Eur. Phys. J. B **61**, 1–24 (2008)
DOI: 10.1140/epjb/e2008-00041-1

THE EUROPEAN PHYSICAL JOURNAL B

Heat and fluctuations from order to chaos*

G. Gallavotti[a]

Dipartimento di Fisica and INFN, Università di Roma, La Sapienza, P.A. Moro 2, 00185 Roma, Italy

Received 19 November 2007
Published online 31 January 2008 –

Abstract. The Heat theorem reveals the second law of equilibrium Thermodynamics (i.e. existence of Entropy) as a manifestation of a general property of Hamiltonian Mechanics and of the Ergodic Hypothesis, valid for 1 as well as 10^{23} degrees of freedom systems, i.e. for simple as well as very complex systems, and reflecting the Hamiltonian nature of the microscopic motion. In Nonequilibrium Thermodynamics theorems of comparable generality do not seem to be available. Yet it is possible to find general, model independent, properties valid even for simple chaotic systems (i.e. the hyperbolic ones), which acquire special interest for large systems: the Chaotic Hypothesis leads to the Fluctuation Theorem which provides general properties of certain very large fluctuations and reflects the time-reversal symmetry. Implications on Fluids and Quantum systems are briefly hinted. The physical meaning of the Chaotic Hypothesis, of SRB distributions and of the Fluctuation Theorem is discussed in the context of their interpretation and relevance in terms of Coarse Grained Partitions of phase space. This review is written taking some care that each section and appendix is readable either independently of the rest or with only few cross references.

PACS. 05.20.-y Classical statistical mechanics – 05.70.Ln Nonequilibrium and irreversible thermodynamics – 05.40.-a Fluctuation phenomena, random processes, noise, and Brownian motion

1 The heat theorem

An important contribution of Boltzmann to Physics as well as to research methods in Physics has been the *Heat theorem*.

Summarizing here an intellectual development, spanning about twenty years of work, the *heat theorem* for systems of particles of positions $\mathbf{q}$ and momenta $\mathbf{p}$, whose dynamics is modeled by a Hamiltonian of the form $H = K(\mathbf{p}) + W(\mathbf{q})$, $K = \frac{1}{2m}\mathbf{p}^2$, can be formulated as follows

Heat theorem. *In a isolated mechanical system, time averages $\langle F\rangle$ of the observables, i.e. of functions F on phase space, are computable as their integrals with respect to probability distributions μ_α which depend on the control parameters α determining the states. It is possible to find four observables, whose averages can be called U, V, T, p, depending on α, so that an infinitesimal change $d\alpha$ implies variations dU, dV of U, V so related that*

$$\frac{dU + p\,dV}{T} = \text{“exact”} \stackrel{def}{=} dS \tag{1.1}$$

where $p = \langle -\partial_V W\rangle$ and V is a(ny) parameter on which W depends, and U, T are the average total energy and the average total kinetic energy.

When the system is large and V is the volume available to the particles the quantity p can be shown to have the interpretation of physical “pressure” on the walls of the available volume.

Remarks. (a) Identification of T with the average kinetic energy had been for Boltzmann a starting point, assumed a priori, from the works of Krönig and Clausius of a few years earlier (all apparently unaware, as everybody else, of the works of Bernoulli et al. [1]).

(b) Connection with observations is made by identifying curves in parameter space, $t \to \alpha(t)$, with *reversible processes*. And in an infinitesimal process, defined by a line element $d\alpha$, the quantity pdV is identified with the work the system performs, dU with the energy variation and $dQ = dU + pdV$ as the heat absorbed. Then relation equation (1.1) implies that Carnot machines have the highest efficiency. The latter is one of the forms of the second law, which leads to the existence of entropy as a function of state in macroscopic Thermodynamics [2].

(c) Equation (1.1), combined with the (independent) assumption that heat extracted at a fixed temperature cannot be fully transformed into work, implies that in any process $\frac{dQ}{T} \le dS$. Hence in isolated systems changing equilibrium state cannot make entropy decrease, or in colorful language the *entropy of the Universe cannot decrease* [3], p. I-44-12. Actually by suitably defining what is meant by irreversible process it is possible to reach the conclusion that, unless the change of equilibrium state is achieved

* Review.

[a] e-mail: giovanni.gallavotti@roma1.infn.it

via a reversible process, the entropy of an isolated system does increase strictly [2]. Conceptually, however, this is an addition to the second law [3], p. I-44-13.

Examples of control parameters are simply U, V, or T, V, or p, V. The theorem holds under some hypotheses which evolved from

(a) all motions are periodic (1866);
(b) aperiodic motions can be considered periodic with infinite period (!) [4];
(c) motion visits all phase space of given total energy: in modern terminology this is the *ergodic hypothesis* (1868–1884) [5].

The guiding idea was that equation (1.1) would be true for all systems described by a Hamiltonian $H = K + W$: *no matter whether having few or many degrees of freedom*, as long as the ergodic hypothesis could be supposed true.

In other words equation (1.1) should be considered as a consequence of the Hamiltonian nature of motions: it is true for all systems whether with one degree of freedom (as in the 1866 paper by Boltzmann) or with 10^{19} degrees of freedom (as in the 1884 paper by Boltzmann).

It is, in a sense, a property of the particular Hamiltonian structure of Newton's equations (Hamiltonian given as sum of kinetc plus potential energy with kinetic energy equal to $\sum_i \frac{1}{2}\mathbf{p}_i^2$ and potential energy purely positional). True for all (ergodic) systems: trivial for 1 degree of freedom, a surprising curiosity for few degrees and an important law of Nature for 10^{19} degrees of freedom (as in 1 cm^3 of H_2).

The aspect of Boltzmann's approach that will be retained here is that some universal laws merely reflect basic properties of the equations of motion which may have deep consequences in large systems: the roots of the second Law can be found [4], in the simple properties of the pendulum motion.

Realizing the mechanical meaning of the second law induced the birth of the theory of ensembles, developed by Boltzmann between 1871 (as recognized by Gibbs in the introduction to his treatise) and 1884, hence of Statistical Mechanics.

Another example of the kind are the reciprocal relations of Onsager, which reflect time reversal symmetry of the Hamiltonian systems considered above. Reciprocity relations are a first step towards understanding non equilibrium properties. They impose strong constraints on transport coefficients, i.e. on the $\mathbf{E}$-derivatives of various average currents induced by external forces of intensities $\mathbf{E} = (E_1, \ldots, E_n)$, which disturb the system from an equilibrium state into a new *stationary state*. The derivation leads to the quantitative form of reciprocity which is expressed by the "Fluctuation-Dissipation Theorems", i.e. by the Green-Kubo formulae, expressing the transport coefficient of a current in terms of the mean square fluctuations of its long time averages.

In the above Boltzmann's papers (as well as in several other of his works) Thermodynamics is derived on the assumption that motions are periodic, hence very regular: see the above mentioned ergodic hypothesis. Nevertheless heat is commonly regarded as associated with the chaotic motions of molecules and thermal phenomena are associated with fluctuations due to chaotic motions at molecular level. A theme that is pursued in this paper it to investigate how to reconcile opposites like order and chaos within a unified approach so general to cover not only equilibrium Statistical Mechanics, but many aspects of nonequilibrium stationary states. An overview is in the first thiteen sections, while the appendices enter into technical details, still keeping at a heuristic level in discusing a matter that is often given little conideration by Physicists because of its widespread reputation of being just abstract Mathematics: hopefully this will help to divulge a theory which is not only simple conceptually nut it seems promising of further developments.

The above comment is meant also to explain the meaning of the title of this paper.

2 Time reversal symmetry

In a way transport coefficients are still equilibrium properties and nothing is implied by reciprocity when $\mathbf{E}$ is strictly $\neq \mathbf{0}$.

It is certainly interesting to investigate whether time reversal has important implications in systems which are really out of equilibrium, i.e. subject to non conservative forces which generate currents (transporting mass, or charge, or heat or several of such quantities).

There have been many attempts in this direction: it is important to quote the reference [6] which summarizes a series of works by a Russian school and completes them. In this paper an extension of the Fluctuation-Dissipation theorem, as a reflection of time reversal, is presented, deriving relations which, after having been further developed, have become known as "work theorems" and/or "transient fluctuation theorems" for transformations of systems out of equilibrium [7–12].

For definiteness it is worth recalling that a dynamical system with equations $\dot{x} = f(x)$ in phase space, whose motions will be given by maps $t \to S_t x$, is called "reversible" if there is a smooth (i.e. continuously differentiable) isometry I of phase space, anticommuting with S_t and involutory, i.e.

$$IS_t = S_{-t}I, \qquad I^2 = 1. \tag{2.1}$$

Usually, if $x = (\mathbf{p}, \mathbf{q})$, time reversal is simply $I(\mathbf{p}, \mathbf{q}) = (-\mathbf{p}, \mathbf{q})$.

The main difficulty in studying nonequilibrium statistical Mechanics is that, after realizing that one should first understand the properties of stationary states, considered as natural extensions of the equilibrium states, it becomes clear that the microscopic description *cannot be Hamiltonian*.

This is because a current arising from the action of a nonconservative force continuously generates "heat" in the system. Heat has to be taken out to allow reaching a steady state. This is empirically done by putting the system in contact with one or more thermostats. In models, thermostats are just forces which act performing work balancing, at least in average, that produced by the external forces, i.e. they "model heat extraction".

It is not obvious how to model a thermostat; and any thermostat model is bound to be considered "unphysical" in some respects. This is not surprising, but it is expected that most models introduced to describe a given physical phenomenon should be "equivalent".

Sometimes it is claimed that the only physically meaningful thermostats for nonequilibrium systems (in stationary states) are made by infinite (3-dimensional) systems which, asymptotically at infinity, are in statistical equilibrium. In the latter cases it is not even necessary to introduce *ad hoc* forces to remove the heat: *motion remains Hamiltonian* and heat flows towards infinity.

Although the latter is certainly a good and interesting model, as underlined already in [13], it should be stressed that it is mathematically intractable unless the infinite systems are "free". i.e. without internal interaction other than linear [13–17].

And one can hardly consider such assumption more physical than the one of finite thermostats. Furthermore it is not really clear whether a linear external dynamics can be faithful to Physics, as shown by the simple one dimensional XY-models, see [18] where a linear thermostat dynamics with a single temperature leads a system to a stationary state, as expected, but the state is not a Gibbs state (at any temperature). The method followed in [18], based on [19], can be used to illustrate some problems which can arise when thermostats are classical free systems, see Appendix D.

3 Point of view

The restriction to finite thermostats, followed here, is not chosen because infinite thermostats should be considered unphysical, but rather because it is a fact that the recent progress in nonequilibrium theory can be traced to

(a) the realization of the interest of restricting attention to *stationary states*, or *steady states*, reached under forcing (rather than discussing approach to equilibrium, or to stationarity);
(b) the simulations on steady states performed in the 80's after the essential role played by *finite thermostats* was fully realized.

Therefore investigating finite thermostat models is still particularly important. This makes in my view interesting to confine attention on them and to review their conceptual role in the developments that took place in the last thirty years or so.

Finite thermostats can be modeled in several ways: but in constructing models it is desirable that the models keep as many features as possible of the dynamics of the infinite thermostats. As realized in [6], p. 452 it is certainly important to maintain the *time reversibility*. Time reversibility expressed by equation (2.1), i.e. existence of a smooth conjugation between past and future, is a fundamental symmetry of nature which (replaced by TCP) even "survives" the so called time reversal violation; hence it is desirable that it is saved in models. An example will be discussed later.

Comment. (1) The second law of equilibrium Thermodynamics, stating existence of the state function entropy, can be derived without reference to the microscopic dynamics by assuming that heat absorbed at a single temperature cannot be cyclically converted into work [2]. In statistical Mechanics equilibrium, states are identified with probability distributions on phase space: they depend on control parameters (usually two, for instance energy and volume) and processes are identified with sequences of equilibrium states, i.e. as curves in the parameters space interpreted as *reversible processes*. The problem of how the situation, in which averages are represented by a probability distribution, develops starting from an initial configuration *is not part of the equilibrium theory*. In this context the second law arises as a theorem in Mechanics (subject to asssumptions) and, again, just says that entropy exists (the heat theorem).

(2) As noted in Section 1, if the scope of the theory is enlarged admitting processes that cannot be represented as sequences of equilibria, called "irreversible processes", then the postulate of impossibility to convert heat into work extracting it from a single thermostat implies, again without involving microscopic dynamics, the inequality often stated as "the entropy of the Universe" cannot decrease in passing from an equilibrium state to another. And, after properly defining what is meant by irreversible process [2], actually strictly increases if in the transformation an irreversible process is involved; however perhaps it is best to acknowledge explicitly that such a strict increase is a further assumption [3], p. I-44-13 leaving aside a lengthy [2], and possibly not exhaustive analysis of how in detail an irreversible transformation looks like. Also this second statement, under suitable assumptions, can become a theorem in Mechanics [20,21], but here this will not be discussed.

(3) Therefore studying macroscopic properties for systems out of equilibrium can be divided into an "easier" problem, which is the proper generalization of equilibrium statistical Mechanics: namely studying stationary states identified with corresponding probability distributions yielding, by integration, the average values of the few observables of relevance. And the problem of approach to a stationary state which is of course more difficult. The recent progress in nonequilibrium has been spurred by restricting research to the easier problem.

4 The chaotic hypothesis (CH)

Following Boltzmann and Onsager we can ask whether there are general relations holding among time averages of selected observables and for all systems that can be modeled by time reversible mechanical equations $\dot{x} = f(x)$.

The difficulty is that in presence of dissipation it is by no means clear which is the probability distribution μ_α

which provides the average values of observables, at given control parameters α.

In finite thermostat models dissipation is manifested by the nonvanishing of the divergence, $\sigma(x) \stackrel{def}{=} -\sum \partial_{x_i} f_i(x)$, of the equations of motion and of its time average σ_+.

If $\sigma_+ > 0$[1], it is not possible that the distributions μ_α be of the form $\rho_\alpha(x)dx$, "absolutely continuous with respect to the phase space volume": since volume contracts, the probability distributions that, by integration, provide the averages of the observables must be concentrated on sets, "attractors", of 0 volume in phase space.

This means that there is no obvious substitute of the ergodic hypothesis: which, however, was essential in equilibrium statistical Mechanics to indicate that the "statistics" μ_α, i.e. the distribution μ_α such that

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T F(S_t x)dt = \int \mu_\alpha(dy)F(y) \qquad (4.1)$$

for all x except a set of zero volume, exists and is given by the Liouville volume (appropriately normalized to 1) on the surfaces of given energy U (which is therefore one of the parameters α on which the averages depend)[2].

It is well-known that identifying μ_α with the Liouville volume does not allow us to derive the values of the averages (aside from a few very simple cases, like the free gas): but it allows us to write the averages as explicit integrals [23], which are well suited to deduce relations holding between certain averages, like the second law equation (1.1) or Onsager reciprocity and the more general Fluctuation Dissipation Theorems.

The problem of finding a useful representation of the statistics of the stationary states in systems which are not in equilibrium arose in the more restricted context of fluid Mechanics earlier than in statistical Mechanics. And through a critique of earlier attempts [24], in 1973 Ruelle proposed that one should take advantage of the empirical fact that motions of turbulent systems are "chaotic" and suppose that their mathematical model should be a "hyperbolic system", in the same spirit in which the ergodic hypothesis should be regarded: namely *while one would be very happy to prove ergodicity because it would justify the use of Gibbs' microcanonical ensemble, real systems perhaps are not ergodic but behave nevertheless in much the same way and are well described by Gibbs' ensemble...* [25].

The idea has been extended in [23,26] to nonequilibrium statistical Mechanics in the form.

Chaotic hypothesis (CH). *Motions on the attracting set of a chaotic system can be regarded as motions of a smooth transitive hyperbolic system*[3].

The hypothesis was formulated to explain the result of the experiment in [27]. In [26] it was remarked that the CH could be adequate for the purpose.

5 "Free" implications of the chaotic hypothesis

Smooth transitive hyperbolic systems share, independently of the number of degrees of freedom, remarkable properties [28].

(1) Their motions can be considered paradigmatic chaotic evolutions, whose theory is, nevertheless, very well understood to the point that they can play for chaotic motions a role alike to the one played by harmonic oscillators for ordered motions [29].

(2) There is a *unique* distribution μ on phase space such that

$$\lim_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau F(S_t x)dt = \int \mu(dy)F(y) \qquad (5.1)$$

for all smooth F and for all but a zero volume set of initial data x [23,28,30,31], see Appendix A. The distribution μ is called the *SRB probability distribution*, see Appendix B.

(3) Averages satisfy a *large deviations rule*: i.e. if the point x in $f = \frac{1}{\tau}\int_0^\tau F(S_t x)\,dt$ is sampled with distribution μ, then

$$\lim_{\tau\to\infty} \frac{1}{\tau}\log \mathrm{Prob}_\mu(f\in\Delta) = \max_{f\in\Delta}\zeta_F(f) \qquad (5.2)$$

is an asymptotic value that controls the probability that the finite time average of F falls in an interval $\Delta = [u,v]$, $u<v$, subset of the interval (a_F, b_F) of definition of ζ_F. In the interval of definition $\zeta_F(f)$ is convex and analytic in f [30,32]. Outside $[a_F,b_F]$ the function $\zeta_F(f)$ can be defined to have value $-\infty$ (which means that values of f in intervals outside $[a_F,b_F]$ can possibly be observed only with a probability tending to 0 faster than exponentially) [30,32].

