

Some problems in nonequilibrium statistical mechanics

Giovanni Gallavotti
Fisica and I.N.F.N. Roma 1

Equilibrium statistical mechanics and classical thermodynamics deal with systems in equilibrium, *i.e.* systems of (newtonian) particles described by Hamiltonian equations, and with the results of their quasi static transformations.

While statistical mechanics of equilibrium systems is a self contained purely mechanical theory, thermodynamics makes use of “thermostats” empirically defined as systems capable of exchanging heat with the system without performing work: but they are not characterized as mechanical systems.

Recently many efforts have been devoted to extending statistical mechanics to nonequilibrium situations. The next more complex class of states should be the theory of *stationary* states in systems out of equilibrium.

These are states of systems on which non conservative forces act performing steadily work, called *thermodynamic forces*, and generating currents, called *thermodynamic fluxes* to follow Onsager’s terminology. One could also introduce, in their study, thermostats: however the thermostats role is much more important than in the equilibrium cases where they entered only in the description of the transformations between equilibrium states but were not necessary to describe the equilibria themselves.

Out of equilibrium the work done by the external forces will have to be continuously dissipated and it seems difficult to develop a mechanical theory of stationary states out of equilibrium without introducing specific thermostat models of mechanical nature.

In the 1980’s this became particularly evident in the experiments on molecular dynamics and thermostats modeled by suitable forces acting on the system were introduced: the Nosé–Hoover thermostats or the Gaussian thermostats of Evans, Morriss, Cohen. Other possible thermostats are stochastic thermostats or thermostats consisting themselves in infinite systems of particles interacting via conservative forces, with both the system particles and between themselves, but asymptotically (in space) in thermal equilibrium (*i.e.* described “far” from the thermostatted system) by a Maxwell–Boltzmann distribution).

The mechanical thermostats approach revealed itself as very useful for computational purposes and (a surprise, perhaps) generated a wealth of ideas and problems allowing even to establish a link with the contemporary theories of chaos in mechanics and fluids.

I illustrate here a few aspects of the “mechanical thermostats” approach as applied to simple systems. A simple system will be described by a differential equation in its phase space: we write it as $\dot{x} = X_E(x)$ where $x = (\dot{\mathbf{q}}, \mathbf{q}) \in R^{6N} \equiv \Omega$ (*phase space*), N =number of particles, m = mass of the particles, with

$$m\ddot{\mathbf{q}} = f(\mathbf{q}) + E \mathbf{g}(\mathbf{q}) - \vartheta_E(\dot{\mathbf{q}}, \mathbf{q}) \equiv X_E(x)$$

where $f(\mathbf{q})$ describes the internal (conservative) forces (*e.g.* hard cores), $E \mathbf{g}(\mathbf{q})$ represents the “external force” (nonconservative) acting on the system, with E being its “strength”: for definiteness we suppose that it is locally conservative (like an electromotive force) but not globally such, and ϑ_E is the force law which models the action exerted by the thermostat on the system to keep it from indefinitely acquiring energy: this is why I shall call it a *mechanical thermostat*.

Prescribing a specific form for $\vartheta_E(\dot{\mathbf{q}}, \mathbf{q})$ will be avoided with care: we expect that the actual mechanism used to take away heat from the system should, to a large extent, be irrelevant. Therefore I will be mainly concerned by properties of the stationary states that are thermostat–independent provided the thermostat fulfills the basic requirement that it forces the system to develop motions in a compact region of phase space.

The most common systems that one wants to study in nonequilibrium statistical mechanics undergo chaotic evolution: I take this as an empirical fact. At the same time the evolution is not Hamiltonian, because of the thermostating forces, and the volume of phase space is not conserved but it contracts at a rate $\sigma(x)$ which depends on the point in phase space

$$\sigma(x) = -\operatorname{div}_x X_E(x) = \sum_{j=1}^{3N} \partial_{\dot{q}_j} \vartheta_{E,i}(x), \quad x = (\dot{\mathbf{q}}, \mathbf{q})$$

A system is called *dissipative* if the time average σ_+ of σ is *strictly positive*. Equilibrium systems, governed by Hamiltonian equations, are necessarily non dissipative.

Dissipativity, $\sigma_+ > 0$, has important consequences: namely the probability distribution that describes the frequency of visit to phase space regions, say small spheres, *cannot be represented by a density on phase space* because it is concentrated on a set of zero volume.

