is the center in E_j. In deducing the second relation we take into account time reversal symmetry I, that the centers $x_j, x_{j'}$ of E_j and $E_{j'} = IE_j$ are such that $x_{j'} = Ix_j$, and (4.2) in order to transform the sum in the denominator of the *l.h.s.* of (4.5) into a sum over the same set of labels that appear in the numerator sum.

It follows then that the ratios between corresponding terms in the ratio (4.5) is equal to $\Lambda_{u,\tau}^{-1}(x)\Lambda_{s,\tau}^{-1}(x)$. This differs from the reciprocal of the total change of phase space volume over the τ time steps (during which the system evolves from the point $S^{-\tau/2}x$ to $S^{\tau/2}x$).

The difference is only due to not taking into account the ratio of the sines of the angles $a(S^{-\tau/2}x)$ and $a(S^{\tau/2}x)$ formed by the stable and unstable manifolds at the points $S^{-\tau/2}x$ and $S^{\tau/2}x$. Therefore $\Lambda_{u,\tau}^{-1}(x)\Lambda_{s,\tau}^{-1}(x)$ will differ from the actual phase space contraction under the action of S^τ, regarded as a map between $S^{-\tau/2}x$ and $S^{\tau/2}x$, by a factor that can be bounded between B^{-1} and B with $B = \max_{x,x'} \frac{|\sin a(x)|}{|\sin a(x')|}$, which is finite and positive, by the linear independence of the stable and unstable manifolds.

But for all the points x_j in (4.5), the reciprocal of the total phase space volume change over a time τ (by the constraint, $\overline{\sigma}_\tau/\langle\overline{\sigma}\rangle_+ = p$, imposed on the summation labels) equals $e^{\tau\langle\overline{\sigma}\rangle_+ p}$ up to a "second" approximation that cannot exceed a factor $B^{\pm 1}$ which is *τ-independent*. Hence the ratio (4.4) will be the exponential: $e^{\tau\langle\overline{\sigma}\rangle_+ p}$, up to a τ–independently bounded factor.

It is important to note that there have been two approximations, as pointed out in the discussion above. They can be estimated, see [3], and imply that the argument of the exponential *is correct up to p, φ, τ independent corrections* making the consideration of the limit $\tau \to \infty$ *necessary* in the formulation of the theorem.

5. CONCLUDING REMARKS

1) One should note that in applications the above theorems will be used through the chaotic hypothesis and therefore other errors may arise because of its approximate validity (the hypothesis in fact essentially states that "things go as if" the system was Anosov): they may depend on the number N of degrees of freedom and we do not control them except for the fact that, if present, their relative value should tend to 0 as $N \to \infty$: there may be (and very likely there are) cases in which the chaotic hypotesis is not reasonable for small N (*e.g.* systems like the Fermi-Pasta-Ulam chains) but it might be correct for large N. We also mention that, on the other hand, for some systems with small N the chaotic hypothesis may be already regarded as valid (*e.g.* for the models in [15], [9], [10]).

2) A frequent remark that is made about the chaotic hypothesis is that it does not seem to keep the right viewpoint on nonequilibrium thermodynamics. In fact it is analogous to the ergodic hypothesis which (*as well known*) cannot be taken as the foundation of equilibrium statistical mechanics, even though it leads to the correct Maxwell Boltzmann statistics, because the latter "holds for other reasons". Namely it holds because in most of phase space (measuring sizes by the Liouville measure) the few interesting macroscopic observable have the same value, [16], see also [17].

An examination of the basic paper of Boltzmann [18], in which the theory of equilibrium ensembles is developed, may offer some arguments for further meditation. The paper starts by illustrating an important, and today almost forgotten, remark of Helmoltz showing that very simple systems ("monocyclic systems") can be used to construct mechanical models of thermodynamics: the example chosen by Boltzmann is *really extreme by all standards*. He shows that a *Saturn ring* of mass m on a Keplerian orbit of major semiaxis a in a gravitational field of strength g can be used to build a model of thermodynamics.

In the sense that one can call "volume" V the gravitational constant g, "temperature" T the average kinetic energy, "energy" U the energy and "pressure" p the average potential energy ma^{-1}: then one infers that by varying, *at fixed eccentricity*, the parameters U, V the relation $(dU + pdV)/T = exact$ holds.

