Dynamical ensembles in stationary states.[!]

G. Gallavotti[#], E. G. D. Cohen[*]

Abstract: We propose as a generalization of an idea of Ruelle's to describe turbulent fluid flow a chaotic hypothesis for reversible dissipative many particle systems in nonequilibrium stationary states in general. This implies an extension of the zeroth law of thermodynamics to non equilibrium states and it leads to the identification of a unique distribution μ describing the asymptotic properties of the time evolution of the system for initial data randomly chosen with respect to a uniform distribution on phase space. For conservative systems in thermal equilibrium the chaotic hypothesis implies the ergodic hypothesis. We outline a procedure to obtain the distribution μ: it leads to a new unifying point of view for the phase space behavior of dissipative and conservative systems. The chaotic hypothesis is confirmed in a non trivial, parameter–free, way by a recent computer experiment on the entropy production fluctuations in a shearing fluid far from equilibrium. Similar applications to other models are proposed, in particular to a model for the Kolmogorov–Obuchov theory for turbulent flow.

§1 *Introduction.*

In a previous paper [GC] we proposed the use of Ruelle's idea (discussed in §2) to obtain the probability distribution for the statistics of turbulent flows in hydro-

[!] Archived in *mp_arc@ math. utexas. edu*, #95-32, in *chao-dyn@xyz.lanl.gov*; #9501015

[#] Fisica, Università di Roma La Sapienza, p.le Moro 2, 00185, Roma, Italia.

[*] The Rockefeller University, New York, N.Y. 10021, USA.

dynamics, as a basis for the study of many particle statistical mechanical systems in nonequilibrium stationary states in general. We did so, by providing a concrete procedure of how to obtain the necessary probability distribution, now called the Sinai–Ruelle–Bowen (SRB) distribution, to compute the statistical properties of the above mentioned systems. The applicability of such a distribution has, so far, only been proved with mathematical rigor for very idealized systems, such as Anosov or Axiom A systems, and it would be impossible at present to give extensions of the proofs for the many particle systems of interest here. Therefore we proposed to use Ruelle's idea as a heuristic principle to obtain the statistical properties of such systems, at least when they are very large, *i.e.* in the thermodynamic limit. This implied that we made a "chaotic hypothesis" that the many particle systems in statistical mechanics are essentially chaotic in the sense of Anosov, *i.e.* they behave, in many respects, *as if they were* Anosov systems as far as their properties of physical interest are concerned. In other words we use the SRB distribution obtained from the strong assumption of chaoticity in the Anosov sense in a *heuristic* way to compute statistical mechanical properties of our system and assume that the corrections due to the possible non–validity of the strong chaoticity assumption become negligible for large systems.

The position that we take here is very similar to that usually taken with respect to the so–called ergodic hypothesis, which has been proven only for very special particle systems, and only very recently for systems with more than one particle, [SiS]. Yet, when used as a principle, it has led to all known results of equilibrium statistical mechanics, beginning with its connection with thermodynamics. It would seem therefore inappropriate, in fact very unfortunate, if the application of the ergodic hypothesis, would have had to wait till it had been proved valid for the many particle statistical mechanical systems in thermal equilibrium of physical interest. Very recently a version of what we shall call the *chaotic hypothesis*, see §2, has been rigorously proved for a single (moving) particle system held in a non equilibrium stationary state and a number of detailed consequences have been derived, which agree with experiment [CELS].

Here we will give a number of possible many particle systems to which the chaotic hypothesis and the ensuing SRB distribution could immediately be applied. So far, only one of those systems: a shearing thermostatted fluid far from equilibrium (see §3, model 2) has been investigated numerically, *viz.* the statistics of the fluctuations of the pressure tensor – or equivalently of the entropy production rate – in this system have been determined numerically and found to be in very good agreement with what one obtains by applying the chaotic hypothesis. Although corresponding experiments have not been done as yet for the other systems we mention, they should provide further checks on the validity of Ruelle's ideas and the chaotic hypothesis as proposed here.

We want to emphasize that the application of the chaotic hypothesis is not restricted to stationary states near equilibrium, *i.e.* to the linear regime of small deviations from thermal equilibrium, as the above mentioned example of a shearing flow shows. The precise limitations of its applicability are unknown, however.

The way we will present the construction of the SRB distribution from the chaotic hypothesis can also be applied to the theory of equilibrium states. It leads then to a new picture of the behavior in phase space of both equilibrium and stationary nonequilibrium systems, which reveals a much closer analogy in their phase space behavior than considered up till now. Thus a unification of the conservative behavior in equilibrium states and of the dissipative behavior in non equilibrium stationary states emerges.

In §2 we describe some general properties that can help visualizing the general phenomenology of the non equilibrium systems that we consider: the discussion leads then to a formal definition of Ruelle's idea and to the precise formulation of the chaotic hypothesis. In §3 we give a variety of examples of nonequilibrium systems to which the chaotic hypothesis can be straightforwardly applied. In §4 we discuss from a somewhat unusual viewpoint the heuristic ideas behind the hypothesis; this leads, in §5 and §6, to an outline and reinterpretation of the classical, [S2,Bo,R1], construction of the appropriate SRB distribution for this system, using Markov partitions. In §7 we briefly summarize the only concrete application so far available, *viz.* that of a shearing fluid, and we discuss our main result, the fluctuation theorem of §7 (which gives a theoretical interpretation of the experiment). In §8 we give a discussion and outlook.

§2 The SRB picture.

For a convenient discussion of the SRB picture of nonequilibrium stationary states it is important to discuss the time evolution in discrete time, rather than in continuous time. This will be obtained by observing the motion when some timing event happens (this is usually done by describing the motion through a Poincaré section). Therefore we fix a *timing event* and envisage performing our observations at every time the event happens. This will have the effect of reducing by one unit the initial phase space dimension.

The choice of the timing event is essentially arbitrary: for many particle systems a reasonable choice could be the event in which the pair of closest particles (among the N we have) is at a distance r, coming from larger distances. We call such event a "collision" and we use it as our timing event. To avoid trivialities r has to be chosen small compared to the average interparticle distance, but not too small (*i.e.* larger than the "core" of the interaction). In the case of a continuous fluid flow a natural timing event, actually used in many numerical experiments starting with [Lo], is the event in which a given coordinate of the velocity field passes through a prefixed value, or assumes a locally maximum value. To uniformize the nomenclature we shall also call such an event a "collision".

The dynamical systems we consider will be defined now by the phase space $\mathcal{C}$ of the "collisions", with dimension $2D$ and the time evolution, which will be a map $S : \mathcal{C} \to \mathcal{C}$ defined by $Sx = S_{\tau(x)}x$, if S_t is the continuous time evolution operator solving the equations of motion in the full phase space $\mathcal{F}$, which in our cases will coincide with a constant "energy surface" or it will be a manifold in it. In fact, as will be discussed in §3, all models we consider here develop on a manifold $\mathcal{F}$ on which one or more observables have a given value (usually the kinetic or total energy and/or a center of mass momentum component or other smooth quantities: the manifold defined by the values of such observables will be what we shall call the "energy surface"). Here $\tau(x)$ is the time interval between the collision $x \in \mathcal{C}$ and the next one.

We shall make a statistical study (like is done in equilibrium): this means that we shall be interested in the properties of the time evolutions of the motions that can be seen by extracting the initial data at random with the Liouville distribution on $\mathcal{F}$. Since our analysis will be performed on $\mathcal{C}$ rather than on $\mathcal{F}$ we shall need the corresponding probability distribution on $\mathcal{C}$. The Liouville distribution $\mu_{\mathcal{L}} = const\,\delta(H(\underline{p}, \underline{q}) - E)d\underline{p}\,d\underline{q}$, when the energy is the conserved quantity defining $\mathcal{F}$ or a similarly defined distribution (*e.g.* $\mu_{\mathcal{L}} = const\,\delta(\sum |\underline{\gamma}_{\underline{k}}|^2 - E)$ in the case of the fluid motion models that we consider in §3, where the variables $\underline{\gamma}_{\underline{k}}$ are the Fourier components of the velocity field) on the full phase space $\mathcal{F}$ (energy surface) naturally generates a probability distribution μ_0 on $\mathcal{C}$: if E is a set on $\mathcal{C}$ we simply set $\mu_0(E)$ proportional to the Liouville measure of the tube of trajectory segments in $\mathcal{F}$ starting at E and with unit length in time, when evolved with the motion corresponding to no external forcing fields. We shall still call μ_0 the "Liouville distribution" (on $\mathcal{C}$).

The first point of our analysis is a generalization of the *zeroth law of thermodynamics* to nonequilibrium stationary states. As expressed by Uhlenbeck and Ford, [UF], the zeroth law of thermodynamics states that a closed conservative mechanical system consisting of a very large number of particles will, when initially not in equilibrium, approach equilibrium, where all *macroscopic* variables have reached stationary values. By (asymptotic) equilibrium one means here that the time averages have reached the value that can be computed by a probability distribution on the energy surface. This law can be *extended* to nonequilibrium systems as:

Extended zero-th law: A dynamical system $(\mathcal{C}, S)$ *describing a many particle system (or a continuum such as a fluid) generates motions that admit a statistics* μ *in the sense that, given any (piecewise smooth), macroscopic observable* F *defined on the points* x *of the phase space* $\mathcal{C}$*, the time average of* F *exists for all* μ_0*–randomly–chosen initial data* x *and is given by:*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} F(S^k x) = \int_{\mathcal{C}} \mu(dx')F(x') \tag{2.1}$$

where μ *is an* S*–invariant probability distribution on* $\mathcal{C}$*.*

In this form we suppose that it holds for all our models. The notation $\mu(dx)$ rather than $r(x)dx$ expresses the possible fractal nature of the support of the distribution μ, and implies that the probability to find the dynamical system in the infinitesimal volume dx around x may not be proportional to dx, so that it cannot be written as $r(x)dx$ with $r(x)$ a probability density and dx the volume measure on phase space.

The main point of this paper is to use an idea of Ruelle's as a guiding principle to describe nonequilibrium stationary states in general. That is, we propose that for such systems there exists a distribution (usually called the SRB distribution) describing the asymptotic statistics of motions with initial data randomly chosen with respect to a uniform distribution on phase space (the Liouville distribution).

For this to be *realistically implemented we assume that macroscopic systems, consisting of very many particles, behave as transitive Anosov systems, i.e.* are "chaotic" in the sense that each point x in phase space admits an unstable and a stable manifold W_x^u, W_x^s which depend continuously on x, are dense in the phase space $\mathcal{C}$, and on which the expansion and contraction rates, respectively, are everywhere separated by a finite gap from 0 (hence no zero Lyapunov exponents occur).[1] We propose therefore the following chaotic hypothesis, in addition to the extended zeroth law, which in [GC] we called *Ruelle's principle*, as a generalization of Ruelle's idea:

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 of the system.

The transitivity is necessary in the theory of Anosov systems, [S2], because it implies the x–independence of the statistics μ, hence in our analysis it is a necessary requirement if we want the chaoticity hypothesis to be compatible with the zeroth law. We intend to show that this hypothesis, although general, leads to concrete verifiable consequences and may be, in this respect, similar to the ergodic hypothesis for equilibrium states but, unlike the ergodic hypothesis, admits an extension to nonequilibrium stationary states.

One could weaken our form of the chaotic hypothesis by replacing "Anosov system" with "Axiom A system", and refer to the general theory of such systems developed in [Bo],[R2] (rather than relying on the work of Sinai on Anosov systems, [S2]); one could even attempt to weaken it further by trying to make use of the general theory of Pesin of non smooth hyperbolic system, [P]. However, we shall not dwell on such somewhat obvious extensions of our ideas, as they do not seem relevant at present.

Examples of model systems in nonequilibrium stationary states to which the chaotic hypothesis is applicable will be given in §3. Numerical evidence leads us to believe that they seem to share a number of properties which we believe to hold also for more general physical systems and which we now summarize. Not all of them are necessary for the applications we shall discuss: however, they are very helpful for building an intuitive, model independent, picture of the phenomena that we attempt to study. When discussing the models from a technical viewpoint we shall mention which properties have been experimentally checked and which have not (yet) been

---

[1] For convenience we recall that an Anosov system $(\mathcal{C}, S)$ is a *smooth* dynamical system such that *every* point $x \in \mathcal{C}$ possesses stable and unstable manifolds W^s_x, W^u_x which depend continuously on x and on which S^n, S^{-n} respectively contract infinitesimal vectors by a factor bounded by $Ce^{-\lambda n}$, for $n \geq 0$, and likewise for $n \leq 0$ they expand by a factor bounded by $C^{-1}e^{-\lambda n}$. The constant λ is therefore such that all Lyapunov exponents verify $|\lambda_j| \geq \lambda$ and, hence, λ can be called a bound on the *Lyapunov spectrum gap*. Note that the continuity of the W^u_x, W^s_x in x implies the *transversality* of the two manifolds, which therefore form everywhere an angle $\vartheta(x)$ bounded away from 0 and π. An Anosov system is transitive if W^u_x, W^s_x are dense in $\mathcal{C}$ for all x.

checked: our applications will only require properties (A,B,C) below.

(A) Dissipation: the phase space volume undergoes a contraction at a rate, on the average, equal to $D\langle\sigma(x)\rangle_+$ where $2D$ is the phase space $\mathcal{C}$ dimension and $\sigma(x)$ is a model dependent "rate" per degree of freedom. The average here is a time average from time zero to plus infinity and the rate is a generalization of the usual *entropy production rate*, (see §3 for motivation of this remark).

We say that a system is *dissipative* if the contraction rate per degree of freedom, $\langle\sigma\rangle_+$, is positive. We shall assume that the models that we consider here in nonequilibrium situations are all dissipative. The instantaneous contraction rate $\sigma(x)$ is, however, a fluctuating quantity and we note that when we consider in this paper entropy production rates and their fluctuations we identify them, mathematically, with phase space contraction rates and their fluctuations, respectively.

