Irreversibility time scale

G.Gallavotti
I.N.F.N. Roma 1, Fisica Roma1

Abstract: Entropy creation rate is introduced for a system interacting with thermostats (*i.e.*, in the usual language, for a system subject to internal conservative forces interacting with "external" thermostats via conservative forces) and a fluctuation theorem for it is proved. As an application a time scale is introduced, to be interpreted as the time over which irreversibility becomes manifest in a process leading from an initial to a final stationary state of a mechanical system in a general nonequilibrium context. The time scale is evaluated in a few examples, including the classical Joule-Thompson process (gas expansion in a vacuum).

PACS numbers 47.52.+j, 05.45.-a, 05.70.Ln, 05.20.-y

Overview

Processes, *i.e.* transformations, between equilibrium states are considered in equilibium Thermodynamics and a few relations between macroscopic observables are found which remain valid all along the transformation, provided the latter can be considered as a slow sequence of successive equilibrium states ("quasi static"). The relations are summarized by the principles: $dU = 0$ ("energy conservation") and integrability of the differential $(dU + pdV)T^{-1}$ ("existence of entropy"). They can be derived from a microscopic theory in which a system is modeled as an aggregate of molecules interacting via conservative forces.
A corresponding theory would be desirable for stationary systems out of equilibrium: the latter are systems subject to external nonconservative forces and kept in stationary states by the concurrent action of "thermostats" and the first question has been how to model a thermostat.
In fact it turned out to be quite difficult to define the temperature out of equilibrium: here we propose that a thermostat is simply a system of particles external to the system, and kept at constant total kinetic energy by a force phenomenologically defined by Gauss' least constraint principle. The constant value of the total kinetic energy will be identified as proportional to the temperature of the thermostat. This is quite different from what is normally called a "isokinetic thermostat" (often declared unphysical) which keeps the kinetic energy of the system, rather than that of the thermostats, constant.
A further interesting proposal, in the recent literature, has been to identify "entropy creation rate" as the phase space contraction rate; but the notion has been criticized mainly because it is referred to the entire set of particles including the ones in the thermostats. This is a serious drawback because it is quite clear that entropy creation, as well as any macroscopic property of the system, should be a well defined property independent on the thermostats size.
On the other hand a general law, "Fluctuation Theorem", has been derived governing fluctuations of phase space contraction rate provided the full system (thermostats included) behaves very chaotically ("Chaotic Hypothesis").
Here the rate of entropy generation by the system will be, differenly, identified with the rate of entropy increase of the thermostats, regarded as equilibrium systems with fixed temperature. It is then shown to satisfy exactly the same asymptotic symmetry property that is expected to hold for the total phase space contraction, *i.e.* the Fluctuation Theorem, with the important feature that the property becomes observable on a time scale substantially shorter than the time scale necessary for observing it for all particles (for which it had been so far proved under the Chaotic Hypothesis), including those in the thermostats.
With the entropy creation rate for the system and the temperature of the thermostats it becomes possible to study processes obtained by varying the external forces and the volume and inducing an evolution of the state of the system from an initial stationary state to a final one. A time scale marking the waiting time necessary to realize the irreversible nature of a process is introduced here. Classical quasi static processes through equilibrium states turn out to have an irreversibility time scale $+\infty$ but the time scale makes sense also for non quasi static processes between stationary states and is discussed here.
Entropy of a stationary state is not needed nor defined, in agreement with the doubts raised elsewhere, [18], on the possibility of defining such a notion in stationary out of equilibrium systems.

1. Forces and thermostats

In studying stationary states in nonequilibrium statistical mechanics, [2, 3], it is common to consider systems of particles in a (finite) container $\mathcal{C}_0$ forced by non conservative forces whose work is controlled by *thermostats* consisting of other particles moving outside $\mathcal{C}_0$ in containers $\mathcal{C}_a$ and interacting with the particles of $\mathcal{C}_0$ through interactions across the walls of $\mathcal{C}_0$, [4]. The positions of the $N \equiv N_0$ particles in $\mathcal{C}_0$ and of the N_a ones in $\mathcal{C}_a$ will be denoted $\mathbf{X}_a$, $a = 0, \ldots, n$, and $\mathbf{X} \stackrel{def}{=} (\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_n)$. Interactions will be described by a potential energy

$$W(\mathbf{X}) = \sum_{a=0}^{m} U_a(\mathbf{X}_a) + \sum_{a=1}^{m} W_a(\mathbf{X}_0, \mathbf{X}_a) \tag{1.1}$$

i.e. thermostats particles only interact indirectly, via the system. All masses will be $m=1$, for simplicity.

