Nonequilibrium Thermodynamics

Giovanni Gallavotti

November 15, 2018

I. EQUILIBRIUM

Personne n'ignore que la chaleur peut être la cause du mouvement, qu'elle possède même une grande puissance motrice: les machines à vapeur, aujourd'hui si répandues, en sont une preuve parlant à tous les yeux.

Equilibrium thermodynamics was born in the early 1800's. On the basis of the experimental results on the properties of rarefied gases (like the Boyle-Mariotte law) Carnot develops, [1], his theorem on the ideal efficiency of machines operating by extracting or damping heat in two reservoirs and converting it into work. His long, very careful and detailed, analysis is presented after a brief introduction (10 pages), starting with the above words, in which present and future usefulness and importance of the vapor machines is enthusiastically (and optimistically) outlined (1824).

The theorem shows that the most efficient machines must operate running in a reversible cycle in which the vapor (of water, air, alcohol or other gases) evolves through a sequence $\mathcal{P}$ of equilibrium states;[1] and this remains true even if the vapor is replaced by a liquid or a solid.

It is an example of what today we call a "universal law", *i.e.* a law that applies to a very large class of systems isolating, among their properties, a few that they all verify in a quantitative form *without adjustable parameters.*

An immediate consequence is the possibility of defining the absolute temperature of a heat reservoir: it is simply defined in terms of the maximum efficiency of a machine operating between the reservoir of interest and a fixed reservoir to which a conventional temperature value is attributed.[2]

A few years later Krönig, [2], established the proportionality of the absolute temperature to the average kinetic energy and Clausius, [3], wrote the first of the works leading to *entropy*, 1850, whose existence constitutes the *second law* of thermodynamics and is implied by Carnot's theorem.

The meaning of the word "entropy" was explained by Clausius himself [4, p.390]:
"*I propose to name the quantity S the entropy of the system, after the Greek word ἡ τροπή "the transformation",* [in German, Verwandlung]. *I have deliberately chosen the word entropy to be as similar as possible to the word energy: the two quantities to be named by these words are so closely related in physical significance that a certain similarity in their names appears to be appropriate.*"

The notion of entropy became quickly fundamental for the theory and applications of equilibrium thermodynamics and almost identified with it: soon it was accompanied by the question of which would be its definition in terms of the atomistic representation of matter.

And in 1866 Boltzmann, [5], proposed to link it to the purely mechanical Maupertuis's *Action Principle*, imagining that atoms were moving on periodic orbits spanning all phase space points compatible with the mechanical conservation laws.[3]

This was a first attempt towards the formulation of the *ergodic hypothesis* and two years later, [8], led to the description of the statistical properties of a system in equilibrium via what are now called *microanonical and canonical ensembles.* The consequences developed by Boltzmann in 1868 were recognized by Maxwell (1879) who commented them [9],[4] and by Gibbs, [10].

Entropy S_A is defined for every equilibrium state A of a given system and is useful to establish the balance between heat exchanged by a system with the surrounding thermostats[5] and work performed on the surrounding mechanical devices.

A key property is that if a transformation through a "path" $\mathcal{P}$ of successive equilibrium states is *reversible* and leads from an equilibrium state A to an equilibrium state B then the the entropy variation is $S_B - S_A = \int_{\mathcal{P}} \frac{dQ}{T}$, if dQ is the amount of heat that the system receives while in contact with a heat reservoir at absolute temperature T, whatever the reversible path $\mathcal{P}$ is. This allows to set the value of the entropy of a generic equilibrium state, determining it up to an additive constant.

However entropy is important also in irreversible processes $\mathcal{I}$: it was established by Clausius that in such a process $S_B - S_A \ge \int_{\mathcal{I}} \frac{dQ}{T}$, where T is the absolute temperature of the reservoir from which the system receives the amount dQ of heat.

In particular if the transformation $A \to A$ is along an irreversible path $\mathcal{I}$ it is $\oint_{\mathcal{I}} \frac{dQ}{T} \le 0$. Also if $\mathcal{I}$ is an irreversible adiabatic path (*i.e.* a sequence of transformations with no heat exchange) it cannot lead from A to

---

[1] Reversible transformations were essential in Carnot's analysis, and he carefully insists to make clear the subtle argument that permits to avoid regarding their definition, requiring for instance a difference in temperature which "can be considered as vanishing", an oxymoron: "À la vérité, les choses ne peuvent pas se passer rigoureusement comme nons l'avons supposé ..., [1, p.13-14].

[2] *e.g.* if the reservoir is water at its triple point then in the Kelvin scale the absolute temperature is fixed to be $T_0 = 273.16^o K$. The temperature T_1 of another reservoir at higher temperature (say) is then given, in principle, by running a reversible machine between the two reservoirs and deriving T_1 so that the efficiency is $1 - \frac{T_0}{T_1}$.

[3] The idea was again proposed four years later by Clausius, [6, 7].

[4] Interestingly *without even mentioning* the H-theorem, discovered in the meantime by Boltzmann.

[5] *i.e.* large bodies whose temperature and volume variations are neglegible during the observation time

A, and necessarily leads from A to a state B with larger entropy.

If a system evolves first from a state A to B at constant temperature T_2 receiving a quantity of heat Q_2, then evolves adiabatically from B to B' at temperature T_1, then to A' at the temperature T_1 ejecting a quantity of heat Q_1 and then adiabatically back to A it will be

$$\frac{Q_2}{T_2} - \frac{Q_1}{T_1} \le 0, \quad \text{which means :} \ \frac{Q_2 - Q_1}{Q_2} \le \frac{T_2 - T_1}{T_2}$$

Since $Q_2 - Q_1$ must be, by energy conservation, the work performed by the system in the cycle, the inequality means that the efficiency in the transformation of heat into work cannot exceed $\frac{T_2-T_1}{T_2}$, and it can equal it if the cycle is a reversible path: which is Carnot's theorem formulated in terms of Clausius' entropy properties.

