Condensed Matter Physics 2006, Vol. 9, No 2(46), pp. 219–236

# Steps toward the foundations of statistical mechanics: in and out of equilibrium

D.Karevski

Laboratoire de Physique des Matériaux, UMR CNRS 7556,
Université Henri Poincaré, Nancy 1,
F–54506 Vandœuvre les Nancy Cedex, France

Received September 6, 2005, in final form April 17, 2006

The first part of the paper is devoted to the foundations, that is the mathematical and physical justification, of equilibrium statistical mechanics. It is a pedagogical attempt, mostly based on Khinchin's presentation, which purpose is to clarify some aspects of the development of statistical mechanics. In the second part, we discuss some recent developments that appeared out of equilibrium, such as fluctuation theorem and Jarzynski equality.

**Key words:** *foundations of statistical mechanics, fluctuation theorem, jarzynski equality*

**PACS:** *05.20.Gb, 05.30.Ch, 05.70.Ln*

## Contents

**1 Introduction** **219**

**2 Foundations of statistical mechanics** **220**
2.1 Interpretation of physical quantities . . . . . 220
2.2 Microcanonical principle . . . . . 221
2.3 Suitability of the microcanonical principle . . . . . 224
2.4 Canonical distribution . . . . . 224

**3 Steps toward non-equilibrium** **225**
3.1 Jarzynski equality . . . . . 225
3.2 Fluctuation theorem . . . . . 227

**4 Work on the $2d$-Ising model** **229**

## 1. Introduction

The main goal of statistical mechanics, at least from the viewpoint of the initiators like Boltzmann, Maxwell, Gibbs[1], Einstein, was to derive thermodynamic laws from the microscopic (atomistic) structure of matter. All these attempts are subjected to start from some models of the structure of matter. However, it is a well known fact that thermodynamics was constructed independently or at least following a parallel road based on the few foundamental laws that are viewed as empirical facts. Quite recently, Lieb and Yngvason tried to clarify some aspects of the second law (entropy) of thermodynamics based on the concept of adiabatic accessibility [4,5]. This work

[1] Gibbs had a pragmatic point of view which was somewhat different from Boltzmann's view. Gibbs viewed statistical mechanics as a branch of rational mechanics, no matter which physical process generates the distribution in phase space. Contrary to that, Boltzmann's viewpoint was to really reduce thermodynamics to mechanics and consequently he necessitated an explanation of the mechanism that lead a mechanical system to equilibrium which was initially in a nonequilibrium state. To skech the differences one can say that the Boltzmann approach is more physical whereas the Gibbsian is more rigourous.

 

was mainly motivated by the fact that usual formulations of the second law, such as Kelvin or Clausius, use concepts such as hot, cold or heat that are intuitive rather than really well defined nor precise before the theory is fully developed. Their basic derivation of the second law (that is the existence of the entropy state function) is based on some abstract postulates of a certain kind of ordering on a set of states.

From our point of view, the problem of the foundations of statistical mechanics is two-fold. One is: given a statistical theory, one has to extract quantities (averages of phase functions) and laws that can be identified with thermodynamical quantities and fundamental laws, the identification itself being of analogy type[2]. Once those fundamental laws are recognized, one can logically develop the entire consequences of these laws. This logical enterprise was perfectly achieved by Gibbs in his celebrated treaty [3]. The other, less easy task, is to justify the use of the statistical theory (precepts) itself from a realistic[3] point of view, that is, so to say to justify the very use of ensembles. Differently stated, why canonical or microcanonical ensembles are suitable to the description of real physical systems? This question arises since it is generally believed that the system to be studied is in a definite state and not distributed over a continuum of states. It is clear that this second task is more physically related to the very structure of matter, and it is from this perspective, that the work of Boltzmann has to be viewed. The (partial) answer to this question is related to the fact that real thermodynamic systems are constituted, at least approximately, of a huge collection of particles. The discussion of these points will be largely developed in the next section. Section 3 deals with non-equilibrium aspects and we present their relations, such as the fluctuation theorem and Jarzynski equality, which seem to many physicists to be of fundamental interest. In the last section we present some results obtained on the Ising model in the fluctuation relation context.

## 2. Foundations of statistical mechanics

### 2.1. Interpretation of physical quantities

The state of our system (classically a point in the phase space or a Hilbert spectral ray quantum mechanically) fully determines the physical (dynamic) quantities which caracterize the given system. We shall generally call such a quantity a phase function (classical case $f(q,p)$, quantum case $f(\psi) = (\psi, \mathbf{Q}\psi)$ where $\mathbf{Q}$ is the operator associated to the quantity $f$). In order to have a suitable theoretical description, one has to identify such phase functions with the various physical quantities obtained experimentally from measurement processes and compare their respective value. However, in order to compare the empirical data with the theoretical predictions, one has to know the actual state of the system, that is, for example classically, to determine $2s$ ($\sim 10^{23}$) coordinates. But in general, the empirical (macroscopic) description, of what is called a (equilibrium) thermodynamic state, is fully specified by a very small set of independent variables, such as the energy, volume, pressure, and so on. So that the question that arises is which state should we choose in order to evaluate the relevant phase functions and compare their value with their experimental counterpart? Obviously, no one has or can have any reasonable answer to such a question. Nevertheless, if one realizes that the measurement of a physical quantity is performed during a finite time, which in general is very large compared to some internal time scale, one realizes that the actual empirical data are given as averages of the quantities over long time periods. But the initial question still remains, that is, which (part of the)[4] trajectory the system is actually following? In order to answer such a question one has to know $2s-1$ independant integrals of motion and it seems that a very small path has been done toward the solution since our starting

[2] As it is very explicitly emphasised in Gibbs treaty [3].

[3] Given that the basic ontology is a single mechanical system composed of many subsystems (particles).

[4] One problem that arises is the fact that time average of a phase function on a given trajectory may have very different values for different time intervals. This difficulty is overcome thanks to a theorem due to Birkhoff, which states that for almost all trajectories, the time averages of a given phase function tend to a definite limit when the time interval tends to infinity. It means in particular that the averages over finite time intervals on a given trajectory (a typical one) will take approximately the same value for sufficiently large time-periods. This remark, basically, is at the heart of the time average procedure used widely to start an exposition of statistical mechanics, see for example [6].

puzzling problem of finding the $2s$ coordinates. At this step, as it is well known, in order to avoid the average over an unknown trajectory, normally one invokes ergodic theorems or hypothesis to replace time averages by phase averages. However, in general very few systems are known to be ergodic and it seems really improbable that in a realistic case one will ever prove ergodicity. But the requirement of ergodicity is too strong. For instance, one has simply to require that only few (corresponding to the empirical ones) phase functions should have time average equal to their phase average. So, it would be an unnecessary hypothesis to demand the validity of such an equality for all phase functions. Another objection that has to be emphasized is the fact that ergodicity is a requirement that involves average over recurrence time, which is too long to have any physical relevance (several astronomical orders of magnitude), but in real experiment, the times involved to obtain the averages are by far shorter than the reccurence time, see for example a discussion of this point in [7]. The reason for this lies in the fact that the majority of phase functions describing physical quantities exhibit a very peculiar behaviour. They are approximately constant on almost all the points of the constant energy manifold (since we are talking here of an isolated system). Why it is so is linked to the fact that the mechanical systems, considered here, are broaken up into a large number of components and the fact that the phase functions of interest are sum functions, that is sum of functions depending on the dynamic coordinates of the component subsystems alone.