(B) Reversibility: there is an isometry, *i.e.* a metric preserving map, i in phase space, which is a map $i : x \to ix$ such that if $t \to x(t)$ is a solution then $i(x(-t))$ is also a solution and furthermore i^2 is the identity.

(C) Chaoticity: the above chaotic hypothesis holds and we can treat the system $(\mathcal{C}, S)$ as a transitive Anosov systems.

We realize that (C) *cannot* hold strictly, even in the case of smooth interaction potentials (in the presence of hard cores the Anosov property, which requires smoothness of the dynamics as a prerequisite, is in fact obviously false). What we mean here is that we assume that the system behaves *as if it was a transitive Anosov system* and that the errors made become negligible (even when there are hard core collisions) at least in large systems. The "long time tails" in the correlations in macroscopic systems are not inconsistent with (C): in fact although the system may be regarded as Anosov for finite N, the size of the maximum time scale (*i.e.* the Lyapunov exponent with minimum absolute value) may be a property that is not uniform in the size of the system (*e.g.* exponential decay of correlations): thus some "predictions" of (C) may become trivial for large systems (this is a familiar event since the Poincaré's recurrence discussions). The predictions of (C) of interest here are the ones that can, at least, be formulated independently of the size of the system. See also the comment 6) in §8 below.

In support of (A,B,C) the following two properties (D) and (E) also are relevant and appear to hold at least for some of the models that we shall treat:

(D) Pairing of Lyapunov exponents: half of the $2D$ Lyapunov exponents are ≥ 0 and half are < 0. If they are ordered so that $0 \leq \lambda_1^+ \leq \lambda_2^+ \leq \ldots \leq \lambda_D^+$ and $0 > \lambda_1^- \geq \lambda_2^- > \ldots > \lambda_D^-$ so that $\lambda_{\max} = \lambda_D^+$ the following *pairing rule* holds:

$$\lambda_j^+ + \lambda_j^- = \frac{1}{D} \sum_{k=1}^{2D} \lambda_k \equiv - \langle \sigma \rangle_+, \qquad j = 1, \ldots, D \tag{2.2}$$

which has been proved for some special cases where the system is non reversible (*i.e.* $\sigma(x)$ is constant) and the pairs do not necessarily consist of exponents with opposite sign; it has been found numerically, in the case of model 2 and related models in the form (2.2) in [ECM1], [SEM] (where it was formulated in the above form).

On the basis of what is presently known, one can conjecture that even if the pairing rule does not hold in the above form it could still hold in the form of an inequality: $-\langle \sigma \rangle_+ \leq \lambda_j^+ + \lambda_j^- \leq 0$ (*weak pairing rule*).

(E) Smoothness of the Lyapunov spectrum: the Lyapunov exponents become for large N a smooth function of their index. This means that, with the labeling of the exponents as in (D) above, if one draws a graph of $x = \frac{j}{D} \to \lambda_j^+ \equiv f_N(\frac{j}{D})$, then in the "thermodynamic limit" ($N \to \infty$ with constant density for particle systems; in the case of fluid systems the role of N will be taken by the Reynolds number) $f_N(x) \xrightarrow[N\to\infty]{} f_\infty(x)$ where $f_\infty(x)$ is a smooth increasing function of $x \in [0, 1]$.

Evidence for the generality of such property comes from [LPR], and [ECM1], [SEM], and quite likely it holds for all the models we consider in §3.

We make the following remarks on the properties (A) to (E).

1) First we note that the irreversible entropy production $\langle \sigma \rangle_+$ in (A) results in a phase space volume contraction. This implies in turn that the attractor which we denote A_0 (and which by property (C) is just the full phase space $\mathcal{C}$) will contain an invariant set A of zero Liouville measure and dimension equal to the *fractal dimension of the motions* (and strictly less than that of the phase space, see below) *but of probability* 1 with respect to the statistics of the motions generated by the dynamics S from initial data chosen randomly with respect to the Liouville distribution μ_0.
It is convenient, therefore, to distinguish between the attractor A_0 as it is usually defined in the literature (which is a closed set for virtually all adopted definitions) and our sets A. The latter are not uniquely defined, but they are in an obvious

sense more intrinsically related to the motions. It can very well be that A_0 is smooth and even coincides with the full phase space, as is the case when (C) holds, while A is much smaller (and is a *fractal*). Thus in this paper, unlike in most established conventions, *we shall call A the attractor*: it will not matter which particular A one considers.

We adopt, as definition of the *fractal dimension of the motions* (*i.e.* of A), the Kaplan Yorke definition (also called the *Lyapunov dimension*, [ER]) The latter is, probably, [ER], quite generally equal to the Hausdorff dimension of those sets A which have the smallest Hausdorff dimension and which are visited with frequency 1 by almost all, with respect to the Liouville distribution μ_0, motions ([ER] p. 641).

2) The above properties (D,E) imply that the attractor A for the motions with a given energy is a *fractal set* with a dimension close to the full dimension $2D$: the fractal dimension will be, in fact, of the order of $2D - O(\langle\sigma\rangle_+ \lambda_{\max}^{-1})D$ as immediately follows if one adopts, as above, the Kaplan Yorke definition of fractal dimension. Note that (E) and the above weaker pairing rule are sufficient for this conclusion. *Systems for which smoothness and the (weak) pairing rule hold do show dimension reduction*, *i.e.* the attractor in phase space has a dimension which is *macroscopically different* from that of the phase space itself.

3) Reversibility implies an important property of the attractor A: if A is an attractor for the forward motion then $A_- = iA$ is an attractor for the backwards motion and, more generally, the statistical properties as $t \to \pm\infty$, of the motions generated by initial data randomly chosen with respect to the Liouville distribution μ_0 are trivially related.

4) Very recently a firm theoretical basis has been given to the smoothness of the spectrum in the thermodynamical limit, [S3].

5) The basic properties for the validity of our results for nonequilibrium stationary states are chaoticity (C) and reversibility (B). If (C) holds the existence of the SRB distribution – *i.e.* of a probability distribution describing the asymptotic statistics of the motions of a system evolving with a dynamics S, whose initial data are chosen randomly with respect to the Liouville distribution μ_0 in the phase space $\mathcal{C}$ – can be proved as a theorem:

Theorem: if a system $(\mathcal{C}, S)$ is a transitive Anosov system then it admits an SRB distribution, [S2].

In conservative systems in equilibrium, verifying (C), the distribution μ is the same as the Liouville distribution μ_0 itself (which is invariant by the Liouville theorem): see [S1,S2],[AA]. Hence (C) implies the ergodic hypothesis in this case and the attractor A can be taken to be the full phase space $\mathcal{C}$.

In this paper we are interested in dissipative systems satisfying (A) where new phenomena occur and μ *is not* the Liouville distribution μ_0. For systems verifying (A),(B),(C) we prove a *fluctuation theorem*, see §7, which is our main technical result.

§*3 Models.*

We now list a number of models to which our theory can conceivably be applied. All these models contain thermostat mechanisms in order to enable the systems to reach a non equilibrium stationary state in the presence of an imposed external field. Model 1 is a model related to electrical conductivity, models 2 and 3 are related to shear flow, model 4 to heat conduction and model 5 to a fluid mechanics model for turbulent flow.

We distinguish, as in §2, between the phase space $\mathcal{F}$ over which the system evolves according to the equations of motion and the collision phase space $\mathcal{C}$ consisting of the timing events ("Poincaré section of $\mathcal{F}$").

The details of the models described here will not be used in the following since our main point is the generality of the derivation of a fluctuation formula from the chaotic hypothesis and its (ensuing) model independence. However, we include them for concreteness and reference.

Model 1: a gas of N identical particles with mass m, interacting via a stable short range spherically symmetric pair potential φ and with an external potential $\varphi^e \neq 0$, enclosed in a box $[-\frac{1}{2}L, \frac{1}{2}L]^2$ and subject to periodic boundary conditions and a horizontal constant external field $E\,\underline{i}$ ($\underline{i}$ is a unit vector in the x–direction). The external potential will be just a hard core interaction which excludes access to a number of obstacles (hard disks, to fix the ideas) so situated that *every* trajectory must suffer collisions with them. The system is in contact with a "thermostat" adding (or subtracting) energy so that the total internal energy stays rigorously constant. The equations of motion are:

$$\begin{aligned}
&\dot{\underline{q}}_j = \frac{1}{m}\underline{p}_j, \qquad \dot{\underline{p}}_j = \underline{F}_j + E\,\underline{i} - \alpha(\underline{p})\underline{p}_j\\
&\underline{F}_j \equiv -\sum_{i\neq j} \underline{\partial}_{\underline{q}_j}\varphi(\underline{q}_j - \underline{q}_i) - \underline{\partial}_{\underline{q}_j}\varphi^e(\underline{q}_j)
\end{aligned} \tag{3.1}$$

with $j=1,\ldots,N$; $\alpha(\underline{p})=E\,\underline{i}\cdot\sum_j \underline{p}_j/\big(\sum_j \underline{p}_j^2\big)$ and $\underline{F}_j$ is the force acting on particle j. The α–term incorporates the coupling to a "gaussian thermostat" and follows from Gauss' "principle of least constraint". The constraint here is the constancy of the internal energy:

$$H(\underline{p},\underline{q})=\sum_{i=1}^{N}\frac{\underline{p}_i^2}{2m}+\sum_{i<j}\varphi(\underline{q}_i-\underline{q}_j)+\sum_i\varphi^e(\underline{q}_i) \tag{3.2}$$

a typical nonholonomic constraint; it follows then from Gauss' principle that the force corresponding to the constraint is proportional to the gradient with respect to $\underline{p}_j$ of H. This model has been studied in great detail, in [CELS], in the case $N=1$ and $\varphi=0$ and φ^e a hard core potential as above, making it a Lorentz model for electrical conductivity if E is an electric field; a similar model has been investigated numerically in [BEC]. It is part of a wide class of models, together with the following models 2,3, whose interest for the theory of non equilibrium stationary states was pointed out in [HHP], [PH], where one can find the first studies performed in the context in which we are interested. The dimension of the phase space $\mathcal{F}$ of this system is $d_0=4N-1$ and that of $\mathcal{C}$ is $d_0-1=2D$ with $D=2N-1$. *The phase space "contraction" rate, i.e. the divergence of the right hand side of (3.1), is* $D\sigma(x)=D\alpha(x)$, which can be written in the form:

$$D\sigma(x)=D\alpha(x)=D\frac{\varepsilon(x)}{DkT(x)} \tag{3.3}$$

where $\varepsilon(x)$ is the work done on the system per unit time by the external field and $kT(x)$ is $\frac{1}{D}\sum_j\frac{\underline{p}_j^2}{m}$ which, if k is Boltzmann's constant, defines a *kind of kinetic temperature*: hence the name of *entropy production rate* per (kinetic) degree of freedom that will be occasionally be given to $\sigma(x)$.[2] Note that $\sigma(x)$ does not have a definite sign.

It has been proved, [CELS], that for $N=1$ and small $E>0$ the average $\langle\sigma\rangle_+$ is positive, *i.e.* the system is dissipative in the sense of §2. There seems to be no reason

[2] If $\bar p$ is the average of $\underline{i}\cdot\underline{p}_j$, then $\frac{1}{N}\sum_j\langle\underline{p}_j^2\rangle=\bar p^2+mkT$ and we see that $T(x)$ cannot be identified with the temperature unless we neglect $\bar p^2$ compared to mkT, *i.e.* we identify the peculiar momentum, needed for the proper definition of the temperature with the ordinary momentum (which should not be done at large E where one cannot identify $\frac{1}{2}\langle\sum_j\underline{p}_j^2\rangle$ with mkT, with T being the usual temperature).

to think that $\langle\sigma\rangle_+$ is not positive. For the above model with $N>1$ no experiments are available yet on the pairing rule or the Lyapunov spectrum smoothness. Nevertheless one can present an argument for the validity of the pairing rule, which may sound convincing but that we have been unable to substantiate mathematically.[3] In this case the time reversal map i is just the usual $i:(\underline{q}\,,\,\underline{p}\,)\to(\,\underline{q}\,,-\,\underline{p}\,)$.

Model 2: a shear flow in a two dimensional container $[-\frac12 L,\frac12 L]^2$ where the particles evolve on a moving background running with velocity $\underline{i}\,\gamma\tilde y$ in the x direction proportional to the height $\tilde y$ in the y direction, is measured with respect to that of the center of mass of the particles. The background exercises a drag on the j–th particle located at height y_j proportional to its peculiar velocity: $\underline{\dot q}_{\,j}-\underline{i}\,\gamma\tilde y_j$ with respect to the background. The introduction of $\tilde y_j$ instead of the usual y_j is due to the boundary conditions we choose (see below), which do not keep the height of the center of mass of the particles fixed. For large N the difference between y_j and $\tilde y_j$ will become negligible (see comment 6) in §8). Similarly for large N (and large L with $n=NL^{-2}$ fixed), the forcing γ should be the shear rate in the fluid, *i.e.* $\gamma=\frac{\partial u_x}{\partial y}$, where $u_x(y)=\underline{i}\,\gamma y$ is the average (local) velocity of the particles in the fluid at height y, as is indeed found in the computer experiments for this system, [ECM2]. If $\underline{F}_{\,j}$ is as in the second equation of (3.1) and $\varphi^e=0$ and $\underline{q}_{\,j}\equiv(x_j,y_j)$, the equations of motion are, using Gauss' principle of least constraint to keep the internal energy fixed:

$$\begin{aligned}&\frac{d^2}{dt^2}\,\underline{q}_{\,j}=\frac1m\,\underline{F}_{\,j}-(\,\underline{\dot q}_{\,j}-\underline{i}\,\gamma\tilde y_j)\,\alpha\\&\alpha(x)\equiv\,-\gamma\big(\textstyle\sum_{j=1}^N\frac{p_{j\,x}p_{j\,y}}{m}-\frac12\sum_{i,j}F_{x\,j,i}\,(y_i-y_j)\big)/\sum_j\,\underline{p}_{\,j}^{\,2}/m\end{aligned}\tag{3.4}$$

with $j=1,\ldots,N$. Here $\underline{p}_{\,j}=m(\,\underline{\dot q}_{\,j}-\underline{i}\,\gamma\tilde y_j)$ is the peculiar momentum relative to the background flow; $\underline{F}_{\,j,i}$ is the force on particle j due to particle i and α is again defined so that (3.2) is a constant of the motion; finally γ plays here the role of a forcing field as E did in model 1. One imposes periodic boundary conditions on the horizontal direction; on the vertical direction a natural boundary condition is perfect

[3] Suppose that the equation (3.1) is modified into the same equation with $\alpha(\,\underline{p}\,)$ constant. Then [Dr],[ECM1] prove that the $4N$ Lyapunov exponents can be paired so that the sum of the corresponding pairs is just $-\alpha$. In model 1 α is *not constant*: however it has an average value $\langle\alpha\rangle$ which is constant on the attractor, with probability 1 with respect to the choice of the initial data (with distribution μ_0): therefore we may hope that "things go as if" α was constant: hence the Lyapunov exponents should be so paired that their sum is $-\langle\alpha\rangle$. *This is not yet the above full pairing rule* because there we assert in addition that half of the exponents are positive and half are negative: and this only "follows" if reversibility (B) is also used.

reflection against the walls at $y = \pm\frac{1}{2}L$. This model has been extensively studied, numerically, in [ECM1], [ECM2] with somewhat different boundary conditions.