This is a major conceptual problem because at first sight it seems to imply that it will not be possible to make use of a “natural” probability distribution playing the role of the Liouville distribution in equilibrium. In other words it seems that in nonequilibrium stationary systems there might be nothing like the *ergodic hypothesis* to guide us.

However in the early 1970’s Ruelle had proposed that chaotic systems would behave in a rather universal way and that the only really relevant feature would be that they behaved as hyperbolic systems. This suggestion generated vast theoretical and experimental work, particularly in turbulence theory. Theoretically, because it allowed to link the theory of fluids to the theory of dynamical systems, and experimentally, because it led to many new experiments showing universal properties of the onset of turbulence.

His suggestion however was not taken too literally, until in 1995, it was interpreted, considering time evolution to be given by a map S (*e.g.* via a Poincaré section of the flow), as the

Chaotic hypothesis: *A chaotic multiparticle system can be regarded, on each of its attractors, as a transitive Anosov system.*

Anosov systems are a class of mathematically defined dynamical systems which although behaving chaotically are, nevertheless, “completely understood”. They play the role for chaotic dynamics that the harmonic oscillators play in regular dynamics. As familiar in mechanics a problem reduced to a harmonic oscillator problem is considered solved because we understand a lot about harmonic motions. The same is essentially true in chaotic dynamics: if a problem can be reduced to an Anosov system then we understand a lot about it.

I am not going to give the definition of Anosov system in the mathematical form. I only give its physical interpretation. One can say that

a system is Anosov if it admits “coarse graining” of any prefixed precision.

This means that one can find a partition of phase space (in fact many) (P_1, \dots, P_n) (where P_i has to be thought of as representing the collection of phase space points for which the outcome of a measurement with n possible results gives the result i , so that given a resolution γ there is a time T such that

(1) the sets $P = P \begin{pmatrix} -T & -T+1 & \dots & T \\ i_{-T} & i_{-T+1} & \dots & i_T \end{pmatrix}$ of points x such that $S^t x \in P_{i_t}$ for $t \in [-T, T]$, *i.e.* the set of points which between $-T$ and T visit the elements of the partition following a prescribed pattern i_{-T}, \dots, i_T , have diameter $\leq \gamma$.

(2) the pattern can be arbitrarily prescribed to be any sequence i_{-T}, \dots, i_T of possible “transitions” *i.e.* such that it is dynamically possible to go from the interior $P_{i_k}^0$ of P_{i_k} to the interior of $P_{i_{k+1}}$ ($SP_{i_k}^0 \cap P_{i_{k+1}}^0 \neq \emptyset$) and in such case P is not empty.

In other words if we are interested only in observables that resolve phase space with precision γ we can represent faithfully phase space as a set of sequences $\{i_t\}_{t=-T}^T$ which can be arbitrarily assigned provided they fulfill a nearest neighbor compatibility condition.

We can say that *the microscopic states of the system can be represented as biinfinite sequences of symbols fulfilling a nearest neighbor compatibility*: such sequences are *very familiar* in statistical mechanics of lattice systems being the states of a Markov chain or of an Ising model with nearest neighbor hard core (expressed by the compatibility of the symbolic patterns).

This remark together with the theorem that the Liouville distribution can, for such systems, be represented as a Gibbs distribution with a short range potential which is not translation invariant (due to the non translation invariance of the Liouville measure) but which has different translation invariant potentials to the right (*i.e.* future) and to the left (*i.e.* past) of the origin, leads to the surprising consequence that *the statistical properties of the system are described by Gibbs states of a one dimensional lattice gas with nearest neighbor hard core and short range potential*.

Such kinds of Gibbs states are well known and considered “trivial” in statistical mechanics of lattice systems: so that by saying that the system is Anosov we are giving a very deep description of its properties in terms of well understood objects and for this reason Anosov systems can be considered the paradigm of chaotic evolution like the harmonic oscillators are the paradigm of regular motions.

What can be done with the hypothesis? first it guarantees the existence of time averages of smooth observables (as a theorem for transitive Anosov maps)

Proposition: *(SRB distribution) if initial data x are randomly selected in the basin of an attractor all of them, apart from a set of zero volume, are such that for all smooth observables F on phase space Ω*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^T F(S^k x) = \int_{\Omega} F(y) \mu(dy)$$

and μ is uniquely associated with the attractor.