(4) A more precise form of equation (5.2) yields also the rate at which the limit is reached: $\mathrm{Prob}_\mu(f\in\Delta) = e^{\tau\max_{f\in\Delta}\zeta_F(f)+O(1)}$ with $O(1)$ bounded uniformly in τ, at fixed distance of Δ from the extremes a_F, b_F. This is ofteen written in a not very precise but mnemocnically convenient form, as long as its real meaning is kept in mind, as

$$P_\mu(f) = e^{\tau\,\zeta_F(f)+O(1)}. \qquad (5.3)$$

(5) The fluctuations described by (5.2) are very large fluctuations as they have size of order τ rather than $O(\sqrt{\tau})$: in fact if the maximum of $\zeta_F(f)$ is at a point $f_0\in(a_F,b_F)$ and is a nondegenerate quadratic maximum, then equation (5.2) implies that $\sqrt{\tau}(f-f_0)$ has an asymptotically Gaussian distribution. This means that the motion can be regarded to be so chaotic that the values of

---

[1] As intuition suggests σ_+ cannot be <0 [22], when motion takes place in a bounded region of phase space, as it is supposed here.

[2] By Liouville volume we mean the measure $\delta(K(\mathbf{p})+W(\mathbf{q})-U)d\mathbf{p}d\mathbf{q}$, on the manifold of constant energy or, in dissipative cases discussed later, the measure $d\mathbf{p}d\mathbf{q}$.

[3] Transitive means "having a dense orbit". Note that here this is a property of the attracting set, which is often not at all dense in the full phase space. Such systems are also called "Anosov systems".

$F(S_t x)$ are independent enough so that the finite time average deviations from the mean value f_0 are Gaussian on the scale of $\sqrt{\tau}$.

(6) A natural extension to (5.2) in which several observables $F_1, \ldots, F_n$ are simultaneously considered is obtained by defining $f_i = \frac{1}{\tau}\int_0^\tau F_i(S_t x)dt$. Then there exists a convex closed set $C \subset \mathcal{R}^n$ and function $\zeta_{\mathbf{F}}(\mathbf{f})$ analytic in $\mathbf{f} = (f_1, \ldots, f_n)$ in the interior of C and, given an open set $\Delta \subset C$,

$$\lim_{\tau\to\infty} \frac{1}{\tau} \log \mathrm{Prob}_\mu(\mathbf{f} \in \Delta) = \max_{\mathbf{f}\in\Delta} \zeta_{\mathbf{F}}(\mathbf{f}) \qquad (5.4)$$

and $\zeta_{\mathbf{F}}(\mathbf{f})$ could be defined as $-\infty$ outside C, with the meaning mentioned in remark (2). If the function $\zeta_{\mathbf{F}}(\mathbf{f})$ attains its maximum in a point $\mathbf{f}_0$ in the interior of C and the maximum is quadratic and nondegenerate, then the *joint fluctuations* of $\phi = \sqrt{\tau}(\mathbf{f} - \mathbf{f}_0)$ are asymptotically Gaussian, which means that have a probability density $\frac{1}{\sqrt{\pi^n \det \mathcal{D}}} e^{-\frac{1}{2}(\phi\cdot\mathcal{D}^{-1}\phi)}$ with $\mathcal{D}$ a positive definite $n \times n$ matrix.

(7) The probability distribution μ depends on the control parameters α of the initial data and therefore as α varies one obtains a collection of probability distributions: this leads to a natural *extension of the ensembles of equilibrium statistical Mechanics* [23].

(8) The most remarkable property, root of all the above, is that the SRB probability distribution μ, can be given a concrete formal representation, in spite of being a distribution concentrated on a set of zero volume [30,32], see Appendix A, B. This raises hopes to use it to derive general relations between averages of observables. As in equilibrium, the averages with respect to μ are destined to remain not computable except, possibly, under approximations (aside very few exactly soluble cases): their formal expressions could nevertheless be used to establish general mutual relations and properties.

(9) Given the importance of the existence and representability of the SRB distribution, Appendix A, B will be entirely devoted to the formulation (A1) and to the physical interpretation of the derivation of its expression: this could be useful for readers who want to understand the technical aspects of what follows, because some may find not satisfactory skipping the technical details even at a heuristic level. The aim of the non technical discussion that follows, preceding the appendices, is to make it worth to invest some time on the technical details.

(10) Applied to a system in equilibrium the CH implies the ergodic hypothesis so that it is a genuine extension of the latter and any results that follow from it will be necessarily compatible with those of equilibrium statistical Mechanics [23].

(11) For very simple systems the distribution μ can be constructed explicitly and time averages of some observables computed. The systems are the discrete time evolutions corresponding to linear hyperbolic maps of tori [28], or the continuous time geodesic motion on a surface of constant negative curvature. The latter systems are rigorously hyperbolic and the SRB distribution can be effectively computed for them *as well as for their small perturbations*.

(12) A frequent remark about the chaotic hypothesis is that it does not seem to keep the right viewpoint on nonequilibrium Thermodynamics. It should be stressed that the hypothesis is analogous to the ergodic hypothesis, which (*as well-known*) cannot be taken as the foundation of equilibrium statistical Mechanics, even though it leads to the correct Maxwell Boltzmann statistics, because the latter "holds for other reasons". Namely it holds because in most of phase space (measuring sizes by the Liouville measure) the few interesting macroscopic observables have the same value [33], see also [20].

6 Paradigms of statistical mechanics and CH

In relation to the last comment is useful to go back to the Heat Theorem of Section 1 and to a closer examination of the basic paper of Boltzmann [5], in which the theory of equilibrium ensembles is developed and may offer arguments for further meditation. The paper starts by illustrating an important, and today almost forgotten, remark by Helmoltz showing that very simple systems ("monocyclic systems") can be used to construct mechanical models of Thermodynamics: and the example chosen by Boltzmann is *really extreme by all standards*.

He shows that the motion of a *Saturn ring* of mass m on Keplerian orbits of major semiaxis a in a gravitational field of strength g can be used to build a model of Thermodynamics. In the sense that one can call

- "volume" V the gravitational constant g,
- "temperature" T the average kinetic energy,
- "energy" U the energy and
- "pressure" p the average potential energy mka^{-1},

then one infers that by varying, *at fixed eccentricity*, the parameters U, V the relation $(dU + pdV)/T =$ *exact* holds. Clearly this *could* be regarded as a curiosity, see [23], Appendix 1.A1, 9.A3.

However Boltzmann (following Helmoltz?[4]) took it seriously and proceeded to infer that under the ergodic hypothesis *any system* small or large provides us with a model of Thermodynamics (being "monocyclic" in the sense of Helmoltz): for instance he showed that the canonical ensemble verifies exactly the second law of equilibrium Thermodynamics (in the form $(dU + p\,dV)/T =$ *exact*) *without any need to take thermodynamic limits* [5,23].

[4] The relation between the two on this subject should be more studied. Boltzmann's paper of 1884 [5], is a natural follow up and completion of his earlier work [34] which followed [4,35]. It seems that the four extremely long papers by Helmoltz, also dated 1884 [36,37], might have at most just stimulated Boltzmann to revisit his earlier works and led him achieve the completion of the mechanical explanation of the second law. Certainly Boltzmann attributes a strong credit to Helmoltz, and one wonders if this might be partly due to the failed project that Boltzmann had to move to Berlin under the auspices of Helmoltz.

The same could be said of the microcanonical ensemble (here, however, he had to change "slightly" the definition of heat to make things work without finite size corrections).

He realized that the Ergodic Hypothesis could not possibly account for the correctness of the canonical (or microcanonical) ensembles; this is clear at least from his (later) paper in response to Zermelo's criticism [38]. Nor it could account for the observed time scales of approach to equilibrium. Nevertheless he called the theorem he had proved the *heat theorem* and never seemed to doubt that it provided evidence for the correctness of the use of the equilibrium ensembles for equilibrium statistical mechanics.

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

The first point has a close analogy in that the consequences of the Chaotic Hypothesis stem from the properties of small dimension hyperbolic systems (the best understood) which play here the role of Helmoltz' monocyclic systems of which Boltzmann's Saturn ring [5] is a special case. They are remarkable consequences because they provide us with *parameter free relations* (namely the Fluctuation theorem, to be discussed below, and its consequences): but clearly it cannot be hoped that a theory of nonequilibrium statistical Mechanics be founded solely upon them, by the same reasons why the validity of the second law for monocyclic systems had in principle no reason to imply the theory of ensembles.

Thus what is missing are arguments similar to those used by Boltzmann to justify the use of ensembles, *independently* of the ergodic hypothesis: an hypothesis which in the end may appear (and still does appear to many) as having only suggested them "by accident". The missing arguments should justify the CH on the basis of an extreme likelihood of its predictions in systems that are very large and that may be not hyperbolic in the mathematical sense. I see no reason, now, why this should prove impossible a priori or in the future. See Section 12 for some of the difficulties that can be met in experiments testing the CH through its consequence discussed in Section 7.

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

7 The fluctuation theorem (FT)

The idea of looking into time reversibility to explain the experimental results of [27] is clearly expressed in the same paper. The CH allows us to use effectively time reversal symmetry to obtain what has been called in [26,39,40] the "*Fluctuation theorem*". In fact a simple property holds for all transitive hyperbolic systems which admit a time reversal symmetry.

The property deals with the key observable $\sigma(x)$, which is the above introduced divergence of the equations of motion, or "phase space contraction rate". Assuming the average phase space contraction to be positive, $\sigma_+ > 0$, let $p = \frac{1}{\tau}\int_0^\tau \frac{\sigma(S_t x)}{\sigma_+} dt$ be the "dimensionless phase space contraction"; let $\zeta(p)$ be the large deviation rate function introduced in Section 5, see equation (5.2), for $F(x) = \frac{\sigma(x)}{\sigma_+}$. By time reversal symmetry the interval of analyticity of $\zeta(p)$ is centered at the origin and will be denoted $(-p^*, p^*)$; furthermore $p^* \geq 1$, because the average of p is 1. Then [26],

Fluctuation theorem (FT). *The probabilities of the large deviations of* p *satisfy, for all transitive time reversible hyperbolic systems,*

$$\zeta(-p) = \zeta(p) - p\sigma_+ \tag{7.1}$$

for all $|p| < p^*$*: this will be called a "fluctuation relation", (FR).*

Remarks. (1) In terms of the notation in equation (5.3) the FT is

$$\frac{P_\tau(p)}{P_\tau(-p)} = e^{p\,\sigma_+\,\tau + O(1)} \tag{7.2}$$

which is the form in which it is often written.

(2) The theorem has been developed, in [26], to understand the results of a simulation [27], whose authors had correctly pointed out that the SRB distribution together with the time reversibility could possibly explain the observations.

(3) Unfortunately the same name, introduced in [26, 39,40] where FT has been proved, has been *subsequently* picked up and attributed to other statements, superficially related to the above FT. Enormous confusion ensued (and sometimes even errors), see [11,41,42]. A more appropriate name for such other, and different, statements has been suggested to be "transient fluctuation theorems". The above FT should be distinguished also from the results in [6] which were the first *transient* fluctuations results, later extended and successfully applied, see [7,8]. It is claimed that the difference between the above FT and the transient statements is just an exchange of limits: the point is that it is a nontrivial one, see counterexamples in [11], and assumptions are needed, which have a physical meaning; the CH is the simplest.

(4) The FT theorem has been proved first for discrete time evolutions, i.e. for maps: in this case the averages over time are expressed by sums rather than by integrals. Hyperbolic maps are simpler to study than the corresponding

continuous time systems, which we consider here, because smooth hyperbolic maps do not have a trivial Lyapunov exponent (the vanishing one associated with the phase space flow direction); but the techniques to extend the analysis to continuous time systems are the same as those developed in [43] for proving the FT for hyperbolic flows and in this review I shall not distinguish between the two kinds of evolutions since the properties considered here do not really differ in the two cases.

(5) The condition $\sigma_+ > 0$, i.e. dissipativity, is *essential* even to define p itself. When the forcing intensity E vanishes also $\sigma_+ \to 0$ and the FR loses meaning because p does. Nevertheless by appropriately dividing both sides of equation (7.1) by σ_+, and then taking the limit, a nontrivial limit can be found and it can be shown, at least heuristically, to give the Green-Kubo relation for the "current" $J \stackrel{def}{=} \langle\frac{\partial\sigma}{\partial E}\rangle_\mu = \langle j\rangle_\mu$ [23,44], generated by the forcing, namely

$$\frac{dJ}{dE}\Big|_{E=0} = \frac{1}{2}\int_{-\infty}^{\infty} \langle j(S_\tau x)\, j(x)\rangle_{E=0} dt \qquad (7.3)$$

which is a general Fluctuation-Dissipation theorem.

(6) The necessity of a bound p^* in FT has attracted undue attention: it is *obvious* that it is there since $\sigma(x)$ is bounded, if CH holds. It also true that the role of p^* is discussed in the paper [39], which is a formal and *contemporary* version of the earlier [26] and of part of the later [40] written for a different audience in mind.

It is therefore surprising that this is sometimes ignored in the literature and the original papers are faulted for not mentioning this (obvious) point, which in any event is fully discussed in [39]. A proof which also discusses p^* is in [45]. It is also obvious that for $p \ge p^*$ the function $\zeta(p)$ can be naturally set to be $-\infty$, as commented in remark (6) to the CH in Section 4, and for this reason equation (7.1) is often written without any restriction on p. This is another point whose misunderstanding has led to errors. For readers familiar with statistical Mechanics there is nothing mysterious about p^*. It is analogous the "close packing density" in systems with hard cores: it is clear that there is a well defined maximum density but its value is not always explicitly computable; and for higher density many thermodynamic functions may be considered defined but as having an infinite value.

Corollary. [23,46], *Under the same assumptions of FT, if* $F_1 = \frac{\sigma(x)}{\sigma_+}$, $F_2, \ldots, F_n$ *are* n *observables of parity* $\varepsilon_i = \pm$ *under time reversal,* $F_i(Ix) = \varepsilon_i F_i(x)$, *the large deviations rate* $\zeta_{\mathbf{F}}(\mathbf{f})$, *defined in equation (5.4), satisfies*

$$\zeta_{\mathbf{F}}(\mathbf{f}^*) = \zeta_{\mathbf{F}}(\mathbf{f}) - \sigma_+ f_1 \qquad (7.4)$$

where $\mathbf{f}^* = (-f_1, \varepsilon_2 f_2, \ldots, \varepsilon_n f_n)$, *in its domain of definition* $C \subset \mathcal{R}^n$.

Remark. Note that the *r.h.s.* of equation (7.4) *does not depend on* $f_2, \ldots, f_n$. The independence has been exploited in [44] to show that when the forcing on the system is due to several forces of respective intensities $E_1, \ldots, E_s$ then by taking $F_1 = \frac{\sigma(x)}{\sigma_+}$, $F_2 = \partial_{E_k}\sigma(x)$, equation (7.4) implies, setting $j_k(x) = \partial_{E_k}\sigma(x)$ and $J_k = \langle j_k\rangle_\mu$, the Green Kubo relations (hence Onsager reciprocity)

$$L_{hk} = \partial_{E_h} J_k\big|_{\mathbf{E}=\mathbf{0}} = \frac{1}{2}\int_{-\infty}^{\infty} \langle j_h(S_\tau x)\, j_k(x)\rangle_{E=0} dt = L_{kh}. \qquad (7.5)$$

Therefore FT can be regarded as an *extension* to a nonlinear regime of Onsager reciprocity and of the Fluctuation-Dissipation theorems. Such a relation was pointed out in the context of volume preserving dynamics (hence in absence of dissipation), see comments in [6], p. 452 in particular. But it is not clear how to obtain from [6] the dissipative case results in equations (7.1), (7.4), (7.5) without the CH.

8 Fluctuation patterns, Onsager-Machlup theory

The last comment makes it natural to inquire whether there are more direct and physical interpretations of the FT (hence of the meaning of CH) when the external forcing is really different from the value 0 (the value always assumed in Onsager's theory).

The proof of the FT allows, as well, to deduce [47], an apparently more general statement (closely related to a relation recently found in the theory of the Kraichnan model of 2-dimensional turbulence and called "multiplicative" fluctuation theorem [48]) which can be regarded as an extension to nonequilibrium of the Onsager-Machlup theory of fluctuation patterns.

Consider observables $\mathbf{F} = (F_1 \stackrel{def}{=} \sigma/\sigma_+, \ldots, F_n)$ which have a well defined time reversal parity: $F_i(Ix) = \varepsilon_{F_i} F_i(x)$, with $\varepsilon_{F_i} = \pm 1$. Let F_{i+} be their time average (i.e. their SRB average) and let $t \to \phi(t) = (\varphi_1(t), \ldots, \varphi_n(t))$ be a smooth bounded function. Look at the probability, relative to the SRB distribution (i.e. in the "natural stationary state") that $F_i(S_t x)$ is $\varphi_i(t)$ for $t \in [-\frac{\tau}{2}, \frac{\tau}{2}]$: we say that $\mathbf{F}$ "follows the fluctuation pattern" ϕ in the time interval $t \in [-\frac{\tau}{2}, \frac{\tau}{2}]$.

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

We denote $\zeta(p, \phi)$ the *large deviation function* for observing in the time interval $[-\frac{\tau}{2}, \frac{\tau}{2}]$ an average phase space contraction $\sigma_\tau \stackrel{def}{=} \frac{1}{\tau}\int_{-\tau/2}^{\tau/2} \sigma(S_t x)dt = p\sigma_+$ and at the same time a fluctuation pattern $\mathbf{F}(S_t x) = \phi(t)$. This means that the probability that the *dimensionless phase space contraction rate* p is in a closed set Δ and F is in a closed neighborhood of an assigned ψ[5], denoted $U_{\psi,\varepsilon}$, is given by:

$$\exp\left(\sup_{p\in\Delta,\phi\in U_{\psi,\varepsilon}} \tau\,\zeta(p,\phi)\right) \tag{8.1}$$

to leading order as $\tau\to\infty$ (i.e. the logarithm of the mentioned probability divided by τ converges as $\tau\to\infty$ to $\sup_{p\in\Delta,\phi\in U_{\psi,\varepsilon}}\zeta(p,\phi)$). Needless to say p and ϕ have to be "possible" otherwise ζ has to be set $-\infty$, as in the FT case in Section 6, Comment (6).

