Séminaire Poincaré 2 (2003) 29 – 62

Séminaire Poincaré

On the Origin and the Use of Fluctuation Relations for the Entropy

Christian Maes*
Instituut voor Theoretische Fysica
K.U.Leuven, Belgium

Abstract. Since about a decade various fluctuation relations for the entropy production have been derived and analyzed. These relations deal with symmetries of the entropy production under time-reversal and have been proposed as a non-perturbative generalization of fluctuation–dissipation relations. I describe a unifying framework for understanding these relations and I present an algorithm to derive them. The fluctuation relations all follow from the main observation that in great generality the path-dependent entropy production is the source-term of time-reversal breaking in the Lagrangian over space-time histories. That is illustrated via a number of examples as well as via a general theoretical argument. I move these relations away from the strict dynamical background in which they originated and take them back to the context of statistical mechanics where entropy is understood in the sense of Boltzmann, as measuring the typicality of a manifest condition. I discuss how a relation between work and free energy is naturally put in that framework and how the transient and steady state fluctuation theorems are simple consequences. The fact that fluctuation symmetries for the entropy production are in general only valid asymptotically for large times, makes them mostly inaccessible for experimental verification, in contrast with a recent claim that they would usefully quantify second law violations. Part of the interest in the resulting fluctuation symmetries is that they are so universally valid, a rare occasion in nonequilibrium statistical mechanics. However they do not provide a systematic perturbation expansion for response functions. For that one needs to go back to the full Lagrangian and also consider the nonequilibrium modifications to its time-symmetric part.

1 Scales

Our first look at the world does not invite to think of nature as one. Depending on scales of length, speed, energy and time, new worlds appear that to a large extent can be explored independently. As a matter of fact, at least when sufficiently excited, man feels quite decoupled from underlying microscopic laws and realities. It is also largely due to that decoupling between and the apparent independence of various scales of description that progress in different scientific domains, in particular in physics, is at all possible. Ignoring microscopic details offers often fast and reliable access to questions belonging to the mesoscopic or macroscopic domain. Nevertheless, part of scientific progress is also in the unification of phenomena and in the convergence of explanations into a limited set of more elementary mechanisms. The extent to which a microscopic theory corresponds to reality depends on the success of that application of reductionism to a variety of natural phenomena of which we all share the same experiences. From these successes we have learnt to speak not of independent scales but rather of a hierarchy of scales with interconnected theories through which we can move and derive one description from another finer level.

Statistical mechanics plays the role of a transfer mechanics between different levels of description. The microscopic laws get connected with the macroscopic behavior. Obviously, the microscopic dynamics is important; perhaps not all details but many features of it, like symmetries and conservation laws, remain visible even on much larger scales. Additional considerations come from counting; these are the statistical aspects. It is there that entropy appears. Entropy throws a bridge between the microscopic motion and the thermal phenomena by introducing the notion of typicality: what can one reasonably expect from a system containing lots of particles and how big are the fluctuations. Not only energy considerations but also entropy considerations make the world work. In the words of Boltzmann: *The general struggle for existence of living creatures is therefore not one for the basic elements - the elements of all organisms are present in abundance*

*http://itf.fys.kuleuven.ac.be/∼christ/

in air, water, and the soil - or for energy, which is unfortunately contained in abundance in any body in the form of unconvertible heat, but it is a struggle for the entropy that is available through the passage of energy from the hot Sun to the cold Earth.

In equilibrium statistical mechanics, one usually forgets about the microscopic dynamics and the central object is the microscopic Hamiltonian. The Gibbs ensembles are constructed and one shows how thermodynamics is obtained from the resulting formalism. Entropy gets translated in various other thermodynamic potentials or free energies. The Gibbs formalism also allows studies of fluctuation and response relations and numerous techniques, computational, perturbative or variational allow detailed studies of the equilibrium system. The equilibrium phases with the resulting phase diagram are constructed from weighing the various possible macroscopic regimes on the energy-entropy balance.
The power of the Gibbs formalism has so far no real counterpart in nonequilibrium physics which to all appearances, shows an even greater variety of phenomena. Most people would agree that there is yet no such thing as nonequilibrium statistical mechanics. In recent years however we have witnessed a renewed interest both in foundational and computational issues from which attempts have been formulated and results have been obtained which are not restricted to the close to equilibrium regime. One guiding idea of the present contribution is to exploit the Gibbsian aspects of the space-time distribution for nonequilibrium systems. The main object of study is now the Lagrangian defined on time-dependent reduced variables and the entropy production will be seen to coincide with the source term for time-reversal breaking. From there it presides over the system's fluctuations and response behavior.