The particles in $\mathcal{C}_0$ will also be subject to external, possibly nonconservative, forces $\mathbf{F}(\mathbf{X}_0,\boldsymbol{\Phi})$ depending on a few strength parameters $\boldsymbol{\Phi}=(\varphi_1,\varphi_2,\ldots)$. It is convenient to imagine that the force due to the confining potential determining the region $\mathcal{C}_0$ is included in $\mathbf{F}$ so that one of the parameters is the volume $V=|\mathcal{C}_0|$.

Furthermore the thermostats particles will be also subject to forces $\boldsymbol{\vartheta}$ with the property that the motions take place at constant total kinetic energy K_a that (if k_B is Boltzmann's constant) will be written as

$$K_a=\sum_{j=1}^{N_a}\frac12\,(\dot{\mathbf{X}}^a_j)^2\stackrel{def}{=}\frac32 N_a k_B T_a\stackrel{def}{=}\frac32 N_a\beta_a^{-1} \tag{1.2}$$

and the parameters T_a will define the *temperatures* of the thermostats. The exact form of the forces that have to be added in order to insure constancy of the kinetic energies should not really matter, within wide limits.

A choice of the thermostatting forces that has been employed in numerical simulations has often been to define them according to Gauss' principle of *least effort* for the constraints $K_a=const$, [5]: the simple application of the principle yields a force on the particles of the a-th thermostat, $a=1,2,\ldots$, to be, see (1.1),(1.2),

$$-\boldsymbol{\vartheta}_a=-\frac{L_a-\dot U_a}{3N_ak_BT_a}\,\dot{\mathbf{X}}_a\stackrel{def}{=}-\,\alpha_a\,\dot{\mathbf{X}}_a \tag{1.3}$$

where $L_a=-\partial_{\mathbf{X}_a}W_a(\mathbf{X}_0,\mathbf{X}_a)\cdot\dot{\mathbf{X}}_a$ is the work done per unit time by the forces that the particles in $\mathcal{C}_0$ exert on the particles in $\mathcal{C}_a$. Independently of Gauss' principle it is immediate to check that (1.3) implies that the motion of the particles in $\mathcal{C}_a$ conserves the kinetic energy K_a.

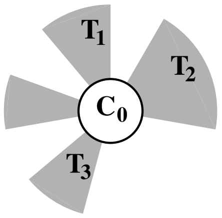

Fig1

Fig.1 The reservoirs occupy finite regions outside C_0, *e.g.* sectors $C_a\subset R^3$, $a=1,2\ldots$. Their particles are constrained to have a *total* kinetic energy K_a constant, by suitable forces $\boldsymbol{\vartheta}_a$, so that the reservoirs "temperatures" T_a, see (1.2), are well defined.

Forces and potentials will be supposed smooth, *i.e.* analytic, in their variables aside from impulsive elastic forces describing shocks, allowed here to model shocks with the containers walls and possible shocks between hard core particles. In conclusion the equations of motion of the system in its *phase space* $\mathcal{F}$, [6],

$$\mathcal{F}\stackrel{def}{=}(\mathcal{C}_0^{N_0}\times R^{3N_0})\times\prod_{a\ge1}(\mathcal{C}_a^{N_a}\times B_a)\equiv\mathcal{F}_0\times\prod_a\mathcal{F}_a \tag{1.4}$$

where B_a is the sphere or radius $(3N_ak_BT_a)^{\frac12}$ in velocity space R^{3N_a}, can be written as

$$\begin{aligned}\ddot{\mathbf{X}}_{0,j}=&-\partial_{\mathbf{X}_{0,j}}W+\mathbf{F}_j, && j=1,\ldots,N_0\\ \ddot{\mathbf{X}}_{a,j}=&-\partial_{\mathbf{X}_{a,j}}W-\boldsymbol{\vartheta}_{a,j}, && j=1,\ldots,N_a\end{aligned} \tag{1.5}$$

with $\boldsymbol{\vartheta}_a$ given by (1.3). Denoting $S_t(\dot{\mathbf{X}},\mathbf{X})$ the configuration into which initial data $(\dot{\mathbf{X}},\mathbf{X})\stackrel{def}{=}(\dot{\mathbf{X}}_a,\mathbf{X}_a)_{a=0,\ldots,n}$ evolve in time t, a further condition has to be added that expresses the physical property that thermostats are *efficient*. Namely for all initial data $x=(\dot{\mathbf{X}},\mathbf{X})$, *except possibly for a set of data x of zero volume* in $\mathcal{F}$, the limit

$$\lim_{T\to\infty}\frac1T\int_0^T f(S_t(\dot{\mathbf{X}},\mathbf{X}))\,dt=\int_{\mathcal{F}}f(y)\mu(dy) \tag{1.6}$$

exists for *all* smooth functions f and is independent of $(\dot{\mathbf{X}},\mathbf{X})$, thus defining a probability distribution on $\mathcal{F}$.