Thus it makes sense to talk about *statistics of the motions* and we see that *the chaotic hypothesis extends the ergodic hypothesis* because in the Hamiltonian cases it insures that the Liouville distribution is the statistics of all motions (aside a set of zero volume), hence it is ergodic.

In particular the time average σ_+ of the phase space contraction is well defined and we have a simple definition of dissipativity. Under the chaotic hypothesis (*i.e.* for Anosov systems) it can be proved (Ruelle) that $\sigma_+ \geq 0$.

Note that μ is not representable by a density if $\sigma_+ > 0$. Nevertheless there is a unique probability distribution which is invariant and which describes the statistics of all data but a set of zero volume.

Can one go beyond the above result which should be regarded as a tautological consequence of the assumptions? Following the path traced by Boltzmann we look for general consequences that can be in principle tested experimentally (either numerically or *in vivo*).

We recall that Boltzmann derived from the ergodic hypothesis the *heat theorem*: namely given a gas in a container and assuming ergodicity and given the control parameters, *e.g.* volume V and energy U , of the equilibria we form a family of probability distributions $\mu_{U,V}$ which are the microcanonical distributions of our system and provide a microscopic statistical description of the equilibrium states.

Then we can compute the time averages that are relevant for thermodynamics like the pressure P , average force per unit surface, and $\frac{3}{2}N\Theta$, average total kinetic energy. The

heat theorem, derived by Boltzmann from the ergodic hypothesis, says that if we change the control parameters by dU and dV respectively then

$$\frac{dU + PdV}{\Theta} = \text{exact differential}$$

which is clearly a remarkable achievement: a universal parameter free relation between observable quantities.

In chaotic dynamics, essentially in all cases examined, it is either natural or at least consistent to interpret the phase space contraction rate σ_+ as the average entropy production rate of the system.

Here I do not want to take this for granted: I warn that it is highly controversial, as anything related to entropy always is. However I think that this is a proper *universal definition* of a notion that still needs to be defined and therefore I proceed to show that it has some remarkable properties.

We imagine to study the quantity

$$p_\tau(x) = \frac{1}{\tau} \sum_{-\frac{\tau}{2}}^{\frac{\tau}{2}-1} \sigma(S^k x)$$

i.e. the average phase space contraction observed on the evolution of x during a time interval τ . Since it depends on x it is a random variable which, in the stationary state, inherits its probability distribution from the SRB distribution μ . We then look for the probability that $p_\tau(x)$ has a value in an interval Δ .

Since p is an average of many quantities which take essentially independent values we expect a *large deviation* property to hold for the distribution of p in the SRB state μ ; *i.e.* the probability $\Pi_\tau(\Delta)$ that $p \in \Delta$ should be for suitable *finite* p_1^*, p_2^*

$$\log \Pi_\tau(\Delta) \simeq \max_{p \in \Delta} \tau \zeta(p), \quad \text{for all } p \in (p_1^* \sigma_+, p_2^* \sigma_+)$$

and the \simeq means that the error is uniform in p in any closed subinterval of $(p_1^* \sigma_+, p_2^* \sigma_+)$ and for τ arbitrarily large. Furthermore if $\sigma_+ > 0$ $p_1^* < p_2^*$ (a property sometimes called “multifractality” of the distribution of p) and the function $\zeta(p)$ is analytic in this interval (a consequence of the fact that the SRB distribution can be represented as a Gibbs state of a one dimensional Ising system with short range interaction hence with no phase transitions).

Suppose that the system is “reversible”: *i.e.* there is an isometry on phase space I such that

$$IS = S^{-1}I, \quad I^2 = 1$$

this could be the velocity reversal or more generally any map with the above property (*e.g.* $I = PT$ where P is the parity isometry and T the velocity reversal or PCT or much more involved symmetries).

Such systems arose naturally after the Nosé–Hoover thermostats were introduced and they are related to Gauss’ least constraint principle which is remarkably related to mechanical thermostats.

In reversible Anosov systems time reversal symmetry implies that $-p_1^* = p_2^* = p^*$.