One can suppose that the total horizontal component of the peculiar momentum and the horizontal position of the center of mass denoted, respectively, P_x, X_x, are 0: this is consistent with the equations of motion. We shall refer to P_x and X_x as *conserved quantities*: but one should bear in mind that they are such in an "improper" way because they are conserved only if their initial values are 0: P_x in fact relaxes to 0 with a Lyapunov exponent which in general is not zero (and equals the time average $\langle\alpha\rangle_+$ of α).

If it is assumed that $P_x = X_x = 0$ and if one recalls that also H is a constant of the motion and one imposes *a priori* its value, then the dimension of the phase space $\mathcal{F}$ is $d_0 = 4N - 3$, which we write $d_0 = 2D + 1$ for uniformity with the notation in model 1, so that $D = 2N - 2$. The phase space contraction rate, *i.e.* the divergence of the r.h.s. of the equation of motion regarded as first order equations for $\underline{p}$, $\underline{q}$ is $D\sigma(x)$ with:

$$\sigma(x) = \alpha(x) + \gamma \frac{\sum_j p_{xj} p_{yj}}{D \sum_j \underline{p}_j^2} = \alpha(x) + \gamma O(N^{-1}) = \frac{\varepsilon(x)}{DkT(x)} \tag{3.5}$$

where $T(x)$ can *actually* be interpreted as a kinetic temperature, so that $\sigma(x)$ can also be called the entropy production rate.

Numerical experiments with N up to 864 show that $\langle\sigma\rangle_+ > 0$, [ECM1], ECM2], [SEM]. Such papers also provide (strong) evidence for the pairing rule and some (weak) evidence for the smoothness. The time reversal map in this case is *not* the usual velocity reversal but $i : (x, y, p_x . p_y) \to (x, -y, -p_x, p_y)$.

Model 3: this is a model for a shear flow, produced by boundary forces, in contrast to model 2 where the shear is produced by body shear forces.

The flow proceeds in a two–dimensional container $[-\frac{L}{2}, \frac{L}{2}]$ and the equations of motion are simply:

$$\dot{\underline{q}}_j = \frac{1}{m} \underline{p}_j, \qquad \dot{\underline{p}}_j = \underline{F}_j \tag{3.6}$$

supplemented by periodic boundary conditions on the horizontal direction and shear generating boundary conditions in the vertical direction:

$$\omega' = f(\omega) \tag{3.7}$$

where ω is the collision angle formed by the incoming velocity with the x-axis, counted counterclockwise for collisions at $y = \frac{1}{2}L$ and clockwise for collisions at $y = -\frac{1}{2}L$; ω' is the outgoing velocity angle formed with the x-axis, counted clockwise at $y = \frac{1}{2}L$ and counterclockwise at $y = -\frac{1}{2}L$.

With the above angular conventions, $\omega' = \omega$ represents the ordinary elastic collision. We shall consider a "shearing collision rule" $\omega' = f(\omega)$, $\omega' \le \omega$, where f is a reversible collision rule. Reversibility here has the literal meaning: the collision obtained by reversing the particle velocity after a given collision (*i.e.* the incoming collision with an angle $\pi - f(\omega)$) produces afterwards the reverse of the original collision, (*i.e.* $\pi - \omega$). This is:

$$\pi - \omega = f(\pi - f(\omega)), \qquad f(\omega) \le \omega \tag{3.8}$$

where the first condition is the *reversibility* condition and the second is the *shearing* condition. The equations (3.8) can be solved by simply thinking that the graph of $f(\omega)$ as a curve $\omega \to (\omega, f(\omega)) \in [0,\pi]^2$ is a concave arc connecting the point $(0,0)$ with the point (π,π), symmetric by reflection around the secondary diagonal of $[0,\pi]^2$. Furthermore one imposes that the collisions preserve also the horizontal total momentum and the total energy. Since the horizontal momentum of the colliding particle (say particle 1) changes (by $|\underline{p}_1|(\cos\omega' - \cos\omega)$) one can impose the two conservation laws by gaussian minimal constraints, *i.e.* by requiring that the variation of the other momenta is:

$$\underline{p}'_j = (1+\vartheta)\,\underline{p}_j + \beta\,\underline{i} \tag{3.9}$$

for suitable multipliers ϑ and β, which after a brief calculation leads to explicit expressions for β, ϑ, with $\beta = O(N^{-1})$ while $\vartheta = O(N^{-2})$, so that the relations (3.9) generate corrections which, while enforcing the constraints, can be regarded as negligible for large N (see comment 6) in §8).

The symmetry of f also guarantees that the collision rule (3.8) corresponds to a gaussian constraint[4] (forcing $|\underline{p}_1|$ to stay constant in the collision, while changing

[4] For this purpose one has to think, [CL], that the collision rule (3.8) is realized as a limit of a gaussian constraint rule acting on a tiny corridor of width δ expanding vertically the container at the top and bottom, where the particles can enter from the inside but are subject to a horizontal field $\pm E_\delta$ (the sign depending on whether the particles are in the upper or lower corridor): the field will produce a bias in the scattering angle such that the particle will come out with an angle different from the incoming angle. The constraint is that the kinetic energy of the particles inside the corridors does not change; the particles colliding with the external corridor walls are just perfectly reflected. In the limit as $\delta \to 0$ and $E_\delta \to \infty$ (at suitable rates) a reflection rule like (3.8) is realized with a special f. By letting E_δ depend on the distance to the corridors boundaries essentially any

the direction with respect to the elastic collision), see [CL] where, however, the horizontal momentum conservation is not imposed.

For concreteness one can take, following [CL], f such that the above arc is an arc of a circle centered on the secondary diagonal and passing through the indicated points. The circle curvature will be a measure of the shear strength. The dimension of the phase space $\mathcal{F}$ of this system is $d_0 = 4N - 3$, if we fix the energy, the horizontal total momentum and the horizontal position of the center of mass. We write $d_0 = 2D + 1$ as in the previous models so that the dimension of the collision space $\mathcal{C}$ is $2D$ with $D = 2N - 2$. We also suppose, naturally, that the collisions with the walls are among timing events. Then at every collision there is a reduction of phase space volume $\frac{\sin\omega'\, d\omega'}{\sin\omega\, d\omega} = \frac{\sin f(\omega)}{\sin\omega} f'(\omega)$, [CL]. If we define $n(x) = 1$ when x is a collision with the wall and $n(x) = 0$ otherwise, the phase space contraction can be conveniently written as $e^{-D\tau(x)\sigma(x)}$ with $\tau(x)$ equal to the time elapsing between the collision at x and the next at Sx and with the entropy production rate $\sigma(x)$ defined by:

$$D\sigma(x) = -\frac{n(x)}{\tau(x)} \log(\frac{\sin f(\omega)}{\sin\omega} f'(\omega)) \tag{3.10}$$

This model has been studied in detail in [CL].[5]

There is numerical evidence that if $f(\omega) \neq \omega$ then $\langle\sigma\rangle_+ > 0$. The pairing and smoothness properties have not yet been studied. The time reversal operation is the "usual" one (see model 1).

Models 1 and 2 are easier to interpret as dynamical systems than this model, but they are physically somewhat artificial, in that the gaussian thermostat is a rather unconventional model for a thermostat and the shear force is a body force rather than the usual boundary force. In this respect model 3 is better as its unphysical reflection laws are a boundary effect which only produces the net effect of generating a shear on the system.

Model 4) is a model for heat conduction considered in [HHP], [PH]. In a box, $[-L - H, L + H] \times [-\frac{L}{2}, \frac{L}{2}]$, N particles move interacting via a short range pair potential; the boundary conditions are perfect reflection horizontally and periodic vertically;

---

f can be realized, in the limit $\delta \to 0$.

[5] It is not difficult to see, by thinking of the constraint as a (limit of) gaussian constraints as above, that *also* in this model $D\sigma(x)$ is an "entropy production rate".

the particles are subject to the nonholonomic constraint that the total kinetic energy in the left part of the box $[-L-H,-H]\times[-\frac{L}{2},\frac{L}{2}]$ and in the right part of the box $[H,L+H]\times[-\frac{L}{2},\frac{L}{2}]$ have constant values, denoted, respectively by $\frac{L}{2L+2H}NkT_-$ and $\frac{L}{2L+2H}kT_+$, $T_+\ge T_-$, *i.e.* obey:

$$\Phi_\pm=\sum_{j=1}^N \chi_\pm(x_j)\frac{\underline{p}_j^{\,2}}{2m}=\frac{N}{2}\frac{L}{H+L}kT_\pm \tag{3.11}$$

Here $\chi_\pm$ are the characteristic functions of the left and right parts of the box which have to be interpreted as *plates* of thickness L *at temperatures* T_+ *and* T_- respectively; $\underline{q}\equiv(x,y)$. In addition we impose that the total energy (given by (3.2) with $\varphi^e=0$) of the gas $\Phi_0=H$ is exactly conserved.

The constraints are implemented using Gauss' principle, *i.e.* by a force proportional to the gradients with respect to the $\underline{p}_j$'s of $\Phi_\pm$ and Φ_0 (which are simply $\chi_\pm(x_j)\frac{\underline{p}_j}{m}$, $\frac{\underline{p}_j}{m}$, respectively), leading to the equations of motion:

$$\dot{\underline{q}}_j=\frac{1}{m}\underline{p}_j,\qquad \dot{\underline{p}}_j=\underline{F}_j-\alpha_+\chi_+(x_j)\underline{p}_j-\alpha_-\chi_-(x_j)\underline{p}_j-\alpha_0\,\underline{p}_j \tag{3.12}$$

with $\alpha_\pm,\alpha_0$ are defined so that $\Phi_\pm,\Phi_0$ are exact constants of motion. The values of $\alpha_\pm,\alpha_0$ can be easily computed; their general expression will not be needed here. For the purpose of illustrating once more that the resulting forces will lead to a reversible dynamics we give their expression in the simple case in which only $\Phi_\pm$ are imposed: in this case the values of $\alpha_\pm$ are relatively simple and α_0 is not present. One finds:

$$\alpha_\pm=\frac{\sum_j\left(\frac{\underline{p}_j}{m}\cdot\underline{\partial}\chi_\pm(x_j)\frac{\underline{p}_j^{\,2}}{2m}+\underline{F}_j\cdot\frac{\underline{p}_j}{m}\chi_\pm(x_j)\right)}{\sum_j\chi_\pm(x_j)^2\,\underline{p}_j^{\,2}} \tag{3.13}$$

Going back to (3.12) we note that, with the three mentioned constraints, model 4 should be a quite realistic model for heat conduction. The dimension of the phase space $\mathcal{F}$ is $d_0=4N-3$, which again we write as $d_0=2D+1$, with $D=2N-2$, so that the dimension of the collision space $\mathcal{C}$ is again $2D$. The phase space contraction rate is in this case:

$$D\sigma(x)=\alpha_+(x)2N_++\alpha_-(x)2N_-+\alpha_0(x)2N+O(1) \tag{3.14}$$

if $N_\pm$ denote the number of particles in the right and left "plates". Eq. (3.14) could also be interpreted as in the previous models as an entropy production rate.

Some numerical evidence that $\langle\sigma\rangle_+ > 0$ if $T_+ > T_-$ can be found in [PH],[HHP]. No evidence for pairing or smoothness rules seems available. The time reversal map is the "usual" one, see model 1.

The above model equations can be made smoother by replacing χ by a smoothed version of the characteristic functions of the plates; say by functions which are $\equiv 1$ except within a distance of the order of the interaction range from the inner boundaries of the plates: in this region $\chi_\pm$ decrease gently from 1 to zero.