This of course puts restrictions on the kind of forces acting on particles: it has to be stressed that the condition that thermostat forces be "efficient" enough (impeding, for instance, an indefinite build up of the energy) is strong but it has to be assumed. It imposes on the potentials conditions that are not well understood, although they seem empirically (*i.e.* in simulations) verified with the simplest choices of molecular potentials, [7].

The notion of thermostat just described has evolved in the last decades and, while it will continue to evolve, it seems to me that it has already reached a point in which one can lay down the above precise notion on which to base a few general considerations, [6]. The class of thermostats just considered is general enough for our purposes and is amply used in numerical simulations.

The probability distribution μ in (1.6) is called the *SRB distribution* or *SRB statistics*. Typically one is interested in studying time averages of observables depending only on the system coordinates $(\dot{\mathbf{X}}_0,\mathbf{X}_0)$ and not on the thermostats coordinates: in such case one needs only to know the distribution $\mu^{\mathcal{C}_0}$ obtained by restricting the distribution μ to such observables: also $\mu^{\mathcal{C}_0}$ will be naturally called the SRB distribution for the system.

Remarks: (1) The time evolution maps S_t, acting on the *full* phase space $\mathcal{F}$, will have the group property $S_t\cdot S_{t'}=S_{t+t'}$ and the SRB distribution μ will be invariant under time evolution. The SRB distribution μ is said to describe, and is identified with, a *stationary state* of the system; it depends on the parameters on which the forces acting on the system depend, *e.g.* $|\mathcal{C}_0|$ (volume) and more generally $\boldsymbol{\Phi}$ (strength of the forcings), on the parameters that characterize the thermostats like their phase space surfaces $\mathcal{F}_a$, *i.e.* $\beta_a\stackrel{def}{=}(k_BT_a)^{-1}$, *etc.* The collection of SRB distributions obtained by letting the parameters vary defines a *nonequilibrium ensemble.*

(2) Finite thermostats are an idealization of the apparently more appealing infinite thermostats in which the

particles have a density ρ_a and an empirical distribution which asymptotically (at infinity) is Gibbsian with temperatures T_a, [6]: and in such infinite thermostats the temperature should stay necessarily constant because the local interaction between the particles in $\mathcal{C}_0$ and those in the thermostats cannot alter the velocity distribution of the thermostat particles at large distance, at least not if the thermostats act as physically expected.

(3) One could imagine replacing the forces, (1.3), that keep the temperature constant, with other thermostats with which the particles in the containers $\mathcal{C}_a$ are in turn in contact: this is a matter of where we decide to stop in our investigation of the properties of what we call "system". Should we be interested in what really happens in the containers $\mathcal{C}_a$ we would have to add them to what we call "system" and model their interactions with other thermostats. In this respect the Gaussian forces that act on the particles in the containers $\mathcal{C}_a$ are a model of further thermostats that act on the particles in $\mathcal{C}_a$. If one wants to avoid completely the forces (Gaussian as in (1.3) or of some other form) that fix the temperatures of the thermostats then one would be forced to consider infinitely extended $\mathcal{C}_a$'s. But unless the infinitely extended systems consist of non interacting particles, an important exception, see [8], not only this would lead to several conceptual problems (like existence of solutions to the equations of motion, phase transitions in the thermostats, nature of the statistical properties of the motions inside the thermostats) but it would also be possibly not needed, as in the end we are interested only in the statistical properties of the motions in the region $\mathcal{C}_0$, *i.e.* on the finite system we started with. In fact the idea of using finite thermostats to study nonequilibrium statistical mechanics, [5], has been the source of the progress in the field in the last decades.

2. Processes

A *process*, denoted Γ, transforming an initial stationary state $\mu_{ini}\equiv\mu_0$ under forcing $\boldsymbol{\Phi}_{ini}\equiv\boldsymbol{\Phi}_0$ into a final stationary state $\mu_{fin}\equiv\mu_\infty$ under forcing $\boldsymbol{\Phi}_{fin}\equiv\boldsymbol{\Phi}_\infty$ will be defined by a piecewise smooth function $t\to\boldsymbol{\Phi}(t)$, $t\in[0,+\infty)$, varying between $\boldsymbol{\Phi}(0)=\boldsymbol{\Phi}_0$ to $\boldsymbol{\Phi}(+\infty)=\boldsymbol{\Phi}_\infty$.

For intermediate times $0<t<\infty$ the time evolution $x=(\dot{\mathbf{X}},\mathbf{X})\to x(t)=S_{0,t}x$ is generated by the equations (1.5) with initial stte in $\mathcal{F}$ and $\boldsymbol{\Phi}(t)$ replacing $\boldsymbol{\Phi}$: it is a non autonomous equation.