Clearly this *could* be regarded as a curiosity.

However Boltzmann (following Helmoltz) took it seriously and proceeded to infer that under the ergodic hypothesis *any system* small or large provided us with a model of thermodynamics (being monocyclic in the sense of Helmoltz): for instance he showed that the canonical ensemble verifies exactly the second principle of thermodynamics (in the form $(dU + pdV)/T = \ exact$) *without any need to take thermodynamic limits*, [18], [19]. The same could be said of the microcanonical ensemble (here, however, he had to change "slightly" the definition of heat to make things work without finite size corrections).

He realized that the ergodic hypothesis could not possibly account for the correctness of the canonical (or microcanonical) ensembles: this is clear at least from his (later) paper in defense against Zermelo's criticism, [20], nor for the observed time scales of approach to equilibrium. Nevertheless he called the theorem he had proved the *heat theorem* and never seemed to doubt that it provided evidence for the correctness of the use of the equilibrium ensembles for equilibrium statistical mechanics.

Hence we see that there are two aspects to consider: first certain relations among mechanical quantities *hold no matter how large* is the size of the system and, secondly, they can be seen and tested not only in small systems, by direct measurments, but even in large systems, because in large systems such mechanical quantities acquire a macroscopic thermodynamic meaning and their relations are "typical" *i.e.* they hold in most of phase space.

The consequences of the chaotic hypothesis stem from the properties of small Anosov systems (the best understood) which play here the role of Helmoltz's monocyclic systems of which Boltzmann's Saturn ring is the paradigm. They are remarkable consequences because they provide us with *parameter free relations* (namely the fluctuation theorem and its consequences): but clearly we cannot hope to found solely upon them a theory of nonequilibrium statistical mnechanics because of the same reasons the validity of the second law for monocyclic systems had in priciple no reason to imply the theory of ensembles.

Thus what is missing are arguments similar to those used by Boltzmann to justify the use of the ensembles *independently* of the ergodic hypothesis: an hypothesis which in the end may appear (and still does appear to many) as having only suggested them "by accident". The missing arguments should justify the fluctuation theorem on the basis of the extreme likelihood of its predictions in systems that are very large and that may be not Anosov systems in the mathematical sense. I see no reason, now, why this should prove impossible *a priori* or in the future.

In the meantime it seems interesting to take the same philosophical viewpoint adopted by Boltzmann: not to consider a chance that *all* chaotic systems share some selected properties and try to see if such properties help us achieving a better understanding of nonequilibrium. After all it seems that Boltmann himself took a rather long time to realize the interplay of the above two basic mechanisms behind the equilibrium ensembles and to propose a solution harmonizing them. "All it remains to do" is to explore if the hypothesis has implications more interesting or deeper than the fluctuation theorem, see [8], [6].

ACKNOWLEDGMENTS

I are indebted to F. Bonetto, E.G.D. Cohen, J. Lebowitz, D. Ruelle for stimulating discussions and comments.

REFERENCES

[1] R. BOWEN and D. RUELLE, The ergodic theory of axiom A flows, *Inventiones Mathematicae*, Vol. **29**, 1975, pp. 181–205.

[2] G. GALLAVOTTI, Reversible Anosov maps and large deviations, *Mathematical Physics Electronic Journal, MPEJ*, (http:// mpej.unige.ch), Vol. **1**, 1995, 12 pp.

[3] G. GALLAVOTTI and E. G. D. COHEN, Dynamical ensembles in nonequilibrium statistical mechanics, *Physical Review Letters*, Vol. **74**, 1995, pp. 2694–2697; Dynamical ensembles in stationary states, *Journal of Statistical Physics*, Vol. **80**, 1995, pp. 931–970. See also: [2].

[4] G. GENTILE, Large deviation rule for Anosov flows, *IHES* preprint, *Forum Mathematicum*, Vol. **10**, 1998, pp. 89–118.

[5] G. GALLAVOTTI, Extension of Onsager's reciprocity to large fields and the chaotic hypothesis, *Physical Review Letters*, Vol. **78**, 1996, pp. 4334–4337. See also: *Chaotic hypothesis: Onsager reciprocity and fluctuation–dissipation theorem*, Journal of Statistical Physics, **84**, 899–926, 1996. And: *Chaotic principle: some applications to developed turbulence*, Journal of Statistical Physics, **86**, 907–934, 1997.