If we suppose, for some reasons, that we can replace time-averages by phase-averages, then the remaining problem is to determine the suitable phase average procedure. In the case of an isolated mechanical system, it is usually argued that one has to restrict the phase average to the constant energy manifold since the actual trajectory is taking place into this subset of the phase space. Indeed, if one considers the energy phase function for such an isolated system and one takes its phase average over several constant energy manifolds, it is clear that the average value will not give the real energy value. To overcome this discrepency, one has to restrict the phase average to that constant energy subset corresponding to the real value. The argument is clear and easy to conceive. However, one has to realize that everything that has been said concerning the energy should also be true for the $2s - 2$ other integrals of motion. So in particular, if we consider some other conserved quantity to be given as the energy is given, we have to restrict the average procedure to the intersection of the two corresponding manifolds, where the real trajectory is taking place. Continuing on these lines, we shall finally arrive at the fact that we have to average over the intersection of the $2s-1$ conserved quantities manifolds, that is over the actual trajectory and this is precisely what we wanted to avoid. In order to escape this vicious circle, let us consider a certain integral of motion defined by a phase function $I$. If $I$ over the constant energy manifold almost always takes the same value, then its time average over almost all trajectories will give almost the same (physical) value. Consequently, its phase average on the constant energy manifold will give a definite value which can be compared to the actual physical value. On the other hand, if this is not true and the phase function $I$ varies widely on the constant energy manifold, no definite average for different trajectories can be affected to it and it will not have a (macroscopic) physical interpretation. However, if such a phase function has indeed a physical interpretation (and so a real possibility of measuring it), one has to treat it on exactly the same footing as the energy is treated, and consequently if the value of that integral is known, we should restrict the phase average to the corresponding manifold of constant energy and constant $I$. It is a fact that in ordinary macroscopic systems, usually only the energy integral has to be considered[5].

## 2.2. Microcanonical principle

As we have seen from the arguments given above, one of the goals of a suitable statistical theory is to give for some phase functions the same values as those obtained experimentally and

[5]One can refer to [6] for some more arguments relating the question to additive integrals of motion, which are explicitely, the energy, the momentum and angular momentum components, that is integrals that are related to the space-time symmetries. If one encloses the system into a rigid vessel at rest, translation and rotation invariance are broken and only the energy integral remains. The additivity argument relies on the fact that the equilibrium statistical distribution over the restricted phase space should have a product structure related to the statistical independence of the different components of the system, that is related to the small interaction between the different components.

compatible with phenomenological thermodynamics. But, as it was argued, the values of the phase functions over a subset of the phase space present generally fluctuations that could be very large. In the following we give some more arguments in the direction of a phase space average.

A thermodynamic state is specified completly by a small set of quantities

$$\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_k \ . \tag{1}$$

This means in particular that all physical quantities $B$ are given as functions of these variables:

$$B = f_B(\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_k) \ . \tag{2}$$

To be more precise, there is one such function, named the fundamental relation:

$$\mathcal{F} = f_{\mathcal{F}}(\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_k) \tag{3}$$

from which all the other quantities can be obtained by appropriate derivations. This is basically thermodynamics, at least if one specify some properties of that fundamental function (second principle). [9] From the microscopic viewpoint, it is clear that the specification of this very small set of quantities is not at all sufficient to completely determine the state of the system. In general, for a given set $(\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_k)$ we have many compatible microscopic states (let us call them, as usual, microstate). To be more specific, we shall continue the discussion in the quantum case. So for the given set $(\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_k)$ we have some set[6]

$$\psi_1, \psi_2, \ldots, \psi_i, \ldots$$

of microstates to which the corresponding values are associated

$$B_1, B_2, \ldots, B_i, \ldots$$

of the phase function such that $B_i = f_B(\psi_i)$.

Let us imagine that we have prepared a collection of $N$ copies partially specified by the set $\mathcal{Q}$. Measuring on each copy the quantity $B$ we obtain with some frequency $n_i/N$ the corresponding value $B_i$. If we want to compare some theoretical value associated to the macrostate specified by the set $\mathcal{Q}$, we have to consider the average of the quantity $B$ over our experimental data. This is given by

$$\langle B \rangle = \frac{1}{N} \left( n_1 B_1 + n_2 B_2 + \cdots + n_i B_i + \cdots \right) \ . \tag{4}$$

The values $B_i$ are in principle known since they are quantum mechanical expectations over the states $\psi_i$. The real problem comes from the fact that no theory can give the values of the frequencies $n_i/N$ since they are related to the actual experimental setup. Different devices, preparation protocols in the experimental setup will lead to different occurences of the microstates and so of the values of $B$. To solve the difficulty one could think of stating an average principle, that is of specifying the set $\{n_i/N\}$. But again, in general different average principles will lead to different $\langle B \rangle$. In order to reconcile this observation with the uniqueness of empirical observations, one is naturally lead to the fact that almost all values $B_i$ should take almost the same value:

$$B_i \approx \langle B \rangle = B^{\mathrm{Ther}} \, .$$

The measure of the set where this equality does not hold should vanish or be reasonnably small. By calling upon the argument of simplicity, in order to calculate the average $\langle B \rangle$, we can choose the microcanonical average principle, that is specifying equal weights to all microstates. One has to realize here that from the given argument the microcanonical average principle is not unique. But, nevertheless, one can argue that, given all the previous (more or less of euristic type) arguments and the fact that the microcanonical principle generates the same macroscopic relations as

[6]Usually the set has the power of the continuum, so that the integer indexation used here should not be taken too seriously.

thermodynamics does, it is legitimate to postulate it by the Laplacien "principle of insufficient reason"[7]. According to an other guideline [8], one can try to base the foundations on the ground set by information theory, that is to have a constructive criterion (maximum-entropy principle) for selecting a probability distribution based on the partial knowledge. In the case of constant energy systems, the maximum-entropy principle leads to the microcanonical distribution. To end this discussion, one can state again that not all phase functions should satisfy the average principle requirements, only those having a macroscopic thermodynamic interpretation. Indeed, no one is expecting that the averaged (microcanonical or whatever) one particle velocity of a gas should give the actual measured velocity of a particle of that gas.

Let us consider for an isolated system the set of all the quantum states for which the energy is precisely fixed to the value $E$. These states are stationary states, that is eigenstates of the Hamiltonian operator $H$ with eigenvalue $E$. The dimension of this Hilbert sub-space $\mathcal{H}_E$ is given by the degeneracy of the eigenvalue $E$:

$$\dim \mathcal{H}_E = \Omega(E) = m\ . \tag{5}$$

Let $\psi_1, \ldots, \psi_m$ be a complete orthonormal set of eigenvectors of $H$ spanning the entire subset $\mathcal{H}_E$. One has for all $\Psi \in \mathcal{H}_E$ the unique decomposition

$$\Psi = \sum_{i=1}^{m} \alpha_i \psi_i \tag{6}$$

with the scalar coefficients $\alpha$ sitting on the complex hypersphere $S^*$

$$\sum_{i=1}^{m} |\alpha_i|^2 = 1\ . \tag{7}$$

The microcanonical average of the phase function $f_Q(\Psi) = (\Psi, \mathbf{Q}\Psi)$ is given by the integral over the complex hypersphere with uniform measure. One has

$$\langle Q \rangle \equiv \langle f_Q \rangle = \int_{S^*} (\Psi, \mathbf{Q}\Psi) \mathrm{d}S^*\ . \tag{8}$$

Together with the decomposition on the base $\psi_1, \ldots, \psi_m$ and with the parametrization $\alpha_k = r_k \mathrm{e}^{\mathrm{i}\varphi_k}$, one arrives at

$$\langle Q \rangle = \sum_{k,l} (\psi_l, \mathbf{Q}\psi_k) \int_{S^*} \alpha_k \alpha_l^* \mathrm{d}S^*\ , \tag{9}$$

$$\langle Q \rangle = \sum_{k,l} (\psi_l, \mathbf{Q}\psi_k) \frac{1}{m} \delta_{k,l}\ , \tag{10}$$

so that finally

$$\langle Q \rangle = \frac{1}{\Omega(E)} \sum_{k} (\psi_k, \mathbf{Q}\psi_k) = Tr_{\mathcal{H}_E} \left\{ \frac{1}{\Omega(E)} \mathbf{Q} \right\}\ . \tag{11}$$

Due to the invariance of the trace with respect to the change of orthonormal bases, the average $\langle Q \rangle$ is also invariant. This final rule is the basic starting point of a microcanonical calculation[8].