The time dependence of $\boldsymbol{\Phi}(t)$ could for instance be due to a motion of the container walls which changes the volume $V_t=|\mathcal{C}_0|$: hence the points $x=(\dot{\mathbf{X}},\mathbf{X})$ evolve at time t in a space $\mathcal{F}(t)$ which also may depend on t.

During the process the initial state evolves into a state μ_t attributing to an observable $F_t(x)$ defined on $\mathcal{F}(t)$ an average value given by

$$\langle F_t\rangle=\int_{\mathcal{F}(t)}\mu_t(dx)F_t(x)\stackrel{def}{=}\int_{\mathcal{F}(0)}\mu_0(dx)F_t(S_tx)\qquad(2.1)$$

We shall also consider the probability distribution $\mu_{SRB,t}$ which is defined as the SRB distribution of the dynamical system obtained by "freezing" $V_t,\boldsymbol{\Phi}(t)$ at the value that they take at time t and imagining the time to further evolve until the stationary state $\mu_{SRB,t}$ is reached: *in general* $\mu_t\ne\mu_{SRB,t}$. Of course the existence of $\mu_{SRB,t}$ has to be assumed *a priori* and this can be implied by the general *Chaotic Hypothesis*, [9]: which states that

Motions of a chaotic system, developing on its attracting set, can be assumed to follow an evolution with the properties of an Anosov system.

This means that in the physical problems just posed, (1.5) with $\mathbf{F},\boldsymbol{\vartheta}$ time independent, the motions are so chaotic that the attracting sets on which their long time motion takes place can be regarded as smooth surfaces on which motion is highly unstable. More precisely

(i) around every point of $\mathcal{F}$ three planes $M_s(x)$, $M_u(x)$, $M_m(x)$ can be identified which vary continuously with x, which are *covariant* (*i.e.* the planes at a point x are are mapped, by the evolution flow S_t, into the corresponding planes around S_tx) and

(ii) the planes, called *stable, unstable and marginal*, with respective positive dimensions $d_s>0,d_u>0$ and $d_m=1$ adding up to the dimension of $\mathcal{F}$, have the property that infinitesimal lengths on the stable plane and on the unstable plane of any point contract at exponential rate as time proceeds towards the future or towards the past, respectively. The length along the marginal direction neither contracts nor expands (*i.e.* it varies relative to the initial value staying bounded away from 0 and ∞): its tangent vector is parallel to the velocity in phase space. In cases in which time evolution is discrete, and determined by a map S, the marginal direction is missing.

(iii) over a time t, positive for the stable plane and negative for the unstable plane, the lengths contraction is exponential, *i.e.* lengths contract by a factor uniformly bounded by $Ce^{-\kappa|t|}$ with $C,\kappa>0$.

(iv) there is a dense trajectory.

The Ergodic Hypothesis provides us with an expression for the equilibrium averages (as integrals over the normalized Liouville distribution on the energy surface): likewise the Chaotic Hypothesis provides us with the *existence* and a *formal expression* for the averages (*i.e.* for the SRB distribution), [10].

Of course the hypothesis is only a "beginning" and one has to learn how to extract information from it, as it was the case with the Liouville distribution once the Ergodic Hypothesis guaranteed it as an appropriate distribution for the study of the statistics of motions, [11].

3. Heat generation

The work L_a in (1.3) will be interpreted as *heat* $\dot{Q}_a$ ceded, per unit time, by the particles in $\mathcal{C}_0$ to the a-th thermostat (because the "temperature" of $\mathcal{C}_a$ remains constant). The entropy creation rate due to heat exchanges between the system and the thermostats can, therefore, be naturally defined as

$$\sigma^0(\dot{\mathbf{X}},\mathbf{X}) \stackrel{def}{=} \sum_{a=1}^{N_a} \frac{\dot{Q}_a}{k_B T_a} \tag{3.1}$$

Remarks: (1) It is natural to define (3.1) as "heat-exchanges entropy creation" rate because it is the amount of work per unit time that the system performs on the thermostats which nevertheless keep their temperature (identified with the kinetic energy) constant.
(2) The above definition of "heat exchanges entropy creation" *does not* require any definition of entropy itself for the system: it only requires the notion of entropy variation of a reservoir considered as a system able to absorb work without changing it into mechanical work nor into its temperature variations. The latter is the "classical" definition of entropy variation of a reservoir.
(3) If the thermostats temperatures are equal, *i.e.* $K_a \sim \frac{3}{2}N_a\beta^{-1}$ and $\boldsymbol{\Phi} = \mathbf{0}$, or $\boldsymbol{\Phi}$ is conservative so that its potential can be imagined merged in U_0, it is possible to check that the SRB distribution, if existing, is necessarily