Model 5): this is a model related to turbulent flow, obtained from the Navier Stokes (NS) equations. We consider the NS equations in a box $[-\frac{L}{2},\frac{L}{2}]^3$, with periodic boundary conditions and for an incompressible fluid. If the velocity field is written in a Fourier series as:

$$\underline{u}(\underline{x}) = \sum_{\underline{k}\neq\underline{0}} e^{i\underline{k}\cdot\underline{x}}\,\underline{\gamma}_{\underline{k}} \tag{3.15}$$

with $\underline{\gamma}_{\underline{k}}$ complex vectors with $\underline{\gamma}_{\underline{k}} = \overline{\underline{\gamma}}_{-\underline{k}}$ (reality of the velocity field) and $\gamma_{\underline{k}} \perp \underline{k}$ (incompressibility) then the NS equations become:

$$\dot{\underline{\gamma}}_{\underline{k}} = -i \sum_{\underline{k}_x+\underline{k}_2=\underline{k}} (\underline{\gamma}_{\underline{k}_1}\cdot\underline{k}_2)\,\Pi_{\underline{k}}\,\underline{\gamma}_{\underline{k}_1} + R\,\underline{f}_{\underline{k}} - \nu\,\underline{k}^2\,\underline{\gamma}_{\underline{k}} \tag{3.16}$$

$\Pi_{\underline{k}}$ is the orthogonal projection over the plane orthogonal to $\underline{k}$; ν is the kinematic viscosity and $R\,\underline{f}_{\underline{k}}$ is the forcing (of course orthogonal to $\underline{k}$) which will be taken to be non zero only for a few components with small $\underline{k}$. Since $\underline{k} = \frac{2\pi}{L}\underline{n}$ with $\underline{n}$ integer this means that the force acts only on the high length scale components. For simplicity we may think that the forcing has only two non vanishing components $R\,\underline{f}_{\underline{k}_1^0}, R\,\underline{f}_{\underline{k}_2^0}$, corresponding to two linearly independent wave numbers $\underline{k}_1^0, \underline{k}_2^0$.[6] The number R therefore determines the forcing strength and will be identified with the Reynolds number (we keep the container size L and the viscosity ν fixed). We take $\underline{\gamma}_{\underline{0}} \equiv \underline{0}$ since it is the conserved center of mass velocity.

In order to obtain equations in the framework of this paper from the phenomenological theory of Kolmogorov–Obuchov, [LL], we shall assume that the above equations can be replaced by the following simpler ones:

[6] The simpler case of only one non zero component can be trivial (*e.g.* in 2–dimensional fluids if the forcing acts on the smallest $\underline{k}$, $|\underline{k}| = k_0$) and is therefore discarded here in favor of the next to the simplest, [Ma].

$$
\begin{aligned}
\dot{\underline{\gamma}}_{\underline{k}} &= -i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} (\underline{\gamma}_{\underline{k}_1} \cdot \underline{k}_2)\Pi_{\underline{k}}\gamma_{\underline{k}_1} + \underline{f}_{\underline{k}} && k_0 \le |\underline{k}| < k_R \\
\dot{\underline{\gamma}}_{\underline{k}} &= -\alpha\underline{\gamma}_{\underline{k}} - i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} (\underline{\gamma}_{\underline{k}_1} \cdot \underline{k}_2)\Pi_{\underline{k}}\gamma_{\underline{k}_1} && k_R \le |\underline{k}| < k_R + (LRf)^{1/2}\nu^{-1}
\end{aligned}
\tag{3.17}
$$

here, if $k_0 = \frac{2\pi}{L}$, and $f = \max|\underline{f}_{\underline{k}}|$, the wave vector k_R is the Kolmogorov momentum scale $k_R = k_0 R^{3/4}$, ([LL] p. 122, (32.6)), so that if N_R is the number of wave vectors ("modes") $\underline{k}, -\underline{k}$ such that when $k_0 \le |\underline{k}| \le k_R + (LRf)^{1/2}\nu^{-1}$ then $N_R \approx (\frac{k_R}{k_0})^3 \approx R^{9/4}$; then the phase space $\mathcal{C}$ has dimension $2N_R - 2 = 2D$ with $D \approx R^{9/4}$, while $\mathcal{F}$ has dimension $d_0 = 2N_R - 1$.[7]

This means that the equations for the amplitudes $\underline{\gamma}_{\underline{k}}$ corresponding to $\underline{k}$'s in the *inertial range*, $k_0 \le |\underline{k}| \le k_R$, are "governed" by the reversible Euler equations. In the *viscous range*, $|\underline{k}| > k_R$ the dissipation phenomena will be idealized by saying that the equations are simply such that only the modes $\underline{k}$ with $k_R < |\underline{k}| < k_R + \nu^{-1/2}$ have a non zero amplitude and evolve in such a way as to keep the total energy constant. This means that the parameter α is an effective thermostat (or viscosity), which has to be chosen so that the total energy is constant, *i.e.* so that $\frac{d}{dt}\sum_{\underline{k}} |\underline{\gamma}_{\underline{k}}|^2 = 0$:

$$
\alpha(x) = \frac{\sum_{\underline{k}} \underline{f}_{\underline{k}} \cdot \underline{\gamma}_{-\underline{k}}}{\sum_{|\underline{k}|>k_R} |\underline{\gamma}_{\underline{k}}|^2} \stackrel{def}{=} \frac{\varepsilon(x)}{D_v kT(x)}
\tag{3.18}
$$

Here $\varepsilon(x)$ and $D_v kT(x)$ are simply the numerator and denominator of the fraction defining $\alpha(x)$, if $2N_{vR}$ is the number of modes in the viscous range and one defines $2D_v = 2N_{vR}$.

The Kolmogorov length k_R^{-1} is introduced here phenomenologically and we do not attempt at a fundamental derivation of (3.17), (3.18). Therefore (3.17) has to be regarded as a phenomenological equation. A class of similar models was introduced in [ZJ].

Note that α is proportional to the work $\varepsilon(x)$ per unit time and per viscous degree of freedom performed on the system, which is dissipated into heat, in order to keep the total energy constant: the proportionality constant is $2D_v kT(x)$

[7] Taking into account the reality and incompressibility conditions, forces the $\underline{\gamma}_{\underline{k}}, \underline{\gamma}_{-\underline{k}}$ to have only two linearly independent components.

with $kT(x) \equiv \frac{1}{2D_v} \sum_{|\underline{k}|<k_R} |\underline{\gamma}_{\underline{k}}|^2$ (which, however, is *not* a constant of motion for (3.17),(3.18) because of the imposed constraint that $\sum_{\underline{k}} |\underline{\gamma}_{\underline{k}}|^2$ is constant rather than $\sum_{|\underline{k}|>k_R} |\underline{\gamma}_{\underline{k}}|^2$). The phase space contraction rate is in this case:

$$D_v \sigma(x) = D_v \alpha(x) = D_v \frac{\varepsilon(x)}{D_v kT(x)} \tag{3.19}$$

Hence $D_v \langle\sigma\rangle_+$ can be thought of as the average amount of *energy dissipation* per unit time by the flow divided by the *kinetic energy* contained in the *viscous modes*. The first quantity plays a major role in Kolmogorov's theory, see [LL] p.119, and its average is usually called ε, (see [LL], (31.1)). Since the kinetic energy contained in the viscous modes can be thought of as a kind of "temperature" we see that $2D_v\langle\sigma\rangle_+$, is proportional to the entropy "production rate". More appropriately we can say that, for R large, $2D_v\langle\sigma\rangle_+$ is proportional, once more, to the "energy dissipation rate" over a kinetic quantity equal to the average kinetic energy contained in the viscous modes *if*, for large R, the two quantities can be regarded as independent random variables.

Note, however, that for the above model (3.17) (introduced, we believe, for the first time here) there is no evidence for $\langle\sigma\rangle_+ > 0$ or for the pairing and smoothness rules. The time reversal map is simply $i : \{\underline{\gamma}_{\underline{k}}\} \to \{-\underline{\gamma}_{\underline{k}}\}$.

§4 The introduction of the SRB distribution.

We now present a heuristic argument providing, in our opinion, a useful characterization of the SRB distribution: this point of view is important for the applications of §7. Our purpose is to look at it from a somewhat different perspective than usual and to show that it leads to a new interpretation of the ergodic hypothesis and to a unification of equilibrium and nonequilibrium statistical mechanics.

We deal with systems of N particles verifying the properties (A,B,C) of §2 and we observe their motions $x \to S^n x$ at discrete times $n = 0, \pm 1, \ldots$ in the collision space $\mathcal{C}$ of dimension $2D$, see §2. For the geometrical arguments used below (see, *eg*, Fig. 1) it will be very useful to keep in mind the paradigm of hyperbolic systems: namely the Anosov map of the 2-dimensional torus T^2:

$$S \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} \mod 2\pi \tag{4.1}$$

which plays a role analogous to that of harmonic oscillators in classical mechanics. This example is not only enlightening, but it is really the main source of intuition. Note that this is a reversible map if i is defined as $i:(\varphi_1,\varphi_2)\to(\varphi_2,-\varphi_1)$, because $iS\equiv S^{-1}i$.

Let O be a fixed point on the attractor A and let W^u_O be the unstable manifold W^u_O of O (dense on A, see (C), §2).[8] The dimension of W^u_O is D, half that of the phase space $\mathcal{C}$, see §2.

For simplicity we shall suppose that O is a time reversal invariant fixed point $O=iO$; this assumption could be easily relaxed.[9]

The key idea on which we base our analysis is that the attractor A should be considered to consist of the *smooth* D–dimensional unstable manifold W^u_O of O (or of any other fixed point or periodic orbit in A with dense stable and unstable manifolds). Of course the manifold W^u_O can only fill A densely: we "lose" the accumulation points. But all the information needed to perform time averages should be already contained in W^u_O itself, as we are only interested in the averages of rather regular observables (*e.g.* piecewise smooth). We think that W^u_O coincides with A in the same sense in which the rationals can be regarded as coinciding with the reals in integration theory (which works only if one considers integrals of smooth functions) and can be used to compute numerically the integrals of smooth functions. In the same way statistical averages with the distribution (2.1) should be computable by simply approximating them with integrals over finite parts of W^u_O, like the sets $S^T\Delta$ obtained by "blowing up" with a large time iterate S^T a small connected surface element Δ of W^u_O containing O.

In other words we want to regard the possible fractality of A as a rather irrelevant accident. We want to think of A as *essentially* identical to W^u_O: the latter surface folds over and over again, being enclosed in the *bounded* phase space $\mathcal{C}$. It therefore

---

[8] There might be *no such point* fixed point O. However a periodic orbit starting at a point P and with period n would be a fixed point for S^n and we could get all the following conclusions by replacing S with S^n, since the statistics of S^n and that of S coincide, when S is chaotic enough. Thus assuming the existence of a fixed point is not restrictive.

[9] It is not difficult to realize that in the models 1,2,3,4 there are always periodic orbits which are time reversal invariant, *i.e.* such that iO is also on the orbit; at least if one is willing to limit the particle density in some interval (whose size may depend on the range of the interaction). Also for model 5 it is very likely that periodic motions (unstable, of course) do exist. Note that since we are assuming (C), §2, it is automatically true that there are periodic orbits (*i.e.* chaotic systems always have a lot of periodic unstable orbits), [S2].

folds itself in $\mathcal{C}$ just as an uncut folio is folded into a book, thereby generating an almost three dimensional fractal set out of a two dimensional smooth manifold. But thinking of A as an unfolded manifold of half the space dimension (D in our notation) leads to a change in the usual point of view, which regards A as a fractal set with dimension close to $2D$.

Introducing forcing and friction (*i.e.* passing from an equilibrium to a stationary non equilibrium problem) should then *not* be thought of as a real "discontinuity": which would be the case if one took the viewpoint that one is passing from a nice smooth $2D$ dimensional attractor $A = \mathcal{C}$ to a nasty, strange, fractal $A \subset \mathcal{C}$ with dimension $2D - O(\alpha\lambda_{\rm max}^{-1})D$, *macroscopically different* from $2D$ (as implied by the pairing (D) together with the smoothness (E), §2).

Rather it should be viewed as an insignificant deformation of the unstable manifold W_O^u which will fold itself in $\mathcal{C}$ not *exactly* as in the conservative case, but leave a few holes between the "pages" to account for its global fractality. This is a change with respect to the conventional point of view for the case of conservative systems: these are no longer really different from the dissipative ones. Their attractor has, in the new (*unconventional*) sense, *exactly* half the dimension of the full phase space (the dimension the conventional point of view attributed to them is that of the *full* phase space, *i.e.* twice as large).

The main consequence of such a viewpoint, besides the mentioned unification of conservative and dissipative dynamics, is that it allows us to think of the attractor as "unfoldable', with the consequence that our intuition about the motion on the attractor is greatly enhanced.

This unfolded attractor, imagined as a flat infinite surface, attracts exponentially fast nearby points: the approach to the attractor follows the stable manifolds associated with the attractor points, which can be thought of as *needles* sticking out of the attractor itself. The motion essentially consists, therefore, of an expanding (*i.e.* as unstable as possible) motion on the unstable manifold W_O^u.

We can now easily understand the statistics μ, (2.1), on the attractor, *i.e.* the SRB statistics, as follows.

Let U be a ball with small radius h, centered at the fixed point O; and let us ask how to compute the time average of an observable F, if the initial data are chosen in U with uniform distribution, say, with a distribution absolutely continuous with respect to the Liouville distribution.

Clearly the average of F over a large time T will be computable by looking at the image S^TU under S^T for large T and by imagining S^TU covered by the density into which the initial uniform density in U evolves in T time steps. If we call Δ *the connected part* of $W^u_O \cap U$ *this also means that we can regard its* S^T *image,* $S^T\Delta$*, of the connected part of* $W^u_O \cap U$ *as a good finite approximation* $S^T\Delta$ *to our attractor.*

The set S^TU will be extremely thin and it will "*coat*" the extremely large portion of W^u_O defined by $S^T\Delta$, (*i.e.* by our "good finite approximation" of the attractor), if we regard the attractor as unfolded.