[6] G. GALLAVOTTI and D. RUELLE, SRB states and nonequilibrium statistical mechanics close to equilibrium, mp_arc@math. utexas. edu, #96-645, *Communications in Mathematical Physics*, Vol. **190**, 1997, pp. 279–285.

[7] D. RUELLE, Positivity of entropy production in nonequilibrium statistical mechanics, *Journal of Statistical Physics*, Vol. **85**, 1996, 1996.3, pp. 1–25.

[8] G. GALLAVOTTI, Equivalence of dynamical ensembles and Navier-Stokes equations, *Physics Letters*, Vol. **223**, 1996, pp. 91–95. And *Dynamical ensembles equivalence in statistical mechanics*, in mp_arc@ math. utexas. edu #96-182, chao-dyn@xyz.lanl.gov #9605006, *Physica D*. Vol. **105**, 1977, pp. 163–184.

[9] D. J. EVANS, E. G. D. COHEN and G. P. MORRISS, Probability of second law violations in shearing steady flows, *Physical Review Letters*, Vol. **71**, 1993, pp. 2401–2404.

[10] F. BONETTO, G. GALLAVOTTI and P. GARRIDO, Chaotic principle: an experimental test, mp_arc @math. utexas. edu, # 96-154, *Physica D*.Vol. **105**, 1997, pp. 226–252.

[11] Y. G. SINAI, Gibbs measures in ergodic theory, *Russian Mathematical Surveys*, Vol. **27**, 1972, pp. 21–69. Also: *Introduction to ergodic theory*, Princeton U. Press, Princeton, 1977.

[12] D. RUELLE, Positivity of entropy production in nonequilibrium statistical mechanics, *Journal of Statistical Physics*, Vol. **85**, 1996, 1996.1, pp. 1–25.

[13] D. RUELLE, *Elements of differentiable dynamics and bifurcation theory*, Academic Press, 1989.

[14] G. GALLAVOTTI, *New methods in nonequilibrium gases and fluids*, in mp_arc@ math. utexas.edu #96-533, and chao-dyn #9610018, in print in open systems.

[15] N. I. CHERNOV, G. L. EYINK, J. L. LEBOWITZ and Y. SINAI, Steady state electric conductivity in the periodic Lorentz gas, *Communications in Mathematical Physics*, Vol. **154**, 1993, pp. 569–601.

[16] W. TAYLOR, *The kinetic theory of the dissipation of energy*, reprinted (in english) in: S. Brush, *Kinetic theory*, Vol. **2**, 1966, pp. 176, Pergamon Press.

[17] J. L. LEBOWITZ, *Microscopic reversibility*, To appear in *Enciclopedia del Novecento*, Enciclopedia Italiana, Roma, 1997, english preprint: Rutgers University, 1997.

[18] L. BOLTZMANN, *Über die eigenshaften monzyklischer und anderer damit verwandter Systeme*, in "Wissenshafltliche Abhandlungen", ed. F.P. Hasenhörl, vol. III, Chelsea, New York, 1968, (reprint of 1884 original).

[19] G. GALLAVOTTI, Trattatello di Meccanica Statistica, *Quaderni CNR-GNFM*, Vol. **50**, 1995, pp. 1–350, Firenze. Free copies can be obtained by requesting them to CNR–GNFM, via S. Marta 13/a, 50139 Firenze, Italy, or by fax at CNR-GNFM -39-55-475915. English version in print, Springer-Verlag, 1999.

[20] L. BOLTZMANN, *Entgegnung auf die wärmetheoretischen Betrachtungen des Hrn. E. Zermelo*, engl. transl.: S. Brush, "Kinetic Theory", Vol. **2**, 1966, pp. 218, Pergamon Press. And: *Zu Hrn. Zermelo's Abhandlung "Ueber die mechanische Erklärung irreversibler Vorgänge*, engl. trans. in S. Brush, "Kinetic Theory", Vol. **2**, pp. 238, (reprints of 1896 and 1897 originals).

(Manuscript received April 7, 1997.)