$$\begin{aligned}\mu_{SRB}(dx) =& Z^{-1} e^{-\beta(K_0+U_0+\sum_{a>0}U_a+\sum_{a>0}W_a)}\cdot \\ &\cdot \prod_{a>0}\delta(K_a - \tfrac{3N_a-1}{2\beta})\cdot d\mathbf{V}\,d\mathbf{X}\end{aligned} \tag{3.2}$$

where $dx = d\mathbf{V}\,d\mathbf{X}$ = phase space volume element, K_a are the kinetic energies of the various subsystems, $a = 0, 1, \ldots$ and Z is a normalization factor. Hence this is a Gibbs distribution in one of the equivalent, if unusual, *equilibrium* ensembles at temperature $(k_B\beta)^{-1}$ and densities $\rho_a = \frac{N_a}{|\mathcal{C}_a|}$, $a = 0, \ldots, n$. This is an important consistency check, following (and extending) [5].

4. Irreversibility degree

It is one of the basic tenets in Thermodynamics that all (nontrivial) processes between equilibrium states are "irreversible": only idealized (strictly speaking nonexistent) "quasi static" processes can be reversible. The question that is addressed in the following is whether irreversibility can be made a quantitative notion at least in models based on microscopic evolution, (1.5).

Write (1.5) as $\ddot{\mathbf{X}} = \boldsymbol{\Xi}(\dot{\mathbf{X}},\mathbf{X}) = \boldsymbol{\Xi}(x)$ with $x = (x_1, \ldots, \dot{x}_1, \ldots, \dot{x}_{3N_{tot}})$, $N_{tot} = \sum_a N_a$. Given a process Γ let

$$\sigma_t^\Gamma(x) \stackrel{def}{=} -\,\mathrm{divergence}(x) \equiv -\sum_{i=1}^{3N_{tot}} \boldsymbol{\partial}_{\dot{x}_i}\boldsymbol{\Xi}_i(x) \tag{4.1}$$

which has the meaning of rate at which the phase volume around x contracts because of the forces action.

Phase space volume can also change because new regions become accessible (or inaccessible) so that the total phase space contraction rate, denoted $\sigma_{tot,t}$, in general will be different from σ_t^Γ.

It is reasonable to suppose, and in some cases it can even be proved (*e.g.* in the example considered in remark (3) of Sect.5), that at every time t the configuration $S_{0,t}x$ is a "typical" configuration of the "frozen" system if the initial x was typical for the initial distribution μ_0: *i.e.* it will be a point in $\mathcal{F}(t) = V_t^N \times R^N \times \prod \mathcal{F}_a$ whose statistics under forces imagined frozen at their values at time t will be $\mu_{SRB,t}$, see comments following (2.1). Therefore phase space volume $|\mathcal{F}_t|$ contracts at rate

$$\sigma_t^v(x) = -\frac{1}{|\mathcal{F}_t|}\frac{d\,|\mathcal{F}_t|}{dt} = -N\frac{\dot{V}_t}{V_t} \tag{4.2}$$

which does not depend on x as it is a property of the phase space available to (any typical) x. A calculation of the divergence $-\sigma^\Gamma$ of (1.5), see also (1.3), shows that

$$\sigma^\Gamma(\dot{\mathbf{X}},\mathbf{X}) = \sum_{a>0} \tfrac{3N_a-1}{3N_a}\tfrac{\dot{Q}_a-\dot{W}_a}{k_B T_a} \simeq \sum_{a>0}\tfrac{\dot{Q}_a}{k_B T_a} - \dot{U} \tag{4.3}$$

if $U \stackrel{def}{=} \sum_a \frac{(3N_a-1)}{3N_a}\frac{W_a}{k_B T_a}$; in the last step $O(N_a^{-1})$ has been neglected (for simplicity and for consistence with macroscopic thermodynamics where N_a are extremely large: it would not be difficult to keep track of it, however). Therefore the total phase space contraction per unit time can be expressed as, see (1.3),(3.1),

$$\sigma_{tot}(\dot{\mathbf{X}},\mathbf{X}) = \sum_a \frac{\dot{Q}_a}{k_B T_a} - N\frac{\dot{V}_t}{V_t} - \dot{U} \tag{4.4}$$

i.e. there is a simple and direct relation between phase space contraction and entropy creation rate, [6]. Eq.(4.3) shows that their difference is a "total time derivative".

In studying stationary states $N\dot{V}_t/V_t = 0$, and the interpretation of $\dot{U}$ is of "reversible" heat exchange between system and thermostats. In this case, in some respects, the difference $\dot{U}$ can be ignored. For instance in the study of the fluctuations of the average over time of entropy creation rate in a *stationary state*: the term $\dot{U}$ gives no contribution, or it affects only the very large fluctuations, [11, 12] if the containers $\mathcal{C}_a$ are very large (or if the forces between their particles can be unbounded, which we are excluding here for simplicity, [11]).