In the next section I specify the key-message of the paper. Sections 3 to 6 contain my personal look at textbook material regarding entropies and the second law. Sections 4.4 and 6 are more original but except perhaps for some notation, much of Sections 3–6 can be skipped when one enjoys running. The rest of the paper substantiates and illustrates the main observation, which is next.

2 Main observations

Here I summarize what is going to come. To get a quick start, I avoid here the many definitions and notations and I write about quantities yet to be defined more precisely later.

There are two main observations.
The first one is that in very great generality the physical entropy production can be identified with the source term of time-reversal breaking in the action governing the space-time distribution. By that I mean the following. Given some type of dynamics on phase space, we can be interested in the probability $P_{\rho_0}(\omega)$ of a trajectory ω of reduced variables. Think about a time-sequence of values that some macroscopic observable can take, or more generally, about some type of contracted description of the system's evolution. The subscript ρ_0 gives the initial distribution on phase space. There is a natural notion of time-reversal which transforms ω into a new trajectory $\Theta\omega$. At some later time τ, the state of the system is described by the distribution ρ_τ, its time-reversal is denoted by $\rho_\tau\pi$, and we are interested in the logarithmic ratio

$$R(\omega) = k_B \log \frac{P_{\rho_0}(\omega)}{P_{\rho_\tau\pi}(\Theta\omega)} \tag{2.1}$$

The constant k_B is Boltzmann's constant which I will forget when convenient. I claim that, whenever a physical interpretation is possible, $R(\omega)$ is the physical entropy production over the time-interval $[0,\tau]$, up to a total time-difference. In other words, writing formally

$$P_{\rho_0}(\omega) = \rho_0(\omega_0)\, e^{-\mathcal{L}(\omega)} \tag{2.2}$$

under time-reversal, the antisymmetric term in the Lagrangian $\mathcal{L}$ gives

$$S(\omega) = \mathcal{L}(\Theta\omega) - \mathcal{L}(\omega) = R(\omega) + \log \rho_t(\omega_t) - \log \rho_0(\omega_0) \tag{2.3}$$

and is the variable entropy production modulo a temporal boundary term. That is true for time-dependent or homogeneous dynamics, stochastic or deterministic, in the transient or in the steady state regime (for which $\rho_0 = \rho_\tau$) but of course we need a physical context to verify it.
The statement is non-trivial and should be argued. I do that first by giving four examples in the next subsection. From it, the reader will also better understand what is meant by the statement. A general argument takes more space and some of it will be provided in Section 7 and further.

The second main observation is that the first observation is useful, basically because of two reasons. First, from that general relation (2.1) between time-reversal and entropy production, a unification follows for a variety of recently obtained results concerning nonequilibrium fluctuations. I have in mind the transient and stationary fluctuation theorems that have appeared first in the study of dynamical systems, but also the so called Jarzynski relation relating equilibrium free energies with irreversible work done. I will discuss these in Sections 8–9. An algorithm appears for deriving fluctuation symmetries and they reformulate (2.1) in terms of probabilities and expectations.
A second reason why (2.1) can be useful is that it gives a general way of departure for studying response relations and for obtaining extensions of fluctuation-dissipation relations. Section 10 gives a glimpse at it. We will see there how the observation (2.1) and hence the fluctuation symmetries, are insufficient to go beyond linear order in perturbation theory. One needs also information about how the time-symmetric term in the Lagrangian gets modified under nonequilibrium conditions. The broader perspective is thus that of a Lagrangian statistical mechanics, see below, which would also include insights about the thermodynamic nature of the time-symmetric part of the Lagrangian.

2.1 Examples

2.1.1 Heat conduction

Consider a finite graph $G = (V, \sim)$ with vertex set V. Every site $i \in V$ carries a momentum and position coordinate $(p_i, q_i) \in \boldsymbol{R}^2$. These oscillators are coupled for a Hamiltonian

$$H(p,q) = \sum_{i \in V} \frac{p_i^2}{2} + U(q) \tag{2.4}$$

for some symmetric nearest neighbor potential

$$U(q) = \sum_i U_i(q_i) + \sum_{i \sim j} \lambda_{ij} \Phi(q_i - q_j) \tag{2.5}$$

where $\lambda_{ij} = \lambda_{ji} \neq 0$ whenever $i \sim j$ and Φ is even. I refer to [62] for a recent review.
Select a non-empty subset $\partial V \subset V$ of boundary sites, that are imagined connected to thermal baths at possibly different temperatures. The dynamics is Hamiltonian except at the boundary ∂V where the interaction with the reservoirs has taken the form of Langevin forces as expressed by the Itô stochastic differential equations

$$\begin{aligned} dq_i &= p_i dt, \quad i \in V \\ dp_i &= -\frac{\partial U}{\partial q_i}(q) dt, \quad i \in V \setminus \partial V \\ dp_i &= -\frac{\partial U}{\partial q_i}(q) dt - \gamma p_i + \sqrt{\frac{2\gamma}{\beta_i}} dW_i(t), \quad i \in \partial V \end{aligned} \tag{2.6}$$