Let dx be a surface element on W^u_O and let us regard S as a map of W^u_O into itself. *We shall call* $\Lambda_u(x)$ *the absolute value of the jacobian determinant* $\partial_u S(x)$ *of* S*, as a map of* W^u_x *into itself, at the point* x. In this way $\Lambda_u(x)$ will be the absolute value of the determinant of a matrix with a dimension equal to that of W^u_x, *i.e.* D. Then we are interested in computing the integral:

$$\int_{S^T\Delta} \rho_T(x)\, F(x)\, dx \tag{4.2}$$

where $\rho_T(x)dx$ is the amount of mass in the cylinder with base dx, which is the image of the cylinder in U with base $S^{-T}dx$ and height equal to the height h of the initial "cloud of data" U. Denoting d_s (resp. d_u) the dimension of the stable (unstable) manifold of O (which in our case are $d_s = d_u = D$), this means that $\rho_T(x)dx$ is proportional to $h^{d_s}|S^{-T}dx|$, where $|S^{-T}dx|$ is the surface area of $S^{-T}dx$. By the above definition of the local expansion rate $\Lambda_u(x)$, one has then:

$$\rho_T(x)dx = const\, \Lambda_u^{-1}(S^{-T}x)\ldots\Lambda_u^{-1}(S^{-1}x)dx \xrightarrow[T\to\infty]{} const \prod_{j=-\infty}^{-1} \Lambda_u^{-1}(S^j x)\, dx \tag{4.3}$$

which is, clearly, a formal relation because ρ_T tends to 0 as $T \to \infty$. Note, however, that (4.3) implies that the *ratios* between $\rho_T(x)$ and $\rho_T(x')$, with x, x' in the surface elements dx, dx', *are well defined*, even in the limit as $T \to \infty$.

The equation (4.2) is for T large already a "good approximation" for the SRB distribution. It shows that the statistical averages should be computable by looking at a large part of W^u_O, namely at the finite approximation of the attractor A called $S^T\Delta$ above, and by imagining it *coated* with a density $\rho_T(x)$, and then using (4.2).

The existence of the limit as $T \to \infty$ of (4.2) can be seen by remarking that the limit can in fact be written as an integral over phase space, in spite of the fact that $\rho_T(x)$ tends manifestly to 0 as $T \to \infty$. For, when $T \to \infty$ what really matters is the amount of mass ending up inside a generic little square E in the phase space $\mathcal{C}$, with center x_E. Since E will be cut many times by $S^T\Delta$ we can imagine that the various "pieces" of $S^T\Delta$ intersecting E are piled up "vertically" in E: fig. 1 shows a picture for the simple case (4.1).

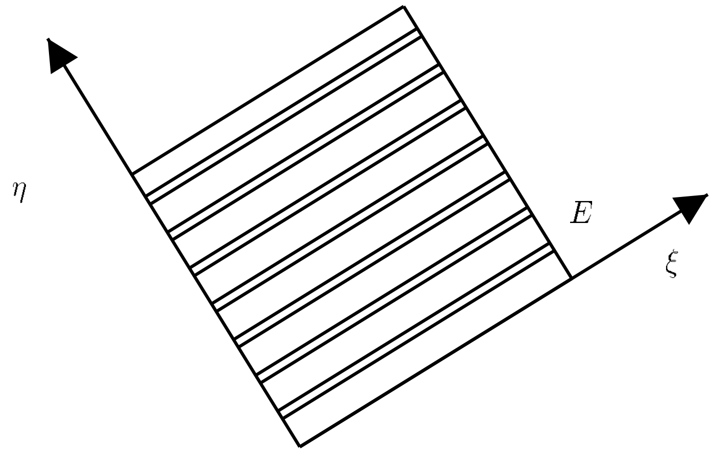

Fig. 1: The parallel lines are intersections of the finite approximation for the attractor, $S^T\Delta$, with the set E, represented by a square. The ξ,η axes are "parallel" to the unstable and stable manifolds W^u_O, W^s_O respectively. Each of them is coated, eventually, by the image S^TU of U which gives them a thickness, (not shown). In the case of the map (4.1) the parallel lines are generated, if one moves on W^u_O away from O, in the following typical order: from bottom to top first one draws, successively, the lower line of each pair; then one draws the second, then one should draw a third series of lines above the second and keep going in this way until the endpoints of $S^T\Delta$ are reached. For $T\to\infty$ the parallel lines fill densely E.

If $\mu_T(E)$ is the total mass initially in U ending up in E after time T, we can rewrite (4.2) as:

$$\sum_E \mu_T(E)F(x_E) \tag{4.4}$$

provided E is so small that we can neglect the variation of F inside E, and that the E's pave the phase space. Suppose we set up a coordinate system in the small box E (which is a box with full dimension $2D$) so that the "horizontal" coordinates are called ξ and the "vertical" ones η. A point in $E\cap A$ is denoted $x(\xi,\eta)$ and the surfaces of constant η are connected surface elements of the unstable foliation W^u in E (a foliation of a set E is a family of disjoint connected surfaces whose union is E), while those of constant ξ are connected surface elements of the stable foliation W^s.

Then we see that (4.2), before the limit as $T \to \infty$ is taken, can be expressed as a sum over the connected parts of the surface[10] $S^T\Delta$ that fall in E (the parallel lines in the above figure)). If $\{\eta_j\}$ represent the j-th line then we can write (4.2) as:

$$\sum_{\eta_j} \int_{\{\eta_j\}} \rho_T(\xi,\eta_j) F(x(\xi,\eta_j))\, d\xi \xrightarrow[T\to\infty]{} \int_E \overline{\rho}^{\eta}(\xi) d\xi\, \nu(d\eta)\, F(x(\xi,\eta)) \tag{4.5}$$

where $d\xi$ denotes an area element on $\{\eta_j\}$. In the limit $T \to \infty$, while $\rho_T(\xi,\eta_j)$ tends to 0 the number of lines $\{\eta_j\}$ tends to infinity and the sum over the surface elements of $S^T\Delta$ that cross E *should converge* to an integral over η and ξ with respect to some measure $\overline{\rho}^{\eta}(\xi)\nu(d\eta)d\xi$ with both the density $\overline{\rho}^{\eta}(\xi)$ along the unstable manifold and the measure ν well defined. The measure ν will give us the detailed information on how the various pieces (*layers*, or lines in fig. 1) of U intersecting E pile up and the distribution of the gaps between them in E, [11] hence on its fractal nature. On the other hand $\overline{\rho}^{\eta}(\xi)$ will be a function such that the ratios $\int \overline{\rho}^{\eta}(\xi) d\xi \nu(\varepsilon) / \int \overline{\rho}^{\eta'}(\xi) d\xi \nu(\varepsilon')$ should tell us the ratio of the masses of U ending up in connected layers that we can denote ε and ε', coating the pieces of unstable manifold passing through η and η', inside E, which should be well defined in the limit $T \to +\infty$ as argued above. We regard as natural, in the above context in which E is fixed, to define ν so that $\nu(\varepsilon)$ is the mass of U ending in a layer ε around η: which implies $\int \overline{\rho}^{\eta}(\xi) d\xi = 1$.

§5 *The thermodynamic analogy.*

In this section we describe theoretical difficulties with the heuristic analysis of §4 and with a mathematical proof of the existence of the limit of (4.2). The solution to the difficulties that will be pointed out necessitates the introduction of more refined ideas and eventually the use of Markov partitions. We first point out the difficulty.

If the analysis of the previous section is correct, *i.e.* , if $\overline{\rho}^{\eta}(\xi)$ really exists, we should be able to "calculate" it, at least formally. While it is evident that the function $\overline{\rho}^{\eta}(\xi)$ is defined up to a constant for each η such that $(\xi,\eta) \in W^u_O$, (because of formula (4.3)

[10] Which, we recall, represents the finite approximation to the attractor defined above.

[11] Note that "gap" here does not mean an actually empty region: since W^u_O is dense in E (by (C), §2) there can be no open regions in E which are not crossed by one connected part of W^u_O. In general, one should think of ν as supported by a dense "Cantor set". The limit (4.5) only defines, of course, the product $\overline{\rho}(\xi) d\xi \nu(d\eta)$.

and the comment following it), it is much less evident that $\overline{\rho}^{\eta}(\xi)$ behaves reasonably regularly in $(\xi,\eta)\in E\cap W^u_O$.

Clearly the right hand side of equation (4.3) can be used to compare the values of $\overline{\rho}^{\eta}(\xi)$ and of $\overline{\rho}^{\eta}(\xi')$ if $x=(\xi,\eta)$ and $x'=(\xi',\eta)$ are points of $E\cap W^u_O$ with the same η (note that in such case only few of the factors in (4.3) have ratios really different from 1, because $S^{-j}x$ and $S^{-j}x'$ approach O exponentially fast and start close on a connected part of W^u_O). However if we compare $\overline{\rho}^{\eta}(\xi)$ and $\overline{\rho}^{\eta'}(\xi)$ with the same ξ and $(\xi,\eta),(\xi,\eta')\in E\cap W^u_O$ we run into the difficulty that the distance between $(\xi,\eta),(\xi,\eta')$ *measured along* W^u_O *may be extremely long*, in fact as long as we please by varying the two points on the surface $\xi=constant$. Therefore the function $\overline{\rho}^{\eta}(\xi)$ *might vary quite irregularly in* E. Hence we see that the existence of a limit in (4.5) is not so obvious even though (4.3) provides immediately an expression for the ratios of the limit density on the set $W^u_O\cap E$.

We *must* find an alternative way to control the variations of such a function in $E\cap W^u_O$ in order to argue that $\overline{\rho}^{\eta}(\xi)$ is well defined. The number of connected components of W^u_O in E is denumerable and we cannot expect that the SRB distribution is supported by a denumerable set of d_u–dimensional surfaces. Hence we are in a position similar to when attempting to define the integral of a continuous function over a segment from the knowledge of the function at the rational points on the segment: this is possible only if the function is not too wildly changing from point to point.

The resolution of this difficulty proceeds in two steps. First we push the analysis of the variability of $\rho_T(\xi,\eta)$ just given somewhat further to arrive at the equations (5.3), (5.4) below which are useful to illustrate the development of the thermodynamic analogy that gave rise to the *thermodynamic formalism*. This will enable us, in §6, to discuss the proper solution to the problem of the existence of the limit (4.5) and, implicitly, of the function $\overline{\rho}^{\eta}(\xi)$, based on this thermodynamic analogy.

Let $x=(\xi,\eta), x'=(\xi,\eta')$ be points of $E\cap W^u_O$ and let $d\xi$ and $d\xi'$ be two infinitesimal surface elements in $E\cap W^u_O$ at different heights η and η', *corresponding* to each other, in the sense that the stable manifolds through $d\xi\subset W^u_x$ intersect the unstable manifold $W^u_{x'}$ exactly on $d\xi'$, see fig. 2 below. Then the masses $\rho_T(\xi,\eta))d\xi$ coating the segment $d\xi$ and $\rho_T(\xi,\eta')d\xi'$ coating the segment $d\xi'$ have a ratio that can be computed by using (4.3) and by remarking that the ratio of the areas $d\xi/d\xi'$ is:

$$\frac{d\xi}{d\xi'}=\frac{|S^{-1}(Sd\xi)|}{|S^{-1}(Sd\xi')|}=\frac{|S^{-2}(S^2d\xi)|}{|S^{-2}(S^2d\xi')|}=\ldots=\frac{|S^{-M}(S^Md\xi)|}{|S^{-M}(S^Md\xi')|}\tag{5.1}$$

hence:

$$\frac{d\xi}{d\xi'} = \Big(\prod_{j=0}^{M-1} \frac{\Lambda_u^{-1}(S^j x)}{\Lambda_u^{-1}(S^j x')} \Big) \frac{|S^M d\xi|}{|S^M d\xi'|} \tag{5.2}$$

See Fig. 2 for an illustration:

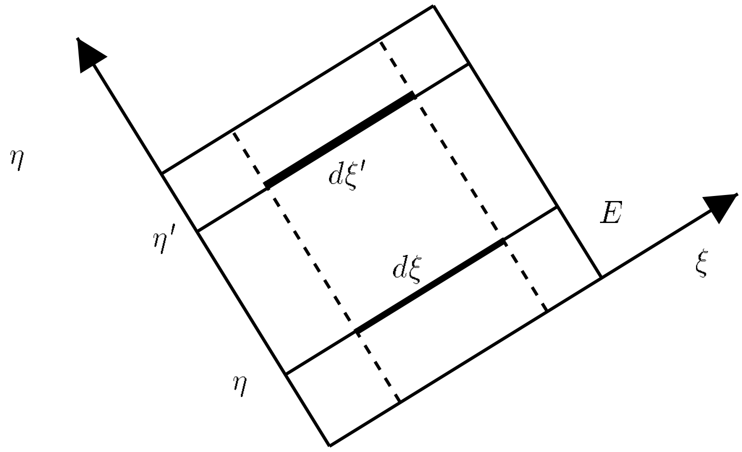

Fig. 2: The two lines at constant η are at height η and η' respectively. The two infinitesimal segments $d\xi$ and $d\xi'$ correspond to each other as they are crossed by the same set of stable manifolds, the extreme two of which are drawn as dashed lines.

But $\frac{|S^M d\xi|}{|S^M d\xi'|} \xrightarrow[M\to\infty]{} 1$ because the two segments approach each other, while greatly and chaotically erring towards ∞ on W_O^u at the exponential speed of the expansion rates.

Hence by combining (5.2) for $M \to \infty$ and (4.3), we see that the ratio of the masses $\rho_T(\xi,\eta)d\xi/\rho_T(\xi,\eta')d\xi'$ in corresponding intervals $d\xi, d\xi'$ near the corresponding points $x, x' \in E$ with local coordinates (ξ,η) and (ξ,η'), is simply:

$$\prod_{-\infty}^{+\infty} \frac{\Lambda_u^{-1}(S^j x)}{\Lambda_u^{-1}(S^j x')} \tag{5.3}$$

This shows that the SRB distribution μ can be formally given by attributing to the "points" on the unstable manifold of A a weight given by:

$$const \prod_{-\infty}^{+\infty} \Lambda_u^{-1}(S^j x) = const\, e^{-\sum_{-\infty}^{+\infty} \log \Lambda_u(S^j x)} \tag{5.4}$$

or by a density on W_O^u given formally by the product in (4.3). Such statements should be interpreted in the same way in which one interprets statements like: "the one dimensional Ising model with nearest neighbor interaction attributes to the spin configuration $(\sigma_i)_{-\infty}^{\infty}$ the probability":

$$const\, e^{-\sum_{-\infty}^{\infty} J\sigma_i\sigma_{i+1}} \tag{5.5}$$

The formal expression (5.4) must, therefore, be interpreted as a limiting statement. The function $h(x) \equiv \log\Lambda_u(x)$ in (5.4) plays the same role as $J\sigma_0\sigma_1$ in the Ising model in (5.5) and the appropriate way of understanding (5.4) is, as we have in fact discussed, as a limit of (4.3). The important realization of Sinai, [S2], was that the equation (5.4), via (5.5), had a close analogy in statistical mechanics.[12]

This remark led Sinai to his general theory of Markov partitions (see below) which is the main technical tool that is used to show mathematically that the limit of (4.2) or (more precisely) the limit in (4.5) really exist and for describing its general properties in satisfactory detail (by deducing them from the well known theory of one dimensional Gibbs states of spin systems with exponentially decaying interaction potentials; the latter is not really different from the theory of the Gibbs state corresponding to (5.5)).