[7] The principle of insufficient reason or principle of indifference states that if there is no known reason for predicting our subject, one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability.

[8] It is known that the state vectors should statisfy symmetry or antisymmetry principle for bosons and fermions. This implies that in the given microcanonical average, one has just to restrict the subspace $\mathcal{H}_E$ to either symmetric subspace $\mathcal{H}_E^s$ or antisymmetric subspace $\mathcal{H}_E^{as}$.

### 2.3. Suitability of the microcanonical principle

For the microcanonical average of a quantity $Q$ to have a physical significance, it is necessary that it should take a value close to the real experimental value or to the value given by a real dynamic theory. As it will be shown bellow, this requirement will be fulfilled by sum functions for which the average $\langle Q\rangle$ is of the order of the number of components $N$ of the system, that is

$$\langle Q\rangle = Nq, \tag{12}$$

where $q$ referes to the one component quantity (a density like value). To show the proposition, let us consider the measure $M_\delta$ of the set of the vector states $\Psi \in \mathcal{H}_E$ such that the probability

$$P_\Psi \left(|Q - Nq| > N\epsilon\right) > \delta \tag{13}$$

that the quantity $Q$ differs significantly (of the order $\epsilon N$) from the average value $Nq$ and is greater than $\delta > 0$. We have

$$\int_{S^*} P_\Psi \left(|Q - Nq| > N\epsilon\right) \mathrm{d}S^* \geqslant \delta M_\delta \tag{14}$$

from which, together with the Chebyshev inequality $\langle x^2\rangle \geqslant \epsilon^2 P(|x| > \epsilon)$, we arrive at

$$M_\delta \leqslant \frac{1}{\delta\epsilon^2 N^2}\int_{S^*}\Big(\Psi, [\mathbf{Q} - Nq]^2\,\Psi\Big)\,\mathrm{d}S^* \tag{15}$$

that is with the previous result on microcanonical average

$$M_\delta \leqslant \frac{1}{\delta\epsilon^2 N^2} Tr_{\mathcal{H}_E}\left\{\frac{1}{\Omega(E)}[\mathbf{Q} - Nq]^2\right\} = \frac{1}{\delta\epsilon^2 N^2} D(Q)\ , \tag{16}$$

where $D(Q)$ is the microcanonical dispersion of the quantity $Q$. Since in usual physical situations the dispersion is growing linearly (which is the content of the law of large numbers) with the number of components $N$, one arrives at

$$\frac{D(Q)}{N^2} \sim \frac{1}{N} \tag{17}$$

so that in the thermodynamic limit this ratio vanishes. This means that the quantity $Q$ would approximately agree with the phase average with a probability arbitrarily close to unity and this finally demonstrates the suitability of the microcanonical phase space average. We see here that the main reason for the validity of such a principle is the fact that thermodynamic systems are built up from myriades of component subsystems and that real relevant physical quantities are sum functions. [2] Moreover, it is possible to show that the result just obtained remains valid if one affects some arbitrary absolutely continuous probability law, with respect to the measure introduced on the constant energy manifold, to the occurence of the state $\psi$.

### 2.4. Canonical distribution

So far we have always considered isolated mechanical systems, but it is clear that such an idealization is neither realistic nor efficient from a technical point of view. Indeed, a real system is never completely isolated and continuously interacts with its surounding. If the system is initialy prepared in a given (quantum) state, due to the interaction with its environment, it will very rapidly make transitions among its accessible states. Since such transitions are induced by purely random processes, for a sufficient time (it is claimed that) the system will sample out all the permissible states with equal probability [9].

Another idealization is the one in which the system of interest is free to exchange an arbitrary amount of energy with a very huge surrounding, the total system plus surounding being isolated. If the surrounding has good enough properties (it is then called a thermal bath), it basically fixes the temperature of the small (in comparison with the bath) system. It is possible to show rigourously that if we accept the principle of microcanonical average, then the small system is distributed according to the canonical law with a density [1,10]

$$\rho = \frac{1}{Z}\mathrm{e}^{-\beta H}\,, \tag{18}$$

where $H$ is the hamiltonian of the small system and where the unique remainder of the bath is the inverse temperature $\beta$. To arrive at the canonical, much more tractable, description, one can also remark that since the sole role of the bath is to fix the temperature of the system, it does not matter what it is actually composed of. So that one can perfectly imagine a huge (infinite) collection of components all identical in nature to the system under consideration, free to exchange energy. One is naturally led to the canonical law and to the concept of ensemble (at least in the canonical case). [3] Hence, we shall not continue further in the development of the logical consequences of the canonical distribution. The ambition of this first part was to clarify some aspects of equilibrium statistical mechanics.

## 3. Steps toward non-equilibrium

As it is well known and strongly emphasized in the specialized literature, the non-equilibrium situation is not so developed as the equilibrium case is. In particular, most physicists will agree to say that one cannot speak of a non-equilibrium statistical mechanics, in sharp contrast with equilibrium statistical theory. Nevertheless, in recent years there have been several developments that have led to general results not restricted to the vicinity of the equilibrium regime (as linear response theory is). In this section we briefly present some aspects of those results, namely the fluctuation theorem and the Jarzynski equality.

### 3.1. Jarzynski equality

The Jarzinsky equality is an "unexpected" equality relating equilibrium quantities with the average of a nonequilibrium process that can be very far from equilibrium [11]. More specifically, if one takes a system that is initially in a state of equilibrium at inverse temperature $\beta$ with an external (work) parameter denoted by $A$ and then if, within a finite time (so that we drive the system out of equilibrium), one tunes the external work parameter to the new value $B$, we shall perform on the system some amount of work $W$ which is specific to the actual microstate of the system. If we repeat this experiment many times and record the values $W$ then the Jarzynski equality states that

$$\left\langle \mathrm{e}^{-\beta W} \right\rangle \equiv \int \mathrm{d}W\ \rho(W)\mathrm{e}^{-\beta W} = \mathrm{e}^{-\beta \Delta F}\,, \tag{19}$$

where $\rho(W)$ is the distribution of work within the given protocol (how we have tuned the work parameter) and $\Delta F = F_B - F_A$ is the free energy difference between the equilibrium states at temperature $\beta^{-1}$ with respectively external parameters $B$ and $A$. In order to demonstrate the Jarzynski equality, we shall follow the lines developed in [12], and present the case of a classical Hamiltonian dynamic system. But it is necessary to mention here that such an equality was also derived in the quantum case [13–15] and within a stochastic markovian dynamics too [11,16]. The Jarzynski equality was soon tested experimentally. One can see [17] for a recent review.