The continuous use of the second law, hence of Carnot's theorem, led to a long debate on the ergodic hypothesis, on entropy, and on the resolution of the antimony between reversible microscopic dynamics and irreversible macroscopic evolutions, which in some respects is still going on.

II. MICROSCOPIC STATIONARITY

In recent times there has been a widespread interest in extending Thermodynamics to a theory dealing with stationary states: these are states of systems on which steady non-conservative forces may act, keeping currents flowing through the systems: and currents can be of various kinds like transporting matter, heat, electric charge, ...

The stationary states are a natural generalization of the equilibrium states (which are very special cases of stationary states) and, of course, a main question is whether general, system independent, properties can be assigned to such states and be useful in studying their properties.

To understand the recent developments it is essential to keep in mind that the study of a physical system starts from an initial datum X where X is a point in "*phase space*" $\mathcal{F}$, often determined by the $6N$ coordinates of the N molecules or atoms. And the point X is generated via suitable experimental devices (it has become common to say "*following a prefixed protocol*") and is a *random set of coordinates*, because of the many unavoidable uncontrolled actions influencing the devices.

The probability of selecting the initial datum X in $\mathcal{F}$ is essentially always unknown, but it is *assumed to have a density*: which means that there is a function $\rho(X)$ such that $\rho(X)dX$ is the probability that, repeating the experiment many times, the datum X falls in the volume element dX. Such a probability distribution is commonly called "absolutely continuous" (with respect to the volume).

Here attention will be devoted to data X generated randomly with"absolutely continuous" distribution. It will be seen that the assumption of existence, behind the protocol of the observation, of an unknown but absolutely continuous probability distribution for the initial data generation is *far from obvious* and in the end it turns out to be a *far reaching physical law*, that could be called law of "initial chaos".

Given the initial data X, the system will evolve describing in time a trajectory denoted $t \to S_t X = X(t)$. In theoretical studies the evolution $X(t)$ will be supposed to be defined by a solution $X(t)$ of a differential equations $\dot{X} = F(X)$.

The interaction with the thermostats is influential and cannot be ignored. This means that the evolution equations must involve the interaction with the surroundings: which is a difficulty because it leads, eventually, to consider infinite systems.[6]

The same difficulty is met even in equilibrium when it is imagined that the temperature is fixed via the contact with a single thermostat.[7]

The difficulty is bypassed, in theoretical studies, by using dynamical models involving finitely many particles obeying equations of motion in which are introduced phenomenological forces (typically proportional to currents, circulating in the system, via "transport coefficients"). Such forces model dissipative effects controlling the actions of the non conservative forces or act by imposing suitable constraints for the same purpose. This gives rise to models which involve forces that are believed to be *physically equivalent* to more physical infinite thermostats but contain *finitely many particles*, [11, 12].

Given the equations of motion, an initial datum X will evolve into $X(t)$, reaching a *stationary state*.

Stationarity is in the sense that the observables will fluctuate in time but will have well defined time independent "*statistics*", *i.e.* distribution of fluctuations around well defined time averages; and for quite a few observables even their values will be essentially constant in time (no need to average them) if the system is large (*i.e.* it consists of a large number of molecules).

In particular, from a microscopic viewpoint, the phase point representing at time t the microscopic state $S_t X$ of the system started in the configuration X, will spend a fraction of the total observation time θ visiting, in the limit $\theta \to \infty$, any chosen open set[8] $\mathcal{D}$ with a well defined frequency $P_0(\mathcal{D})$.

Therefore imagining phase space divided into small regions $\{\mathcal{D}_i\}_{i=1,\ldots,\mathcal{N}}$ the frequencies of visit $P_0(\mathcal{D}_i)$ will determine the average values of all observables O whose variation in each of the $\mathcal{D}_i$ is negligible, so that their values can be denoted $O(\mathcal{D}_i)$: their time average will be

[6] *i.e.* interaction with the surroundings which interact also with their own surroundings, which interact with their own surroundings, ...

[7] In this respect Boltzmann in his 1868 paper shewed that in a very large system in a microcanonical equilibrium a finite volume subsystem behaves as if in contact with a reservoir with a temperature fixed and is described by a canonical ensemble.

[8] Hence any closed, or just "measurable", set in $\mathcal{F}$.

$\langle O\rangle = \sum_i O(\mathcal{D}_i)P_0(\mathcal{D}_i)$.

Of course the size of the sets $\mathcal{D}_i$ must be adapted to the observables that are studied: therefore the choice of the $\mathcal{D}_i$ can be regarded as a choice of a "coarse graining" of phase space.

The frequencies $P_0(\mathcal{D}_i)$ will arise from a probability distribution P_0 on $\mathcal{F}$ *invariant* with respect to the evolution: in the sense that if X evolves in time τ into $S_\tau X$ then *any* (reasonable, *e.g.* open) set $\mathcal{D}$ evolves into some $S_\tau\mathcal{D}$ and the two sets are visited with the *same frequency* $P_0(\mathcal{D}) = P_0(S_\tau\mathcal{D})$.

A fundamental theorem, by Sinai in its simplest form, states that if the motion of the system is "*chaotic*", as is the case for virtually all systems ultimately constituted by atoms, then the probability is 1 that a protocol, of the above mentioned kind for obtaining the initial datum X, generates a motion S_tX which visits the sets $\mathcal{D}$ in phase space with well defined frequency $P_0(\mathcal{D})$, [13–17].