Then the following *parameter free* relation holds as a theorem for transitive reversible Anosov systems:

Proposition: (*Fluctuation theorem*) the rate function for the large deviations of the entropy creation fluctuations verifies the symmetry relation. called the fluctuation relation,

$$\zeta(-p) = \zeta(p) - p, \quad \text{for all } |p| < p^* \sigma_+$$

and $p^* > 0$ if $\sigma_+ > 0$.

Remark: In particular in the case of Hamiltonian systems the theorem is trivial but it becomes non trivial as soon as $\sigma_+ > 0$. In the Hamiltonian case it is, indeed, easy to see that the variable p_τ even tends to 0 as $\tau \rightarrow \infty$ uniformly in x . For this reason it is better to formulate it as a large deviation symmetry for the quantity $a = \frac{p_\tau(x)}{\sigma_+}$ where it takes the form $\bar{\zeta}(-a) = \bar{\zeta}(a) - a\sigma_+$ for $|a| < p^* < +\infty$ (and $\bar{\zeta}$ and ζ are trivially related). Forgetting the important condition on the range of validity in p is easier if the theorem is written in the form $\zeta(-p) = \zeta(p) - p$ and indeed this has generated a rather surprising confusion in the literature as well as errors and paradoxes.

The above result was observed in a numerical experiment before its subsequent proof for Anosov systems: and the proof was motivated and inspired by the result of the experiment itself. It has since been checked in several numerical experiments and the fact that it has been confirmed can be taken as evidence in favor of the validity of the chaotic hypothesis.

Of course the chaotic hypothesis in the form proposed above is extremely ambitious because of its generality and a more checks are desirable: since the statement checks have been produced rather steadily.

Before trying to connect the above fluctuation relation with experiments I mention another remarkable result that can be proved in the same way. Let F_1, \dots, F_n be n “arbitrary” observables which are odd under time reversal $F_j(Ix) = -F_j(x)$. Let $\varphi_1, \dots, \varphi_m$ be n functions defined in the interval $[-\frac{\tau}{2}, \frac{\tau}{2}]$ and define $I\varphi(t) \stackrel{def}{=} -\varphi(-t)$. Consider the joint probability of the event that $F_j(S^t x) \simeq \varphi_j(t)$ i.e. that F_j evolves following the pattern φ_j and $p_\tau(x) = p$ call this $\Pi_\tau(F_j \simeq \varphi_j, p)$. Then

Proposition: Given a transitive reversible Anosov system and fixed arbitrarily n smooth observables F_1, \dots, F_n

$$\frac{P_\tau(F_j \simeq \varphi_j, p)}{P_\tau(F_j \simeq I\varphi_j, -p)} = e^{\tau p + O(1)}, \quad |p| < p^* \sigma_+$$

This is a result of the kind considered by Onsager–Machlup but without the restriction of being close to equilibrium. One should note its universality and its independence on the choice of the observables F_j . The error term $O(1)$ is uniform in p in closed intervals in $(-p^*, p^*)$, in τ and in the choice of the φ functions provided the patterns $F_j = \varphi_j$ and the value p are simultaneously possible (otherwise one gets 0/0 and the formula should be interpreted appropriately).

The fluctuation theorem becomes trivial if $\sigma_+ = 0$: however one is often interested in systems that depend on parameters \mathbf{E} and which for $\mathbf{E} = \mathbf{0}$ become Hamiltonian. In this case the fluctuation relation given by theorem (i.e. $\bar{\zeta}(-a) = \bar{\zeta}(a) = a\sigma_+$, see above) can be divided by a suitable power of the fields \mathbf{E} and still give a non trivial result in the limit $\mathbf{E} \rightarrow \mathbf{0}$: it has been shown, heuristically, that indeed in this way one gets nontrivial relations between quantities that are naturally interpreted as transport coefficients. namely one obtains Onsager reciprocity and Green–Kubo formulae.

Hence in this sense the fluctuation relation extends to “large” forcing” the Onsager reciprocity.

A theory than can be checked only in numerical experiments leaves us somewhat dissatisfied but there are major obstacles to checking the theory on natural systems.

The first obstacle is that natural systems are large and the rate function will tend to depend on the size of the system making fluctuations of the entropy production virtually unobservable. In fact this is an obstacle already for small systems and it is surprising

that the fluctuation relation can be observed (and confirmed) in numerical experiments in which the number of particles can go up to the order of the hundreds.