The starting point of the demonstration given in [12] is to consider the hamiltonian of the system and its thermal environment, together with an interaction term:

$$\mathcal{H}(\Gamma;\lambda) = \mathcal{H}_s(x,\lambda) + \mathcal{H}_e(y) + \mathcal{H}_{\mathrm{int}}(x,y), \tag{20}$$

where the subscript $s(e)$ refers to the system(environment) hamiltonian with collective dynamic variables represented by $x(y)$. $\Gamma = (x, y)$ is the phase space coordinate of the full system plus environment dynamic system. The interaction term $\mathcal{H}_{\mathrm{int}}(x, y)$ is supposed to be small enough in order that one can interpret $\mathcal{H}_s$ as the internal energy of the system of interest. Let us initially consider that the system and environment are described by the thermal equilibrium Gibbs state with work parameter $\lambda = A$

$$p(\Gamma) = \frac{1}{Z(A)} \exp\big( -\beta \mathcal{H}(\Gamma; A)\big), \tag{21}$$

where $Z(A)$ is the normalization factor. When we vary the work parameter from the initial value $A$ to the final value $B$ within a time $\tau$ and a predefined protocol, the energy change of the system over the given microscopic trajectory $\{\Gamma\}_t$ is given by

$$\mathcal{H}_s(x_\tau; B) - \mathcal{H}_s(x_0; A) = \int_0^\tau \mathrm{d}t\ \dot{\lambda} \frac{\partial \mathcal{H}_s}{\partial \lambda}(x_t; \lambda_t) + \int_0^\tau \mathrm{d}t\ \dot{x} \frac{\partial \mathcal{H}_s}{\partial x}(x_t; \lambda_t), \tag{22}$$

where the first integral is interpreted as the work $W$ performed on the system and so the second integral is the heat absorbed during the process. One may notice that, since the dynamics is hamiltonian, the work $W$ is given by the total energy difference

$$W = \mathcal{H}(\Gamma_\tau; B) - \mathcal{H}(\Gamma_0; A)\ . \tag{23}$$

To compute the average $\langle \mathrm{e}^{-\beta W}\rangle$ over the initial Gibbs state, one can use for the work the previous expression and write

$$\langle \mathrm{e}^{-\beta W}\rangle = \int \mathrm{d}\Gamma_0\ p(\Gamma_0) \mathrm{e}^{-\beta[\mathcal{H}(\Gamma_\tau; B) - \mathcal{H}(\Gamma_0; A)]} \tag{24}$$

and using the expression (21), one arrives at

$$\langle \mathrm{e}^{-\beta W}\rangle = \frac{1}{Z(A)} \int \mathrm{d}\Gamma_0\ \mathrm{e}^{-\beta \mathcal{H}(\Gamma_\tau; B)}\ . \tag{25}$$

Now, by the canonical change of variables $\Gamma_0 \to \Gamma_\tau$, and using Liouville theorem on the invariance of the measure under canonical transformations, one finally arrives at

$$\langle \mathrm{e}^{-\beta W}\rangle = \frac{1}{Z(A)} \int \mathrm{d}\Gamma_\tau\ \mathrm{e}^{-\beta \mathcal{H}(\Gamma_\tau; B)} = \frac{Z(B)}{Z(A)}\,. \tag{26}$$

This last relation is looking like the Jarzynski equality (19) but the partition functions entering here are those of the full system and environment. Nevertheless, one can arrive at the Jarzynski equality (19) simply by noting that if we are able to neglect the interaction term (which has to be small enough), then the partition functions factorize into an environment term independent of the work parameter and a system term depending on $\lambda$. The ratio $Z(B)/Z(A)$ can then be rewritten as the Jarzynski equality

$$\langle \mathrm{e}^{-\beta W}\rangle = \frac{Z(B)}{Z(A)} \approx \frac{\int \mathrm{d}x\ \mathrm{e}^{-\beta \mathcal{H}_s(x;B)}}{\int \mathrm{d}x\ \mathrm{e}^{-\beta \mathcal{H}_s(x;A)}} = \frac{Z_s(B)}{Z_s(A)} = \mathrm{e}^{-\beta \Delta F}\,, \tag{27}$$

where this last form refers only to quantities pertaining to the system of interest.

Let us somehow discuss on a physical ground the Jarzynski equality. Indeed, if we perform on the system a reversible process by varying slowly enough the work parameter $\lambda$ from $A$ to $B$, then it is clear that

$$\Delta F = \widetilde{W},$$

where $\widetilde{W}$ is the thermodynamic work performed on the system. If the switching is fast enough such that the system has no time to equilibrate with the new work parameter, then the second law of thermodynamics states that the work $\widetilde{W}$ has to be larger so that one has for a general process the least work principle:

$$\widetilde{W} \geqslant \Delta F \tag{28}$$

which can be written

$$\widetilde{W}_{\mathrm{dis}} = \widetilde{W} - \Delta F \geqslant 0, \tag{29}$$

where the equality holds in the reversible case. If we rewrite the Jarzynski equality with $W_{\text{dis}} = W - \Delta F$, we have simply

$$\langle \exp(-\beta W_{\text{dis}}) \rangle = 1 \; . \tag{30}$$

Using Jensen's relation $\langle \exp(x) \rangle \geqslant \exp(\langle x \rangle)$, we see that the Jarzynski equality implies

$$\langle W_{\text{dis}} \rangle \geqslant 0 \tag{31}$$

which is the least work principle if one identifies the average work $\langle W \rangle$ with the thermodynamic work $\widetilde{W}$. From this, roughly speaking, one sees that if $\sigma$ is a parameter that controls irreversibility (one may think of the entropy production), the density distribution $\rho_\sigma(W)$

$$\lim_{\sigma \to 0} \rho_\sigma(W) = \delta(W - \Delta F)$$

in the reversible case $\sigma = 0$. Moreover, in the linear response regime, one expects Gaussian fluctuations of the work $W$ such that the density takes the form

$$\rho(W) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(W - W_0)^2}{2\theta}\right)$$

near equilibrium, where $\theta = \langle (W - \langle W \rangle)^2 \rangle$ is the variance of the distribution and $W_0 = \langle W \rangle$ the mean value of the work. Using this expression together with the Jarzynski equality, one can relate the fluctuation of the work to the dissipated work $\langle W_{\text{dis}} \rangle$, or to the entropy production if one notes that $T\sigma = \langle W_{\text{dis}} \rangle = T\Delta S - Q = \langle W \rangle - \Delta F$:

$$\langle W_{\text{dis}} \rangle = \frac{1}{2} \beta\theta \; ,$$

or equivalently

$$\sigma = k_{\text{B}} \frac{1}{2} \left\langle (w - \langle w \rangle)^2 \right\rangle$$

with $w \equiv W/k_{\text{B}}T$. We recover here a fluctuation dissipation theorem, relating the dissipated work to the fluctuation of it. Within this reduced work variable $w$, the density of the reduced dissipated work $w_d$ is rewritten

$$\pi_\sigma(w_d) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(w_d - \sigma/2)^2}{2\sigma}\right) ,$$

where the Boltzmann constant $k_{\text{B}}$ has been absorbed into the definition of the entropy production. In the reversible limit, one has

$$\lim_{\sigma \to 0} \pi_\sigma(w_d) = \delta(w_d) \; ,$$

that is no dissipation at all.

The very interesting feature of the Jarzynski equality is that it provides a method to recover free-energy differences from non-equilibrium experiments. This has been recently done for example on single molecule stretching experiments [17–19]. The possibility of recording free-energy differences was also experienced on simple theoretical models as one can see [20–22] for some examples and [17,23] for recent reviews on the experimental and theoretical aspects respectively.