The β_i are the inverse temperatures of the heat baths coupled to the boundary sites $i \in \partial V$; $dW_i(t)$ are mutually independent, one-dimensional white noises.
Consider a time-interval $[0, \tau]$; we start the dynamics from the distribution ρ_0 on the (p_i, q_i) obtaining a distribution ρ_τ at time τ. The kinematical time-reversal is $\pi(p,q) = (-p, q)$. A trajectory

$\omega = (p_i(t), q_i(t))$ has time-reversal $\Theta\omega = (-p_i(\tau - t), q_i(\tau - t))$. The density (2.1) was computed in [47] and was found to be (with $k_B = 1$)

$$R(\omega) = \sum_{i \in \partial V} \beta_i J_i(\omega) + \log \rho_0(p(0), q(0)) - \log \rho_\tau(p(\tau), q(\tau)) \tag{2.7}$$

where the J_i are the time-integrated heat currents at the boundary. The definition of these heat currents follows from applying a global energy balance to the equations (2.6). The change of entropy in the reservoirs corresponds to that energy dissipated in the environment divided by the temperature of the reservoirs: $S(\omega) = \sum_i \beta_i J_i(\omega)$.
When we take the average of R in the steady state $\rho_0 = \rho_\tau$ we immediately have from (2.1) that

$$\langle e^{-R} \rangle = 1 \text{ and hence } \langle R \rangle \geq 0$$

or

$$\sum_i \beta_i \langle J_i \rangle \geq 0$$

If the volume V is a linear chain $[0, N]$ with $\partial V = \{0, N\}$ the left and right endpoint of the chain, then by the steady state condition $\langle J_N \rangle = -\langle J_0 \rangle$. Therefore

$$(\beta_0 - \beta_N) \langle J_0 \rangle \geq 0$$

which shows that the heat current into the coldest reservoir is positive. By applying the inequalities at the end of Section 9.2.4 strict inequalities can be obtained. I know of no other method which shows so easily and in such great generality that thermodynamically elementary result.

The computation that leads to (2.7) is the generalization of a Langevin-type calculation starting from:

$$dx(t) = -F(x(t))\, dt + dW(t)$$

for which the path-space measure (2.2) is formally given by

$$P(\omega) = \exp -\frac{1}{2} \int_0^\tau dt \big[\big(\dot{x}(t) + F(x(t))\big)^2 - \nabla F(x(t)) \big]$$

The antisymmetric part under time-reversal in the action $\log P$ indeed gives the dissipated power in terms of a stochastic Stratonovich integral of the "force" $F(x(t))$ times the "velocity" $\dot{x}(t)$.

2.1.2 Asymmetric exclusion process

I now look at a bulk driven diffusive lattice gas where charged particles, subject to an on-site exclusion, hop on a ring in the presence of an electric field and in contact with a heat bath at inverse temperature β. Write $\xi(i) = 0$ or 1 depending on whether the site $i \in \mathcal{T}$ is empty or occupied; $\mathcal{T} = \{1, \ldots, \ell\}$ with periodic boundary conditions. The electric field does work on the system and the 'probability per unit time' to interchange the occupations of i and $i + 1$ is given by the exchange rate

$$c(i, i+1, \xi) = e^{\beta E/2} \xi(i)(1 - \xi(i+1)) + e^{-\beta E/2} \xi(i+1)(1 - \xi(i)) \tag{2.8}$$

Consider a path ω of the process in which at a certain time, when the configuration is ξ, a particle hops from site i to $i+1$, obtaining the new configuration $\xi^{i,i+1}$. Then, the time-reversed trajectory shows a particle jumping from $i + 1$ to i. Therefore, our (2.1) is given by summing over all jump times, contributions of the form

$$\log \frac{c(i, i+1, \xi)}{c(i, i+1, \xi^{i,i+1})} = \beta E \left[\xi(i)(1 - \xi(i+1)) - \xi(i+1)(1 - \xi(i)) \right] \tag{2.9}$$

The right-hand side reconstructs the particle current because we get $+1$ or -1 depending on whether the particle has jumped at time t in the direction of E or opposite to it:

$$S(\omega) = \beta E \sum_{i,t} J_{i,t}(\omega)$$

with $J_{i,t}(\omega) = \pm 1$ depending on whether a particle has passed $i \to i+1$ or $i+1 \to i$. Multiplying the current with the force E (with charge and distance taken to be 1) we obtain the variable Joule heating.