Without entering into the details of Sinai's work we shall use (5.4) in §6 to give a heuristic justification of formula (6.3) below, which is the basis of our analysis in §7.

§6 *Coarse graining and Markov partitions.*

The above discussion has, as already said at the beginning of §5, just heuristic value as it has not led to a really usable formula for $\overline{\rho}^{\eta}(\xi)\, d\xi\nu(d\eta)$, but just to a few relations that such function must obey when evaluated on W^u_O.

The solution lies in a stricter interpretation of the thermodynamic analogy (*i.e.* the similarity between (5.4) and (5.5)). To understand the rigorous solution (given in [S2]) to the problem of showing the existence of the limit (4.5) and the existence of $\overline{\rho}^{\eta}(\xi)$ and $\nu(d\eta)$, one has to introduce the concept of a "parallelogram" and of a *Markov partition* $\mathcal{E}$ of the phase space $\mathcal{C}$ into parallelograms. This can be ultimately related to the problem of constructing a good division of the phase space in cells (*i.e.* a "good" coarse graining) so that the evolution can be correctly represented as a cell permutation, without "distorting" the hyperbolic nature of the motion (for such an interpretation of what follows see [Ga4]).

[12] Hence the name of "thermodynamic formalism" given by Ruelle to the mathematical theory based on the above point of view, [R1].

A parallelogram will be a small set with a boundary consisting of pieces of the stable and unstable manifolds joined together as described below. The smallness has to be such that the parts of the manifolds involved look essentially "straight": *i.e.* the sizes of the sides have to be small compared to the smallest radii of curvature of the manifolds W_x^u and W_x^s, as x varies in $\mathcal{C}$.

Therefore let δ be a length scale small compared to the minimal (among all x) curvature radii of the stable and unstable manifolds. Let $W_x^{u,\delta}, W_x^{s,\delta}$ be the connected parts of W_x^u, W_x^s containing x and contained in a sphere of radius δ.

Let us first define a *parallelogram* E in the phase space $\mathcal{C}$, to be denoted by $\Delta^u \times \Delta^s$, with center x and axes Δ^u, Δ^s with Δ^u and Δ^s small connected surface elements on W_x^u and W_x^s containing x. Then E is defined as follows. Consider $\xi \in \Delta^u$ and $\eta \in \Delta^s$ and suppose that the intersection $\xi \times \eta \equiv W_\xi^{s,\delta} \cap W_\eta^{u,\delta}$ is a unique point (this will be so if δ is small enough and if Δ^u, Δ^s are small enough compared to δ as we can assume, because the stable and unstable manifolds are "smooth"[13] and transversal, see footnote [1]).

The set $E = \Delta^u \times \Delta^s$ of all the points generated in this way when ξ, η vary arbitrarily in Δ^u, Δ^s will be called a parallelogram (or rectangle), if the boundaries $\partial\Delta^u, \partial\Delta^s$ of Δ^u and Δ^s as subsets of W_x^u and W_x^s, respectively, have zero surface area on the manifolds on which they lie. The sets $\partial_u E \equiv \Delta^u \times \partial\Delta^s$ and $\partial_s E = \partial\Delta^u \times \Delta^s$ will be called the *unstable* or *horizontal* and *stable* or *vertical* sides of the parallelogram E.

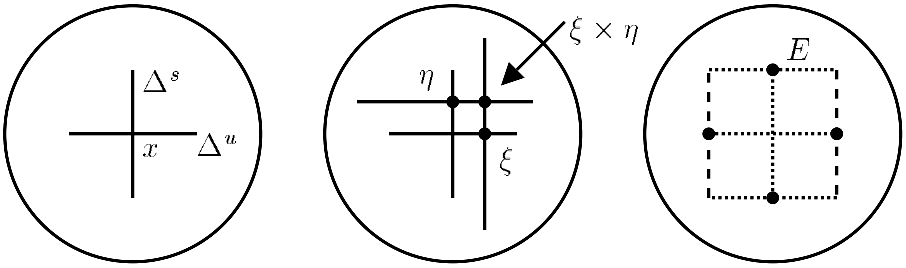

Fig. 3: The circles are a neighborhood of x of size very small compared to the curvature of the manifolds; the first picture shows the axes; the intermediate picture shows the $\times$ operation and $W_\eta^{u,\delta}, W_\xi^{s,\delta}$ (the horizontal and vertical segments through η and ξ, respectively, have size δ); the third picture shows the rectangle E with the axes and the four marked points are the boundaries $\partial\Delta^u$ and $\partial\Delta^s$. The picture refers to a two dimensional case and the stable and unstable manifolds are drawn as flat, i.e. the Δ's is very small compared to the curvature of the manifolds.

[13] This is only approximately true because the tangent planes of the stable and unstable manifolds are only Hölder continuous with some positive exponent, related to the gap between the positive or negative Lyapunov exponents and 0.

Consider now a partition $\mathcal{E} = (E_1, \ldots, E_{\mathcal{N}})$ of $\mathcal{C}$ into $\mathcal{N}$ rectangles E_j with pairwise disjoint interiors. We call $\partial_u \mathcal{E} \equiv \cup_j \partial_u E_j$ and $\partial_s \mathcal{E} \equiv \cup_j \partial_s E_j$: these are called respectively the *unstable boundary* of $\mathcal{E}$ and the *stable boundary* of $\mathcal{E}$, or also the horizontal and vertical boundaries of $\mathcal{E}$, respectively.

We say that $\mathcal{E}$ is a *Markov partition* if the transformation S acting on the stable boundary of $\mathcal{E}$ maps it into itself (in formula this is: $S\partial_s\mathcal{E} \subset \partial_s\mathcal{E}$) and if likewise the map S^{-1} acting on the unstable boundary maps it into itself ($S^{-1}\partial_u\mathcal{E} \subset \partial_u\mathcal{E}$).

The actual construction of the SRB distribution then proceeds from the important result of the theory of Anosov systems expressed by what we shall call "Sinai's first theorem":

Theorem: every transitive Anosov system admits a Markov partition $\mathcal{E}$, [S2].

The above theorem is the first step towards a controlled version of the heuristic arguments given above and towards a usable form of equation (4.5) based on a suitable interpretation of (5.3). It can be extended to imply the existence of more special Markov partitions: for instance to show the existence of Markov partitions with any one of the following three properties (the last shows that the first two can be realized simultaneouly and will play a key role in our analysis):

1) The construction of $\mathcal{E}$ can be done, [Ga3], so that the horizontal axes of E_j *all lie on* W^u_O (and the vertical on W^s_O) and their union is a set that can be obtained from a single small connected surface element $\overline{\Delta}$ of W^u_O (resp. $\overline{\Delta}'$ of W^s_O) containing O by dilating it with a high iterate S^Q of the time evolution S. In other words the union $\cup_j \Delta^u_j$ of the horizontal axes of the parallelograms $E_j \in \mathcal{E}$ can be regarded as a good finite approximation to our attractor A, because it has the form $S^Q\overline{\Delta}$ with $\overline{\Delta}$ a connected surface element of the unstable manifold W^u_O, containing O. Likewise the union of the stable axes can be regarded as a large connected part of the stable manifold W^s_O.

2) If the reversibility property holds it is clear that $i\mathcal{E}$ is also a a Markov partition. This follows from the definition of Markov partition and from the fact that reversibility implies:

$$W^s_x = iW^u_{ix} \tag{6.1}$$

The definition of a Markov partition also implies that the intersection of two Markov partitions is a Markov partition, hence it is clear that there are Markov partitions $\mathcal{E}$

that are reversible in the sense that $\mathcal{E} = i\mathcal{E}$.

3) Furthermore one can construct a Markov partition $\mathcal{E}$ which is reversible and at the same time verifies the property 1) above, [Ga3].

Here we shall use Markov partitions that verify property 3) above.

In order to formulate Sinai's second theorem, which gives an expression for a controlled approximation to the SRB distribution, we consider the partition $\mathcal{E}_T = \cap_{-T}^{T} S^{-j}\mathcal{E}$ obtained by intersecting the images under S^j, $j = -T, \ldots, T$ of $\mathcal{E}$. Then $\mathcal{E}_T$ is still a Markov partition and it is time reversal invariant if $\mathcal{E}$ is. We now construct a probability distribution that we regard concentrated on the finite approximation A_T to the attractor, consisting of the union of the horizontal axes of $\mathcal{E}_T$, (see remarks 1),2) above), and equal, with the notations of remark 1 above, to $S^{T+Q}\overline{\Delta} = A_T$.

We note that the parallelograms of $\mathcal{E}_T$ can be labeled by the strings of symbols $j_{-T}, \ldots, j_T$ and they consist of the points x such that $S^k x \in E_{j_k}$ for $-T \le k \le T$. In other words the parallelograms consist of those points x which in their time evolution visit at time k the parallelogram j_k (one usually says that these are the points whose *symbolic dynamics* on $\mathcal{E}$ coincide, at the times ("sites") k between $-T$ and T). One can convince oneself that the transformation properties of the boundaries of the parallelograms of $\mathcal{E}$ imply that the strings $j_{-T}, \ldots, j_T$ that can be generated by points in phase space can be identified with those which are "compatible with nearest neighbors", *i.e.* such that $SE_{j_k} \cap E_{j_{k+1}}$ do have interior points. It is the latter property that allows us to regard (5.4) as essentially similar to (5.5) and the SRB distribution as a Gibbs state on a system of spins (the labels j of the parallelograms in $\mathcal{E}$) subject to a hard core short range potential (*i.e.* with a compatibility condition between nearest neighbor spins).

In a parallelogram any point can be regarded as center. However there are special points that are usually taken as centers because they play a special role. Accordingly we shall take as center $x_{j_{-T},\ldots,j_T}$ of a parallelogram $E_{j_{-T},\ldots,j_T}$ a point whose symbolic dynamics string $\underline{j}$ at the times k with $|k| > T$ is fixed in a standard way; *i.e.* by defining j_k for $k > T$ (respectively $k < -T$) as a compatible sequence depending only on j_T (respectively j_{-T}). We cannot in general make the choice of continuing $j_{-T}, \ldots, j_T$ at times $k > T$ or $k < -T$ with a fixed symbol because this may lead to an incompatible sequence (hence the choice that we make is "the simplest" possible: it leaves some arbitrariness because there are many strings of symbols that start with a given symbol. The arbitrariness is however irrelevant for what follows).

We can visualize the small parallelograms forming $\mathcal{E}_T$ as a lattice of parallelograms: two parallelograms adjacent and on the same vertical strip will have horizontal axes that correspond to each other in the same sense that the close horizontal surface elements used in deriving (5.3) correspond. Therefore if we attribute to the horizontal axis of the parallelogram E_j in $\mathcal{E}_T$ with center x_j a weight equal to $\overline{\Lambda}_{u,\tau}^{-1}(x_j)=\prod_{h=-\tau/2}^{\tau/2-1}\Lambda_u^{-1}(S^h x_j)$ for some large number of time steps $\tau\le 2T$ *we see that the ratios of the weights of corresponding surface elements automatically realize an approximation of the product ("with $\tau=\infty$") in (5.3)*, at least if we take however $\tau\ll T$ so that the size of each parallelogram is so small that the weight we attribute to each does not depend on which point of E_j we regard as a center and that no essential ambiguity arises as to which weight to attibute to a parallelogram. [14] Note that the above weight is the inverse of the expansion coefficient of the map S^τ as a map of
$W^u_{S^{-\tau/2}x}$ to $W^u_{S^{\tau/2}x}$ (between $S^{-\tau/2}x$ and $S^{\tau/2}x$), *i.e.* :

$$\overline{\Lambda}_{u,\tau}(x)=\prod_{j=-\tau/2}^{\tau/2-1}\Lambda_u(S^j x) \tag{6.2}$$

A similar quantity $\overline{\Lambda}_{s,\tau}(x)$ can be defined by regarding S^τ as a map of $W^s_{S^{-\tau/2}x}$ to $W^s_{S^{\tau/2}x}$.

The construction thus generates a probability distribution which, by the above analysis, verifies (5.3) more and more exactly as $\tau\to\infty$. Hence this analysis suggests the following theorem which we shall call "Sinai's second theorem", (see [S2]; for details about the connection of what can actually be found in [S2], [Bo], [R2] and the interpretation that we quote below see [Ga3] Sec. 3 and Eq. (3.12) in particular, or [Ga2] Eq. (1.10)):

Theorem: If $(\mathcal{C},S)$ is a transitive Anosov system the SRB distribution μ exists and the μ average of a smooth function F is:

$$\int_{\mathcal{C}}\mu(dx)F(x)=\lim_{T\ge\tau/2\to\infty}\frac{\sum_j\overline{\Lambda}_{u,\tau}^{-1}(x_j)F(x_j)}{\sum_j\overline{\Lambda}_{u,\tau}^{-1}(x_j)}\stackrel{def}{=}\lim_{T\ge\tau/2\to\infty}\int_{\mathcal{C}}\mu_{T,\tau}(dx)F(x) \tag{6.3}$$

[14] The size of the parallelograms of $\mathcal{E}_T$ is clearly decreasing as $e^{-\lambda T}$, at least, if λ is the spectral gap, see footnote [1] above. However our "natural choice" of the center x_j dispenses us from this restriction in the formulation of the theorem below.

where, with the above notations, x_j is the point chosen in $E_j \in \mathcal{E}_T$.