### 3.2. Fluctuation theorem

Soon after the derivation of the Jarzynski relation it was demonstrated [36] that such result can be derived from a much more general theorem, namely the fluctuation relation which is a statement on the fluctuations of the entropy production. The first example of such a relation was obtained numerically in [24] for steady states systems and was soon derived for driven thermostated deterministic systems [26], thermostated steady state systems with deterministic [27] and stochastic dynamics [28–30]. The theorem can be written in the following form:

$$\frac{P(\sigma_\tau)}{P(-\sigma_\tau)} \simeq \exp(\tau\sigma_\tau), \tag{32}$$

where $P(\sigma_\tau)$ is the probability of having the entropy production rate $\sigma_\tau$ over a time $\tau$. This relation is in fact an asymptotic statement which is valid in the $\tau \to \infty$ limit. The precise statement of the theorem could be put in the following way: [7,23] Let $\sigma(\Gamma)$ be the entropy production rate in the stationary non-equilibrium state of the dynamic system represented at the initial time by the phase space point $\Gamma$. Let us define the "dimensionless average entropy creation rate"

$$p = \frac{1}{\sigma_+ \tau} \int_0^\tau \mathrm{d}t \; \sigma(\Gamma_t),$$

where $\sigma_+ \geqslant 0$ is the average entropy creation rate in the steady state and where the + subscript emphasizes the positivity of it. The fluctuation theorem states that

$$\zeta(p) = \zeta(-p) + p\sigma_+ \; , \tag{33}$$

where $\zeta(p)$ is the large deviation function of $p$ defined as

$$\zeta(p) = \lim_{\tau\to\infty} \frac{1}{\tau} \ln \pi_\tau(p)$$

with $\pi_\tau(p)$ the probability distribution of the variable $p$. The validity of the fluctuation theorem is very broad since it was proved within several different dynamic contexts, such as reversible hyperbolic dynamic systems [27] or nondeterministic dynamics [26,28,29]. It was also proved that the fluctuation theorem in the limit of a vanishing entropy production rate implies the linear (near-equilibrium) fluctuation-dissipation theorem [28,31,32]. The theorem was also tested in a number of numerical investigations [33] and recently on granular materials and turbulent flows [34]. We may also mention some recent extension for stochastic systems which are not capable of equilibrating with their environment in the limit of zero external drive [35].

Whereas the fluctuation theorem is an asymptotic relation, Crooks derived a very interesting identity in the case of stochastic microscopically reversible dynamics [16,36]. This identity reads

$$\frac{P_{\mathrm{F}}(\Omega)}{P_{\mathrm{R}}(-\Omega)} = \mathrm{e}^{\Omega}, \tag{34}$$

where $\Omega$ is the entropy production of the system driven for some time $\tau$ within a forward protocol $\lambda_{\mathrm{F}}(t)$ to a new state. $P_{\mathrm{F}}$ is the distribution of entropy production in the forward process, while $P_{\mathrm{R}}$ is the distribution of the entropy production in the backward or reverse process, that is when the system is driven in a time-reversed manner.

A particularly intersting case is the one in which the system is initially in contact with a thermal heat bath. The forward protocol is such that the system being initially (by convention we shall take the far past $t = -\infty$ to label the initial time) at equilibrium with an external parameter $\lambda_-$, is driven out of equilibrium by varying the control parameter to a new value $\lambda_+$. The variation of $\lambda$ takes a finite time $\tau$, let us say within the interval $[0, \tau]$. For times larger than $\tau$, the system is equilibrating with the heat bath and terminates in an equilibrium state characterized by the new external parameter value $\lambda_+$. Since the equilibrium entropy is given by $S = \langle -\ln \rho(\Gamma)\rangle = -\sum_\Gamma \rho(\Gamma) \ln \rho(\Gamma)$, it is quite natural to identify

$$s(\Gamma) = -\ln \rho(\Gamma)$$

as the entropy of the microstate $\Gamma$. The entropy production for a given trajectory in phase space is then

$$\Omega = \ln \rho(\Gamma_{-\infty}) - \ln \rho(\Gamma_\infty) - \beta Q, \tag{35}$$

where $\rho(\Gamma_{\pm\infty}) = \rho(\Gamma, \lambda_\pm)$ are the canonical equilibrium densities associated to the initial and final equilibrium states and where $Q$ is the heat exchanged during the process with the heat bath. Introducing into the previous expression the explicit equilibrium density

$$\rho(\Gamma, \lambda) = \mathrm{e}^{\beta F(\lambda) - \beta E(\Gamma,\lambda)} \; ,$$

one arrives, with $\Delta E = W + Q$, for the entropy production at

$$\Omega = \beta[W - \Delta F]\ .$$

So that the Crooks fluctuation identity can be written as a relation for the work performed on the system:

$$\frac{P_{\mathrm{F}}(W)}{P_{\mathrm{R}}(-W)} = \mathrm{e}^{\beta(W-\Delta F)}\ . \tag{36}$$

From this last expression, one can see that the Jarzynski identy is trivially recovered since

$$\left\langle \mathrm{e}^{-\beta W} \right\rangle = \int P_{\mathrm{F}}(W)\mathrm{e}^{-\beta W}\ \mathrm{d}W = \mathrm{e}^{-\beta\Delta F} \int P_{\mathrm{R}}(W)\ \mathrm{d}W = \mathrm{e}^{-\beta\Delta F}\ , \tag{37}$$

where the normalization condition of $P_{\mathrm{R}}$ has been used.

An interesting point is when one considers a cyclic transformation. In this case the final state is identical to the initial state and $\Delta F = 0$ so that the entropy production is given by $-Q/T$. Utilizing again Jensen's inequality we obtain that in average the entropy production is positive, which means that since work is performed on the system, heat is tranfered to the bath (and not the converse which is Clausius's basic statement).

## 4. Work on the $2d$-Ising model

In this paper, we discussed some aspects of equilibrium statistical mechanics and focused some attention on non-equilibrium fluctuation relations. In order to give a concrete example in the out of equilibrium situation, we give some preleminary results for work distributions obtained on the $2d$-Ising model [37]. The zero field Hamiltonian is given by

$$\mathcal{H} = -\sum_{(ij)} S_i S_j\ ,$$

where the $S = \pm 1$ are classical Ising variables and where the sum extends over nearest neighbors. In the numerical simulations, we use a Metropolis dynamics. The protocol we use is such that at the intial time $t = 0$, the system beeing in an equilibrium state at inverse temperature $\beta$ and zero field, we switch on the field for a time $\tau$ with a linear law:

$$h(t) = \dot{H}t = \frac{H}{\tau}t\ .$$

For each initial realization, we compute the work W performed on the system by the external field with

$$W = -\dot{H}\int_0^\tau \mathrm{d}t\ M(t),$$

where $M(t) = \sum_i S_i(t)$ is the instantaneous magnetization of the system.