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

Conditional reversibility relation. *If* $\mathbf{F}=(F_1,\ldots,F_n)$ *are n observables with defined time reversal parity* $\varepsilon_{F_i}=\pm1$ *and if* τ *is large the fluctuation pattern* $\phi(t)$ *and its time reversal* $I\varphi_i(t)\stackrel{def}{=}\varepsilon_{F_i}\varphi_i(-t)$ *will be followed with equal likelihood if the first is conditioned to a contraction rate p and the second to the opposite −p. This holds because:*

$$\frac{\zeta(p,\phi)-\zeta(-p,I\phi)}{p\sigma_+}=1 \qquad \text{for } |p|\le p^* \tag{8.2}$$

with ζ *introduced in equation (8.1) and a suitable* $p^*\ge1$.

It will appear, in Section 9, that the phase space contraction rate should be identified with a macroscopic quantity, the *entropy creation rate*. Then the last theorem can be interpreted as saying, in other words, that while it is very difficult, in the considered systems, to see an "anomalous" average entropy creation rate during a time τ (e.g. $p=-1$), it is also true that "that is the hardest thing to see". Once we see it, *all the observables will behave strangely* and the relative probabilities of time reversed patterns will become as likely as those of the corresponding direct patterns under "normal" average entropy creation regime.

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

The result in equation (8.2) is a "*relation*" rather than a theorem because, even in the hyperbolic cases, the precise restrictions on the "allowed" test functions $\varphi_i(t)$ have not been discussed in [47] from a strict mathematical viewpoint and it would be interesting to formulate them explicitly and investigate their generality[6].

The result can be informally stated in a only apparently stronger form, for $|p|<p^*$, and with the warnings in Remark (4) preceding the analogous equation (5.3), as

$$\frac{P_\tau(\text{for all } j, \text{and } t\in[-\frac12\tau,\frac12\tau]\,:\,F_j(S_tx)\sim\varphi_j(t))}{P_\tau(\text{for all } j, \text{and } t\in[-\frac12\tau,\frac12\tau]\,:\,F_j(S_tx)\sim-\varphi_j(-t))}= e^{\,p\,\sigma_+\,\tau+O(1)}, \tag{8.3}$$

where P_τ is the SRB probability, *provided the phase space contraction* $\sigma(x)$ *is a function of the observables* $\mathbf{F}$. This is certainly the case if σ is one of the F_i, for instance if $\sigma=F_1$. Here $F_j(S_tx)\sim\varphi_j(t)$ means $|F_j(S_tx)-\varphi_j(t)|$ small for $t\in[-\frac{\tau}{2},\frac{\tau}{2}]$.

Remarks. (1) A relation of this type has been remarked recently in the context of the theory of Lagrangian trajectories in the Kraichnan flow [48].

(2) One should note that in applications results like equation (8.3) will be used under the CH and therefore other errors may arise because of its approximate validity (the hypothesis in fact essentially states that "things go as if" the system was hyperbolic): they may depend on the number N of degrees of freedom and we do not control them except for the fact that, if present, their relative value should tend to 0 as $N\to\infty$: there may be (and there are) cases in which the chaotic hypothesis is not reasonable for small N (e.g. systems like the Fermi-Pasta-Ulam chains) but it might be correct for large N. We also mention that, on the other hand, for some systems with small N the CH may be already regarded as valid (e.g. for the models in [27,49,50]).

(3) The proofs of FT and the corollaries are not difficult. Once their meaning in terms of coarse graining is understood, the a priori rather mysterious SRB distribution μ is represented, surprisingly, as a Gibbs distribution for a 1–dimensional spin system, which is elementary and well understood. In Appendix A, B some details are given about the nature of coarse graining and in Appendix C the steps of the proof of FT are illustrated.

In conclusion the FT is a general parameterless relation valid, in time reversible systems, independently of the number of degrees of freedom: the CH allows us to consider it as a manifestation of time reversal symmetry.

9 Reversible thermostats and entropy creation

Recalling that kinetic theory developed soon after the time average of a mechanical quantity, namely kinetic energy, was understood to have the meaning of absolute temperature, it is tempting to consider quite important that, from the last three decades of research on nonequilibrium statistical Mechanics, an interpretation emerged of the physical meaning of the mechanical quantity σ = phase space contraction.

A system in contact with thermostats can generate entropy in the sense that it can send amounts of heat into

---

[5] By "closed neighborhood" $U_{\psi,\varepsilon}$, $\varepsilon>0$, around ψ, we mean that $|F_i(S_tx)-\psi_i(t)|\le\varepsilon$ for $t\in[-\frac{\tau}{2},\frac{\tau}{2}]$.

[6] A sufficient condition should be that $\varphi_i(t)$ are bounded and smooth.

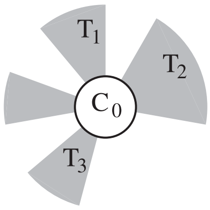

Fig. 1. Particles in $\mathcal{C}_0$ ("system particle") interact with the particles in the shaded regions ("thermostats particles"); the latter are constrained to have a fixed total kinetic energy.

the thermostats thus increasing their entropy by the ratio of the heat to the temperature, because the thermostats must be considered in thermal equilibrium.

Furthermore if phase space contraction can be identified with a physical quantity, accessible by means of calorimetric/thermometric measurements, then the FT prediction becomes relevant and observable and the CH can be subjected to tests, *independently on the microscopic model that one may decide to assume*, which therefore become possible also in real experiments.

It turns out that in very general thermostat models entropy production rate can be identified with phase space contraction *up to a "total time derivative"*: and since additive total time derivatives (as we shall see) do not affect the asymptotic distribution of time averages, one can derive a FR for the entropy production (a quantity accessible to measurement) from a FR for phase space contraction (a quantity, in general, *not accessible* except in numerical simulations, because it requires a precise model for the system, as a rule not available).

As an example, of rather general nature, consider the following one, obtained by imagining a system which is in contact with thermostats that are "external" to it. The particles of the system $\mathcal{C}_0$ are enclosed in a container, also called $\mathcal{C}_0$, with elastic boundary conditions surrounded by a few thermostats which consist of particles, all of unit mass for simplicity, interacting with the system via short range interactions, through a portion $\partial_i\mathcal{C}_0$ of the surface of $\mathcal{C}_0$, and subject to the constraint that the total kinetic energy of the N_i particles in the ith thermostat is $K_i = \frac12\dot{\mathbf{X}}_i^2 = \frac32 N_i k_B T_i$. A symbolic illustration is in Figure 1.

The equations of motion will be (all masses equal for simplicity)

$$\begin{aligned} m\ddot{\mathbf{X}}_0 =& -\partial_{\mathbf{X}_0}\Big(U_0(\mathbf{X}_0)+\sum_{j>0}W_{0,j}(\mathbf{X}_0,\mathbf{X}_j)\Big)+\mathbf{E}(\mathbf{X}_0),\\ m\ddot{\mathbf{X}}_i =& -\partial_{\mathbf{X}_i}\Big(U_i(\mathbf{X}_i)+W_{0,i}(\mathbf{X}_0,\mathbf{X}_i)\Big)-\alpha_i\dot{\mathbf{X}}_i \end{aligned} \tag{9.1}$$

with α_i such that K_i is a constant. Here $W_{0,i}$ is the interaction potential between particles in $\mathcal{C}_i$ and in $\mathcal{C}_0$, while U_0, U_i are the internal energies of the particles in $\mathcal{C}_0, \mathcal{C}_i$ respectively. We imagine that the energies $W_{0,j}, U_j$ are due to *smooth* translation invariant pair potentials; repulsion from the boundaries of the containers will be elastic reflection.

It is assumed, in equation (9.1), that there is no direct interaction between different thermostats: their particles interact directly only with the ones in $\mathcal{C}_0$. Here $\mathbf{E}(\mathbf{X}_0)$ denotes possibly present external positional forces stirring the particles in $\mathcal{C}_0$. The constraints on the thermostats kinetic energies give

$$\alpha_i \equiv \frac{Q_i-\dot U_i}{3N_ik_BT_i} \qquad\longleftrightarrow\qquad K_i\equiv \text{const.} \stackrel{def}{=} \frac32 N_ik_BT_i \tag{9.2}$$

where Q_i is the work per unit time that particles outside the thermostat $\mathcal{C}_i$ (hence in $\mathcal{C}_0$) exercise on the particles in it, namely

$$Q_i \stackrel{def}{=} -\dot{\mathbf{X}}_i\cdot\partial_{\mathbf{X}_i}W_{0,i}(\mathbf{X}_0,\mathbf{X}_i) \tag{9.3}$$

and it will be interpreted as the "*amount of heat*" Q_i entering the thermostat $\mathcal{C}_i$ per unit time.

The main feature of the model is that the thermostats are external to the system proper: this makes the model suitable for the study of situations in which no dissipation occurs in the interior of a system but it occurs only on the boundary.

The *divergence* $-\sigma(\dot{\mathbf{X}},\mathbf{X})$ of the equations of motion, which gives the rate of contraction of volume elements around $d\dot{\mathbf{X}}d\mathbf{X}$, does not vanish and can be computed in the model in Figure 1; simple algebra yields, remarkably,

$$\begin{aligned} \sigma(\dot{\mathbf{X}},\mathbf{X}) =& \varepsilon(\dot{\mathbf{X}},\mathbf{X})+\dot R(\mathbf{X}),\\ \varepsilon(\dot{\mathbf{X}},\mathbf{X}) =& \sum_{j>0}\frac{Q_j}{k_BT_j}, \qquad R(\mathbf{X})=\sum_{j>0}\frac{U_j}{k_BT_j} \end{aligned} \tag{9.4}$$

where $\varepsilon(\dot{\mathbf{X}},\mathbf{X})$ can be interpreted as the *entropy production rate*, because of the meaning of Q_i in equation (9.3) [7].

This is an interesting result because of its generality: it has implications for the thermostated system considered in Figure 1, for instance. It is remarkable that the quantity p has a simple physical interpretation: equation (9.1) shows that the functions $\zeta_\sigma(p)$ and $\zeta_\varepsilon(p)$ are *identical* because, since R is bounded by our assumption of smoothness, equations (9.2) and (9.3) imply

$$\frac1\tau\int_0^\tau \sigma(S_t(\dot{\mathbf{X}},\mathbf{X}))dt \equiv \frac1\tau\int_0^\tau \varepsilon(S_t(\dot{\mathbf{X}},\mathbf{X}))dt+\frac{R(\tau)-R(0)}{\tau}, \tag{9.5}$$

so that

$$\sigma_+ = \lim_{\tau\to\infty}\frac1\tau\int_0^\tau \sigma(S_t(\dot{\mathbf{X}},\mathbf{X}))dt \equiv \lim_{\tau\to\infty}\frac1\tau\int_0^\tau \varepsilon(S_t(\dot{\mathbf{X}},\mathbf{X}))dt=\varepsilon_+ \tag{9.6}$$

[7] Equation (9.4) are correct up to $O(N^{-1})$ if $N=\min N_j$ because the addends should contain also a factor $(1-\frac{1}{3N_j})$ to be exact: for simplicity $O(1/N)$ corrections will be ignored here and in he following (their inclusion would imply trivial changes without affecting the physical interpretation) [51].

and the asymptotic distributions of

$$p' = \frac{1}{\tau}\int_0^\tau \frac{\sigma(S_t(\dot{\mathbf{X}},\mathbf{X}))}{\sigma_+}dt,$$

and of
$$p = \frac{1}{\tau}\int_0^\tau \frac{\varepsilon(S_t(\dot{\mathbf{X}},\mathbf{X}))}{\varepsilon_+}dt \qquad (9.7)$$

are the same.

Equation (9.1) are time reversible (with $I(\dot{\mathbf{X}},\mathbf{X}) = (-\dot{\mathbf{X}},\mathbf{X})$): then under the CH the large deviations rate $\zeta(p)$ for the observable $\frac{\sigma}{\sigma_+}$ satisfies the "*fluctuation relation*", equation (7.1). It also follows that the large deviations rate for $\frac{\varepsilon}{\varepsilon_+}$, identical to $\zeta(p)$, satisfies it as well.

The point is that ε is measurable by "calorimetric and thermometric measurements", given its interpretation of entropy increase of the thermostats. Therefore the CH can be subjected to test or it can be used to "predict" the frequency of occurrence of unlikely fluctuations.

Comment. This is a rather general example of thermostats action, but it is just an example. For instance it can be generalized further by imagining that the system is thermostated in its interior. A situation that arises naturally in the theory of electric conduction. In the latter case the electrons move across the lattice of the metal atoms and the lattice oscillations, i.e. the phonons, absorb or give energy. This can be modeled by adding a "inner" thermostat force $-\alpha_0\dot{\mathbf{x}}_i$, acting on the particles in $\mathcal{C}_0$, which fixes the temperature of the electron gas. Actually a very similar model appeared in the early days of Statistical Mechanics, in Drude's theory of electric conductivity [52]. Other examples can be found in [51].

10 Fluids

The attempt to put fluids and turbulence within the context provided by the ideas exposed in the previous sections forces to consider cases in which dissipation takes place irreversibly. This leads us to a few conjectures and remarks.

To bypass the obstacle due to the nonreversibility of the fluid equations which, therefore, seem quite far from the equations controlling the thermostated systems just considered, the following "equivalence conjecture" [53], has been formulated. Consider the two equations for an incompressible flow with velocity field $\mathbf{u}(\mathbf{x},t)$, $\boldsymbol{\partial}\cdot\mathbf{u}=0$, in periodic boundary condition for simplicity,

$$\dot{\mathbf{u}} + \underset{\sim}{\mathbf{u}}\cdot\underset{\sim}{\partial}\,\mathbf{u} = \nu\Delta\mathbf{u} - \partial p + \mathbf{g},$$
$$\dot{\mathbf{u}} + \underset{\sim}{\mathbf{u}}\cdot\underset{\sim}{\partial}\mathbf{u} = \alpha(\mathbf{u})\Delta\mathbf{u} - \partial p + \mathbf{g}, \qquad (10.1)$$

where $\alpha(\mathbf{u}) = \frac{\int \mathbf{u}\cdot\mathbf{g}\,d\mathbf{x}}{\int(\partial\mathbf{u})^2\,d\mathbf{x}}$ is a "Lagrange multiplier" determined so that the total energy $\mathcal{E} \overset{def}{=} \int \mathbf{u}^2\,d\mathbf{x}$ is exactly constant.

Note that velocity reversal $I:\ \mathbf{u}(\mathbf{x}) \to -\mathbf{u}(\mathbf{x})$ anticommutes, in the sense of equation (2.1), with the time evolution generated by the second equation (because $\alpha(I\mathbf{u}) = -\alpha(\mathbf{u})$), which means that "fluid elements" retrace their paths with opposite velocity.

Introduce the "local observables" $F(\mathbf{u})$ as functions depending only upon finitely many Fourier components of $\mathbf{u}$, i.e. on the "large scale" properties of the velocity field $\mathbf{u}$. Then, *conjecture* [54], the two equations should have "same large scale statistics" in the limit $R\to+\infty$. If μ_ν and $\widetilde{\mu}_{\mathcal{E}}$ denote the respective SRB distributions of the first and the second equations in equation (10.2), by *"same statistics"* as $R\to\infty$ it is meant that

(1) if the total energy $\mathcal{E}$ of the initial datum $\mathbf{u}(0)$ for the second equation is chosen equal to the average $\langle\int\mathbf{u}^2\,d\mathbf{x}\rangle_{\mu_\nu}$ for the SRB distribution μ_ν of the first equation, then
(2) the two SRB distributions μ_ν and $\widetilde{\mu}_{\mathcal{E}}$ are such that, in the limit $R\to\infty$, the difference $\langle F\rangle_{\mu_\nu} - \langle F\rangle_{\widetilde{\mu}_{\mathcal{E}}} \xrightarrow[R\to+\infty]{} 0$.

So far *only numerical tests* of the conjecture, in strongly cut off 2-dimensional equations, have been attempted [55].

An analogy with the thermodynamic limit appears naturally: namely the Reynolds number plays the role of the volume, locality of observables becomes locality in $\mathbf{k}$-space, and $\nu,\mathcal{E}$ play the role of canonical temperature and microcanonical energy of the SRB distributions of the two different equations in (10.1), respectively μ_ν and $\widetilde{\mu}_{\mathcal{E}}$.

The analogy suggests to question whether reversibility of the second equation in equation (10.1) can be detected. In fact to be able to see for a large time a viscosity opposite to the value ν would be very unphysical and would be against the spirit of the conjecture.

If the CH is supposed to hold it is possible to use the FT, which is a consequence of reversibility, to estimate the probability that, say, the value of α equals $-\nu$. For this purpose we have to first determine the attracting set.

Assuming the K41 [53], theory of turbulence the attracting set will be taken to be the set of fields with Fourier components $\mathbf{u}_\mathbf{k}=0$ unless $|\mathbf{k}|\le R^{\frac{3}{4}}$.