Theorem: there is a unique probability distribution P_0 (called "SRB distribution") determining the frequencies $P_0(\mathcal{D})$ and therefore the statistical properties of the motion. Furthermore:
(a) the frequencies $P_0(\mathcal{D})$ are independent on the protocol generating X, provided it generates random X's with an absolutely continuous distribution,
(b) and given the protocol the frequencies are X-independent with probability 1,
(c) and if there is one invariant distribution which is absolutely continuous then it must be P_0.

The requirement, aside from the mentioned law of initial chaos, is that the equations of motion generate *chaotic motions in a mathematically precise sense* that will be briefly called "*chaotic hypothesis*", or CH, essentially meaning that

Chaotic Hypothesis: any datum X evolves into $S_tX \equiv X(t)$ for $t > 0$,
(a) never stopping (i.e. $\min|\dot X| > 0$) and,
(b) after a transient time t_0, an observer following $S_tX = X(t)$ and oriented as $\dot X(t)$ will see $X(t)$ as a hyperbolic fixed point, i.e. as a saddle point, and
(c) data $X \in \mathcal{A}$ evolve so that S_tX covers densely a smooth surface $\mathcal{A}$ ("attracting set").

The theorem applies to any system with this property, whether isolated (as in studying equilibrium) or in contact with external thermostats and under action of non conservative forces (as in general stationary states).[9]

The result eliminates an essential difficulty. Leaving aside the need to adapt the cells $\mathcal{D}_i$ to the observables selected for analysis, if the equations of motion generate chaotic motions[10] then there are, as a mathematical theorem, many (*actually infinitely many*) stationary probability distributions $P \neq P_0$ that are invariant: *i.e.* in the cases covered by the above theorem there are data which visit the sets $\mathcal{D}$ with frequencies $P(\mathcal{D})$ *different* from $P_0(\mathcal{D})$, *i.e.* different from the SRB distribution.

Hence, without applying the theorem, requiring stationarity does not provide a recipe to determine the statistical properties of the motion.[11]

The theorem allows to select unambiguously *which is the distribution* that determines the statistics P_0 observed in a given case, hence which is the *physically important statistics* associated with the equations of motion: it is valid both in equilibrium cases and in nonequilibrium ones.

Of course this implies accepting as a *physical law* that any protocol generating the system initial configurations X is "absolutely continuous" and, furthermore, the evolution of such X's is chaotic (leaving aside evident exceptions): this law (proposed by Ruelle, [18, 19], together with its main implications), cannot be proved and it has to be accepted as a law of Physics.

The (idealized)[12] case of isolated systems, *i.e.* the equilibrium cases for systems obeying Hamiltonian equations of motion with a fixed energy, have the property that one, P_0, *among the many stationary distributions* compatible with the constraints (usually just with a fixed value of the energy), is such that $P_0(\mathcal{D})$ can be expressed simply as an integral over phase space $P_0(\mathcal{D}) \equiv \int_{\mathcal{D}} \rho_0(X)dX$ of a density function $\rho_0(X)$, which exists as a consequence of Liouville's theorem.[13]

Hence P_0, under the chaotic hypothesis, determines the statistical properties of the motions X. It can be said that isolated systems not only conserve the total energy but also admit an *invariant* absolutely continuous way of measuring the phase space elements of the energy surface (here invariant means that $P_0(\mathcal{D}) = P_0(S_t\mathcal{D})$ for all $t, \mathcal{D}$). This also shows that the CH implies, for isolated systems, the ergodic hypothesis.

III. DISCRETE REPRESENTATION OF MOTION. EQUILIBRIUM.

Accepting the chaotic hypothesis, the problem of identifying the probability distribution controlling the statis-

[9] The requirement that the attracting set is a smooth surface can be weakened replacing "smooth" with "closed": in the first case the motion on $\mathcal{A}$ is a "Anosov system" while in the second it satisfies "Axiom A", [18].

[10] As it is the case apart from very few remarkable exceptions, like the arrays of elastic oscillators with interactions linear in the relative distances.

[11] In equilibrium the ergodic hypothesis, for instance, has to be added to select the microcanonical ensemble distributions as the ones describing the statistics of the motions.

[12] Because it is only possible to contemplate a protocol that generates initial data constrained to have a prefixed energy

[13] Which states that any Hamiltonian system admits an invariant distribution with density over the surface of fixed energy (which is usually the only constraint).

tics of the observations, is solved in general by the theorem in Sec.II, for stationary states of systems, whether *in equilibrium or not.* And the question of how to *extend thermodynamics to nonequilibrium stationary states* can be posed, and one among the first questions is whether an extension of the entropy as a state function is possible.

It is important to review first how the microscopic interpretation of entropy arises in equilibrium thermodynamics.

Via the H-theorem of Boltzmann (1872), the entropy of an isolated rarefied gas is identified to be proportional to a quantity called H, [20]. Later (1884) Boltzmann proposed, *for general* equilibrium cases (including liquids and solid materials), the entropy H to be proportional to the logarithm of the phase space volume W available to the N molecules (*i.e.* the phase space points X with total energy E and with positions in a fixed volume V), namely: $S = k_B \log W$ with k_B being the Boltzmann's constant. This formula is also consistent with the arbitrariness of the additive constant in the entropy definition: the latter simply reflects the arbitrariness of the unit employed to measure the volume W, which has the dimension of the $3N$-th power of an action.[14]

At this point it is convenient to go back to the ergodic hypothesis in its original formulation and see whether it can be applied also to stationary nonequilibria and to the extension of thermodynamics to general stationary states.

For some time it was apparently believed that, if the motion was so chaotic that the trajectory $S_t X$ would become dense in the region of phase space compatible with the constraints, then an invariant distribution would necessarily have the above absolute continuity property, hence it would have the form $\rho_0(X)dX$ suggested by Liouville's theorem.