The above $\mu_{T,\tau}$ as defined by the middle ratio in (6.3), can be taken as a "concrete" procedure to follow in approximating the SRB distribution.

In the case of equilibrium under assumption (C) in §2 the distribution μ in equation (6.3) can be shown to coincide with the microcanonical ensemble, as already mentioned, [S1].[15]

§7 *Application.*

The chaotic hypothesis can be taken as an extension of the ergodic hypothesis for equilibrium statistical mechanics to systems in nonequilibrium stationary states (conservative or dissipative). In the equilibrium (*i.e.* conservative) case it implies the ergodic hypothesis (*but it is stronger*) and hence the microcanonical distribution, which we know how to use in order to draw physical consequences.

It is therefore legitimate to ask whether the chaotic hypothesis and the ensuing SRB distribution have any predictive value of their own. Just as the ergodic hypothesis implies the well tested classical thermodynamics, the new hypothesis should imply, for example, irreversible thermodynamics of nonequilibrium stationary states, without the necessity of solving the equations of motion. It is not clear that this is so.

However there are already some experimental results that offer support to the chaotic hypothesis, since one can understand their outcome by using it.

Here we examine, in particular, one experimental result, [ECM2], which the authors already attempted to explain by relating it to our chaotic hypothesis. Some of the data in [ECM2] require, to be unambigously understood, the discussion in [ES]. We shall take the viewpoint of the preceding sections to make more precise, and detailed, the argument in [ECM2], modifying it to some extent in order to put it on a more mathematical basis.

[15] One should not be disturbed by the fact that this is a rigorous mathematical theorem only for Anosov systems or, for somewhat more general systems *e.g.* "axiom A" attractors, [R1]: one should not forget that the microcanonical ensemble is also lacking a mathematical justification in equilibrium theory. In fact the only equilibrium case in which one can prove the ergodic hypothesis is for the hard sphere gas, [S1], [BSC], [SiS], a major piece of work. In the case of only one hard sphere moving in a lattice of obstacles (model 1 with $N=1$, $\varphi=0$ and φ^u a suitable hard core potential, *i.e.* a triangular lattice of hard disks) the present point of view can also be shown to hold in the presence of dissipation, [CELS].

In the context of this paper the experiment [ECM2] deals with model 2 in §3 and measures the entropy production rate (*i.e.* phase space contraction rate, see §3, (3.5)) as seen on a stretch of time τ, short compared to the duration of the experiment T, and repeating the measurement T/τ times.[16] We emphasize that this is an experiment on a system *far* from equilibrium. Calling $D\sigma_\tau(x)$ the entropy production rate measured on the motion originating at $S^{-\tau/2}x$ and observed τ units of time (we take τ even for simplicity), we define it, see (3.4),(3.5), by:

$$D\sigma_\tau(x) = D\,\frac{1}{\tau}\sum_{j=-\tau/2}^{\tau/2-1}\sigma(S^j x) \stackrel{def}{=} D\,\langle\sigma\rangle_+\,a_\tau(x) \tag{7.1}$$

where $\langle\sigma\rangle_+$ is the average in the future of $\sigma(S^j x)$, which is a constant almost everywhere in phase space with respect to μ_0–random choices of initial data.[17]

The total entropy production while the phase space point x evolves between $S^{-\tau/2}x$ and $S^{\tau/2}x$ is obtained by multiplying (7.1) with the time elapsed during τ such timing collisions. For simplicity we think that the *time interval* t_0 between the timing collisions is constant. Note that x is the middle point of the segment of a trajectory of (discrete) time length τ, defining the fluctuation $a_\tau(x)$ in (7.1).

It is perhaps important to stress that $\langle\sigma\rangle_+$ is very different from the limit as $\tau\to+\infty$ of $\sigma_\tau(x)$ in (7.1): in fact the latter (by the time reversal symmetry) vanishes, while the former is positive as follows from numerical evidence, see §3, and as assumed in (A) in §2.

The experiment divided the a_τ–axis into small intervals $I_0, I_{\pm1}, \ldots$ and measured the quantity $a_\tau(x)$ building a histogram counting how many times the a_τ fell into the interval I_p (where $a_\tau(x) = p$). Obviously we expect a distribution $\pi_\tau(p)$ centered around a (forward) average which is, (see (7.1)), exactly 1. The result for $\pi_\tau(p)$ can be found in fig. 1 of [ECM2], for one (rather large) value of τ and one of γ and $N = 56$.

A second experimental result is for $\Pi_\tau(p) = -\frac{1}{2N\tau t_0\langle\sigma\rangle_+}\log\frac{\pi_\tau(p)}{\pi_\tau(-p)}$, *i.e.* essentially for the logarithm of the ratio of the probability that $a_\tau(x) = p$ to that of $a_\tau(x) = -p$.

---

16 The reader should not mind that the symbol for the integer T is sometimes also used for the absolute temperature.

17 This because the SRB distribution verifies the extended zeroth law, (2.1), which says that the averages are, with μ_0–probability 1, independent of the initial data.

The result, fig. 2 of [ECM2] is, for the rather large value of τ considered, a remarkably precise straight line for $\Pi_\tau(p)$ as a function of p, *i.e.* $\Pi_\tau(p)$ is a linear function of p.

A third experiment shows that the slope of this line as a function of τ verifies, even for large deviations, the relation of proportionality to τ for τ large (fig. 3 of [ECM2]).

The results are rather precise with apparently little margin for errors, hence one has to find a theoretical reason that the probability distribution of $D\,\sigma_\tau(x)$ has the form:

$$\begin{aligned} \pi_\tau(p)\,dp \stackrel{def}{=} & P(a_\tau \in (p, p+dp)) = e^{-\tau\zeta(p)+\tau Cp} dp \qquad \text{or} \\ \frac{\pi_\tau(p)}{\pi_\tau(-p)} = & e^{2\tau Cp} \end{aligned} \tag{7.2}$$

for a suitably chosen constant C and a suitable even function $\zeta(p)$ with minimum at $p = 1$ and with the argument of the exponential correct up to, apparently, p, τ independent corrections (see fig. 3 of [ECM2]). In [ECM2] a theoretical argument is presented which leads to $2C = Dt_0\langle\sigma\rangle_+$, if t_0 is the average time between timing events.

We are now going to show, and this is our main technical result (and a theorem under assumptions (A,B,C) of §2), what we call a *fluctuation theorem*:

Fluctuation theorem: Let $(\mathcal{C}, S)$ verify the properties (A,B,C) of §2, (dissipativity, reversibility and chaoticity). Then the probability $\pi_\tau(p)$ that the total entropy production $D\tau t_0\sigma_\tau(x)$, (7.1), over a time interval $t = \tau t_0$ (with t_0 equal to the average time between timing events) has a value $Dt\langle\sigma\rangle_+ p$ verifies the large deviation relation:

$$\frac{\pi_\tau(p)}{\pi_\tau(-p)} = e^{Dt\langle\sigma\rangle_+ p} \tag{7.3}$$

with an error in the argument of the exponential which can be estimated to be $p,\tau-$independent.

This means that if one plots the logarithm of the left hand side of (7.3) as a function of p one observes a straight line with more and more precision as τ becomes large (in agreement with figure 3 in [ECM2]).

Remark: since the above theorem is deduced under the assumptions (A,B,C) only, the result (7.2) will apply as well to the models 1,3,4,5. This gives a parameterless prediction of the outcome of several numerical experiments similar to the one described above.

The main ideas for the proof, [GC], [Ga3], [Ga2], of the above theorem are the following.

The probability that $a_\tau(x) \in I_p$ over the probability that $a_\tau(x) \in I_{-p}$ is, if one uses the notations and the approximation $\mu_{T,\tau}$ to μ described at the end of §4 (see (6.3)) with $F(x) = a_\tau(x)$:

$$\frac{\sum_{j,\, a_\tau(x_j)=p} \overline{\Lambda}_{u,\tau}^{-1}(x_j)}{\sum_{j,\, a_\tau(x_j)=-p} \overline{\Lambda}_{u,\tau}^{-1}(x_j)} \tag{7.4}$$

where $\overline{\Lambda}_{u,\tau}(x)$ is the absolute value of the jacobian determinant of S^τ as a map of W^u_O into itself, evaluated at the point $S^{-\tau/2}x \in S^T\Delta$, (*i.e.* as a map between $S^{-\tau/2}x$ and $S^{\tau/2}x$, see (6.2), (7.1)).

Since $\mu_{T,\tau}$ in (6.3) is only an approximation at fixed T, τ *an error is involved in using (7.4).* It can be shown that this error can be estimated to affect the result only by a factor bounded above and below uniformly in τ, p, [GC], [Ga3], [Ga2]. This is a remark technically based on the thermodynamic analogy pointed out in (5.4), (5.5).

We now try to establish a one to one correspondence between the addends in the numerator of (7.4) 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 (7.4).

This is possible because of the reversibility property (B), §2. Let $x \in A$, then $ix \in iA$. By using the identity $S^{-\tau}(S^\tau x) = x$, the identity $S^{-\tau}(iS^{-\tau}x) = ix$ (time reversal) and (6.1), we deduce the relations:[18]

$$a_\tau(x) = -a_\tau(ix), \qquad \overline{\Lambda}_{u,\tau}(ix) = \overline{\Lambda}_{s,\tau}^{-1}(x) \tag{7.5}$$

which are identities, see [ECM2],[GC],[Ga3]. The first equality in (7.5) is obvious as in all the cases considered the i operation changes the sign to $\sigma(x)$, the rate of change of the phase space volume. The second equality in (7.5) is also easy to check: in fact let β be a surface element on W^u_x around $S^{-\tau/2}x$ and let $\beta' = S^\tau\beta$ be its S^τ image around $S^{\tau/2}x$: then $\overline{\Lambda}_{u,\tau}(x) = \frac{|\beta'|}{|\beta|}$. Applying i to β and β' one obtains surface elements $i\beta$ and $i\beta'$ on W^s_{ix} with the same area as β and β' (because i is an isometry) around respectively $S^{\tau/2}ix$ and $S^{-\tau/2}ix$ so that the expansion rate $\overline{\Lambda}_{s,\tau}(ix)$ is $\frac{|i\beta|}{|i\beta'|} = \frac{|\beta|}{|\beta'|} = \overline{\Lambda}_{u,\tau}^{-1}(x)$.

[18] The key remark is that time reversal i maps E_j into iE_j and at the same time *changes* the horizontal surface elements of E_j into the vertical ones of iE_j and the vertical surface elements of E_j into the horizontal of iE_j, see (6.1).

The ratio (7.4) can therefore be written simply as:

$$\frac{\sum_{E_j, a_\tau(x_j)=p} \overline{\Lambda}_{u,\tau}^{-1}(x_j)}{\sum_{E_j, a_\tau(x_j)=-p} \overline{\Lambda}_{u,\tau}^{-1}(x_j)} \equiv \frac{\sum_{E_j, a_\tau(x_j)=p} \overline{\Lambda}_{u,\tau}^{-1}(x_j)}{\sum_{E_j, a_\tau(x_j)=p} \overline{\Lambda}_{s,\tau}(x_j)} \tag{7.6}$$

where $x_j \in E_j$ is the center in E_j. In deducing the second relation we make us of the existence of the time reversal symmetry i and of (7.5), and assume that the centers $x_j, x_{j'}$ of E_j and $E_{j'} = iE_j$ are chosen so that $x_{j'} = ix_j$.

It follows then that the ratios between corresponding terms in the ratio (7.6) is equal to $\overline{\Lambda}_{u,\tau}^{-1}(x)\overline{\Lambda}_{s,\tau}^{-1}(x)$. This differs from the reciprocal of the total change of phase space volume over the τ time steps between the point $S^{-\tau/2}x$ and $S^{\tau/2}x$ only because it does not take into account the ratio of the sines of the angles $\vartheta(S^{-\tau/2}x)$ and $\vartheta(S^{\tau/2}x)$ formed by the stable and unstable manifolds at the points $S^{-\tau/2}x$ and $S^{\tau/2}x$, see footnote [1]. But $\overline{\Lambda}_{u,\tau}^{-1}(x)\overline{\Lambda}_{s,\tau}^{-1}(x)$ will differ from the actual phase space contraction under the action of S^τ, 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 $B = \max_{x,x'} \frac{|\sin\vartheta(x)|}{|\sin\vartheta(x')|}$ which is finite by the transversality of the stable and unstable manifolds.

Now for all the points x_j in (7.6), the reciprocal of the total phase space volume change over a time τt_0 is $e^{a_\tau(x_j)\langle\sigma\rangle + t_0\tau D}$, which (by the constraint imposed on the summation labels $a_\tau = p$) equals $e^{Dt_0\tau\langle\sigma\rangle_+ p}$. Hence the ratio (7.4) will be $e^{Dt_0\tau\langle\sigma\rangle_+ p}$, to leading order as $D, \tau \to \infty$, proving (7.3), with $2C = D\langle\sigma\rangle_+ t_0$. It is important to note that there are two errors ignored here, as pointed out in the discussion above.[19] They imply that the argument of the exponential *is correct up to p, τ independent corrections* (which are in fact observed in the experiment as fig.3 of [ECM2] shows). One should note that other errors may arise because of the approximate validity of our main chaotic assumption (which states that things go "as if" the system was Anosov): they may depend on D and we do not control them except for the fact that, if present, their relative value should tend to 0 as $D \to \infty$: there may be (and very likely there are) cases in which the chaotic hypotesis is not reasonable for small N (*e.g.* systmes like the Fermi-Pasta-Ulam chains) but it might be correct for large N. We also mention that for some systems with small D for which the chaotic hypothesis may be already regarded as valid (*e.g.* model 1 with $N = 1$ in [CELS]). In such cases care must be taken in not confusing D (in model 1 it is $D = 2N - 1$, in model 2 it

[19] In the previous paragraph and in the paragraph following (7.4).

is $D = 2N - 2$ and so on) with half the dimension of the full phase space $\mathcal{F}$, as the latter might be so much larger (if N is small) that the difference in dimension can actually be observed in numerical experiments.