Then the expected identity $\langle\alpha\rangle=\nu$, between the average friction $\langle\alpha\rangle$ in the second of equation (10.1) and the viscosity ν in the first, implies that the divergence of the evolution in the second of equation (10.1) is in average

$$\sigma \sim \nu \sum_{|\mathbf{k}|\le R^{3/4}} 2|\mathbf{k}|^2 \sim \nu\left(\frac{2\pi}{L}\right)^2 \frac{8\pi}{5}R^{15/4}. \qquad (10.2)$$

By FT the SRB-probability to see, in motions following the second equation in equation (10.2), a "*wrong*" average friction $-\nu$ for a time τ is

$$\mathrm{Prob}_{srb} \sim \exp\left(-\tau\nu\frac{32\pi^3}{5L^2}R^{\frac{15}{4}}\right) \overset{def}{=} e^{-g\tau}. \qquad (10.3)$$

It can be estimated in the situation considered below for a flow in air:

$$\begin{cases} \nu = 1.5 \times 10^{-2}\,\frac{\text{cm}^2}{\text{sec}}, \quad v = 10\,\frac{\text{cm}}{\text{sec}} \quad L = 100 \text{ cm} \\ R = 6.67 \times 10^4, \quad g = 3.66 \times 10^{14}\ \text{sec}^{-1} \\ P \stackrel{def}{=} \text{Prob}_{srb} = e^{-g\tau} = e^{-3.66\times 10^8}, \quad \text{if} \quad \tau = 10^{-6} \end{cases} \tag{10.4}$$

where the first line are data of an example of fluid motion and the other two lines follow from equation (10.3). They show that, by FT, viscosity can be $-\nu$ during 10^{-6} s (*say*) with probability P as in equation (10.4): unlikelihood is similar in spirit to the estimates about Poincaré's recurrences [53].

(2) If we imagine that the particles are so many that the system can be well described by a macroscopic equation, like for instance the NS equation, then there will be two ways of computing the entropy creation rate. The first would be the classic one described for instance in [56], and the second would simply be the divergence of the microscopic equations of motion in the model of Figure 1, under the assumption that the motion is closely described by macroscopic equations for a fluid in local thermodynamic equilibrium, like the NS equations. This can be correct in the limit in which space and time are rescaled by ε and ε^2 and the velocity field by ε, and ε is small. Since local equilibrium is supposed, it will make sense to define a local entropy density $s(\mathbf{x})$ and a total entropy of the fluid $S = \int s(\mathbf{x})\,d\mathbf{x}$.

The evaluation of the expression for the entropy creation rate as a divergence σ of the microscopic equations of motion leads to [57], a value $\langle\varepsilon\rangle$ with average (over a microscopically long time short with respect to the time scale of the fluid evolution) related to the classical entropy creation rate in a NS fluid as

$$\begin{aligned} &k_B\langle\varepsilon\rangle = k_B\varepsilon_{classic} + \dot{S}, \\ &k_B\varepsilon_{classic} = \int_{\mathcal{C}_0}\left(\kappa\left(\frac{\partial T}{T}\right)^2 + \eta\,\frac{1}{T}\underset{\sim}{\tau}'\,\underset{\sim}{\partial}\mathbf{u}\right)d\mathbf{x} \end{aligned} \tag{10.5}$$

where $\underset{\sim}{\tau}'$ is the tensor $(\partial_i u_j + \partial_j u_i)$ and η is the dynamic viscosity, so that the two expressions differ by the time derivative of an observable, which equals the total equilibrium entropy of the fluid $S = \int s(\mathbf{x})\,d\mathbf{x}$ where s is the thermodynamical entropy density in the assumption of local equilibrium; see comment on additive total derivatives preceding Figure 1.

11 Quantum systems

Recent experiments deal with properties on mesoscopic and atomic scale. In such cases the quantum nature of the systems may not always be neglected, particularly at low temperature, and the question is whether a fluctuation analysis parallel to the one just seen in the classical case can be performed in studying quantum phenomena.

Thermostats have, usually, a macroscopic phenomenological nature: in a way they should be regarded as classical macroscopic objects in which no quantum phenomena occur. Therefore it seems natural to model them as such and define their temperature as the average kinetic energy of their constituent particles so that the question of how to define it does not arise.

Consider the system in Figure 1 when the quantum nature of the particles in $\mathcal{C}_0$ cannot be neglected. Suppose for simplicity (see [58]) that the nonconservative force $\mathbf{E}(\mathbf{X}_0)$ acting on $\mathcal{C}_0$ vanishes, i.e. consider the problem of heat flow through $\mathcal{C}_0$. Let H be the operator on $L_2(\mathcal{C}_0^{3N_0})$, space of symmetric or antisymmetric wave functions $\Psi(\mathbf{X}_0)$,

$$H = -\frac{\hbar^2}{2m}\Delta_{\mathbf{X}_0} + U_0(\mathbf{X}_0) + \sum_{j>0}\big(U_{0j}(\mathbf{X}_0,\mathbf{X}_j) + U_j(\mathbf{X}_j) + K_j\big) \tag{11.1}$$

where $\Delta_{\mathbf{X}_0}$ is the Laplacian, and note that its spectrum consists of eigenvalues $E_n = E_n(\{\mathbf{X}_j\}_{j>0})$, for $\mathbf{X}_j$ fixed (because the system in $\mathcal{C}_0$ has finite size).

A system-reservoirs model can be the *dynamical system* on the space of the variables $\big(\Psi, (\{\mathbf{X}_j\},\ \{\dot{\mathbf{X}}_j\})_{j>0}\big)$ defined by the equations (where $\langle\cdot\rangle_\Psi$ = expectation in the state Ψ)

$$\begin{aligned} &-i\hbar\dot{\Psi}(\mathbf{X}_0) = (H\Psi)(\mathbf{X}_0), \qquad \text{and for } j>0 \\ &\ddot{\mathbf{X}}_j = -\big(\partial_j U_j(\mathbf{X}_j) + \langle\partial_j U_j(\mathbf{X}_0,\mathbf{X}_j)\rangle_\Psi\big) - \alpha_j\dot{\mathbf{X}}_j \\ &\alpha_j \stackrel{def}{=} \frac{\langle W_j\rangle_\Psi - \dot{U}_j}{2K_j}, \quad W_j \stackrel{def}{=} -\dot{\mathbf{X}}_j\partial_j U_{0j}(\mathbf{X}_0,\mathbf{X}_j) \end{aligned} \tag{11.2}$$

here the first equation is Schrödinger's equation, the second is an equation of motion for the thermostats particles similar to the one in Figure 1 (whose notation for the particles labels is adopted here too). The model has no pretension of providing a physically correct representation of the motions in the thermostats nor of the interaction system thermostats, see comments at the end of this section.

Evolution maintains the thermostats kinetic energies $K_j \equiv \frac{1}{2}\dot{\mathbf{X}}_j^2$ exactly constant, so that they will be used to define the thermostats temperatures T_j via $K_j = \frac{3}{2}k_B T_j N_j$, as in the classical case.

Let $\mu_0(\{d\Psi\})$ be the *formal* measure on $L_2(\mathcal{C}_0^{3N_0})$

$$\Big(\prod_{\mathbf{X}_0} d\Psi_r(\mathbf{X}_0)\,d\Psi_i(\mathbf{X}_0)\Big)\,\delta\left(\int_{\mathcal{C}_0}|\Psi(\mathbf{Y})|^2\,d\mathbf{Y} - 1\right) \tag{11.3}$$

with Ψ_r, Ψ_i real and imaginary parts of Ψ. The meaning of (11.3) can be understood by imagining to introduce an orthonormal basis in the Hilbert space and to "cut it off" by retaining a large but finite number M of its elements, thus turning the space into a high dimensional space C^M (with $2M$ real dimensions) in which

$d\Psi = d\Psi_r(\mathbf{X}_0)\, d\Psi_i(\mathbf{X}_0)$ is simply interpreted as the normalized Euclidean volume in C^M.

The formal phase space volume element $\mu_0(\{d\Psi\}) \times \nu(d\mathbf{X}\, d\dot{\mathbf{X}})$ with

$$\nu(d\mathbf{X}\, d\dot{\mathbf{X}}) \stackrel{def}{=} \prod_{j>0} \big(\delta(\dot{\mathbf{X}}_j^2 - 3N_j k_B T_j)\, d\mathbf{X}_j\, d\dot{\mathbf{X}}_j\big) \quad (11.4)$$

is conserved, by the unitary property of the wave functions evolution, just as in the classical case, *up to the volume contraction in the thermostats* [51].

If $Q_j \stackrel{def}{=} \langle W_j \rangle_\Psi$ and R is as in equation (9.4), then the contraction rate σ of the volume element in equation (11.4) can be computed and is (again) given by equation (9.4) with ε, that will be called *entropy production rate*: setting $R(\mathbf{X}) \stackrel{def}{=} \sum_{j>0} \frac{U_j(\mathbf{X}_j)}{k_B T_j}$, it is

$$\sigma(\Psi, \dot{\mathbf{X}}, \mathbf{X}) = \varepsilon(\Psi, \dot{\mathbf{X}}, \mathbf{X}) + \dot{R}(\mathbf{X}),$$
$$\varepsilon(\Psi, \dot{\mathbf{X}}, \mathbf{X}) = \sum_{j>0} \frac{Q_j}{k_B T_j}. \quad (11.5)$$

In general solutions of equation (11.2) *will not be quasi periodic* and the chaotic hypothesis [23,40,58], can be assumed: if so the dynamics should select an SRB distribution μ. The distribution μ will give the statistical properties of the stationary states reached starting the motion in a thermostat configuration $(\mathbf{X}_j, \dot{\mathbf{X}}_j)_{j>0}$, randomly chosen with "uniform distribution" ν on the spheres $m\dot{\mathbf{X}}_j^2 = 3N_j k_B T_j$ and in a random eigenstate of H. The distribution μ, if existing and unique, could be named the *SRB distribution* corresponding to the chaotic motions of equation (11.2).

In the case of a system *interacting with a single thermostat* at temperature T_1 the latter distribution should be equivalent to the canonical distribution, up to boundary terms.

Hence an important consistency check, for proposing equation (11.2) as a model of a thermostated quantum system, is that there should exist at least one stationary distribution equivalent to the canonical distribution at the appropriate temperature T_1 associated with the (constant) kinetic energy of the thermostat: $K_1 = \frac{3}{2} k_B T_1\, N_1$. In the corresponding classical case this is an established result [23,51,59].

A natural candidate for a stationary distribution could be to attribute a probability proportional to $d\Psi\, d\mathbf{X}_1\, d\dot{\mathbf{X}}_1$ times

$$\sum_{n=1}^{\infty} e^{-\beta_1 E_n} \delta(\Psi - \Psi_n(\mathbf{X}_1)\, e^{i\varphi_n})\, d\varphi_n\, \delta(\dot{\mathbf{X}}_1^2 - 2K_1) \quad (11.6)$$

where $\beta_1 = 1/k_B T_1$, Ψ are wave functions for the system in $\mathcal{C}_0$, $\dot{\mathbf{X}}_1, \mathbf{X}_1$ are positions and velocities of the thermostat particles and $\varphi_n \in [0, 2\pi]$ is a phase, $E_n = E_n(\mathbf{X}_1)$ is the n-th level of $H(\mathbf{X}_1)$, with $\Psi_n(\mathbf{X}_1)$ the corresponding eigenfunction. The average value of an observable O for the system in $\mathcal{C}_0$ in the distribution μ in (11.6) would be

$$\langle O \rangle_\mu = Z^{-1} \int \mathrm{Tr}\,(e^{-\beta H(\mathbf{X}_1)} O)\, \delta(\dot{\mathbf{X}}_1^2 - 2K_1) d\mathbf{X}_1\, d\dot{\mathbf{X}}_1 \quad (11.7)$$

where Z is the integral in (11.7) with 1 replacing O, (normalization factor). Here one recognizes that μ attributes to observables the average values corresponding to a Gibbs state at temperature T_1 with a random boundary condition $\mathbf{X}_1$.

However equation (11.6) *is not invariant* under the evolution equation (11.2) and it seems difficult to exhibit explicitly an invariant distribution. Therefore one can say that the SRB distribution for the evolution in (11.2) is equivalent to the Gibbs distribution at temperature T_1 only as a conjecture.

Nevertheless it is interesting to remark that under the *adiabatic approximation* the eigenstates of the Hamiltonian at time 0 evolve by simply following the variations of the Hamiltonian $H(\mathbf{X}(t))$ due to the motion of the thermostats particles, without changing quantum numbers (rather than evolving following the Schrödinger equation and becoming, therefore, different from the eigenfunctions of $H(\mathbf{X}(t))$).

In the adiabatic limit in which the classical motion of the thermostat particles takes place on a time scale much slower than the quantum evolution of the system the distribution (11.6) *is invariant*.

This can be checked by first order perturbation analysis which shows that, to first order in t, the variation of the energy levels (supposed non degenerate) is compensated by the phase space contraction in the thermostat [58]. Under time evolution, $\mathbf{X}_1$ changes, at time $t > 0$, into $\mathbf{X}_1 + t\dot{\mathbf{X}}_1 + O(t^2)$ and, assuming non degeneracy, the eigenvalue $E_n(\mathbf{X}_1)$ changes, by perturbation analysis, into $E_n + t\, e_n + O(t^2)$ with

$$e_n \stackrel{def}{=} t\langle \dot{\mathbf{X}}_1 \cdot \partial_{\mathbf{X}_1} U_{01} \rangle_{\Psi_n} + t\dot{\mathbf{X}}_1 \cdot \partial_{\mathbf{X}_1} U_1 =$$
$$- t\,(\langle W_1 \rangle_{\Psi_n} + \dot{R}_1) = -\frac{1}{\beta_1}\alpha_1. \quad (11.8)$$

Hence the Gibbs factor changes by $e^{-\beta t e_n}$ and at the same time phase space contracts by $e^{t\frac{3N_1 e_n}{2K_1}}$, as it follows from the expression of the divergence in equation (11.5). *Therefore if β is chosen such that $\beta = (k_B T_1)^{-1}$ the state with distribution equation (11.6) is stationary*, (recall that for simplicity $O(1/N)$, see footnote[7] on p. 9 is neglected). This shows that, *in the adiabatic approximation*, interaction with only one thermostat at temperature T_1 admits at least one stationary state. The latter is, by construction, a Gibbs state of thermodynamic equilibrium with a special kind (random $\mathbf{X}_1, \dot{\mathbf{X}}_1$) of boundary condition and temperature T_1.

Remarks. (1) The interest of the example is to show that even in quantum systems the chaotic hypothesis makes sense and the interpretation of the phase space contraction in terms of entropy production remains unchanged.

In general, under the chaotic hypothesis, the SRB distribution of (11.2) (which in presence of forcing, or of more than one thermostat is certainly quite not trivial, as in the classical Mechanics cases) will satisfy the fluctuation relation because the fluctuation theorem only depends on reversibility: so the model (11.2) might be suitable (given its chaoticity) to simulate the steady states of a quantum system in contact with thermostats.

(2) It is certainly unsatisfactory that a stationary distribution cannot be explicitly exhibited for the single thermostat case (unless the adiabatic approximation is invoked). However, according to the proposed extension of the CH, the model does have a stationary distribution which should be equivalent (in the sense of ensembles equivalence) to a Gibbs distribution at the same temperature.

(3) The non quantum nature of the thermostat considered here and the specific choice of the interaction term between system and thermostats should not be important: the very notion of thermostat for a quantum system is not at all well defined and it is natural to think that in the end a thermostat is realized by interaction with a reservoir where quantum effects are not important. Therefore what the analysis really suggests is that in experiments in which really microscopic systems are studied the heat exchanges of the system with the external world should fulfill a FR.

(4) The conjecture can probably be tested with present day technology. If verified it could be used to develop a "Fluctuation Thermometer" to perform temperature measurements which are *device independent* in the same sense in which the gas thermometers are device independent (i.e. do not require "calibration" of a scale and "comparison" procedures).

Consider a system in a stationary state, and imagine inducing small currents and measuring the average heat output rate Q_+ and the fluctuations in the finite time average heat output rate, generated by inducing small currents, i.e. fluctuations of $p = \frac{1}{\tau}\int_0^\tau \frac{Q(t)}{Q_+}dt$ obtaining the rate function of $\zeta(p)$.

Then it becomes possible to read from the slope of $\zeta(p)-\zeta(-p)$, equal to $\frac{Q_+}{k_BT}$ by the FR, directly the inverse temperature that the thermostat in contact with the system has: this could be useful particularly in very small systems (classical or quantum). The idea is inspired by a similar earlier proposal for using fluctuation measurements to define temperature in spin glasses [60,61], p. 216.

12 Experiments?

The (partial) test of the chaotic hypothesis via its implication on large fluctuations probabilities (i.e. the fluctuation relation) is quite difficult. The main reason is that if the forcing is small the relation degenerates (because $\varepsilon_+ \to 0$) and it can be shown [44], that to lowest nontrivial order in the size of the forcing it reduces to the Green-Kubo formula, which is (believed to be) well established so that the fluctuation relation will not be significant, being "true for other reasons" [56]. See Section 3.

Hence one has to consider large forcing. However, under large forcing, large fluctuations of p become very rare, hence their statistics is difficult to observe. Furthermore the statistics seems to remain Gaussian for p, in a region around $p = 1$ where the data can be considered reliably unbiased (see below), and until rather large values of the forcing field or values of $|p-1|$ large compared to the root mean square deviation $\frac{D}{\sqrt{\tau}} = \langle (p-1)^2\rangle^{1/2}$ are reached. Hence $\zeta(p) = -\frac{1}{2D^2}(p-1)^2$ and linearity in p of $\zeta(p)-\zeta(-p)$ is trivial. *Nevertheless*, in this regime, it follows that $\frac{2}{D^2} = \sigma_+$ which is a nontrivial relation and therefore a simple test of the fluctuation relation.

The FR was empirically observed first in such a situation in [27], in a simulation, and the first dedicated tests, after recognizing its link with the CH, were still performed in a Gaussian regime, so that they were really only tests of $\frac{2}{D^2} = \sigma_+$ and of the Gaussian nature of the observed fluctuations.