This seems to have been the opinion of Boltzmann and Maxwell and in support of the latter statement the *ergodic hypothesis* was proposed. In the words of Maxwell,[9], on Boltzmann's work, [8], where the canonical and microcanonical ensembles were introduced it can be read: "*The only assumption which is necessary for the direct proof* [of what are now called the microcanonical and canonical distributions] *is that the system, left to itself in its actual state of motion, will, sooner or later, pass through every phase which is consistent with the equation of energy*".

As the above quotation says, a phase space point X evolves in time visiting all X' compatible with the constraints (in isolated systems this meant all X' with the same energy as X and with particles all located in their container V).

This can be consistent only if the phase space $\mathcal{F}$ is not a continuum but consists of a finite number of points on a regular lattice. The evolution *in a fixed time step* t, small with respect to the duration of atomic collisions, can be seen as a map $X \to X'$ in which the X describing initially the system becomes after time t another phase space point X'. The map should be thought as a discretized solution of the equations of motion (now *common in computer simulations*). Thus the evolution map $X \to X'$ having to visit all possible configurations is just a *one cycle permutation* of the finite (very large) number of possible X's.

The ergodic hypothesis simply means that the permutation must be *cyclic*; the motions will be periodic and the uniformity of the lattice forming the discrete phase space points on the energy surface implies that the frequency of visit to a set $\mathcal{D} \subset \mathcal{F}$ equals the fraction $P_0(\mathcal{D})$ of points contained in $\mathcal{D}$ (if $\mathcal{D}$, although small, contains a very large number of discrete points). Hence average values of observables can be computed via a distribution on the energy surface which is uniform (with respect to the surface elements area), *i.e.* a microcanonical distribution.

The assumption admits obvious exceptions (*e.g.* harmonic lattices and integrable systems) but it was believed to be quite generally correct, in the discretized version of phase space, for isolated systems.

Maxwell, as well as Boltzmann and Clausius, worried about the boldness of the assumption in the above form. About it Maxwell writes, [9], "*But if we suppose that the material particles, or some of them, occasionally encounter a fixed obstacle such as the sides of a vessel containing the particles, then, except for special forms of the surface of this obstacle, each encounter will introduce a disturbance into the motion of the system, so that it will pass from one undisturbed path into another. The two paths must both satisfy the equation of energy, and they must intersect each other in the phase for which the conditions of encounter with the fixed obstacle are satisfied, but they are not subject to the equations of momentum. It is difficult in a case of such extreme complexity to arrive at a thoroughly satisfactory conclusion, but we may with considerable confidence assert that except for particular forms of the surface of the fixed obstacle, the system will sooner or later, after a sufficient number of encounters, pass through every phase consistent with the equation of energy*".

The obvious objection is that the time scale for a phase space point X to go through a full orbit, *i.e.* to visit all possible phase space points, must be, even for a small system (Boltzmann considered the example of $1\,\mathrm{cm}^3$ of hydrogen at normal conditions), unimaginably long: therefore the *ergodic hypothesis cannot be directly physically relevant* and has to be accompanied by proving other properties which explain why the equilibrium statistics can be observed within "human time scales".

Furthermore, even if the phase space points are supposed on a regular lattice, it is not determined which would be their spacings. Their number falling in a region $\mathcal{D}$ remains determined up to a factor depending on the lattice meshes δp and δq in the momenta and posi-

[14] This is the form given to entropy by Planck: it is however implicit in several works of Boltzmann, among which [21, 22].

tions, *i.e.* on the precision with which the lattice represents the continuum: the number of points in $\mathcal{D}$ will be $\frac{vol(\mathcal{D})}{h^{3N}}$, where $h = \delta p \delta q$ is a unit of action determined by the precision of the discrete representation of the continuum.

Maxwell and Boltzmann addressed the time scale question, in the case of rarefied gases, developing the Boltzmann's equation, which is an approximation to the evolution, describing properties of a very restricted class of observables. Nevertheless the class is wide enough to contain all the observables obviously interesting for the theory of gas motions.

It leads to estimate the time scale necessary for a gas to relax to equilibrium (described by the microcanonical distribution, as it would be formally predicted by the ergodic hypothesis), or more precisely for achieving a state in which the few observables that possibly are of interest (like density, pressure, temperature, few particles correlations) show a well defined value with relatively small fluctuations.

In other words, the *Founding Fathers* were well aware that the ergodicity assumption could not be directly used to justify the approach to equilibrium, as well as the time scale necessary for attaining the equilibrium distribution.

They used it as a kind of 'symmetry property' that *a priori* implied a description of the equilibrium states in terms of the equilibrium ensembles; however they insisted that what was really interesting and physically necessary was that, for each observable of interest, like the occupation numbers of domains ("cells") Δ, in the 6-dimensional single atom phase space, large enough to contain a sizable fraction of the total number N of atoms, there should be a measurable time scale for reaching the average value.

The analysis can be found already in their attempts, by Maxwell, Boltzmann, Thomson, to obtain macroscopic equations describing the evolution (to equilibrium or stationarity) of several observables ([20, 23–25]).

IV. ENTROPY AND NONEQUILIBRIUM

Turning to nonequilibrium, the simplest case to keep in mind is that of a gas in contact with two reservoirs at *different temperatures*. One can also think to electrically charged particles moving in a lattice (with periodic boundary conditions, modeling a "wire") of obstacles (molecules of a "crystal") and subject to an electric field and to a thermostat. Or to a fluid in a periodic container subject to a stirring force and in contact with a thermostat to dissipate friction heat.

The equations of motion $\dot{X} = F(X)$ may have, and do have for several models of nonequilibrium systems, a "*time reversal symmetry I*": here I is a smooth map of phase space and the solution $S_t X$ with initial datum X has the property that $IS_tX = S_{-t}IX$.