The p independence of the coefficient of C in (7.2) is therefore a key test of the theory (and it should hold with corrections of order $O(\tau^{-1})$).

§*8 Outlook.*

We end with a number of remarks.

1) The interest of our discussion in §7 is not, of course, the fluctuation theorem which is essentially proved there (for a formal proof see [Ga2] and [Ga3]), but in the clarification of the relevance of the properties (A,B,C) mentioned in §2. Furthermore, it is interesting that our chaotic hypothesis, §2 does have some concrete and experimentally verifiable consequences (verified here in the case of model 2): such consequences have the remarkable feature of being predictions *without free parameters*, hinting that the hypothesis might have a quite general validity. One cannot be too demanding on the matter of mathematical rigor: one should not forget that the ergodic hypothesis is far from being proved either, particularly in the generality one would want.

2) The fluctuation formula (7.3) holds also for the models 1,3,4,5 because the fluctuation theorem applies to all such models (see remark following the theorem): but the numerical experiments do not seem to exist yet.

3) The pairing property (D) and the smooth distribution of the Lyapunov exponents (E) have been used here only to get some intuition and to visualize the hyperbolic nature of the attractor and the equality of the dimensions of the stable and unstable manifolds. It seems interesting to perform numerical experiments to try to investigate better, at least in the systems that we are considering, if the density function $f_\infty(x)$ is really positive at $x = 0$, see (E), §2, as the numerical results seem to suggest in some cases, [LPR], [ECM1], [SEM]. The recent work [S3] provides the first rigorous results towards establishing (E) and a discussion of the theoretical aspects of property (E).

4) Note that the fluctuation theorem ((7.3)) applied to model 5 leads to an interesting consequence on the large deviations properties of the magnitude of the energy dissipation ε in turbulent flows. In this case we take, for simplicity, the kinetic energy $D_v kT(x)$ of the viscous modes to be a non fluctuating quantity equal to $D_v kT$. Then

the random variable p associated with $\sigma(x)=\frac{\varepsilon(x)}{D_v kT(x)}$ in the fluctuation theorem of §7 is just proportional to the average over a time interval t of the energy dissipation rate ε. This is a variable that is assumed to be constant in the Kolmogorov–Obuchov theory: what we say here is that it is in fact a fluctuating quantity and we predict (on the basis of the fluctuation theorem of §7) that the time average $\langle\varepsilon\rangle_t$, over a time interval t, of $\varepsilon(x)$,*i.e.* $\langle\varepsilon\rangle_t\equiv tD_v\langle\varepsilon\rangle_+ p$ is such that its probability distribution $\pi_t(p)$ verifies the *linear large deviation law*:

$$\frac{1}{D_v tp\langle\varepsilon\rangle_+/kT}\log\left[\pi_t(p)/\pi_t(-p)\right]=1 \tag{8.1}$$

up to corrections of order $O(t^{-1})$. If $T=T(x)$ has to be regarded as a fluctuating variable then (8.1) must be regarded as a property of the fluctuations of the entropy production rate $\sigma(x)=\frac{\varepsilon(x)}{kT(x)}$ rather than of the energy dissipation, (with some obvious modifications, *e.g.* $\langle\varepsilon\rangle_+/kT\to\langle\varepsilon/kT\rangle_+$).

5) Concerning the particularity of the gaussian thermostat we think, see [Ga1],[C], that there should be, also in nonequilibrium, several physically equivalent ways of describing the same stationary distribution corresponding to different μ and to different physical ways of reaching the stationary state, at least in the thermodynamic limit. Thus it may well be that the gaussian thermostat turns out to be equivalent to other models of thermostats, which could be described by rather different attractors. For instance a stochastic thermostat, in which a particle colliding with the wall is scattered with a maxwellian distribution at a given temperature, will certainly be described by a statistics μ which is absolutely continuous with respect to the Liouville distribution.[20] In the thermodynamic limit this might just give the same result as obtained with a statistics which, for finite N, is on a fractal attractor. This mechanism is like the one realized by the microcanonical and the canonical ensembles (the first is concentrated on a set of configurations which has zero probability with respect to the second, as long as $N<\infty$). This is clearly a question that requires further investigations.

6) One can also regard the gaussian thermostat as a device to eliminate some trivial Lyapunov exponents. For instance in model 4 we could simply *not* introduce the

[20] Note that a stochastic model of thermostat is described by a stochastic differential equation and therefore our discussion does not apply without some major modification.

gaussian constraint that the *total* energy is constant (which led to the $\alpha_0 \underline{p}_j$ terms): we believe that physically the system would then still behave in the same way, *for large* N, L. But without such a constraint we could not assume the phase space to have $4N-3$ dimensions, because H would not be rigorously constant. We expect, however, that in such case the energy H is *approximately* constant, in fact more and more so as $L, N \to \infty$ with $NL^{-2} = n$ constant. Thus the variability of H probably leads to a zero Lyapunov exponent making the chaoticity assumption manifestly invalid. But this would be so only in a somewhat trivial way as its violation is due to a zero Lyapunov exponent associated with a variable which is "almost" a constant of the motion. Hence it is natural to fix the value of the energy rigorously *a priori* by a constraint (realized via a minimum constraint principle, like Gauss' principle) and so dispose of the extra 0 (or very close to 0) Lyapunov exponent, recovering then again a situation in which the system is strictly chaotic. Such a point of view can be extended to cover cases in which hyperbolicity is not valid because of the existence of quasi exact conservation rules. An example is in fact model 2 in which the variables $\tilde{y}_j, \tilde{p}_{yj}$ can be replaced by y_j, p_{yj} thus turning P_x, X_x into *variable* quantities: their variability is however clearly due to the special (vertical) boundary conditons used and it should therefore not matter whether they are kept rigorously constant or not, in the limit of $N, L \to \infty$. The dynamics can be modified by turning such quantities into exact conservation laws and the new dynamics should be indistinguishable from the previous one in the thermodynamic limit. Another example is provided by the constraints imposed on model 3 to achieve that the horizontal momentum is conserved.

7) Like for the ergodic hypothesis in equilibrium, the range of validity of the chaotic hypothesis for nonequilibrium stationary states is not known; the more complicated the nature of the latter states, maintained in the presence of external fields or special boundary conditions, makes this case even more difficult.

Acknowledgements: We are indebted to J.L. Lebowitz and G.L. Eyink for clarifying and patient explanations on their papers [CELS] and [CL], even before publication and to the latter for informing us about the reference [ZJ]. We are also indebted to them as well as to Y. Sinai and especially to N. Chernov for very helpful comments. G.G. is indebted to Rutgers University and to Rockefeller university for partial support and to CNR-GNFM for travel support. E.G.D.C. acknowledges financial support under contract DE-FG02-88-ER13847 of the US Department of Energy.

References.

[AA] Arnold, V., Avez, A.: *Ergodic problems of classical mechanics*, Benjamin, 1966.

[Bo] Bowen, R.: *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture notes in mathematics, vol. **470**, Springer Verlag, 1975.

[BEC] Baranyai, A., Evans, D.T., Cohen, E.G.D.: *Field dependent conductivity and diffusion in a two dimensional Lorentz gas*, Journal of Statistical Physics, **70**, 1085–1098, 1993.

[BSC] Bunimovitch, L., Sinai, Y.G., Chernov, N.I.: *Markov partitions for two-dimensional hyperbolic billiards*, Russian Mathematical Surveys, **45**, n.3, 105-152, 1990. And: *Statistical properties of two dimensional hyperbolic billiards*, Russian Mathematical Surveys, **46**, n.4, 47–106, 1991.

[C] Cohen, E.G.D.: *Boltzmann and statistical mechanics*, Lecture delivered at the "Interbational Conference on Boltzmann and its legacy 150 years after his birth", Accademia Nazionale dei Lincei, Roma 1994, May 21–28, in print.

[CELS] Chernov, N.I., Eyink, G.L., Lebowitz, J.L., Sinai, Y.G.: *Steady state electric conductivity in the periodic Lorentz gas*, Communications in Mathematical Physics, **154**, 569–601, 1993.

[CL] Chernov, N.T., Lebowitz, J.L.: *Stationary shear flow in boundary driven hamiltonian systems*, Rutgers University preprint, may 1994.

[Dr] Dressler, U.: *Symmetry property of the Lyapunov exponents of a class of dissipative dynamical systems with viscous damping*, Physical Review, **38A**, 2103–2109, 1988.

[ECM1] Evans, D.J.,Cohen, E.G.D., Morriss, G.P.: *Viscosity of a simple fluid from its maximal Lyapunov exponents*, Physical Review, **42A**, 5990–5997, 1990.

[ECM2] Evans, D.J.,Cohen, E.G.D., Morriss, G.P.: *Probability of second law violations in shearing steady flows*, Physical Review Letters, **71**, 2401–2404, 1993.

[ER] Eckmann, J.P., Ruelle, D.: *Ergodic theory of strange attractors*, Reviews of Modern Physics, **57**, 617–656, 1985.

[ES] Evans, D.J., Searles, D.J.: *Equilibrium microstates which generate second law violating steady states*, Research school of chemistry, Canberra, ACT, 0200, preprint, 1993.

[Ga1] Gallavotti, G.: *Ergodicity, ensembles, irreversibility in Boltzmann and beyond*, Journal of Statistical Physics. **78**, 1571–1589, 1995.

[Ga2] Gallavotti, G.: *Reversible Anosov diffeomorphisms and large deviations.*, in *mp_arc@ math. utexas. edu*, #95-19.

[Ga3] Gallavotti, G.: *Topics in chaotic dynamics*, Lectures at the Granada school, ed. Garrido–Marro, Lecture Notes in Physics, **448**, 1995.

[Ga4] Gallavotti, G.: *Coarse graining and chaos*, in preparation.

[G] Gibbs, J.W.: *Elementary principles in statistical mechanics*, Ox Bow Press, 1981, (reprint).

[GC] Gallavotti, G., Cohen, E.G.D.: *Dynamical ensembles in nonequilibrium statistical mechanics*, Physical Review Letters, **74**, 2694–2697, 1995.

[HHP] Holian, B.L., Hoover, W.G., Posch. H.A.: *Resolution of Loschmidt's paradox: the origin of irreversible behavior in reversible atomistic dynamics*, Physical Review Letters, **59**, 10–13, 1987.

[LL] Landau, L.D., Lifshitz, E.M.: *Fluid mechanics*, Pergamon press, Paris, 1959.

[Lo] Lorenz, E.: *Deterministic non periodic flow*, J. of the Athmospheric Sciences, **20**, 130- 141, 1963.

[LPR] Livi, R., Politi, A., Ruffo, S.: *Distribution of characteristic exponents in the thermodynamic limit*, Journal of Physics, **19A**, 2033–2040, 1986.

[Ma] Marchioro, C.: *An example of absence of turbulence for any Reynold number*, Communications in Mathematical Physics, **105**, 99–105, 1986.

[P] Pesin, Y.: *Dynamical systems with generalized hyperbolic attractors: hyperbolic, ergodic and topological properties*, Ergodic Theory and Dynamical Systems, **12**, p. 123–151, 1992.

[PH] Posch, H.A., Hoover, W.G.: *Non equilibrium molecular dynamics of a classical fluid*, in "Molecular Liquids: New Perspectives in Physics and Chemistry", ed. J. Teixeira-Dias, Kluwer Academic Publishers, p. 527–547, 1992.

[R1] Ruelle, D.: *Chaotic motions and strange attractors*, Lezioni Lincee, notes by S. Isola, Accademia Nazionale dei Lincei, Cambridge University Press, 1989; see also: Ruelle, D.: *Measures describing a turbulent flow*, Annals of the New York Academy of Sciences, **357**, 1–9, 1980. For more technical expositions see Ruelle, D.: *Ergodic theory of differentiable dynamical systems*, Publications Mathémathiques de l' IHES, **50**, 275–306, 1980.

[R2] Ruelle, D.: *A measure associated with axiom A attractors*, American Journal of Mathematics, **98**, 619–654, 1976.

[S1] Sinai, Y.G.: *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*, Russian Mathematical Surveys, **25**, 137–189, 1970.

[S2] Sinai, Y.G.: *Gibbs measures in ergodic theory*, Russian Mathematical Surveys, **27**, 21–69, 1972. Also: *Lectures in ergodic theory*, Lecture notes in Mathematics, Princeton U. Press, Princeton, 1977. And: *Markov partitions and C-diffeomorphisms*, Functional Analysis and Applications, **2**, 64–89, 1968, n.1; and: *Construction of Markov partitions*, Functional analysis and Applications, **2**, 70–80, 1968, n.2.

[S3] Sinai, Y.G.: *A remark concerning the thermodynamical limit of Lyapunov spectrum*, Princeton University preprint, march 1995.

[SiS] Simányi, N. Szász, D.: *The Boltzmann-Sinai Ergodic Hypothesis for Hard Ball Systems*, preprint, archived in *mp_arc@ math. utexas. edu*, #95–133.

[SEM] Sarman, S., Evans, D.J., Morriss, G.P.: *Conjugate pairing rule and thermal transport coefficients*, Physical Review, **45A**, 2233–2242, 1992.

[UF] Uhlenbeck, G.E., Ford, G.W.: *Lectures in Statistical Mechanics*, American Mathematical society, Providence, R.I., pp. 5,16,30, 1963.

[ZJ] Zhen-Su She, Jackson, E.: *Constrained Euler system for Navier–Stokes turbulence*, Physical Review Letters, **70**, 1255–1258, 1993.