Of course in simulations the forcing can be pushed to "arbitrarily large" values so that the fluctuation relation can, in principle, be tested in a regime in which $\zeta(p)$ is sensibly non Gaussian, see [62]. But far more interesting will be cases in which the distribution $\zeta(p)$ is sensibly not Gaussian and which deal with laboratory experiments rather than simulations. Skepticism towards the CH is mainly based on the supposed non measurability of the function $\zeta(p)$ in the large deviation domain (i.e. $|p-1| \gg \sqrt{\langle (p-1)^2\rangle}$).

In experimental tests several other matters are worrysome, among which:

(a) is reversibility realized? This is a rather stringent and difficult point to understand on a case by case basis, because irreversibility creeps in, inevitably, in dissipative phenomena.

(b) is it allowed to consider R, i.e. the "entropy production remainder" in (9.3), bounded? if not there will be corrections to FR to study (which in some cases [63,64], can be studied quite in detail).

(c) does one introduce any bias in the attempts to see statistically large deviations? For instance in trying to take τ large one may be forced to look at a restricted class of motions, typically the ones that remain observable for so long a time. It is easy to imagine that motions observed by optical means, for instance, will remain within the field of the camera only for a characteristic time τ_0 so that any statistics on motions that are observed for times $\tau > \tau_0$ will be biased (for it would deal with untypical events).

(d) chaotic motions may occur under influence of stochastic perturbations, so that extensions of FT to stochastic systems may need to be considered. This is not really a problem because a random perturbation can be imagined as generated by coupling of the system to another dynamical system (which, for instance, in simulations would be the random number generator from which the noise is drawn), nevertheless it demands careful analysis [65].

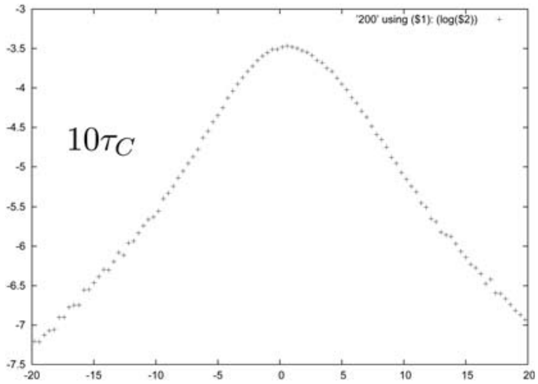

Fig. 2. An histogram of $\log P_\tau(p)$, taken from the data of [66] at time $\tau = 10\tau_C = 200$ ms: it shows the rather typical nonconvexity for $|p - 1| \sim 8$ which is of the order of standard deviation.

(e) Nonconvex shape of $\zeta(p)$, at $|p - 1|$ beyond the root mean square deviation, see Figure 2, is seen often, possibly always, in the experiments that have been attempted to study large deviations. Therefore the interpretation of the nonconvexity, via well understood corrections to FR, seems to be a forced path towards a full test of the FR, beyond the Gaussian regime [64].

All the above questions arise in the recent experiment by Bandi-Cressman-Goldburg [66]. It encounters all the related difficulties and to some extent provides the first evidence for the FR (hence the CH) in a system in which the predictions of the FR are not the result of a theoretical model which can be solved exactly. The interpretation of the results is difficult and further investigations are under way.

The experiment outcome is not incompatible with FR and, in any event, it proves that good statistics can be obtained for fluctuations that extend quite far beyond the root mean square deviation of $p-1$: an asset of the results in view of more refined experiments.

A very promising field for experimental tests of the CH and the FR is granular materials: in granular materials collisions are not elastic, nevertheless an experiment is proposed in [67]. See Comment (6) in Section 13 and Comment (4) to equation (11.8) for other hints at possible experiments and applications.

13 Comments

(1) In the context of the finite thermostats approach, besides systems of particles subject to deterministic evolution, stochastically evolving systems can be considered and the FT can be extended to cover the new situations [48,65,68–70].

(2) Alternative quantum models have also been considered in the literature [71] (stochastic Langevin thermostats), or infinite thermostats (free and interacting, and possibly with further noise sources) [13,16,17,72,73].

(3) Many simulations have been performed, starting with the experiment which showed data that inspired the FT [27], and continuing after the proof of FT and the formulation of the CH, e.g. [50]: a few had the purpose of testing the Fr in a non Gaussian regime for the fluctuations of the variable p [62]. In some cases the results had to be examined closely to understand what was considered at discrepancy with the FT [64] (and was not).

(4) The physical relevance of the particular quantum thermostat model remains an open question and essentially depends on the conjecture that the (unknown) SRB distribution for the model in the single thermostat case is equivalent to the Gibbs distribution at the same temperature (a property valid in the corresponding classical cases). Hence the main interest of the model is that it shows that a FR is in principle possible in finite thermostated quantum systems in stationary state.

(5) Few experiments have so far been performed (besides numerical simulations) to investigate CH and FT: extensions to randomly forced systems are possible [68–70], and can be applied to systems that can be studied in laboratory [66,74]: the first experiment designed to test the FR in a laboratory experiment is the recent work [66]. The results are consistent with the FR and indicate a promising direction of research.

(6) An interesting consequence of the FT is that

$$\langle e^{-\Delta S/k_B}\rangle_{srb} \stackrel{def}{=} \langle e^{-\int_0^\tau \sum_{j>0} \frac{Q_j(t)}{k_B T_j}\,dt}\rangle_{srb} = O(1) \qquad (13.1)$$

in the sense that the logarithms of both sides divided by τ agree in the limit $\tau \to \infty$ (i.e. $\lim_{\tau\to+\infty} \frac{1}{\tau} \log \langle e^{\Delta S/k_B}\rangle = 0$) with corrections of order $O(\frac{1}{\tau})$. This has been pointed out by Bonetto, see [23], and could have applications in the same biophysics contexts in which the work theorems [7,8], have been applied: for instance one could study stationary heat exchanges is systems out of equilibrium (rather than measure free energy differences between equilibrium states at the same temperature as in [7,8]). The boundedness of the l.h.s. of equation (13.1) implied by (13.1) can be used to test whether some heat emissions have gone undetected (which would imply that the l.h.s. of equation (13.1) tends to 0, rather than staying of $O(1)$). This is particularly relevant as in biophysics one often studies systems in stationary states while actively busy at exchanging heat with the surroundings.

(7) Another property, which is not as well known as it deserves, is that for hyperbolic systems, and by the Chaotic Hypothesis of Section 2, virtually for all chaotic evolutions, it is possible to develop a rigorous theory of coarse graining [12,75]. It leads to interpreting the SRB distributions as uniform distributions on the attractor; hence to a variational principle and to the existence of a Lyapunov function describing the approach to the stationary state, i.e. giving a measure of the distance from it [21,57].

However it also seems to lead to the conclusion that *entropy* of a stationary state *cannot be defined* if one requires that it should have properties closely analogous to the equilibrium entropy. For instance once coarse graining has been properly introduced, it is tempting to define the entropy of a stationary state as k_B times the logarithm of the number of "microcells" into which the attractor is decomposed, see Appendix A, B.

This quantity can be used as a Lyapunov function, see [57], but it depends on the size of the microcells in a nontrivial way: changing their size, the variation of the so defined entropy does not change by an additive constant depending only on the scale of the coarse graining (*at difference with respect to the equilibrium case*), but by a quantity that depends also on the control parameters (e.g. temperature, volume, etc.) [21].

Given the interest of coarse graining, in Appendix A mathematical details about it are discussed in the context of the SRB distribution and CH; and a physical interpretation is presented in Appendix B; hopefully they will also clarify the physical meaning of the two.

(8) Finally it is often said that the FR should hold *always* or, if not, it is incorrect. In this respect it has to be stressed that the key assumption is the CH, which implies the FR *exactly* in time reversible situations. However it is clear that CH is an idealization and the correct attitude is to interpret deviations from FR in terms of corrections to the CH. For instance:

– CH implies exponential decay of time correlations. But in some cases there are physical reasons for long range time correlations.
– Or the CH implies that observables have values in a finite range. But there are cases in which phase space is not bounded and observables can take unbounded values (or such for practical purposes).
– Time reversal is necessary. But there are cases in which it is violated.
  The pdf of p should be log-convex: but it is seldom so.

What is interesting is that it appears that starting from CH and examining the features responsible for its violations it may be possible to compute even quantitatively the corrections to FR. Examples of such corrections already exist [63,64,76]. It would be interesting to have a concrete experiment, designed to test FR and try to understand the observed deviations; the BCG experiment in Section 12 offers, if further developed, the possibility of simple tests making use the existing experimental apparatus and of the observations that it has proved to be accessible.

I am grateful to M. Bandi, A. Giuliani, W. Goldburg and F. Zamponi for countless comments and suggestions and to M. Bandi, W. Goldburg for providing their data, partially reported in Figure 2. Partially supported also by Institut des Hautes Etudes Scientifiques, by Institut Henri Poincaré and by Rutgers University.

Appendix A: Coarse graining, SRB and 1*D* Ising models

In equilibrium phase space volume is conserved and it is natural to imagine it divided into tiny "cells", in which all observables of interest are constant. The equilibrium distribution can be constructed simply by imagining to have divided phase space Σ ("energy surface") into cells of equal Liouville volume, small enough so that every interesting physical observable F is constant in each cell. Then the dynamics is a cyclic permutation of the cells (*ergodic hypothesis*) so that the stationary distribution is just the volume distribution.

In a way, this is an "accident", based on what appears to be a fundamentally incorrect premise, which leads to various difficulties as it is often considered in the context of attempts to put on firm grounds the notion of a "coarse grained" description of the dynamics. Confusion is also added by the simulations: the latter are sometimes interpreted as *de facto* coarse grained descriptions. It seems, however, essential to distinguish between coarse graining and representation of the dynamics as a permutation of small but finite cells.

Undoubtedly dynamics can be represented by a permutation of small phase space volumes, as any simulation program effectively does. But it is also clear that the cells used in the simulations are *far too small* (i.e. of the size determined by the computer resolution, typically of double precision reals) to be identified with the coarse cells employed in phenomenological studies of statistical Mechanics.

On the other hand if coarse grain cells are introduced which are not as tiny as needed in simulations the *dynamics will deform* them to an extent that after a short time it will no longer be possible to identify which cell has become which other cell! And this applies even to equilibrium states.

In this respect it looks as an accident the fact that, nevertheless, at least in equilibrium a coarse grained representation of time evolution appears possible. And easily so, with small cells subject to the only condition of having equal volume; but the huge amount of literature on attempts at establishing a theory of coarse graining did not lead to a precise notion, nor to any agreement between different proposals.

Under the CH systems are hyperbolic and a precise analysis of coarse graining seems doable, see [21,29,77]. The key is that it is possible to distinguish between "microcells", so tiny that evolution is well approximated by a permutation on them, and "cells" which are still so small that the (few) interesting observables have constant value on them. The latter cells can be identified with "coarse grain cells"; yet they are very large compared to the microcells and time evolution *cannot* be represented as their permutation. *Neither in equilibrium nor out of equilibrium.*

That SRB distribution cannot be considered a permutation of naively defined coarse cells *seems* to be well known and to have been considered a drawback of the SRB distributions: it partly accounts for the skepticism that often, still now, accompanies them.

The point that will be made, see the review [77]*, is that hyperbolicity provides us with a natural definition of coarse grained cells. At the same time it tells us which is the weight to be given to each cell which, in turn, implies that each cell can be imagined containing many "microcells"* *whose evolution is a simple permutation of them* (just as in numerical simulations).

In this appendix, we consider for simplicity discrete time systems: in this case hyperbolic systems are described by a smooth map S, transitive and smoothly invertible, with the property that every phase space point x is a "saddle point". Out of x emerge the stable and the unstable manifolds $W^s(x), W^u(x)$ of complementary dimension. The expansion and contraction that take place near every point x can be captured by the matrices $\partial S_u(x)$, $\partial S_s(x)$ obtained by restricting the matrix (*Jacobian matrix*) $\partial S(x)$, of the derivatives of S, to its action on the vectors tangent to the unstable and stable manifolds through x: the evolution S maps $W^u(x)$,$W^s(x)$ to $W^u(Sx), W^s(Sx)$, respectively, and its derivative (i.e. its linearization) maps tangent vectors at x into tangent vectors at Sx.

A quantitative expression of the expansion and contraction is given by the "local expansion" or "local contraction" rates defined by

$$\Lambda_1^u(x) \stackrel{def}{=} \log|\det(\partial S)_u(x)|,$$
$$\Lambda_1^s(x) \stackrel{def}{=} -\log|\det(\partial S)_s(x)|. \tag{A.1}$$

Since time is now discrete, phase space contraction is now defined as $\sigma(x) = -\log|\det(\partial S)|$ and related to $\Lambda_1^u(x), \Lambda_1^s(x)$ by

$$\sigma(x) = -\Lambda_1^u(x) + \Lambda_1^s(x) - \log\frac{\sin\delta(Sx)}{\sin\delta(x)}, \tag{A.2}$$

where $\delta(x)$ is the angle (in the metric chosen in phase space) between $W^s(x), W^u(x)$ (which is bounded away from 0 and π by the smoothness of the hyperbolic evolution S).

This suggests to imagine constructing a partition $\mathcal{P}$ of phase space into closed regions $\mathcal{P} = (P_1,\ldots,P_m)$ with pairwise disjoint interiors, each of which is a "rectangle" defined as follows.

The rectangle P_i, see the following Figure 3 for a visual guide, has a center κ_i out of which emerge portions $C\subset W^s(\kappa_i), D\subset W^u(\kappa_i)$ of its stable and unstable manifolds, small compared to their curvature, which form the "axes" of P_i, see Figure 3. The set P_i, then, consists of the points x obtained by taking a point p in the axis D and a point q in the axis C and setting $x \stackrel{c}{=} W^s(p)\cap W^u(q)$, just as in an ordinary rectangle a point is determined by the intersection of the lines through any two points on the axes and perpendicular to them, see Figure 3. The symbol $\stackrel{c}{=}$ means that x is the point closest to p and to q along paths in $W^s(p)$ and, respectively $W^u(q)$ [8].

Note that in a rectangle *anyone* of its points κ could be the center in the above sense with a proper choice of C, D, so that κ_i does not play a special role and essentially serves as a label identifying the rectangle. In dimension higher than 2 the rectangles may (and will) have rather rough (non differentiable) boundaries [78].

[8] This proviso is needed because often, and certainly in transitive hyperbolic maps, the full manifolds $W^s(p), W^u(q)$ are dense in phase space and intersect infinitely many times [30,32].

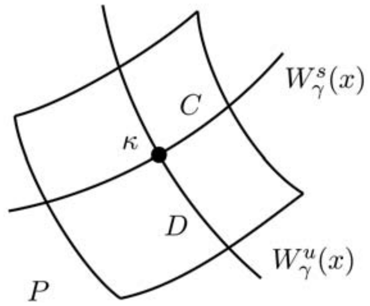

Fig. 3. A rectangle P with a pair of axes C, D crossing at the corresponding center κ.

It is a key property of hyperbolicity (hence of systems for which the CH can be assumed) that the partition $\mathcal{P}$ can be built to enjoy of a very special property.

Consider the sequence, *history of* x, $\xi(x) \stackrel{def}{=} \{\xi_i\}_{i=-\infty}^{\infty}$ of symbols telling into which of the sets of $\mathcal{P}$ the point S^ix is, i.e. where x is found at time i, or $S^ix\in P_{\xi_i}$. This is unambiguous aside from the zero volume set $\mathcal{B}$ of the points that in their evolution fall on the common boundary of two P_ξ's.

Define the matrix Q to be $Q_{\xi,\xi'}=0$, unless there is an interior point in P_ξ whose image is in the interior of $P_{\xi'}$: and in the latter case set $Q_{\xi,\xi'}=1$. Then the history of a point x, which in its evolution does not visit a boundary common to two P_ξ's, must be a sequence ξ verifying the property, called *compatibility*, that, $Q_{\xi_k,\xi_{k+1}}=1$ for all times k.

The matrix Q tells us which sets $P_{\xi'}$ can be reached from points in P_ξ in one time step. Then transitive hyperbolic maps admit a partition (in fact infinitely many) of phase space into rectangles $\mathcal{P}=(P_1,\ldots,P_m)$, so that

(1) if ξ is a compatible sequence then there is a point x such that $S^kx\in P_{\xi_k}$, see (for instance) Chapter 9 in [23], ("*compatibility*"). The points x outside the exceptional set $\mathcal{B}$ (of zero volume) determine uniquely the corresponding sequence ξ;

(2) the diameter of the set of points $E(\xi_{-\frac12 T},\ldots,\xi_{\frac12 T})$ consisting of all points which between time $-\frac12 T$ and $\frac12 T$ visit, in their evolution, the sets P_{ξ_i} is bounded above by $c\,e^{-c'T}$ for some $c, c'>0$ (i.e. the code $\xi\to x$ determines x "*with exponential precision*");

(3) there is a power k of Q such that $Q^k_{\xi\xi'}>0$ for all ξ,ξ' ("*transitivity*").

Hence points x can be identified with sequences of symbols ξ verifying the compatibility property and the sequences of symbols determine, with exponential rapidity, the point x which they represent.

The partitions $\mathcal{P}$ are called *Markov partitions*. Existence of $\mathcal{P}$ is nontrivial and rests on the chaoticity of motions: because the compatibility of all successive pairs implies that the full sequence is actually the history of a point (a clearly false statement for general partitions)[9].

If the map S has a time reversal symmetry I (i.e. a smooth involution I, such that $IS = S^{-1}I$, see Eq. (2.1)) the partition $\mathcal{P}$ can be so built that $I\mathcal{P} = \mathcal{P}$, hence $IP_i = P_{I(i)}$ for some $I(i)$. This is done simply by replacing $\mathcal{P}$ by the finer partition whose elements are $P_i \cap IP_j$, because if $\mathcal{P}, \mathcal{P}_1$ and $\mathcal{P}_2$ are Markovian partitions also the partition $I\mathcal{P}$ is such, as well as the partition $\mathcal{P}_1 \vee \mathcal{P}_2$ formed by intersecting all pairs $P \in \mathcal{P}_1$, $P' \in \mathcal{P}_2$ (this is best seen from the geometric interpretation in footonote[9] and from the time reversal property that $IW_u(x) = W_s(Ix)$).