We first show in figure 1 the work distribution obtained with an initial paramagnetic state for three different protocols with ending field $h = 1.5$ reached after time $\tau = 25$, $\tau = 50$ and $\tau = 100$, that is smoother and smoother protocols. As it can be seen in figure 1, the distributions are Gaussian and behave as predicted in the previous section, with a dissipated work proportional to the variance of the distribution. From the shift to the left (negative value), one can extract the free energy difference, since as $\tau$ is increasing, the transformation is less and less irreversible (the entropy production is smaller and smaller). Since the distribution is Gaussian (near equilibrium like regime), the free energy difference is given by

$$\Delta F = \langle W \rangle - \frac{\beta}{2}\theta,$$

where $\theta$ is the variance of the distribution. We have verified here that the Jarzynski equality holds, since the previous relation is satisfied and gives the same value $\Delta F$ for different $\tau$ values. Moreover,

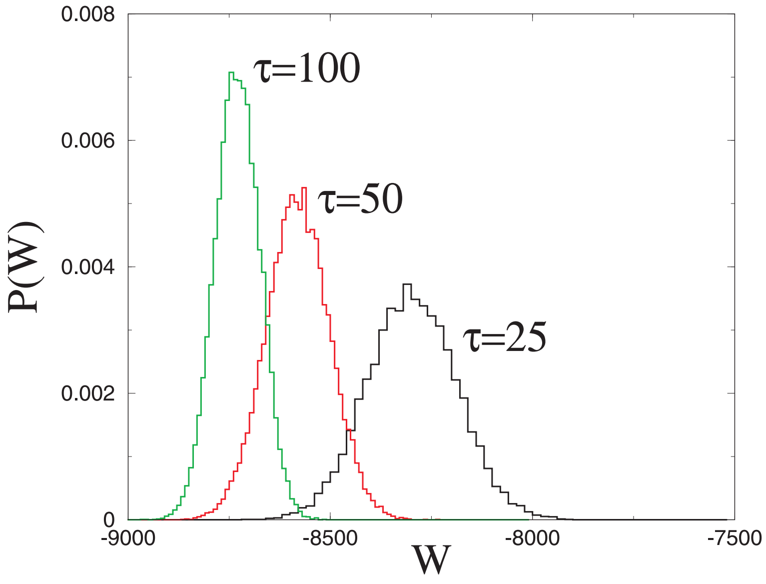


**Figure 1.** Work probablility density distribution of the 2d Ising model at $\beta = 0.2$ (paramagnetic initial phase) with a linear varying field from $\beta h(0) = 0$ to $\beta h(\tau) = 0.3$ for several time intervals $\tau$. The system size is $L \times L = 128^2$.

from our data we find a linear dependence of the dissipated work with the perturbing speed $\dot{H}$, as it is expected near equilibrium or more generally in the gaussian case [17], with

$$\langle W_{\rm dis} \rangle \simeq \alpha(H) L^2 \dot{H}$$

with a field dependent slope $\alpha(H)$. Having analized the work distributions for several ending fields, we arrived numerically at an expression for $\alpha(H)$ which is very well fit by:

$$\alpha(H) = \gamma\mu(H) = \gamma\mu \tanh(\beta\mu H)$$

with $\gamma \simeq 0.43$ and $\mu(\beta = 0.2) \simeq 3.65$, where $\mu(H)$ is the paramagnetic magnetization. Indeed, if one extracts the free energy difference as a function of the field $H$, one obtains a very good fit with the paramagnetic solution

$$\Delta F = F(H) - F(0) = -\frac{L^2}{\kappa\beta} \ln\cosh\frac{\mu H}{k_{\rm B}T}$$

with a magnetic moment $\mu \simeq 3.65$ and volume $\kappa \simeq 4.6$. If one uses the equilibrium magnetization extracted from that expression:

$$M(H, \beta) = -\frac{\partial F}{\partial H} = \frac{L^2}{\kappa}\mu \tanh(\beta\mu H) \ ,$$

the dissipated work can be expressed as

$$\langle W_{\rm dis} \rangle \simeq 2\dot{H} M(H, \beta)$$

so that it depends only on the equilibrium quantities of the paramagnetic gaz.

If we decrease the temperature of the initial equilibrium state, at a moment we shall arrive at the critical point. Of course, since our system is finite, the correlation length will not diverge since it will reach the boundaries of the system. Due to the long range critical correlations, one expects a different behaviour of the work performed on the system along the same linear protocol

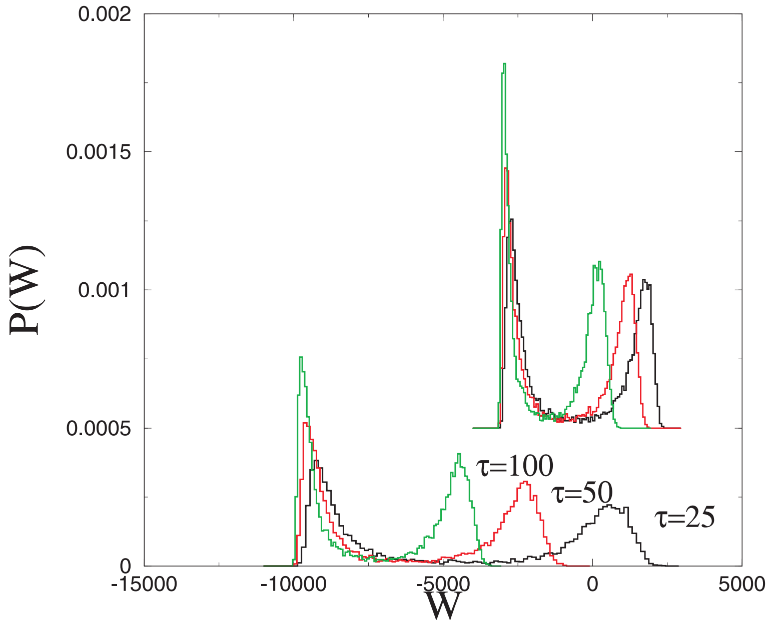


**Figure 2.** Work probablility density distribution of the 2d Ising model at $\beta_c$ (critical initial phase) with ending fields $\beta_c h = 0.1$ and $\beta_c h = 0.3$ for several time intervals $\tau$. The system size is $L \times L = 128^2$. The distribution with $\beta_c h = 0.1$ has been shifted vertically by an amount of $5 \times 10^{-4}$.

as in the paramagnetic phase. Indeed, it is the case as it can be seen in figure 2 where the work distribution is plotted for two ending fields, using three different protocols with $\tau = 25$, $\tau = 50$ and $\tau = 100$. Insted of having one pronouced peak as in the paramagnetic case, we have to face two peak distributions.

What we see is that the left-most peak, sitting close to the reversible work $\Delta F$, is moving very slightly toward negative values for larger time intervals $\tau$. The right-most peak has a strong dependance with the perturbing speed $\dot{H}$ as it is obvious from figure 2 and seems to converge asymptotically toward the reversible value $\Delta F$. We can understand the structure of the distribution in the following way: since our system is finite, even at the critical temperature $\beta_c^{-1}$, there is a finite magnetization of the order $L^{d-\beta/\nu}$. When turning on the field, if the initial magnetization is pointing in the direction of the field, we expect a negative work of the order $-HM$, where $M$ is the initial magnetization. On the contrary, if the initial magnetization is pointing in the opposite direction, at least for very fast protocols, one will have a positive work $+HM$ since the system will not have enough time to react to the variation of the field. For lower field speed $\dot{H}$, during the time interval $\tau$, part of the magnetization will flip in the field direction, leading to a smaller work value. It is clear that the real situation is more complicated than the rough description just sketched above. In fact, one has to take into account the domain growth, since it will be the leading process for the variation of the total magnetization, and so for the work performed on the system. This study is currently under investigation.

Let us finally discuss the ferromagnetic situation. In figure 3, we have presented the work distribution, for the three field speeds $\dot{H}$, obtained with an ending field $\beta h = 0.1$, with $\beta = 0.7$. What is seen is two very well separeted peaks, one sitting at a negative value $W_-$ and the other at a positive value $W_+$ of the same order of magnitude. Moreover, each peak has the same weight $1/2$. For clarity, the right-most peak has been plotted in an insert. The peaks are more or less Gaussian like (in fact the right peak is a little bit asymmetric and so deviates from the gaussian behaviour) and one can see that as the process get slower, the peaks get sharper. What is also manifest is the fact that the left peak is almost no more translated to the left as in the paramagnetic case.