Can one proceed, as done in equilibrium, and imagine the phase space configurations X compatible with the constraints as a discrete set of points located in the usual continuum phase space? This is tempting as it would bring back the idea that the motion of a phase space point wonders visiting successively all other points: it would also explain the existence of a *unique stationary distribution*, which would be simply the distribution giving equal weight to all points.

It would be natural to form a partition $\mathcal{P}$ of the continuum phase space of a system into finitely many sets $\mathcal{D}_i$ and call $P_0(\mathcal{D})$ the probability attributed to each set $\mathcal{D}$ by the invariant SRB distribution, *i.e.* the frequency of visit to $\mathcal{D}$ from an initial datum X obtained with probability 1 via some protocol (see Sec.II). Such probability is well defined *although generally not expressible* by an integral over $\mathcal{D}$. And then *replace* the continuum phase space by a *finite number* $\mathcal{N}$ of points, with $\mathcal{N}P_0(\mathcal{D}) \gg 1$ of them in each $\mathcal{D} \in \mathcal{P}$.

Furthermore the evolution should be a *one cycle permutation* of the phase space points: in this way each cell $\mathcal{D}$ is visited (in a very long time) with a frequency $P_0(\mathcal{D})$ which is, therefore, uniquely determined and is a representation of the SRB distribution. The time necessary needs not be too long if the cells $\mathcal{D}$ are not too small (although small enough so that the observables of thermodynamic interest can be considered constant in each of them) and contain a fraction of the total number of discrete points of order 1 (so that $\mathcal{N}P_0(\mathcal{D}) \gg 1$).

But in the case of nonequilibrium the equations of motion are no longer Hamiltonian and are dissipative. This is manifested by the *divergence* $\sigma(X)$ of the equations of motion:[15] which is not 0 (as it is for the isolated evolutions, *i.e.* in the Hamiltonian cases) and must have a ≥ 0 average $\langle\sigma\rangle$.[16]

If the "chaotic hypothesis" (CH) holds, $X(t)$ evolves towards a surface $\mathcal{A}$[17] and initial data X starting out of it evolve in time with their distance to $\mathcal{A}$ tending to 0 (exponentially fast). The surface $\mathcal{A}$ can possibly be different from the entire phase space compatible with the constraints and have lower dimensionality: in any case if $\langle\sigma\rangle > 0$ the statistics will be a probability distribution which gives probability 1 to a subset in $\mathcal{A}$, the "*attractor*" $\mathcal{A}_o \subset \mathcal{A}$, which has 0 volume, or 0 surface area if $\mathcal{A}$ is a surface of lower dimension.

A discretization of phase space should therefore be a discrete representation of the attractor $\mathcal{A}_o \subset \mathcal{A}$ that can be imagined replacing the continuum phase space by a *regular* lattice in which the position and momentum coordinates are on a grid spaced by δq and δp, and then

[15] Given a general ODE $\dot{x}_j = f_j(x), j = 1, \ldots, n$ the divergence definition is $\sigma(x) = -\sum_j \partial_{x_j} f_j(x)$ and gives the rate of compression of a volume element dx around x; it can be > 0 (compression) or < 0 (expansion) depending on x.

[16] The average $\langle\sigma\rangle$ cannot be < 0, *i.e.* phase space cannot keep expanding forever while a stationary state is reached.

[17] Possibly of dimension lower than that of phase space if the forces that keep the system out of equilibrium are not small enough.

discard the points that are not on the cyclic permutation toward which data generated by the initial protocol evolve.[18]

Under the CH *heuristic arguments* can be developed to *estimate the number $\mathcal{N}$ of discrete points* necessary to give an *accurate description* of the motions of data on the attractor $\mathcal{A}_0 \subset \mathcal{A}$, [12, Sec 3.11], out of the $\mathcal{N}_0$ regular grid points of the discretized phase space. Once the discretization is obtained and $\mathcal{N}$ is estimated, it is will tempting to define entropy as proportional to $\log \mathcal{N}$.

However the result is that $\log \mathcal{N}$ might *not be defined up to an additive constant* depending only on the precision of the discretization: but changing the precision (*i.e.* the size of the discretization meshes) it changes by a quantity which depends *also* on the stationary state considered, in particular it depends on the average phase space contraction $\langle\sigma\rangle$: this is in sharp contrast with the equilibrium result where changing the precision changes $\log \mathcal{N}$ by a constant *independent* of the equilibrium state studied.

This indicates[19] that entropy, as a *function of state*, might not be definable for stationary states out of equilibrium, [12, Sec.3.10,3.11], unless a physical constraint determining the maximum precision of the discretization can be found (but such quantization of phase space would require extra information).

However one of the main features of the *extension* of entropy to rarefied gases not in equilibrium, but isolated and evolving towards equilibrium, is that it is a "Lyapunov function" varying with time and approaching (monotonically) a maximum value as a limit value, namely the equilibrium entropy, [26, 27].

It is conceivable that in the evolution to a stationary state it could be possible to define a Lyapunov function with the same property of evolving (possibly not monotonically) to a maximum which is reached at stationarity, [12, 28].

If an initial non stationary distribution is considered (including possibly a distribution specifying a single phase space point ξ_0) then the fraction $P(\xi,t)$ of times in $[0,t]$ that the point $S_t\xi_0$ visits ξ tends to $\frac{1}{\mathcal{N}}$, as prescribed by the SRB distribution in the above discrete representation, where $\mathcal{N}$ is the number of points in the discretized attractor $\mathcal{A}_0$. Then $S(t) = k_B \sum_\xi -\overline{P}(\xi,t)\log\overline{P}(\xi,t)$ tends, as $t\to\infty$, to:

$$S_\infty = k_B \sum_\xi -\frac{1}{\mathcal{N}}\log\frac{1}{\mathcal{N}} = k_B \log \mathcal{N},$$

hence S_∞ is the maximum value that $S(t)$ can reach[20] and therefore $S(t)$ can play the role of a Lyapunov function.