Even in the case of processes the quantity $\dot{U}$ has to oscillate in time with 0 average on any interval of time (t, ∞) if the system starts and ends in a stationary state.

For the above reasons we define the *entropy creation rate* in a process to be (4.4) *without* the $\dot{U}$ term:

$$\varepsilon(\dot{\mathbf{X}},\mathbf{X}) = \sum_a \frac{\dot{Q}_a}{k_B T_a} - N\frac{\dot{V}_t}{V_t} \tag{4.5}$$

It is interesting, and necessary, to remark that in a stationary state the time averages of ε, denoted ε_+, and of $\sum_a \frac{\dot{Q}_a}{k_B T_a}$, denoted $\sum_a \frac{\dot{Q}_{a+}}{k_B T_a}$, coincide because $N\dot{V}_t/V_t = 0$, as $V_t = const$, and $\dot{U}$ has zero time average being a total derivative. On the other hand under very general assumptions, much weaker than the Chaotic Hypothesis, the time average of the phase space contraction rate is ≥ 0, [13, 14], so that in a stationary state:

$$\varepsilon_+ \equiv \sum_a \frac{\dot{Q}_{a+}}{k_B T_a} \geq 0 \tag{4.6}$$

which is a consistency property that has to be required for any proposal of definition of entropy creation rate.

Remarks: (1) In stationary states the above models are a realization of *Carnot machines*: the machine being the system in $\mathcal{C}_0$ on which external forces $\mathbf{F}$ work leaving the system in the same stationary state (a special "cycle") but achieving a transfer of heat between the various thermostats (in agreement with the second law, see (3.1),(4.6), only if $\varepsilon_+ \geq 0$).
(2) In a stationary state $\sigma_{tot} = \sum_a \frac{\dot{Q}_a}{k_B T_a} + \dot{U}$ will satisfy, under the Chaotic Hypothesis of Sec.2, a fluctuation relation, see [7, 9], since the dynamics (1.5) is *time reversible*: the large fluctuations of $p = \frac{1}{\varepsilon_+}\frac{1}{T}\int_0^T (\sum_{a>0} \frac{\dot{Q}_a}{k_B T_a} + \dot{U})\, dt$ will be controlled by a *rate function* $\zeta(p)$ satisfying $\zeta(-p) = \zeta(p) - p\varepsilon_+$, see [7, 11]. However the term $\dot{U}$ will not contribute to the relation because U is bounded, [11]. Therefore the fluctuations of $p = \frac{1}{\varepsilon_+}\frac{1}{T}\int_0^T \sum_{a>0} \frac{\dot{Q}_a}{k_B T_a}\, dt$ will have exactly the *same* rate function.
(3) Note, however, that U can be very large (being of the order of $\sum_{a>0} N_a$) if the thermostats are large and its contribution to the averages of σ_{tot} tends to 0 as slowly as $T^{-1}O(\sum_{a>0} N_a)$. Hence the time necessary to see the fluctuation relation satisfied with prefixed accuracy when p is defined as the average of σ_{tot} may be enormously larger than the time necessary to see the fluctuation relation satisfied with p defined as the average of $\sum_a \frac{\dot{Q}_a}{k_B T_a}$.
(4) Therefore the Fluctuation Theorem holds for the physically interesting $\sum_a \frac{\dot{Q}_a}{k_B T_a}$ *as a consequence* of its validity for σ_{tot} and it becomes visible over much shorter time scale, independent on the size of the thermostats.

Coming back to the question of defining a irreversibility degree of a process Γ we distinguish between the (non stationary) state μ_t into which the initial state μ_0 evolves in time t, under varying forces and volume, and the state $\mu_{SRB,t}$ obtained by "freezing" forces and volume at time t and letting the system settle to become stationary, see comments following Eq. (2.1). We call ε_t the entropy creation rate (4.5) and ε_t^{srb} the entropy creation rate in the "frozen" state $\mu_{SRB,t}$.

The proposal is to define *irreversibility degree* $\mathcal{I}(\Gamma)$ or *irreversibility time scale* $\mathcal{I}(\Gamma)^{-1}$ of a process Γ by setting:

$$\mathcal{I}(\Gamma) = \frac{1}{N^2}\int_0^\infty \Big(\langle\varepsilon_t\rangle_{\mu_t} - \langle\varepsilon_t^{srb}\rangle_{SRB,t}\Big)^2 dt \tag{4.7}$$

If the Chaotic Hypothesis is assumed then the state μ_t will evolve exponentially fast under the "frozen evolution" to $\mu_{SRB,t}$. Therefore the integral in (4.7) will converge for reasonable t dependences of $\boldsymbol{\Phi}, V$.