A Markov partition such that $I\mathcal{P} = \mathcal{P}$ is called "reversible" and histories on it have the simple property that $(\xi(Ix))_i = (\xi(x))_{-I(i)}$.

Markov partitions, when existing, allow us to think of the phase space points as the configurations of a "1-dimensional spin system", i.e. as sequences of finitely many symbols $\xi \in \{1, 2, \ldots, m\}$ subject to the "hard core" constraint that $Q_{\xi_i,\xi_{i+1}} = 1$. Hence probability distributions on phase space which give 0 probability to the boundaries of the elements of the Markov partitions (where history may be ambiguous) can be regarded as stochastic processes on the configurations of a 1-dimensional Ising model (with finite spin m), and functions on phase space can be regarded as functions on the space of compatible sequences[10].

The remarkable discovery, see reviews in [30,32], is that the SRB distribution not only can be regarded as a stochastic processes, but it *is a short range Gibbs distribution* if considered as a probability on the space of the compatible symbolic sequences ξ on $\mathcal{P}$, and with a potential function $A(\xi) = -\Lambda_1^u(x(\xi))$, see below and [28].

The sequences ξ are therefore much more natural, given the dynamics S, than the sequence of decimal digits that are normally used to identify the points x via their cartesian coordinates[11].

Definition (**Coarse graining**). Given a Markovian partition $\mathcal{P}$ let $\mathcal{P}^T$ be the finer partition of phase space into sets of the form

$$E_\xi = E_{\xi_{-T/2},\ldots,\xi_{T/2}} \stackrel{def}{=} \bigcap_{-T/2}^{T/2} S^k P_{\xi_k}. \tag{A.3}$$

The sets E_ξ will be called "elements of a description of the microscopic states coarse grained to scale γ" if γ is the largest linear dimension of the nonempty sets E_ξ. The elements E_ξ of the "coarse grained partition $\mathcal{P}^T$ of phase space" are labeled by a finite string

$$\xi = (\xi_{-T/2}, \ldots, \xi_{T/2}) \tag{A.4}$$

with $\xi_i = 1, \ldots, m$ and $Q_{\xi_i,\xi_{i+1}} = 1$.

Define the *forward* and *backward* expansion and contraction rates as

$$U_{u,\pm}^{T/2}(x) = \sum_{j=0}^{\pm T/2} \Lambda_1^u(S^j x), \qquad U_{s,\pm}^{T/2}(x) = \sum_{j=0}^{\pm T/2} \Lambda_1^s(S^j x) \tag{A.5}$$

and select a point $\kappa(\xi) \in E_\xi$ for each ξ. Then the SRB distribution μ_{SRB} and the volume distribution μ_L on the phase space Ω, which we suppose to have Liouville volume, footnote p. 4, $V(\Omega)$, attribute to the *nonempty* sets E_ξ the respective probabilities μ and μ_L

$$\mu(\xi) \stackrel{def}{=} \mu_{SRB}(E_\xi) \quad \text{and respectively} \quad \mu_L(\xi) \stackrel{def}{=} \frac{V(E_\xi)}{V(\Omega)} \tag{A.6}$$

if $V(E)$ denotes the Liouville volume of E. The distributions μ, μ_L are shown [23,28], to be defined by

$$\begin{aligned} \mu(\xi) &= h_{u,u}^T(\xi) e^{\left(-U_{u,-}^{T/2}(\kappa(\xi)) - U_{u,+}^{T/2}(\kappa(\xi))\right)} \\ \mu_L(\xi) &= h_{s,u}^T(\xi) e^{\left(U_{s,-}^{T/2}(\kappa(\xi)) - U_{u,+}^{T/2}(\kappa(\xi))\right)} \end{aligned} \tag{A.7}$$

where $\kappa(\xi) \in E_\xi$ is the center of P_{ξ_0} and $h_{u,u}^T(\xi)$, $h_{s,u}^T(\xi)$ are suitable functions of ξ, *uniformly bounded* as ξ and T vary and which are mildly dependent on ξ; so that they can be regarded as constants for the purpose of the present discussion, cf. Chapter 9 in [23].

If γ is a scale below which all interesting observables are (for practical purposes) constant, then choosing $T =$

[9] The Markovian property has a geometrical meaning: imagine each P_i as the "stack" of the connected unstable manifolds portions $\delta(x)$, intersections of P_i with the unstable manifolds of its points x, which will be called unstable "layers" in P_i. Then if $Q_{i,j} = 1$, the expanding layers in each P_i expand under the action of S and their images *fully cover* the layers of P_j which they touch. Formally let $P_i \in \mathcal{P}$ and $x \in P_i$, $\delta(x) \stackrel{c}{=} P_i \cap W_u(x)$: the if $Q_{i,j} = 1$, i.e. if SP_i visits P_j, it is $\delta(Sx) \subset S\delta(x)$.

[10] It is worth also stressing that the ambiguity of the histories for the points which visit the boundaries of the sets of a Markovian partition is very familiar in the decimal representation of coordinates: it corresponds to the ambiguity in representing a decimal number as ending in infinitely many 0's or in infinitely many 9's.

[11] If the phase space points are considered as sequences ξ then the dynamics becomes a "trivial" left shift of histories. This happens always in symbolic dynamics, but in general it is of little interest unless compatibility can be decided by a "hard core condition" involving only nearest neighbors (in general compatibility is a global condition involving all symbols, i.e. as a hard core it is one with infinite range). *Furthermore* also the statistics of the motion becomes very well understood, because short range $1D$ Gibbs distributions are elementary and well understood.

$O(\log \gamma^{-1})$ the sets E_ξ are a coarse graining of phase space suitable for computing time averages as weighted sums over the elements of the partition.

And both in equilibrium and out of equilibrium the SRB distribution *will not attribute equal weight* to the sets E_ξ. The weight will be instead proportional to $e^{\left(-U^{T/2}_{u,-}(\kappa(\xi))-U^{T/2}_{u,+}(\kappa(\xi))\right)}$, i.e. to the inverse of the exponential of the expansion rate of the map S^T along the unstable manifold and as a map of $S^{-\frac{T}{2}}\kappa(\xi)$ to $S^{\frac{T}{2}}\kappa(\xi)$. The more unstable the cells are the less weight they have. Given equation (A.7) the connection with the Gibbs state with potential energy $A(\xi) = \Lambda^u_1(\xi)$ appears, see [28], Section 4.3 and Chapters 5, 6.

The sets E_ξ represent macroscopic states, being just small enough so that the physically interesting observables have a constant value within them; and we would like to think that they provide us with a model for a "*coarse grained*" description of the microscopic states. The notion of coarse graining is, here, precise and, nevertheless, quite flexible because it contains a free "resolution parameter" γ. Should one decide that the resolution γ is not good enough because one wants to study the system with higher accuracy then one simply chooses a smaller γ (and, correspondingly, a larger T).

Appendix B: SRB and coarse graining: a physicist's view

How can the analysis of Appendix A be reconciled with the numerical simulations, and with the naive view of motion, as a permutation of cells? The phase space volume will generally contract with time: yet we want to describe the evolution in terms of an evolution permuting microscopic states. Also because this would allow us to count the microscopic states relevant for a given stationary state of the system and possibly lead to extending to stationary nonequilibria Boltzmann's definition of entropy.

Therefore we divide phase space into *equal* parallelepipedal *microcells* Δ of side size $\varepsilon \ll \gamma$ and try to discuss time evolution in terms of them: we shall call such cells "microscopic" cells, as we do not associate them with any particular observable; they represent the highest microscopic resolution.

The new microcells should be considered as realizations of objects alike to those arising in computer simulations: in simulations the cells Δ are the "digitally represented" points with coordinates given by a set of integers and the evolution S is a *program* or *code* simulating the solution of equations of motion suitable for the model under study. The code operates *exactly* on the coordinates (the deterministic round offs, enforced by the particular computer hardware and software, should be considered part of the program).

The simulation will produce (generically) a chaotic evolution "for all practical purposes", i.e.

(1) if we only look at "macroscopic observables" which are constant on the coarse graining scale $\gamma = e^{-\frac{1}{2}\overline{\lambda}T}\ell_0$ of the partition $\mathcal{P}^T$, where ℓ_0 is the phase space size and $\overline{\lambda} > 0$ is the least contractive line element exponent (which therefore fixes the scale of the coarse graining, by the last definition)[12] and;
(2) if we look at phenomena on time scales far shorter than the recurrence times (always finite in finite representations of motion, but of size usually so long to make the recurrence phenomenon irrelevant)[13].

The latter conclusion can be reached by realizing that

(a) there has to be a small enough division into microcells that allows us to describe evolution as a map (otherwise numerical simulations would not make sense);
(b) however the evolution map cannot be, in general, a permutation. In simulations it will happen, *essentially always*, that it (i.e. the software program) will send two distinct microcells into the same one. It does certainly happen in nonequilibrium systems in which phase space contracts in the average[14];
(c) even though the map will not be one-to-one, nevertheless it will be such *eventually*: because any map on a finite space is a permutation of the points which are recurrent. This set is the *attractor* of the motions, that we call $\mathcal{A}$ and which will be imagined as a the collection of microcells approximating the unstable manifold and intersecting it. All such microcells will be considered taking part in the permutation: but this is not an innocent assumption and in the end is the reason why the SRB is unique, see below;
(d) every permutation can be decomposed into cycles: each cycle will visit each coarse cell with the same frequency (unless there are more than one stationary distributions describing the asymptotics of a set of microcells initially distributed uniformly, a case that we exclude because of the transitivity assumption). Hence it is not restrictive to suppose that there is only one cycle ("ergodicity" on the attractor).

Then consistency between the expansion of the unstable directions and the existence of a cyclic permutation of the microcells in the attractor $\mathcal{A}$ *demands that the number of microcells in each coarse grained cell* E_ξ*, equation (A.3),*

[12] Here it is essential that the CH holds, otherwise if the system has long time tails the analysis becomes much more involved and so far it can be dealt, even if only qualitatively, on a case by case basis.

[13] To get an idea of the orders of magnitude consider a gas of N particles of density ρ at temperature T: the metric on phase space will be $ds^2 = \sum_i (\frac{d\mathbf{p}_i^2}{k_B T} + \frac{d\mathbf{q}_i^2}{\rho^{-2/3}})$; hence the size of a microcell will be $\sqrt{O(N)}\,\delta_0$ if δ_0 is the precision with which the coordinates are imagined determined (in simulations $\delta_0 \simeq 10^{-14}$ in double precision) as all contributions to ds^2 are taken of order $O(1)$. Coarse grained cells contain, in all proposals, many particles, $O(N)$, so that their size will contain a factor δ rather than δ_0 and will be $\delta/\delta_0 = O(N^{1/3})$ larger.

[14] With extreme care it is sometimes, and in equilibrium, possible to represent evolution with a code which is a true permutation: the only example that I know, dealing with a physically relevant model, is in [79].

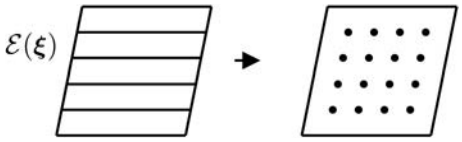

Fig. 4. A very schematic and idealized drawing of the attractor layers $\Delta(\xi)$, remaining after a transient time, inside a coarse cell $\mathcal{E}(\xi)$; the second drawing (indicated by the arrow) represents schematically what the layers really are, if looked closely: namely collections of microcells laying uniformly on the attractor layers, i.e. the discretized attractor intersected with the coarse cell.

must be inversely proportional to the expansion rate, i.e. it has to be given by the first of equation (A.7).

More precisely we imagine, developing a heuristic argument, that the attractor in each coarse cell $\mathcal{E}(\xi)$ will appear as a stack of a few portions of unstable manifolds, the "layers" of footnote[9], whose union form the (disconnected) surface $\Delta(\xi)$ intersection between $\mathcal{E}(\xi)$ and the attractor. Below $\Delta(\xi)$ will be used to denote both the set and its surface, as the context demands. The stack of connected surfaces $\Delta(\xi)$ is imagined covered uniformly by $N(\xi)$ microcells, see Figure 4.

Let $t \stackrel{def}{=} T+1$. Transitivity implies that there is a smallest integer $m \geq 0$ such that $S^{t+m}\mathcal{E}(\xi)$ intersects all other $\mathcal{E}(\xi')$: the integer m is t-independent (and equal to the minimum m such that $Q^m_{\sigma,\sigma'} > 0$). In $t+m$ time steps each coarse cell will have visited all the others and the layers describing the approximate attractor in a single coarse cell will have been expanded *to cover the entire attractor* for the map S^{t+m} [15]. The latter coincides with the attractor for S because S^j is transitive for all j if it is such for $j=1$ and this property has to be reflected by the discretized dynamics at least if j is very small compared to the (enormous) recurrence time on the discrete attractor as t is, being a time on the coarse grain scale.

Suppose first that $m=0$, hence $S^t\Delta(\xi)$ is the entire attractor for all ξ. This is an assumption useful to exhibit the idea but unrealistic for invertible maps: basically this is realized in the closely related SRB theory for a class of non invertible expansive maps of the unit interval).

So the density of microcells will be $\rho(\xi) = \frac{N(\xi)}{\Delta(\xi)}$ and under time evolution S^t the unstable layers $\Delta(\xi')$ in $\mathcal{E}(\xi)$ expand and cover all the layers in the cells $\mathcal{E}(\xi')$. If the coarse cell $\mathcal{E}(\xi)$ is visited, in $t=T+1$ time steps, by points in the coarse cells ξ', a property that will be symbolically denoted $\mathcal{E}_{\xi'} \in S^{-t}\mathcal{E}(\xi)$, a fraction $\nu_{\xi,\xi'}$ of the $N(\xi')$ microcells will end in the coarse cell $\mathcal{E}(\xi)$, and $\sum_\xi \nu_{\xi,\xi'} = 1$. Then consistency with evolution as a cyclic permutation demands

$$N(\xi) = \sum_{\xi'} \frac{N(\xi')}{\Delta(\xi')} \frac{1}{e^{\Lambda_{u,T}(\xi')}} \Delta(\xi) \stackrel{def}{=} \mathcal{L}(N)(\xi), \qquad \text{i.e.} \tag{B.1}$$

because the density of the microcells on the images of $\Delta(\xi')$ decreases by the expansion factor $e^{\Lambda_{u,T}(\xi')}$, so that $\nu_{\xi,\xi'} = \frac{\Delta(\xi)}{\Delta(\xi')}\frac{1}{e^{\Lambda_{u,T}(\xi')}}$.

As a side remark it is interesting to point out that for the density $\rho(\xi)$ equation (B.1) becomes simply $\rho(\xi) = \sum_{\xi'} e^{-\Lambda_{u,T}(\xi')}\rho(\xi')$, closely related to the similar equation for invariant densities of Markovian surjective maps of the unit interval [28].

The matrix $\mathcal{L}$ has all elements > 0 (because $m=0$) and therefore has a simple eigenvector v with positive components to which corresponds the eigenvalue λ with maximum modulus: $v = \lambda\,\mathcal{L}(v)$ (the "Perron-Frobenius theorem") with $\lambda = 1$ (because $\sum_\xi \nu_{\xi,\xi'} = 1$). It follows that the consistency requirement uniquely determines $N(\xi)$ as proportional to v_ξ. Furthermore $S^t\Delta(\xi)$ is the entire attractor; then its surface is ξ independent and equal to $e^{\Lambda_{u,T}(\xi)}\Delta(\xi)$: therefore $N(\xi) = const\, e^{-\Lambda_{u,T}(\xi)}$.

The general case is discussed by considering S^{t+m} instead of S^t: this requires taking advantage of the properties of the ratios $e^{\Lambda_{u,T}(\xi)}/e^{\Lambda_{u,T+m}(\xi)}$. Which are not only uniformly bounded in T but also only dependent on the sequence $\xi = (\xi_{-\frac12 T}, \ldots, \xi_{\frac12 T})$ through a few symbols with labels near $-\frac12 T$ and $\frac12 T$: this correction can be considered part of the factors $h^T_{u,u}$ in the rigorous formula equation (A.7).

Note that $e^{\Lambda_{u,T}(\xi)}\Delta(\xi) = $ constant reflects Pesin's formula [28], for the approximate dynamics considered here.

So the SRB distribution arises naturally from assuming that dynamics can be discretized on a regular array of point ("microcells") and become a one cycle permutation of the microcells on the attractor. This is so under the CH and holds whether the dynamics is conservative (Hamiltonian) or dissipative.

Remark. It is well known that hyperbolic systems admit (uncountably) many invariant probability distributions, besides the SRB. This can be seen by noting that the space of the configurations is identified with a space of compatible sequences. On such a space one can define uncountably many stochastic processes, for instance by assigning an arbitrary short range translation invariant potential, and regarding the corresponding Gibbs state as a probability distribution on phase space. However the analysis just presented apparently singles out SRB as the unique invariant distribution. This is due to our assumption that, in the discretization, microcells are regularly spaced and centered on a regular discrete lattice and evolution eventually permutes them in a (single, by transitivity) cycle consisting of the microcells located on the attractor (and therefore locally evenly spaced, as inherited from the regularity of the phase space discretization).