However we can observe fluctuations in small volumes of a large system: and such a small volume can be regarded as a subsystem thermostatted by the rest of the system: hence we can imagine to apply the theory to the subsystem.

An interesting example can be derived from a classical theory of electric conduction going back to the dawn of statistical mechanics. This is Drude's theory in which N particles of unit mass and charge move in a medium of obstacles, "crystal", under the influence of an electric field E in the direction \mathbf{u} of the x -axis (say) and after each collision with the spheres of the medium their speed is reset to $\sqrt{3k_B\Theta}$ where Θ is the "temperature" of the medium. This is a model of a thermostatted system.

A simpler model for our purposes is to thermostat the system by a force which rather than resetting the speed at each collision makes it strictly constant $\sqrt{k_B\Theta}$: this is

$$\ddot{\mathbf{q}}_i = E\mathbf{u} - \alpha\dot{\mathbf{q}}_i, + \text{elastic collision rule}$$

where $\alpha = \frac{E \sum_i \dot{\mathbf{q}}_{i,u}}{\sum_i \dot{\mathbf{q}}_i^2}$ and the energy $\sum_i \dot{\mathbf{q}}_i^2$ is *constant* and fixed equal to $3Nk_B\Theta$, which is a slightly different thermostat. Note that the divergence is in this case

$$\sigma(x) = \left(1 - \frac{1}{3N}\right) \frac{EJ}{k_B\Theta}$$

where $J = \sum_i \dot{\mathbf{q}}_{i,u}$ is the current in the direction \mathbf{u} of the field $E\mathbf{u}$.

It is a conjecture that the two thermostats are equivalent in the sense that in the limit of large boxes the statistical properties of observables depending only on what can be seen in a fixed finite region ("local observable") are the same for the Drude's and the Gaussian thermostats.

The geometry is very simple and the position space is described in Fig.1:

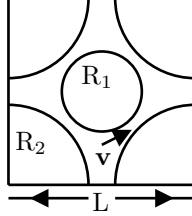


Fig.1: General billiard structure with scatterers of radius R_1 and R_2 in a periodic box with side length L , (case $k \times \ell = 1 \times 1$).

The box $[-\frac{k}{2}L, \frac{k}{2}L] \times [-\frac{\ell}{2}L, \frac{\ell}{2}L]$ consists of $k \cdot \ell$ unit lattice cells with side L joined to form a square box: at the box boundary we impose periodic boundary conditions (pb) or, alternatively, semi periodic boundary conditions ($\frac{1}{2}pb$): in the latter case the "horizontal" box walls are reflecting. Periodicity is essential as it "closes the circuit" and it makes the electric field nonconservative making at the same time necessary a thermostat to avoid that the system energy increases indefinitely.

In modern terms the thermostat is a model for the interaction between the electrons and the phonons of the lattice which provide the physical thermalization mechanism. The number of particles is N equal, say, to the number of obstacles and the particles are not quite independent because the thermostat depends on all the velocities: it is a conjecture that Drude's model with the thermostat acting only at the collision moment.

Note that, *in full generality*, in a stationary state the average work of the field has to be exactly compensated by the work of the thermostat

$$\langle E \mathbf{g}(\mathbf{q}) \cdot \dot{\mathbf{q}} \rangle \equiv \langle \dot{\mathbf{q}} \cdot \vartheta_{\mathbf{E}}(\dot{\mathbf{q}}, \mathbf{q}) \rangle$$

if the inner forces are conservative and the work $\langle \dot{\mathbf{q}} \cdot \vartheta_{\mathbf{E}}(\dot{\mathbf{q}}, \mathbf{q}) \rangle$ has the interpretation of heat absorbed by the thermostat. In the example aboved the heat ceded to the reservoir is

$$\langle \dot{\mathbf{q}} \cdot \vartheta_{\mathbf{E}}(\dot{\mathbf{q}}, \mathbf{q}) \rangle = E \langle J \rangle$$

and friction is uniform in space and we expect an extensive work to be performed by the field and absorbed by the thermostat. This means that the rate function $\zeta(p)$ will be also proportional to the volume V : hence the fluctuations in entropy creation rate, which in this case are proportional to fluctuations of the electric current will be very difficult to observe.