A physical definition of "quasi static" transformation is a transformation that is "very slow". This can be translated mathematically into an evolution in which $\boldsymbol{\Phi}_t$ evolves like, if not exactly, as

$$\boldsymbol{\Phi}_t = \boldsymbol{\Phi}_0 + (1 - e^{-\gamma t})(\boldsymbol{\Phi}_\infty - \boldsymbol{\Phi}_0). \tag{4.8}$$

An evolution Γ close to quasi static, but simpler for computing $\mathcal{I}(\Gamma)$, would proceed changing Φ_0 into $\Phi_\infty = \Phi_0 + \Delta$ by Δ/δ steps of size δ, each of which has a time duration t_δ long enough so that, at the k-th step, the evolving system settles onto its stationary state at field $\Phi_0 + k\delta$. If the corresponding time scale can be taken $= \kappa^{-1}$, independent of the value of the field so that t_δ can be defined by $\delta e^{-\kappa t_\delta} \ll 1$, then $\mathcal{I}(\Gamma) = const\,\delta^{-1}\delta^2 \log\delta^{-1}$ because the variation of $\sigma_{(k+1)\delta,+} - \sigma_{k\delta,+}$ is, in general, of order δ as a consequence of the differentiability of the SRB states with respect to the parameters, [15].

5. Comments and examples

Remarks: **(1)** A drawback of the definition proposed in Sec.4 is that although $\langle\varepsilon_t^{srb}\rangle_{SRB,t}$ is *independent* on the metric that is used to measure volumes in phase space the quantity $\langle\varepsilon_t\rangle_{\mu_t}$ *depends* on it. Hence the irreversibility degree considered here reflects also properties of our ability or method to measure (or to imagine to measure) distances in phase space. One can keep in mind that a metric independent definition can be simply obtained by minimizing over all possible choices of the metric: but the above simpler definition seems, nevertheless, preferable.

(2) It is $\mathcal{I}(\Gamma) \geq 0$, but I have not been able to prove that $\mathcal{I}(\Gamma) > 0$ except in the limit of a quasi static process: this is a natural question to raise.

(3) Suppose that a process takes place because of the variation of an acting conservative force, for instance because a gravitational force changes as a large mass is brought close to the system, while no change in volume occurs and the thermostats have all the same temperature. Then the "frozen" SRB distribution, for all t, is given by (3.2) and $\langle\varepsilon^{srb}\rangle_{SRB,t} = 0$ (because the "frozen equations" admit a SRB distribution which has a density, see Eq. (3.2), in phase space). The isothermal process thus defined has *therefore* $\mathcal{I}(\Gamma) > 0$ except, possibly, if is *isentropic*, *i.e.* if initial and final thermodynamic entropies are equal. In fact the latter are cases in which the initial, intermediate and final states are described by a density function $\rho_t(x)$ on phase space hence, [13], it is $\langle\varepsilon_t^{srb}\rangle_{SRB,t} = 0$ and,[16], $\langle\varepsilon_t\rangle_{\mu_t} = -\frac{d}{dt}\int \rho_t(x)\log\rho_t(x)\, dx$. Since the initial and final states are equilibrium states this means that if the initial and final thermodynamic entropies (*i.e.* Gibbs') are different it must be $\langle\varepsilon_t\rangle_{\mu_t} \neq 0$ on a set of positive measure and $\mathcal{I}(\Gamma) > 0$ and we expect it to be of order δ (see last comment in Sect.4). On the other hand if the thermodynamic entropies are the same and Γ is almost quasi static then the discussion at the end of Sec.4 shows that the irreversibility time scale may tend to $+\infty$ when $\delta \to 0$ much faster than in the cases in which the entropies differ (*e.g.* as δ^{-3} rather than as δ^{-1}).

(4) The comment (3) shows that if a connection between thermodynamic entropy and irreversibility scale can be established at all it will neither be too direct nor too simple because in the considered isothermal process linking two equilibrium states there is, in general, a non zero entropy variation hence $\mathcal{I}(\Gamma)>0$ but, if performed very slowly, it can be close to be reversible both in the classical Thermodynamics and in the irreversibility scale senses.

(5) Consider a typical irreversible process. Imagine a gas in an adiabatic cylinder covered by an adiabatic piston and imagine to move the piston. The simplest situationstions arises if gravity is neglected and the piston is suddenly moved at speed w.

Unlike the cases considered so far the absence of thermostats (adiabaticity of the cylinder) imposes an extension of the analysis. In this case the phase space is the surface of constant energy in $\mathcal{C}_0^{N_0}\times R^{3N_0}$ rather than $\mathcal{C}_0^{N_0}\times R^{3N_0}$. Therefore if the piston moves at a speed which is too slow (*i.e.* slower than the maximum velocity of the particles on the energy surface) it will change the total energy of the gas in a way that depends on the special configuration initially chosen.