Other invariant distributions can be obtained by custom made discretizations of phase space which will not

[15] To see this it is convenient to remark that the S^{t+m}-image of a layer $\delta(x) \subset \Delta(\xi)$ of the attractor will cover some of the layers of $\Delta(\xi)$, because $S^t\mathcal{E}(\xi)$ visits and fully covers all coarse cells $\mathcal{E}(\xi')$, see footnote[9]. Hence $S^{t+m}\Delta(\xi)$ will fully cover at least part of the layers of the attractor in $\mathcal{E}(\xi)$. Actually *it will cover the whole of* $\Delta(\xi)$, because if a layer of $\Delta(\xi)$ was left out then it will be left out of all the iterates of S^{t+m} and a nontrivial invariant subset of the attractor for S^t would exist.

cover the attractor in a regular way. This is what is done when other distributions, "not absolutely continuous with respect to the phase space volume", are to be studied in simulations. A paradigmatic example is given by the map $x \to 3x \bmod 1$: it has an invariant distribution attributing zero probability to the points x that, in base 3, lack the digit 2: to find it one has to write a program in which data have this property and make sure that the round off errors will not destroy it. Almost any "naive" code that simulates this dynamics using double precision reals represented in base 2 will generate, instead, the corresponding SRB distribution which is simply the Lebesgue measure on the unit interval (which is the Bernoulli process on the symbolic dynamics giving equal probability $\frac{1}{3}$ to each digit).

The physical representation of the SRB distribution just obtained, see [23,29], shows that there is no conceptual difference between stationary states in equilibrium and out of equilibrium. In both cases, if motions are chaotic they are permutations of microcells and the *SRB distribution is simply equidistribution over the recurrent microcells.* In equilibrium this gives the Gibbs microcanonical distribution and out of equilibrium it gives the SRB distribution (of which the Gibbs one is a very special case).

The above heuristic argument is an interpretation of the mathematical proofs behind the SRB distribution which can be found in [28,80] (and heuristically is a proof in itself). Once equation (A.7) is given, the expectation values of the observables in the SRB distributions can be formally written as sums over suitably small coarse cells and symmetry properties inherited from symmetries of the dynamic become transparent. The Fluctuation Theorem is a simple consequence of equation (A.7), see Appendix C: however it is conceptually interesting because of the surprising unification of equilibrium and nonequilibrium behind it.

The discrete representation, in terms of coarse grain cells and microcells leads to the possibility of counting the number $\mathcal{N}$ of the microcells and therefore to define a kind of entropy function: see [21] where the detailed analysis of the counting is performed and the difficulties arising in defining an entropy function as $k_B \log \mathcal{N}$ are critically examined.

Appendix C: Why does FT hold?

As mentioned the proof of FT in quite simple [26]. By the first of equations (A.5), (A.7) and by the theory of $1D$-short range Ising models, see [39] for details, the probability that p is in a small interval centered at p compared to the probability that it is in the opposite interval is

$$\frac{P_\tau(p)}{P_\tau(-p)} = \frac{\sum_{i\to p\sigma_+\tau} e^{-\sum_{-\tau/2}^{\tau/2} \Lambda_1^u(S^k\kappa_i)+B(i,\tau)}}{\sum_{i\to -p\sigma_+\tau} e^{-\sum_{-\tau/2}^{\tau/2} \Lambda_1^u(S^k\kappa_i)+B(i,\tau)}} \qquad \text{(C.1)}$$

where $\sum_{i\to p\sigma_+\tau}$ is sum over the centers κ_i of the rectangles E_i labeled by $i \stackrel{def}{=} (\xi_{-\tau/2},\ldots,\xi_{\tau/2})$ with the property

$$\sum_{k=-\tau/2}^{\tau/2} \sigma(S^k\kappa_i) + B(i,\tau) \simeq p\sigma_+\tau \qquad \text{(C.2)}$$

where $\simeq$ means that the left hand side is contained in a very small interval (of size of order $O(1)$ [39], call it b) centered at $p\sigma_+\tau$; the $B(i,\tau)$ is a term of order 1 (a boundary term in the language of the Ising model interpretation of the SRB distribution): $|B(i,\tau)| \le b < +\infty$: and it takes also into account the adjustments to be made because of the arbitrariness of the choice of κ_i[16]. Independence on i,τ of the bound on $B(i,\tau)$ reflects smoothness of S and elementary properties of short range $1D$ Ising chains [39].

Suppose that the symbolic dynamics has been chosen time reversible, i.e. the time reversal map I maps P_i into $IP_i = P_{I(i)}$ for some $I(i)$: this is not a restriction as discussed in Appendix A. Then the above ratio of sums can be rewritten as a ratio of sums over the same set of labels,

$$\frac{P_\tau(p)}{P_\tau(-p)} = \frac{\sum_{i\to p\sigma_+\tau} e^{-\sum_{-\tau/2}^{\tau/2} \Lambda_1^u(S^k\kappa_i)+B(i,\tau)}}{\sum_{i\to p\sigma_+\tau} e^{-\sum_{-\tau/2}^{\tau/2} \Lambda_1^u(S^k I(\kappa_i))+B(I(i),\tau)}}. \qquad \text{(C.3)}$$

Remark that $\Lambda_1^u(Ix) = -\Lambda_1^s(x)$ (by time reversal symmetry) and that (by Eq. (A.3)) $\sum_{k=-\tau/2}^{\tau/2}(\Lambda_1^u(S^k(x)) + \Lambda_1^s(S^{-k}(x)))$ can be written as

$$\sum_{k=-\tau/2}^{\tau/2} (\Lambda_1^u(S^k(x)) + \Lambda_1^s(S^k(x))) = \sum_{k=-\tau/2}^{\tau/2} \sigma(S^k x) + B(x,\tau) \qquad \text{(C.4)}$$

with $B(x,\tau) \le b$ (again by the smoothness of S), possibly redefining b.

Therefore the ratio of corresponding terms in the numerator and denominator (i.e. terms bearing the same summation label i) is precisely $p\sigma_+\tau$ up to $\pm 3b$. Hence

$$e^{\tau\sigma_+ p - 3b} \le \frac{P_\tau(p)}{P_\tau(-p)} < e^{\tau\sigma_+ p + 3b} \qquad \text{(C.5)}$$

so that FT holds for finite τ with an error $\pm\frac{3b}{\tau}$, infinitesimal as $\tau \to +\infty$. For a detailed discussion of the error bounds see [39].

Of course for all this to make sense the value of p must be among those which not only are possible but

[16] Which is taken here $\kappa_i =$ the center of P_{ξ_0}, but which could equivalently made by choosing other points in E_ξ, for instance by continuing the string $i = (\xi_{-\tau/2},\ldots,\xi_{\tau/2})$ to the right and to the left, according to an a priori fixed rule depending only on $\xi_{\tau/2}$ and $\xi_{-\tau/2}$ respectively. Thus turning it to a biinfinite compatible string ξ_i which therefore fixes a new point κ_i'.

also such that the values close enough to possible values are possible. This means that p has to be an internal point to an interval of values that contains limit points of $\lim_{\tau\to+\infty}\frac{1}{\tau}\sum_{k=0}^{\tau}\frac{\sigma(S^k x)}{\sigma_+}$ for a set of x's with positive SRB probability: the value p^* in FT is the supremum among the value of p with this property, [39] (contrary to statements in the literature this physically obvious remark is explicitly present in the original papers: and one should not consider the three contemporary references [26,39,40], has having been influenced by the doubts on this point raised much later.)

The assumptions have been: (a) existence of a Markovian partition, i.e. the possibility of a well controlled symbolic dynamics representation of the motion; (b) smooth evolution S and (c) smooth time reversal symmetry: the properties (a), (b) are equivalent to the CH. Of course positivity of σ_+ is *essential*, in spite of contrary statements; if $\sigma_+ = 0$ the leading terms would come from what has been bounded in the remainder terms and, in any event the analysis world be trivial, with or without chaoticity assumptions [64].

Since Lorenz [81], symbolic dynamics is employed to represent chaos and many simulations make currently use of it; smoothness has always been supposed in studying natural phenomena (lack of it being interpreted as a sign of breakdown of the theory and of necessity of a more accurate one); time reversal is a fundamental symmetry of nature (realized as T or TCP in the Physics notations). Hence in spite of the ease in exhibiting examples of systems which are not smooth, not hyperbolic, not time reversal symmetric (or any subset thereof) the CH still seems a good guide to understand chaos.

Appendix D: Harmonic thermostats

Here the "efficiency" of a harmonic thermostat is discussed. It turns out that in general a thermostat consisting of infinite free systems is a very simple kind of Hamiltonian thermostat, but it has to be considered with caution as it can be inefficient in the sense that it might not drive a system towards equilibrium (i.e. towards a Gibbs distribution). In the example given below a system in interaction with an infinite harmonic reservoir at inverse temperature β is considered. It is shown that the interaction can lead to a stationary state, of the system plus reservoir, which is not the Gibbs state at temperature β^{-1}. The following is a repetition of the analysis in [18], adapting it to the situation considered here.

A simple model is a 1-dimensional harmonic oscillators chain, of bosons or fermions, initially in a Gibbs state at temperature β^{-1}. The Hamiltonian for the equilibrium *initial* state will be

$$H_0 = \sum_{x=1}^{N-1} -\frac{\hbar^2}{2m}\Delta_{q_x} + \sum_{x=1}^{N-1}\frac{m\omega^2}{2}q_x^2 + \sum_{x=1}^{N}\frac{m\mu^2}{2}(q_x-q_{x-1})^2 \qquad \text{(D.1)}$$

with boundary conditions $q_0 = q_N = 0$ and $\hbar, m, \omega^2, \mu^2 > 0$. The initial state will be supposed to have a density matrix $\rho_0 = \frac{e^{-\beta H_0}}{\mathrm{Tr}\, e^{-\beta H_0}}$. Time evolution will be governed by a *different* Hamiltonian

$$H_\lambda = H_0 + \frac{m\lambda}{2}q_1^2, \qquad \lambda+\omega^2>0. \qquad \text{(D.2)}$$

The question of "thermostat efficiency" is: does $\rho_t \overset{def}{=} e^{\frac{i}{\hbar}tH_\lambda}\rho_0 e^{-\frac{i}{\hbar}tH_\lambda}$ converge as $t\to+\infty$ to $\rho_\infty = \frac{e^{-\beta H_\lambda}}{\mathrm{Tr}\, e^{-\beta H_\lambda}}$. Or: does the system consisting in the oscillators labeled $2,3,\ldots$ succeed in bringing up to the new equilibrium state the oscillator labeled 1? Convergence means that the limit $\langle A\rangle_{\rho_t} \xrightarrow[t\to+\infty]{} \langle A\rangle_{\rho_\infty}$ exists, at least for the observables A essentially localized in a finite region.

The Hamiltonian in equation (D.2) can be diagonalized by studying the matrix

$$V_\lambda = m\begin{pmatrix}\omega^2+2\mu^2+\lambda & -\mu^2 & 0 & \ldots\\ -\mu^2 & \omega^2+2\mu^2 & -\mu^2 & \ldots\\ 0 & -\mu^2 & \omega^2+2\mu^2 & \ldots\\ \ldots & \ldots & \ldots & \ldots\end{pmatrix}$$

$$\overset{def}{=} V_0 + \lambda m P_1. \qquad \text{(D.3)}$$

The normalized eigenstates and respective eigenvalues of V_0 are

$$\Psi_k^0(x) \overset{def}{=} \sqrt{\frac{2}{N}}\sin\frac{\pi k}{N}x,$$
$$\Lambda_k^0 = m\left(\omega^2+2\mu^2\left(1-\cos\frac{\pi k}{N}\right)\right) \qquad \text{(D.4)}$$

and the vectors Ψ_k^0 will be also denoted $|k\rangle$ or $|\Psi_k^0\rangle$.

To solve the characteristic equation for V_λ, call Ψ a generic normalized eigenvector with eigenvalue Λ; the eigenvalue equation is

$$\langle k|\Psi\rangle(\Lambda_k^0-\Lambda) + \lambda\, m\,\langle k|\Omega\rangle\,\langle\Omega|\Psi\rangle = 0 \qquad \text{(D.5)}$$

where Ω is the vector $\Omega = (1,0,\ldots,0)\in\mathcal{C}^{N-1}$, so that $P_1 = |\Omega\rangle\langle\Omega|$. Hence, noting that $\langle\Omega|\Psi\rangle$ cannot be 0 because this would imply that $\Lambda=\Lambda_k^0$ for some k and therefore $|\Psi\rangle = |k\rangle$ which contradicts $\langle\Omega|\Psi\rangle = 0$, it is

$$\langle k|\Psi\rangle = -\lambda\, m\,\frac{\langle k|\Omega\rangle\langle\Omega|\Psi\rangle}{\Lambda_k^0-\Lambda} \qquad \text{(D.6)}$$

and the compatibility condition that has to be satisfied is

$$\frac{\langle\Omega|\Psi\rangle}{\lambda\, m} = \sum_{k=1}^{N-1}\frac{|\langle\Omega|k\rangle|^2}{\Lambda-\Lambda_k^0}\langle\Omega|\Psi\rangle = \sum_{k=1}^{N-1}\frac{2\sin^2\frac{\pi k}{N}}{N}\frac{\langle\Omega|\Psi\rangle}{\Lambda-\Lambda_k^0}. \qquad \text{(D.7)}$$

Once equation (D.7) is satisfied, equation (D.6) imply that the eigenvalue equation, equation (D.5), is satisfied, and by a $|\Psi\rangle\neq 0$ (determined up to a factor).

Equation (D.7) has $N-1$ solutions, corresponding to the $N-1$ eigenvalues of V_λ. This follows by comparing the graph of $y(\Lambda) \equiv \frac{1}{\lambda m}$ with the graph of the function of Λ in the intermediate term of equation (D.7). One of the solutions remains isolated in the limit $N \to \infty$, because the equation

$$1 = \frac{2\,\lambda\, m}{\pi} \int_0^\pi \frac{\sin^2 \kappa}{\Lambda - \Lambda^0(\kappa)} d\kappa,$$
$$\Lambda^0(\kappa) \stackrel{def}{=} m\left(\omega^2 + 4\mu^2 \sin^2 \frac{\kappa}{2}\right) \tag{D.8}$$

has, uniformly in N, only one isolated solution for $\Lambda < \inf \Lambda^0(\kappa) = m\omega^2$ if $\lambda < 0$, or for $\Lambda > \sup \Lambda^0(\kappa)$ if $\lambda < 0$. Suppose for definiteness that $\lambda < 0$.

Let $\Psi_k^\lambda(x)$, $k = 1, \ldots, N-1$, be the corresponding eigenfunctions. The matrices $U_{\lambda;k,x} = \Psi_k^\lambda(x)$ are unitary and $(U_\lambda)_{\lambda=0} \equiv U_0$. It is $U_{0;k,x} = \sqrt{\frac{2}{N}} \sin \frac{\pi k}{N} x$ and $\langle \Psi_k^0 | \Psi_{k'}^\lambda \rangle = \frac{\langle k|\Omega\rangle}{Z_N(k')(\Lambda_{k'}^\lambda - \Lambda_k^0)}$ with $Z_N(k')^2 = \sum_k \frac{|\langle k|\Omega\rangle|^2}{(\Lambda_{k'}^\lambda - \Lambda_k^0)^2}$ by equation (D.6). Then setting $\alpha_x^\pm = \frac{p_x \pm i q_x}{\sqrt{2}}$ let

$$a_{\lambda;k}^+ \stackrel{def}{=} (U_\lambda \alpha^+)_k, \qquad a_{\lambda;k}^- = \stackrel{def}{=} (\alpha^- U_\lambda^*)_k \tag{D.9}$$

where U^* is the adjoint of U (so that $UU^* = 1$ if U is unitary). It is

$$\alpha_x^+ = \sum_k \overline{U}_{\lambda;k,x} a_{\lambda;k}^+, \qquad a_{\lambda;k}^+ = \sum_{h,y} U_{\lambda;k,y} \overline{U}_{0;h,y} a_{0;h}^+ \tag{D.10}$$

if the overbars denote complex conjugation.

The operators $a_{\lambda,k}^\pm$ will be creation and annihilation operators for quanta with energy $\hbar\sqrt{\frac{\Lambda_k^\lambda}{m}} \stackrel{def}{=} E_\lambda(k)$. So a state with $n_k = 0, 1, \ldots$ quanta in state k will have energy $\sum_k E_\lambda(k)(n_k + \frac{1}{2})$.

Consider the observable $a_{\lambda,1}^+ a_{\lambda,1}^- = A$. Its average is *time independent*, in the evolution generated by H_λ, and if $W \stackrel{def}{=} U_\lambda U_0^*$ it is equal to

$$\begin{aligned}\langle A \rangle_{\rho_t} &\equiv \langle A \rangle_{\rho_0} \equiv \operatorname{Tr} \rho_0 (W \mathbf{a}_0^+)_1 (W \mathbf{a}_0^-)_1 \\ &= \sum_k \operatorname{Tr} \rho_0 W_{1,k} W_{1,k'} a_{0,k}^+ a_{0,k'}^- \\ &= \sum_{k=1}^{N-1} |W_{1,k}|^2 \frac{\sum_{n=0}^{n_f} e^{-\beta E_0(k) n}\, n}{\sum_{n=0}^{n_f} e^{-\beta E_0(k) n}}\end{aligned} \tag{D.11}$$

where $n_f = 1$ if the statistics of the quanta is fermionic (this was the case in [18]) or $n_f = +\infty$ if it is bosonic. In the two cases the result is

$$\sum_k |W_{1,k}|^2 \frac{1}{e^{\beta E_0(k)} \pm 1}. \tag{D.12}$$

If the system reached thermal equilibrium, setting $\rho_\lambda(k) \stackrel{def}{=} \frac{1}{e^{\beta E_\lambda(k)} \pm 1}$, this should be $\rho_\lambda(1)$, which is impossible, as it can be checked by letting $\beta \to +\infty$ and remarking that it is $E_\lambda(1) < E_0(1)$ with a difference positive uniformly in N. Furthermore the observable A is localized near the site $x = 1$: because the wave function of the lowest eigenvalue is $\frac{1}{Z_N(1)} \sum_h \frac{\langle h|\Omega\rangle}{\Lambda_h^0 - \Lambda_k^\lambda} |\Psi_h^0\rangle$ so that

$$\Psi_1^\lambda(x) = \frac{1}{Z_N(1)} \sum_h \frac{\Psi_h^0(1)\Psi_h^0(x)}{\Lambda_1^\lambda - \Lambda_h^0} \xrightarrow[N\to\infty]{} \frac{1}{Z_\infty} \frac{2}{\pi} \int_0^\pi \frac{\sin\kappa \sin\kappa x}{\Lambda_1^\lambda - \Lambda^0(\kappa)} d\kappa \tag{D.13}$$

and the integral tends to 0 as $x \to \infty$ faster than any power, so that $0 < Z_\infty < \infty$ and Ψ_1^λ is normalizable.