Noting that the above expressions in this example (and in quite a few other examples) have a form that suggests that the average phase space contraction could be naturally identified with the entropy creation rate and noting that this is the slope of the line $\bar{\zeta}(a) - \bar{\zeta}(-a) = a\sigma_+$ it is tempting to define the temperature of the mechanical thermostat as $k_B\Theta$ equal to the ratio between the heat absorbed per unit time by the reservoir (*i.e.* $W = \langle \dot{\mathbf{q}} \cdot \vartheta_{\mathbf{E}}(\dot{\mathbf{q}}, \mathbf{q}) \rangle$) per unit time and the average phase space contraction

$$\Theta = \frac{W}{k_B\sigma_+}$$

which is a definition which makes sense in general. If the system is close to equilibrium then the above mentioned relation between the fluctuation relation and Green–Kubo transport formulae shows that near equilibrium this general definition coincides with the usual notion of temperature.

Given the difficulty (impossibility) of observing global fluctuations of the current or of the dissipation one would like to look at the local fluctuations of the dissipation: however phase space contraction is no longer defined if one looks at a subsystem, not even in the Hamiltonian case. One can then define it thermodynamically to be the amount of energy W_τ entering the region in the time τ and not coming outdivided by τ : this is the dissipation averaged over time τ and is interpreted as heat ceded to the thermostat. It can be measured in numerical experiments as well as in certain natural experiments. Of course we can define the entropy creation rate only if we have a definition of phase space contraction.

The attitude is then to study the fluctuations of $p = W_\tau/w_+$ where w_+ is the average of W_τ (independent of τ by stationarity: if we had been studying these quantities globally the fluctuation relation would identify the function $\bar{\zeta}_{loc}(p) - \bar{\zeta}_{loc}(-p)$ with $p\sigma_+$: therefore if the distribution of p is such that the fluctuation relation $\bar{\zeta}_{loc}(p) - \bar{\zeta}_{loc}(-p) = pc$ for some c it is consistent to define $\sigma_+ \stackrel{def}{=} c$ and the local temperature as $\Theta_{loc} = \frac{W_+}{k_Bc}$.

In other words the local temperature is defined so as to make the fluctuation relation true by definition

of course this is only possible if the linearity in p of $\bar{\zeta}_{loc}(p) - \bar{\zeta}_{loc}(-p)$ holds. Therefore several experiments have been attempted to test this property.

Then the question arises whether in a large system of Anosov type one can prove a local fluctuation relation. Furthermore one would like that reasonable relations should exist between such notions in regions of different sizes. And in molecular dynamics one can even envisage and independent definition of phase space contraction so that all the above notions shold be consistent.

The only large Anosov systems that one can so far study are lattices of chaotic reversible maps reversibly interacting by nearest neighbors: for them a detailed theory is possible and it is independent of the size of the system at least if the coupling is small. In such systems when the dissipation of the unperturbed system is positive also for the interacting

it is positive and $\sigma_+ > 0$ is proportional to the volume of the system (*i.e.* number of lattice sites).

It is also possible to define in a rather natural way the local dissipation. If the state of the system is $\{\varphi_\xi\}_{\xi \in V}$ and if $V_0 \subset V$ is a subvolume of V one simply considers the logarithm $\sigma_{V_0}(\varphi)$ of the determinant of the submatrix $\partial_{\varphi_{V_0}} S(\varphi_V)$ of the derivatives of the components with labels in V_0 with respect to the variables with labels in V_0 .

It can be proven that $\langle \sigma_{V_0} \rangle = V_0 c$ and that the large deviations of the phase space contraction averaged over time τ are controlled by a function $\bar{\zeta}(a)$ which has the form $V_0 \zeta_0(a)$ with $\zeta_0(a)$ independent of V_0 . This means that the large deviation function scales with the volume and therefore one can check the fluctuation relation by looking at small regions. Like for the density fluctuations in a gas they can be measured by looking at a small region and then by scaling the result.

Unfortunately lattices of reversible chaotic maps are rather special systems quite far from the ones that one expects to meet in nature. In fact the reversibility assumption also looks as a strong assumption. This however might be not so serious because many systems are perturbations of Hamiltonian systems which are time reversible and the nonreversibility comes from the thermostat. However there are reversible thermostats and there is a conjecture of equivalence between thermostats.

So for the moment the experimental perspectives are in the direction of checking the linear relation.