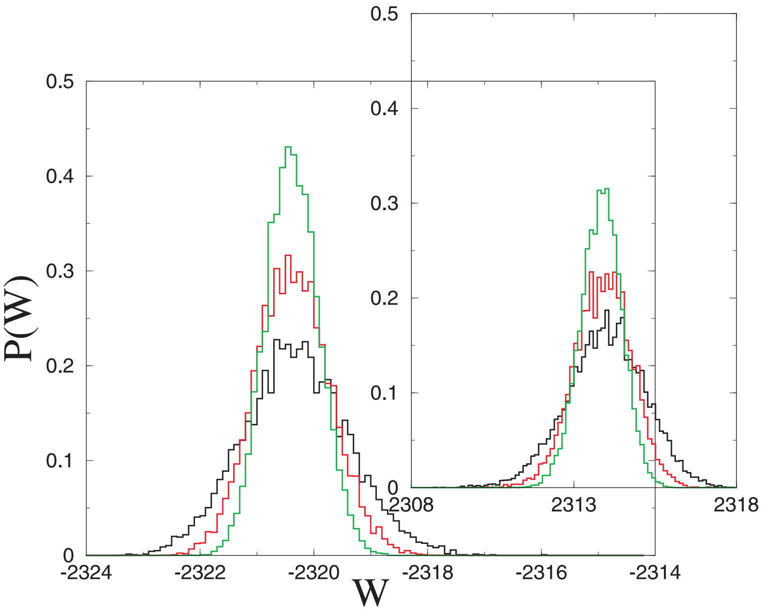


**Figure 3.** Work probablility density distribution of the 2d Ising model at $\beta = 0.7$ (ferromagnetic initial phase) with ending field $\beta h = 0.1$ for several time intervals $\tau$. The system size is $L \times L = 128^2$.

Indeed, it corresponds to the initial equilibrium magnetization pointing in the direction of the field, and since at $\beta = 0.7$, the magnetization is nearly saturated, there is no magnetization gain to be expected during the swiching of the field. The work is then of the order of $-|HM_{\rm eq}(\beta, 0)|$. The right peak corresponds to a microstate with an initial magnetization pointing in the opposite direction of the field. In this case, if the field is not strong enough, the macroscopic magnetization will stay unaltered by the presence of the opposite field and we shall have more or less $W_+ \sim +|HM_{\rm eq}(\beta, 0)|$. In the slow swiching limit, the work distribution takes the form

$$P(W) = \frac{1}{2}\delta_{W_-} + \frac{1}{2}\delta_{W_+}$$

and so the average work $\langle W \rangle$ is given by

$$\langle W \rangle = \frac{W_- + W_+}{2}$$

while, using Jarzynski equality, the free energy difference $\Delta F$ is such that

$$\mathrm{e}^{-\beta\Delta F} \simeq \frac{1}{2}\mathrm{e}^{-\beta W_-} \ ,$$

where the argument of the exponential which is positive dominates largely the other exponential. From that, we have

$$\Delta F \simeq -|W_-| + \frac{1}{\beta}\ln 2 \simeq -|W_-|,$$

where $\ln 2/\beta$ has been neglected since the work $W_-$ is extensive. Together with the previous expression, we have for the dissipated work

$$\langle W_{\rm dis} \rangle = \langle W \rangle - \Delta F \simeq \frac{W_+ - W_-}{2} \simeq -\Delta F$$

that is, the dissipated work is of the order of the reversible work (in magnitude). This striking result is due to the fact that almost all the irreversibility is put in by the reversed magnetized microstates. If one used, as a macrostate only, those microstates with the magnetization pointing in the field direction, the situation would be more usual, that is Gaussian with small dissipation. From the data presented here, the ratio of the dissipated work with the free energy is less than $5.10^{-4}$, which explains why there is no apparent shift toward the left of the curves for decreasing perturbation speed $\dot{H}$.

We close here the discussion of these numerical preliminary results. The point was to give an example where we can extract some equilibrium quantities within non-equilibrium (computer) experiments. However, one may notice that in order to obtain accurate results for the free energy difference from the Jarzynski equality, one has to sample out very efficiently the tails (at least in the left-most part) of the distribution since we need to compute exponentials of extensive quantities (work). But nevertheless, this program seems to be realistic enough and is currently under progress. Finally, one may think of realizing such situations experimentally, measuring nanoparticle's magnetization within a SQUID experiment [38].

## Acknowledgements

Arnaldo Donoso, Bertrando Berches, Ricardo Paredes, and all the people who made Mochima spring/summer school possible are all gratefully acknowledged. I would like to give a special Thank to the students who made the school to be so pleasant. Part of the material presented here was discussed and done in collaboration with Christophe Chatelain. Je remercie le $\mathcal{G}$roupe $\mathcal{M}$ pour son soutien.

## References

1. Khinchin A.Y. Mathematical Foundations of Statistical Mechanics. New York, Ed. Dover, 1949.
2. Khinchin A.Y. Mathematical Foundations of Quantum Statistics. New York, Ed. Dover, 1960.
3. Gibbs J.W. Elementary Principles in Statistical Mechanics. Yale University Press, 1902; French Ed. Hermann, 1998.
4. Lieb E.H., Yngvason J., A guide to entropy and the second law of thermodynamics, Notices of the AMS, 1998, **45**, 571; see also cond–mat/9708200.
5. Giles R. Mathematical Foundations of Thermodynamics. Pergamon, Oxford, 1964.
6. Landau L., Lifchitz E. Physique Statistique. Physique Théorique. **5**, Ed. Mir, 1967.
7. Gallavotti G. Statistical Mechanics, A Short Treatise. Springer, 1999.
8. Jaynes E.T., Information theory and statistical mechanics. I, Phys. Rev., 1957, **106**, 620.
9. Callen H.B. Thermodynamics and an Introduction to thermostatistics. New York, Ed. Wiley, 1985.
10. Darrigol O. The Origins of the Entropy Concept. Poincaré Séminaire, 2003, Ed. Birkhäuser Verlag, 2004.
11. Jarzynski C., Phys. Rev. Lett., 1997, **78**, 2690; Jarzynski C., Phys. Rev. E, 1997, **56**, 5018.
12. Jarzynski C., J. Stat. Mech.: Theor. Exp., 2004, P09005.
13. Yukawa S., J. Phys. Soc. Jpn., 2000, **69**, 2367.
14. Mukamel S., Phys. Rev. Lett., 2003, **90**, 170604.
15. Monnai T. ArXiv: cond–mat/0410623 v3.
16. Crooks G.E., J. Stat. Phys., 1998, **90**, 1481.
17. Ritort F. Work Fluctuations, Transient Violations of the Second Law and Free-Energy Recovery Methods: Perspectives in Theory and Experiments. Poincaré Séminaire 2003, Ed. Birkhäuser Verlag, 2004.
18. Hummer G., Szabo A. Proc. Nat. Acad. Sci. USA, 2001, **98**, 3658.
19. Liphardt J., Dumont S., Smith S.B., Tinoco Jr. I., Bustamante C., Science, 2002, **296**, 1832.
20. Mazonka O., Jarzynski C. ArXiv: cond–mat/9912121.
21. Douarche F., Ciliberto S., Petrosyan A. ArXiv: cond–mat/0504465.
22. Ritort F., Bustamante C., Tinoco Jr. I., Proc. Nat. Acad. Sci. USA, 2002, **99**, 13544.
23. Maes C. On the Origin and the Use of Fluctuation Relations for the Entropy. Poincaré Séminaire 2003, Ed. Birkhäuser Verlag, 2004.
24. Evans D.J., Cohen E.G.D., Morriss G.P., Phys. Rev. Lett., 1993, **71**, 2401.
25. Crooks G.E., Phys. Rev. E, 1999, **60**, 2721.
26. Evans D.J., Searles D., Phys. Rev. E, 1994, **50**, 1645.
27. Gallavotti G., Cohen E.G.D., J. Stat. Phys., 1995, **80**, 931.