The function S_∞ depends *non trivially* on the precision of the discretisation (as just discussed); and changing just the precision, *i.e.* the discretization mesh, S_∞ will change depending nontrivially on the particular stationary state. So, when studying the stationary states of a system, as functions of the parameters entering into its equations of motion, it does not seem that S_∞ can defined just up to an additive constant independent of the particular stationary state, hence *cannot be considered a function of state.*

The question on whether it is possible to define a function of state generalizing the entropy function to nonequilibrium steady states remains an interesting open question.

Still for all choices of the discretisation $S(t)$ will have the property that it reaches the maximum value on the SRB distribution, *i.e.* on the natural stationary state. Entropy, as a function of state, may not defined in general stationary states although the approach to stationarity may admit a Lyapunov function (possibly related to the above $S(t)$): the latter would extend the role plaid in the approach to equilibrium by the entropy function as recently defined beyond the rarefied gases by the wider interpretation of the formula $S = k_B \log W$, [26, 27].

V. QUEST FOR UNIVERSALITY

The remarkable general validity of the distributions giving the statistical properties of the equilibrium states of very general systems is extended to general stationary states via the associated SRB distributions, as described by the theorem in Sec.II. The latter however do not have the same character as they may seem, at first, too theoretical.[21] However the problem may simply be due to the still recent introduction of the SRB distributions.[22]

Universal relations very often reflect symmetry properties: the Onsager reciprocity relations are a manifestation of the basic time reversibility of the equations of motion and a first example of general and universal property for nonequilibrium systems. However they are a property that holds only as a first order approximation in terms of the size of the active forces determining the nonequilibrium dynamics.

[18] The regularity of the lattice representing the discretized phase space reflects the special relation (called absolute continuity in Sec.1) between the protocols that generate initial data and the volume measure. In the equilibrium cases (*i.e.* Hamiltonian evolutions) all grid points are thought to be part of the evolution cycle, as a literal interpretation of the ergodic hypothesis.

[19] But does not "prove", even under the CH, because the estimates are heuristic.

[20] because the maximum of $-\sum_{i=1}^M p_i \log p_i$ is $\log M$ and is achieved when $p_i = M^{-1}$.

[21] As opposed to the equilibrium cases where the ergodic hypothesis yields a concrete universal prescription to determine the statistics of the motions to be given by the "Gibbs states".

[22] The complaints, that can sometimes be found in the literature, on the lack of explicit examples of many degrees of freedom systems are not well founded because examples which are chaotic, reversible (or not) and admit SRB distributions do exist, [29–32].

It is natural, as a first step in searching for universal properties, to wonder whether the reciprocity relations have an extension to general non equilibria, as a consequence of the time reversibility, that is *always valid* in the basic equations, even in presence of dissipation.

The question can be studied for systems in contact with thermostats and subject to non conservative forces so that their stationary states (if existing) will be out of equilibrium. Models of thermostats have been introduced which involve forces that are believed to be equivalent to purely mechanical infinite thermostats but *involve finitely many particles* and are described by *reversible equations*, [11, 12].[23]

For such systems, particularly important in numerical simulations, time reversal is a valid symmetry and a preliminary problem is to see how comes that a *reversible equation leads to irreversible evolution.*

Dissipation is controlled by the phase space contraction, *i.e.* the divergence $\sigma(X)$ of the equations of motion: and $\sigma(X)$ is a particularly interesting observable (which often has the interpretation of thermostats entropy increase rate). Time reversal symmetry implies that for each configuration X with $\sigma(X) > 0$ there is another IX with $\sigma(IX) < 0$: reversibility implies that dissipation (*i.e.* phase space contraction) has to coexist with phase space expansion.

In dissipative systems motions evolve towards an attracting set $\mathcal{A}$ which is not necessarily time reversal symmetric. Followed backwards in time they also evolve towards an attracting set $\mathcal{B} = I\mathcal{A}$, which however attracts points only in the backward evolution: and $\mathcal{B}$ is a repelling sets for the evolution forward in time if $\mathcal{B} \neq \mathcal{A}$.

However if the forces driving the system to a stationary nonequilibrium state are small enough, although not infinitesimal as they are in the theory of Onsager reciprocity, and if the "chaotic hypothesis" can be assumed it follows that both $\mathcal{A}$ and $\mathcal{B}$ *are the same* because motions with any initial data are dense on phase space.[24] Nevertheless the SRB distribution on $\mathcal{A} = \mathcal{B}$ for the forward motion and the SRB distribution for the backward motion are *mutually singular.*

In other words there is a fractal set $\mathcal{A}_+ \subset \mathcal{A}$ of data (a "forward attractor") which have probability 1 to generate the SRB distribution μ_+ for the forward motions and probability 0 to generate the SRB μ_- for the backward motions, and vice-versa.[25]

Actually data in $\mathcal{A}_+$ run forward or backward in time generate the same SRB statistics μ_+ and the corresponding statement holds for the SRB statistics μ_-: it is $\mu_\pm(\mathcal{A}_\pm) = 1, \mu_\pm(\mathcal{A}_\mp) = 0$. So the irreversibility is made manifest, in such time reversible systems, by the different statistics obeyed with probability 1 by motions generated by a protocol as discussed in Sec.1 and observed as time tends to $+\infty$ or to $-\infty$.