Therefore the thermostatic action of the system in the sites $2, 3, \ldots$ on the site 1 is *not efficient* and the state does not evolve towards the Gibbs state at temperature β^{-1}, not even in the limit $N \to +\infty$.

This result should be contrasted with the closely related case in which the system oscillator at 1 plus the others is started in a equilibrium state for H_λ and at time 0 is evolved with Hamiltonian H_0. In this case the system thermalizes properly, see the analogous analysis in [18], see also [14] for a large class of related examples.

Of course the question of effectiveness of a thermostat could be discussed also for non linear theormostats, finite or infinite. It seems that, under mild assumptions, non linear thermostat models should be efficient, i.e. generate proper heat exchanges even when acting only at the boundary as in the case of the thermostats considered in Section 9. The analysis in [82] gives some preliminary evidence in this direction.

Harmonic thermostats are nevertheless very interesting, provided the above pathologies are excluded by a careful formulation of the models: see for instance [14], see also [17]. It is also clear that the pathologies seem to be related to the fact that the thermostats constituents are "not interacting" or "linearly interacting": their origin in the above analysis is shown to be related to the existence of isolated eigenvalues of the Hamiltonian at the bottom of the spectrum and this is the property that should be excluded. The pathologies are likely to be absent in models in which there is nonlinear interaction within the thermostats constituents so that such models should be perfectly well behaving (i.e. efficient in the sense of this paper). However the latter models are also highly nontrivial even at a purely mathematical level.

Appendix E: Bohmian quantum systems

Consider the system in Figure 1 and suppose, as in Section 10, that the nonconservative force $\mathbf{E}(\mathbf{X}_0)$ acting on the system vanishes, i.e. consider the problem of heat flow through $\mathcal{C}_0$. Let H be the operator on $L_2(\mathcal{C}_0^{3N_0})$, space of symmetric or antisymmetric wave functions Ψ,

$$H_{\mathbf{X}} = -\frac{\hbar^2}{2m}\Delta_{\mathbf{X}_0} + U_0(\mathbf{X}_0) + \sum_{j>0} \left(U_{0j}(\mathbf{X}_0, \mathbf{X}_j) + U_j(\mathbf{X}_j) + K_j\right) \tag{E.1}$$

where $\Delta_{\mathbf{X}_0}$ is the Laplacian, and note that its spectrum consists of eigenvalues $E_n = E_n(\{\mathbf{X}_j\}_{j>0})$, depending on the configuration $\mathbf{X} \stackrel{def}{=} \{\mathbf{X}_j\}_{j>0}$,

Thermostats will be modeled as assemblies of classical particles as in Section 9: thus their temperature can be defined as the average kinetic energy of their particles and the question of how to define it does not arise.

The viewpoint of Bohm on quantum theory seems quite well adapted to the kind of systems considered here. A system–reservoirs model can be the *dynamical system* on the variables $\big(\Psi, \mathbf{X}_0, (\{\mathbf{X}_j\}, \{\dot{\mathbf{X}}_j\})_{j>0}\big)$ defined by

$$\begin{aligned}
-i\hbar\dot{\Psi}(\mathbf{X}_0) &= (H_{\mathbf{X}}\Psi)(\mathbf{X}_0),\\
\dot{\mathbf{X}}_0 &= \hbar \mathrm{Im}\ \frac{\boldsymbol{\partial}_{\mathbf{X}_0}\Psi(\mathbf{X}_0)}{\Psi(\mathbf{X}_0)}, \qquad \text{and for } j>0\\
\ddot{\mathbf{X}}_j &= -\Big(\partial_j U_j(\mathbf{X}_j) + \partial_j U_j(\mathbf{X}_0,\mathbf{X}_j)\Big) - \alpha_j \dot{\mathbf{X}}_j\\
\alpha_j &\stackrel{def}{=} \frac{W_j - \dot{U}_j}{2K_j}, \qquad W_j \stackrel{def}{=} -\dot{\mathbf{X}}_j \cdot \partial_j U_{0j}(\mathbf{X}_0,\mathbf{X}_j)
\end{aligned} \tag{E.2}$$

here the first equation is Schrödinger's equation, the second is the velocity of the Bohmian particles carried by the wave Ψ, the others are equations of motion for the thermostats particles analogous to the one in equation (9.1), (whose notation for the particles labels is adopted here too). Evolution maintains the thermostats kinetic energies $K_j \equiv \frac{1}{2}\dot{\mathbf{X}}_j^2$ exactly constant so that they will be used to define the thermostats temperatures T_j via $K_j = \frac{3}{2}k_B T_j N_j$, as in the classical case.

Note that if there is no coupling between system and thermostats, i.e. the system is "isolated", then there are many invariant distributions: e.g. the probability distributions μ proportional to

$$\sum_{n=1}^{\infty} e^{-\beta_0 E_n}\delta(\Psi - \Psi_n\, e^{i\varphi_n})\, |\Psi(\mathbf{X}_0)|^2 d\varphi_n d\mathbf{X}_0 \times \prod_j \delta(\dot{\mathbf{X}}_j^2 - 2K_j)d\dot{\mathbf{X}}_j d\mathbf{X}_j \tag{E.3}$$

where E_n and Ψ_n are time independent, under the assumed absence of interaction between system and thermostats, and are the eigenvalues and the corresponding eigenvectors of H. Then the distributions μ are invariant under the time evolution.

Time invariance of this kind of distributions is discussed in [83], Section 4, where it appears as an instance of what is called there a "quantum equilibrium". The average value of an observable $O(\mathbf{X}_0)$, which depends only on position $\mathbf{X}_0$, will be the "usual" Gibbs average

$$\langle O\rangle_\mu = Z^{-1}\int \mathrm{Tr}\left(e^{-\beta_0 H} O\right). \tag{E.4}$$

For studying nonequilibrium stationary states consider several thermostats with interaction energy with $\mathcal{C}_0$, $W_j(\mathbf{X}_0,\mathbf{X}_j)$, as in equation (9.1). The equations of motion should be equation (E.2)

In general solutions of equation (E.2) *will not be quasi periodic* and the Chaotic Hypothesis [23,40,58], can be assumed: if so the dynamics should select an invariant distribution μ. The distribution μ will give the statistical properties of the stationary states reached starting the motion in a thermostat configuration $(\mathbf{X}_j, \dot{\mathbf{X}}_j)_{j>0}$, randomly chosen with "uniform distribution" ν on the spheres $m\dot{\mathbf{X}}_j^2 = 3N_j k_B T_j$ and in a random eigenstate of H. The distribution μ, if existing and unique, could be named the *SRB distribution* corresponding to the chaotic motions of equation (E.2).

In the case of a system *interacting with a single thermostat* the latter distribution should be equivalent to the canonical distribution. As in Section 11 an important consistency check for the model just proposed in equation (E.2) is that there should exist at least one stationary distribution μ equivalent to the canonical distribution at the appropriate temperature T_1 associated with the (constant) kinetic energy of the thermostat: $K_1 = \frac{3}{2}k_B T_1 N_1$. However also in this case, as already in Section 11, it does not seem possible to define a simple invariant distribution, not even in the adiabatic approximation. As in Section 11, equivalence between μ and a Gibbs distribution at temperature T_1 can only be conjectured.

Furthermore it is not clear how to define phase space contraction, hence how to formulate a FT, although the equations are reversible.

References

1. S.G. Brush, *History of modern physical sciences: The kinetic theory of gases* (Imperial College Press, London, 2003)
2. M.W. Zemansky, *Heat and thermodynamics* (McGraw-Hill, New-York, 1957)
3. R.P. Feynman, R.B. Leighton, M. Sands, *The Feynman lectures in Physics* (Addison-Wesley, New York, 1963), Vol. I, II, III
4. L. Boltzmann, *Über die mechanische Bedeutung des zweiten Haupsatzes der Wärmetheorie*, Volume 1, p. 9 of *Wissenschaftliche Abhandlungen*, edited by F. Hasenöhrl (Chelsea, New York, 1968)
5. L. Boltzmann, *Über die Eigenshaften monozyklischer und anderer damit verwandter Systeme*, Volume 3, p. 122 of *Wissenschaftliche Abhandlungen* (Chelsea, New-York, 1968)
6. G.N. Bochkov, Yu.E. Kuzovlev, Physica A **106**, 443 (1981)
7. C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997)
8. C. Jarzynski, J. Stat. Phys. **98**, 77 (1999)
9. D.J. Evans, D.J. Searles, Phys. Rev. E **50**, 1645 (1994).
10. G.E. Crooks, Phys. Rev. E **60**, 2721 (1999)
11. E.G.D. Cohen, G. Gallavotti, J. Stat. Phys. **96**, 1343 (1999)
12. G. Gallavotti, *Entropy, nonequilibrium, chaos and infinitesimals*, e-print `arXiv:cond-mat/0606477` (2006)
13. R.P. Feynman, F.L. Vernon, Ann. Phys. **24**, 118 (1963)
14. K. Hepp, E. Lieb, Phys. Rev. A **8**, 2517 (1973)
15. J.P. Eckmann, C.A. Pillet, L.R. Bellet, Comm. Math. Phys. **201**, 657 (1999)

16. W. Aschbacher, Y. Pautrat, V. Jakšić, C.A. Pillet, Lect. Notes Math. **1882**, 1 (2006)
17. P. Hänggi, G.L. Ingold, Chaos **15**, 026105 (2005)
18. D. Abraham, E. Baruch, G. Gallavotti, A. Martin-Löf, Studies Appl. Math. **51**, 211 (1972)
19. E. Barouch, M. Dresden, Phys. Rev. Lett. **23**, 114 (1969)
20. J.L. Lebowitz, Physics Today 32 (1993)
21. G. Gallavotti, Comm. Math. Phys. **224**, 107 (2001)
22. D. Ruelle, Comm. Math. Phys. **189**, 365 (1997)
23. G. Gallavotti, *Statistical Mechanics. A short treatise* (Springer Verlag, Berlin, 2000)
24. D. Ruelle, Progress in Theoretical Physics Supplement **64**, 339 (1978)
25. D. Ruelle, *Ergodic theory*, Volume Suppl X of *The Boltzmann equation*, edited by E.G.D Cohen, W. Thirring, Acta Physica Austriaca (Springer, New York, 1973)
26. G. Gallavotti, E.G.D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995)
27. D.J. Evans, E.G.D. Cohen, G.P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993)
28. G. Gallavotti, F. Bonetto, G. Gentile, *Aspects of the ergodic, qualitative and statistical theory of motion* (Springer Verlag, Berlin, 2004)
29. G. Gallavotti, J. Stat. Phys. **78**, 1571 (1995)
30. Ya.G. Sinai, Russian Math. Surveys **27**, 21 (1972)
31. R. Bowen, D. Ruelle, Invent. Math. **29**, 181 (1975)
32. Ya.G. Sinai, *Lectures in ergodic theory,* Lecture notes in Mathematics (Princeton University Press, Princeton, 1977)
33. W. Thomson, in Proceedings of the Royal Society of Edinburgh **8**, 325 (1874)
34. L. Boltzmann, *Einige allgemeine sätze über Wärmegleichgewicht*, Volume 1, p. 259 of *Wissenschaftliche Abhandlungen*, edited by F. Hasenöhrl (Chelsea, New York, 1968)
35. L. Boltzmann, *Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten*, Volume 1, p. 49 of *Wissenschaftliche Abhandlungen*, edited by F. Hasenöhrl (Chelsea, New York, 1968)
36. H. Helmholtz, *Prinzipien der Statistik monocyklischer Systeme*, Volume III of *Wissenschaftliche Abhandlungen* (Barth, Leipzig, 1895)
37. H. Helmholtz, *Studien zur Statistik monocyklischer Systeme*, Volume III of *Wissenschaftliche Abhandlungen* (Barth, Leipzig, 1895)
38. L. Boltzmann, *Reply to Zermelo's Remarks on the theory of heat*, Volume 1, p. 392 of *History of modern physical sciences: The kinetic theory of gases*, edited by S. Brush (Imperial College Press, London, 2003)
39. G. Gallavotti, Math. Phys. Electronic J. (MPEJ) **1**, 1 (1995)
40. G. Gallavotti, E.G.D. Cohen, J. Stat. Phys. **80**, 931 (1995)
41. G. Gallavotti, E.G.D. Cohen, Phys. Rev. E **69**, 035104 (2004)
42. A. Giuliani, F. Zamponi, G. Gallavotti, J. Stat. Phys. **119**, 909 (2005)
43. G. Gentile, Forum Math. **10**, 89 (1998)
44. G. Gallavotti, Phys. Rev. Lett. **77**, 4334 (1996)
45. D. Ruelle, J. Stat. Phys. **95**, 393 (1999)
46. G. Gallavotti, Annales de l'Institut H. Poincaré **70**, 429 (1999) and `chao-dyn/9703007`
47. G. Gallavotti, Open Systems and Information Dynamics **6**, 101 (1999)
48. R. Chetrite, J.Y. Delannoy, K. Gawedzki, J. Stat. Phys. **126**, 1165 (2007)
49. N.I. Chernov, G.L. Eyink, J.L. Lebowitz, Ya.G. Sinai, Comm. Math. Phys. **154**, 569 (1993)
50. F. Bonetto, G. Gallavotti, P. Garrido, Physica D **105**, 226 (1997)
51. G. Gallavotti, Chaos **16**, 043114 (2006)
52. R. Becker, *Electromagnetic fields and interactions* (Blaisdell, New-York, 1964)
53. G. Gallavotti, *Foundations of Fluid Dynamics* (second printing) (Springer Verlag, Berlin, 2005)
54. G. Gallavotti, Physica D **105**, 163 (1997)
55. L. Rondoni, E. Segre, Nonlinearity **12**, 1471 (1999)
56. S. de Groot, P. Mazur, *Non equilibrium thermodynamics* (Dover, Mineola, NY, 1984)
57. G. Gallavotti, Chaos **16**, 023130 (2006)
58. G. Gallavotti, *Quantum nonequilibrium and entropy creation*, e-print `arXiv:cond-mat/0701124`, Unpublished, 2007
59. D.J. Evans, G.P. Morriss, *Statistical Mechanics of Nonequilibrium Fluids* (Academic Press, New-York, 1990)
60. L.F. Cugliandolo, J. Kurchan, L. Peliti, Phys. Rev. E 2898 (1997)
61. A. Crisanti, F. Ritort, J. Phys. A R181 (2003)
62. S. Lepri, R. Livi, A. Politi, Physica D **119**, 140 (1998)
63. R. Van Zon, E.G.D. Cohen, Phys. Rev. Lett. **91**, 110601 (2003)
64. F. Bonetto, G. Gallavotti, A. Giuliani, F. Zamponi, J. Stat. Phys. **123**, 39 (2006)
65. F. Bonetto, G. Gallavotti, G. Gentile, Ergodic Theory and Dynamical Systems, **28**, 21–47 (2008); `doi: 10.1017/S0143385707000417`
66. M. Bandi, J.R. Cressman, W. Goldburg, J. Stat. Phys. **130**, 27–38 (2008); `doi: 10.1007/s10955-007-9355-4`
67. F. Bonetto, G. Gallavotti, A. Giuliani, F. Zamponi, J. Stat. Mech. P05009 (2006)
68. J. Kurchan, J. Phys. A **31**, 3719 (1998).
69. J. Lebowitz, H. Spohn, J. Stat. Phys. **95**, 333 (1999)
70. C. Maes, J. Stat. Phys. **95**, 367 (1999)
71. J. Kurchan, A quantum fluctuation theorem, e-print `arXiv:cond-mat/0007360`, unpublished (2000)
72. V. Jakšić, C.A. Pillet, Comm. Math. Phys. **226**, 131 (2002)
73. V. Jakšić, Y. Ogata, C.A. Pillet, Ann. Henri Poincaré (2007)
74. J.R. Cressman, J. Davoudi, W.I. Goldburg, J. Schumacher, New J. Phys. **6**, 53 (2004)
75. F. Bonetto, G. Gallavotti, Comm. Math. Phys. **189**, 263 (1997)
76. F. Zamponi, *Is it possible to experimentally verify the fluctuation relation? A review of theoretical motivations and numerical evidence*, e-print `arXiv:cond-mat/0612019` (2006)
77. G. Gallavotti, Chaos **14**, 680 (2004)
78. R. Bowen, in *Proceedings of the American Mathematical Society* **71**, 130 (1978)
79. D. Levesque, L. Verlet, J. Stat. Phys. **72**, 519 (1993)
80. R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeormorphisms*, Volume 470 of Lecture Notes in Mathematics (Springer-Verlag, Berlin-Heidelberg, 1975)
81. E. Lorenz, J. Atmospheric Sci. **20**, 130 (1963)
82. P. Garrido, G. Gallavotti, J. Stat. Phys. **126**, 1201 (2007)
83. D. Dürr, S. Goldstein, N. Zanghì. J. Stat. Phys. **68**, 259 (1992)