28. Kurchan J., J. Phys. A, 1998, **31**, 3719.
29. Lebowitz J.L., Spohn H., J. Stat. Phys., 1999, **95**, 333.
30. Maes C., J. Stat. Phys., 1999, **95**, 367.
31. Gallavotti G., Phys. Rev. Lett., 1996, **78**, 4334.
32. Gallavotti G., Annales de l'institut H. Poincaré, 1999, **70**, 429.
33. Bonetto F., Gallavotti G., Garrido P.L., Physica D, 1997, **105**, 226; Gallavotti G., Rondoni L., Segre E., Physica D, 2004, **187**, 338; Zamponi F., Ruocco G., Angelani L., J. Stat. Phys., 2004, **115**, 1655.
34. Ciliberto S., Laroche C., J. Phys. IV, 1998, **8**, Pr6; Goldburg W.I., Goldschmidt Y.Y., Kellay H., Phys. Rev. Lett., 2001, **87**, 245502; Ciliberto S. et al., Physica A, 2004, **340**, 240; Feitosa K., Menon N., Phys. Rev. Lett., 2004, **92**, 164301.
35. Zamponi F., Cugliandolo L.F., Kurchan J., ArXiv: cond–mat/0504750.
36. Crooks G.E., Phys. Rev. E, 2000, **61**, 2361.
37. Chatelain C., Karevski D., in preparation.
38. In principle such experiment is possible in our laboratory. The challenge is to persuade colleagues that it is indeed interesting enough.

## Questions and answers

$\mathcal{Q}$ (Rafael Rangel, Alexander López): What is the status of the ergodic hypothesis concerning the foundations of statistical mechanics?

$\mathcal{A}$ As it is known, statistical mechanics can be built upon the assumption of ergodicity and several systems have been proved to be indeed ergodic, that is, their time average, where the averaging time should be long enough, equals its statistical average with respect to some unique invariant measure. This hypothesis was first introduced by Boltzmann in 1887 and is linked to the mixing properties, introduced by Gibbs in 1902, which supposes that the two-time correlations decay at long times as $\langle A(t)B(0)\rangle \rightarrow_{\lim t\rightarrow\infty} \langle A\rangle\langle B\rangle$. Mixing implies ergodicity. The argumentation goes as follows: in a real experiment, the measurement of a physical quantity is not instanteneous but rather takes place over a relatively long time period with respect to the relevant microscopic time-scales. This justifies the use of a time average over the dynamic trajectory. However, many systems are not ergodic as for exemple is the Fermi, Pasta and Ulam model. One may think about spin glasses as well. For the use of statistical methods, on the one hand, one can argue that demanding ergodicity to hold is a too strong requirement since what is usually physicaly relevant is a very small subset of the set of all dynamic quantities that one can define. Usually those physically relevant quantities are sum-functions. Then, one can restrict the ergodicity demand only to those physical quantities and then ergodicity is just a signature of typicallity, which is linked to the large number of degrees of freedom in a thermodynamic system. More recently, it is to be notted that several authors, see for example Goldstein S., Lebowitz J.L., Tumulka R., Zanghi N., Phys. Rev. Lett. 96, 050403 (2006) have proved that even if the state of a quantum composite system is given by a single wave function, the reduced density matrix of a subsystem is canonical for the overwhelming majority of wave functions compatible with the given constraints. In this respect, it follows that the canonical ensemble arises in quantum mechanical systems without any genuine randomness. The essential idea leading to this result is due to E. Schrödinger himself.

On the other hand, one can argue that even for a nominally isolated system, they are always random external perturbations that allow the system to make stochastic transitions among permissible states, thus sampling randomly the many-dimensional state space.Then from the time-inversion symmetry, detailed balance among permissible microscopic states holds which finally leads to equal probabilities of states in equilibrium.

$\mathcal{Q}$ (Bertrand Berche): You said that the jusitification of the suitability of the microcanonical average principle resides essentially in the fact that the sum-functions take almost the same value over the microstates: but this is only for microcanonical ensemble.

$\mathcal{A}$ If one has more than the energy integral, for example momentum conservation as well, then in the average principle one has to take that new symmetry into account in order to have a suitable statistical description. It will reflect in the thermodynamic description by its associated extensive parameter which must be included in the description in order to have suitable thermodynamics.

$\mathcal{Q}$ (Alexander López): How would your picture change while passing from a classical description to the quantum one?

$\mathcal{A}$ I would say that the change from classical to quantum description is basically a technical detail. The general picture of the justification of the statistical average suitability remains unchanged.

$\mathcal{Q}$ (Monica Garcia): If you take into account quantum effects, what changes in the Jarzynski equality?

$\mathcal{A}$ Jarzynski equalities have been demonstrated for quantum systems as well. But the main differences with the classical situation are related to the proper definition of work, or measurement protocol, that one will use in the quantum case. In some cases, one has quantum ($\hbar$) corrections to the classical Jarzynski equality.

$\mathcal{Q}$ (Juan Luis Cabrera): If the initial phase space is a subset of the final one, does the Jarzynski equality still hold?

$\mathcal{A}$ One of the criticisms to the Jarzynski equality was precisely based on this question. For exemple, one may think about an isolated piston separeted by a removable wall. Initially, only one cell is filled with an equilibrated gas and then the separating wall is removed. This is the free expansion experiment and consequently $W = 0$. But as it is well known the free energy difference is $\Delta F = -k_{\mathrm{B}}TN\ln(V_f/V_i)$, where $V_f$ and $V_i$ are respectively the final and initial volumes occupied by the gas, which is clearly in contradiction with the jarzynski equality. The discrepancy arises from the fact that the initial configuration of the gas is not the proper canonical distribution as obtained from the initial Hamiltonian with the proper wall boundaries. It is necessary that the initial phase space condition be sampled from the initial canonical distribution associated to the initial Hamiltonian defined on the whole accessible phase space for the given protocol. That is, in the free expansion situation distribution with gas particles on both sides of the separating wall.

# Основи статистичної механіки

Д.Каревскі

Група М, Лабораторія фізики матеріалів, UMR CNRS 7556,
Університет Анрі Пуанкаре, Нансі 1, 54506 Вандувр лє Нансі Седекс, Франція

Отримано 6 вересня 2005 р., в остаточному вигляді – 17 квітня 2006 р.

Перша частина статті присвячена основам, тобто математичним і фізичним підтвердженням, рівноважної статистичної механіки. Це педагогічна спроба, що здебільшого базується на представленнях Хінчина, і вона ставить собі за мету прояснити певні аспекти розвитку статистичної механіки. У другій частині ми обговорюємо недавні роботи в нерівноважній ділянці, такі як флюктуаційна теорема чи рівність Яжинського.

**Ключові слова:** *основи статистичної механіки, флюктуаційна теорема, рівність Яжинського*

**PACS:** *05.20.Gb, 05.30.Ch, 05.70.Ln*