Therefore it is interesting to study the effect of time reversal symmetry on systems in which the attracting set $\mathcal{A}$ is equal to its time reversal $I\mathcal{A}$ and both coincide with the phase space available (although the forward and backward SRB statistics are mutually singular). In general such motions *taking place on* $\mathcal{A}$ will have a surface contraction rate $-\sigma$ with a *positive time average* $\langle\sigma\rangle > 0$ of σ, in the forward evolution (*as well as* in the backward evolution).

For such systems the fluctuations of the dissipation $\sigma(S_tX)$, as the initial data X evolve into S_tX and $t \to +\infty$, have remarkable properties because, under the chaotic hypothesis, the motions on $\mathcal{A}$ are in several aspects are well understood.[26]

Define the function: "probability $P(p, \tau, \delta)$ of the set of X such that $\frac{1}{\tau}\int_0^\tau \frac{\sigma(S_tX)}{\langle\sigma\rangle} dt \in [p, p+\delta]$": this function is interesting because often $\sigma(X)$ has the physical interpretation of *entropy generated* (in the thermostats) per unit time by the stationary state considered.

Then a general theorem applies to the motions on $\mathcal{A}$, essentially comparing the probability that p has a given value to that of having the opposite value, *i.e.* the probability of entropy production in time τ equal to $\sim p\langle\sigma\rangle\tau$ to that of $\sim -p\langle\sigma\rangle\tau$: *if CH holds and the motions are time reversible and their trajectories are dense on phase space*[27] the probabilities can be shown to be expressed, to leading order for large τ, via a density $\sim e^{s(p)\tau}$ and :

Fluctuation Theorem: $s(-p) = s(p) - p\,\langle\sigma\rangle\,\tau$

with $s(p)$ convex and maximal at $p = 1$: which is a *universal* relation in the sense that it contains no free parameters, [33, 34].

It is remarkable, in particular, that it establishes a relation between the "normal", *i.e.* most probable, value of the average entropy production in a time τ, corresponding to $p = 1$, to the highly unlikely "opposite event" with $p = -1$, and the relative probability depends on the entropy production rate $\langle\sigma\rangle$ but is otherwise independent on the system considered.

The above relation is called *fluctuation theorem*, FT, when the mathematical assumptions are satisfied, or *fluctuation relation*, FR, if the hypotheses are considered phenomenologically valid. There are few examples in which time reversibility holds, the models have

[23] Here, as in the previous section, time reversal is a *smooth map* of phase space such that $IS_tX = S_{-t}IX$.

[24] This is consequence of the "structural stability" property of the CH.

[25] the "attractor" $\mathcal{A}_+$ on $\mathcal{A}$ is not uniquely defined: usually it is an invariant set which has μ_+-probability 1 and has maximal fractal dimension: typically it can be modified by subtracting from it the trajectory of one or more data X.

[26] Under the CH the evolution on $\mathcal{A}$ is called a "Anosov system": such systems can be considered as the paradigm of chaotic motions, and in the theory of chaotic dynamics play a role similar to that payed by harmonic oscillators for regular dynamics.

[27] for instance, under the CH, at small forcing of a thermostatted Hamiltonian system.

a nonequilibrium stationary state, and the FR holds, although the FT hypotheses, would be to difficult to check (if at all holding), [35, 36].

It shows that in such systems the entropy of the thermostats grows by $\sim \langle\sigma\rangle\tau$ in time τ with a probability $e^{\langle\sigma\rangle\tau}$ larger than the probability that it decreases by the same amount, to leading order in $\tau \to \infty$.

It has been shown that, under the chaotic hypothesis, the fluctuation theorem implies Onsager reciprocity so the FR provides an extension to Onsager's theorem, [37, 38], at small but not necessarily infinitesimal forcing.

The fluctuation relation is a first universal relation found, under the CH, for general nonequilibrium systems at small forcing. And the question is whether it is of any interest for systems which are not described by reversible equations of motion: since most macroscopic equations involve frictional forces the applicability of the FR may seem restricted, at most, to microscopic and small systems. But time reversal is even "behind" macroscopic equations derived phenomenologically and containing forces explicitly violating time reversal (like friction).

This suggests that the same phenomena could be equally described by equations which are time reversible: such equations would show a variable phase space contraction $\sigma(X)$ (unlike the familiar time irreversible ones, which typically show a constant phase space contraction, proportional to some transport coefficients, *e.g.* friction). The divergence $\sigma(X)$ arising in the reversible models can be regarded as a special observable. As such it can be studied in the model with irreversible evolution $S_t^{irr}X$: if the two descriptions are equivalent it can be expected that the fluctuations of $p = \frac{1}{\tau}\int_0^\tau \sigma(S_t^{irr}X)dt$ also satisfy the FR. If so the FR could be observable even in irreversible models (if they derive from microscopic models which satisfies time reversal, as they should), [39].

The possibility of having the same system described by irreversible or reversible equations indicates interesting analogies with the equivalence of ensembles in equilibrium statistical mechanics: a subject to which some attention is currently devoted, [39].

[1] S. Carnot. *Réflections sur la puissance motrice du feu et sur les machines propres à développer cette puissance.* Online in https://gallica.bnf.fr; original Bachelier, 1824; reprinted Gabay, 1990., Paris, 1824.

[2] A. Krönig. Grundzüge einer Theorie der Gase. *Annalen der Physik und Chemie*, XCIX:315–322, 1856.

[3] R. Clausius. On the motive power of heat, and on the laws which can be deduced from it for the theory of heat. *Philosophical Magazine*, 2:1–102, 1851.

[4] R. Clausius. Über einige für Anwendung bequeme formen der Hauptgleichungen der mechanischen Wärmetheorie. *Annalen der Physik und Chemie*, 125:353–401, 1865.

[5] L. Boltzmann. Über die mechanische Bedeutung des zweiten Hauptsatzes der Wärmetheorie. *Wiener Berichte*, 53, (W.A.,#2):195–220, (9–33), 1866.