Therefore the simplest situation arises when the piston is moved at speed so large that no energy is gained or lost by the particles because of the collisions with the moving wall (there are no such collisions). This is an extreme idealization of the classic Joule-Thomson experiment.

Let S be the section of the cylinder and $H_t=H_0+w\,t$ be the distance between the moving lid and the opposite base. Let $\Omega=S\,H_t$ be the cylinder volume. In this case the volume of phase space changes only because the boundary moves and it increases by $N\,w\,S\,\Omega^{N-1}$ per unit time, *i.e.* its rate of increase is $N\frac{w}{H_t}$.

Hence $\langle\varepsilon_t\rangle_t$ is $-N\frac{w}{H_t}$, while $\varepsilon_t^{srb}\equiv0$ if $T=\frac{L}{w}$ is the duration of the transformation ("Joule-Thomson" process) increasing the cylinder length by L, then

$$\mathcal{I}(\Gamma)=N^{-2}\int_0^T N^2\big(\frac{w}{H_t}\big)^2\,dt\xrightarrow[T\to\infty]{}w\frac{L}{H_0(H_0+L)}\qquad(5.1)$$

and the transformation is irreversible. The irreversibility time scale approaches 0 as $w\to\infty$, as possibly expected. If $H_0=L$, i.e. if the volume of the container is doubled, then $I(\Gamma)=\frac{w}{L}$ and the irreversibility time scale of the process coincides with its "duration".

(6) A different situations arises if in the context of (5) the piston is replaced by a sliding lid which divides the cylinder in two halves of height L each: one empty at time zero and the other containing the gas in equilibrium. At time 0 the lid is lifted and a process Γ' takes place. In this case $\dot V_t=V\delta(t)$ because the volume $V=S\,L$ becomes suddenly double. Therefore the evaluation of the irreversibility scale yields

$$\mathcal{I}(\Gamma')=N^{-2}\int_0^\infty N^2\delta(t)^2\,dt\equiv+\infty\qquad(5.2)$$

so that the irreversibility becomes immediately manifest, $I(\Gamma')=+\infty$, $\mathcal{I}(\Gamma')^{-1}=0$. This idealized experiment is rather close to the actual Joule-Thomson experiment.

In the latter example it is customary to estimate the degree of irreversibility at the lift of the lid by the *thermodynamic equilibrium entropy* variation between initial and final states. It would of course be interesting to have a general definition of entropy of a non stationary state (like the states μ_t at times $(t\in(0,\infty)$ in the example just discussed) that would allow connecting the degree of irreversibility to the thermodynamic entropy variation in processes leading from an initial equilibrium state to a final equilibrium state, see [17].

Acknowledgements: I am indebted to F. Bonetto, A. Giuliani and F. Zamponi for many ideas, discussions and for earlier collaborations that preceded and stimulated the development of this work.

[1] G. Gallavotti, Chaos **14**, 680 (2004).
[2] G. Gallavotti, Physica D **112**, 250 (1998).
[3] G. Gallavotti, cond-mat/0402676 (2004).
[4] D. Ruelle, Journal of Statistical Physics **95**, 393 (1999).
[5] D. Evans and G. Morriss, *Statistical Mechanics of Nonequilibrium Fluids* (Academic Press, New-York, 1990).
[6] G. Gallavotti, cond-mat/0510027 (2005).
[7] G.Gallavotti, Open Systems and Information Dynamics **6**, 101 (1999).
[8] J.P.Eckmann, C. Pillet, and L. R. Bellet, Communications in Mathematical Physics **201**, 657 (1999).
[9] G. Gallavotti and E. Cohen, Physical Review Letters **74**, 2694 (1995).
[10] G. Gallavotti, F. Bonetto, and G. Gentile, *Aspects of the ergodic, qualitative and statistical theory of motion* (Springer Verlag, Berlin, 2004).
[11] F. Bonetto, G. Gallavotti, A. Giuliani, and F. Zamponi, cond-mat/0507672 (2005).
[12] R. V. Zon and E. Cohen, Physical Review Letters **91**, 110601 (+4) (2003).
[13] D. Ruelle, Journal of Statistical Physics **85**, 1 (1996).
[14] D. Ruelle, Communications in Mathematical Physics **189**, 365 (1997).
[15] D. Ruelle, Communications in Mathematical Physics **187**, 227 (1997).
[16] L. Andrej, Physics Letters **111A**, 45 (1982).
[17] S. Goldstein and J. Lebowitz, Physica D **193**, 53 (2004).

REVTEX & EQALIGNNO & FANCYHDR