[6] R. Clausius. Ueber die Zurückführung des zweites Hauptsatzes der mechanischen Wärmetheorie und allgemeine mechanische Prinzipien. *Annalen der Physik*, 142:433–461, 1871.

[7] R. Clausius. Bemerkungen zu der priorität reclamation des hrn. boltzmann. *Annalen der Physik*, 144:265–280, 1872.

[8] L. Boltzmann. Studien über das gleichgewicht der lebendigen kraft zwischen bewegten materiellen punkten. *Wiener Berichte*, 58, (W.A.,#5):517–560, (49–96), 1868.

[9] J. C. Maxwell. On Boltzmann's theorem on the average distribution of energy in a system of material points. *In: The Scientific Papers of J.C. Maxwell, Cambridge University Press, Ed. W.D. Niven, Vol.2*, pages 713–741, 1879.

[10] J. Gibbs. *Elementary principles in statistical mechanics.* Schribner, Cambridge, 1902.

[11] D. J. Evans and G. P. Morriss. *Statistical Mechanics of Nonequilibrium Fluids.* Academic Press, New-York, 1990.

[12] G. Gallavotti. *Nonequilibrium and irreversibility.* Theoretical and Mathematical Physics. Springer-Verlag and http://ipparco.roma1.infn.it & arXiv 1311.6448, Heidelberg, 2014.

[13] Ya. G. Sinai. Markov partitions and C-diffeomorphisms. *Functional Analysis and Applications*, 2(1):64–89, 1968.

[14] Ya. G. Sinai. Construction of Markov partitions. *Functional analysis and Applications*, 2(2):70–80, 1968.

[15] Ya.G. Sinai. *Lectures in ergodic theory.* Lecture notes in Mathematics. Princeton University Press, Princeton, 1977.

[16] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeormorphisms*, volume 470 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin-Heidelberg, 1975.

[17] R. Bowen and D. Ruelle. The ergodic theory of axiom A flows. *Inventiones Mathematicae*, 29:181–205, 1975.

[18] D. Ruelle. *Chaotic motions and strange attractors.* Accademia Nazionale dei Lincei, Cambridge University Press, Cambridge, 1989.

[19] D. Ruelle. *Elements of differentiable dynamics and bifurcation theory.* Academic Press, New-York, 1989.

[20] L. Boltzmann. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Wiener Berichte*, 63, (W.A.,#22):275–370, (316–402), 1872.

[21] L. Boltzmann. Analytischer Beweis des zweiten Hauptsatzes der mechanischen Wärmetheorie aus den Sätzen über das Gleichgewicht des lebendigen Kraft. *Wiener Berichte*, 63, (W.A.,#20):712–732,(288–308), 1871.

[22] L. Boltzmann. Über die Eigenshaften monozyklischer und anderer damit verwandter Systeme. *Crelles Journal*, 98, (W.A.,#73):68–94, (122–152), 1884.

[23] J.C. Maxwell. On the dynamical theory of gases. *In: The Scientific Papers of J.C. Maxwell, Cambridge University Press, Ed. W.D. Niven, Vol.2*, pages 26–78, 1866.

[24] W. Thomson. Kinetic theory of dissipation of energy. *Nature*, IX:441–444, 1874.

[25] L. Boltzmann. *Lectures on gas theory, English edition annotated by S. Brush.* University of California Press, Berkeley, 1964.

[26] P. L. Garrido, S. Goldstein, and J. L. Lebowitz. Boltzmann entropy for dense fluids not in local equilibrium. *Physical Review Letters*, 92:050602 (+4), 2005.

[27] S. Goldstein, D. Huse, J. Lebowitz, and P. Sartori. On the nonequilibrium entropy of large and small systems. *arXiv:1712.08961*, pages 1–20, 2018.

[28] G. Gallavotti. Counting phase space cells in statistical mechanics. *Communication in Mathematical Physics*, 224:107–112, 2001.

[29] J. Bricmont and A. Kupiainen. Coupled analytic maps. *Nonlinearity*, 8:379–396, 1995.

[30] G. Gallavotti, F. Bonetto, and G. Gentile. *Aspects of the ergodic, qualitative and statistical theory of motion.* Springer Verlag, Berlin, 2004.

[31] F. Bonetto, P. Falco, and A. Giuliani. Analyticity of the SRB measure of a lattice of coupled Anosov diffeomorphisms of the torus. *Journal of Mathematical Physics*, 45:3282–3309, 2004.

[32] F. Bonetto, G. Gallavotti, and G. Gentile. A fluctuation theorem in a random environment. *Ergodic Theory and Dynamical Systems*, 28:21–47, 2008.

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

[34] G. Gentile. A large deviation theorem for Anosov flows. *Forum Mathematicum*, 10:89–118, 1998.

[35] B. Derrida, J. L. Lebowitz, and E. R. Speer. Exact free energy functional for a driven diffusive open stationary nonequilibrium system. *Physical Review Letters*, 89:030601, 2002.

[36] A. Gerschenfeld, B. Derrida, and J. L. Lebowitz. Anomalous Fouriers law and long range correlations in a 1D non-momentum conserving mechanical model. *Journal of Statistical Physics*, 141:757–766, 2010.

[37] G. Gallavotti. Chaotic hypothesis: Onsager reciprocity and fluctuation–dissipation theorem. *Journal of Statistical Physics*, 84:899–926, 1996.

[38] G. Gallavotti and D. Ruelle. SRB states and nonequilibrium statistical mechanics close to equilibrium. *Communications in Mathematical Physics*, 190:279–285, 1997.

[39] G. Gallavotti. Finite thermostats in classical and quantum nonequilibrium. *European Physics Journal Special Topics*, 227:217–229, 2018.