The above is not surprising as it is already implicit in the very definition of the model. The same structure will always be recovered for Markov chains. Here is how it goes:

Suppose that K is a finite set and let $(x(t), t \in [0,\tau])$ be a $K-$valued stationary Markov process. The transition rate to go from a to b is denoted by $k(a,b), a, b \in K$ and we assume that $k(b,a) = 0$ whenever $k(a,b) = 0$ (dynamic reversibility). Suppose now that

$$\frac{k(a,b)}{k(b,a)} = \frac{k_o(a,b)}{k_o(b,a)}\, e^{E\,J(a,b)} \tag{2.10}$$

with the detailed balance condition

$$\frac{k_o(a,b)}{k_o(b,a)} = e^{U(b)-U(a)}$$

for the reference (equilibrium) process with (driving) $E = 0$. The time-reversal of a trajectory $\omega = (x(t))$ is $\Theta\omega = (x(\tau - t))$. We can compute (2.1) or (2.3) directly via a so called Girsanov formula:

$$S(\omega) = U(x(\tau)) - U(x(0)) + E \sum_t J(x(t), x(t^+))$$

where we sum over all the jump times t at which $x(t) \to x(t^+)$. If the model has a physical interpretation with (2.10) expressing local detailed balance for a driving force E and a current J, then $S(\omega)$ indeed reproduces the entropy production. One can extend this to spatially extended systems, so called interacting particle systems, see [49, 50]. One shows there can be no current without heat, meaning that detailed balance is equivalent with zero mean entropy production, see also [60, 61, 68].

2.1.3 Strongly chaotic dynamical systems

Here there is *a priori* no notion of physical entropy production but sometimes, making analogies with systems described via effective thermostated dynamics, one thinks of the phase space contraction as entropy production, see also [18, 1]. The trajectory ω is given by an orbit $(x, \varphi(x), \ldots, \varphi_\tau(x))$ in phase space. For so called Anosov diffeomorphisms φ, there is a natural stationary distribution $P(x)$ which turns out to be a Gibbs measure for the potential

$$U(x) = -\log ||D\varphi|_{E^u(x)}||$$

where $E^u(x)$ is the unstable subspace of the tangent space at the point x ($U(x)$ is the sum of the positive local Lyapunov exponents). Over a trajectory, the Lagrangian equals

$$\mathcal{L}(\omega) = \sum_{t=0}^{\tau} U(\varphi_t x)$$

There is a time-reversal π for which φ is dynamically reversible: $\pi \circ \varphi = \varphi^{-1} \circ \pi$ and we write $\Theta = \pi \circ \varphi, \Theta^2 = \mathrm{id}$.
For (2.3) we need

$$U(\Theta x) = \log ||D\varphi|_{E^s(x)}||$$

where the stable subspace $E^s(x)$ of the tangent space at the point x appears. The expansions and contractions in the unstable and stable directions give together rise to a net contraction rate (positive or negative) $U(\Theta x) - U(x)$. Therefore,

$$S(\omega) \simeq \sum_0^\tau J(\varphi_t x)$$

with $J(x) = -\log ||D\varphi(x)||$, the phase space contraction rate. That is again a version of (2.1) but now obtaining the phase space contraction rate for S. I refer to [18, 66, 52] for more details and for an extension, see also further under Section 7.3.

2.1.4 Diffusion in local thermodynamic equilibrium

I consider diffusion processes whose space-time fluctuations in the history ω of the particle density $n_t(r)$ at space-time (r,t) in the volume V are governed by a Lagrangian

$$\mathcal{L}(\omega) = \frac{1}{2}\int_0^\tau dt \int_V dr \big(\vec{w}, \frac{1}{\chi(n_s(r))}\,\vec{w}\big) \tag{2.11}$$

where $(\cdot,\cdot)$ denotes the scalar product in $\boldsymbol{R}^3$ and

$$\vec{w} = \nabla^{-1}\big(\frac{\partial n_t(r)}{\partial t} - \frac{1}{2}\nabla\cdot(D(n_t(r))\nabla n_t(r))\big)$$

is the vector whose divergence equals the difference between left-hand side and right-hand side in the hydrodynamic equation

$$\frac{\partial n_t(r)}{\partial t} = \nabla\cdot J_r(n_t), \quad \begin{aligned} J_r(n_t) &= \frac{1}{2}\chi(n_t(r))\nabla(-s'(n_t(r))) \\ &= \frac{1}{2}D(n_t(r))\nabla n_t(r)) \end{aligned}$$

$D(n_t(r))$ is the diffusion matrix, connected with the mobility matrix χ via

$$\chi(n_t(r))^{-1}D(n_t(r)) = -s^{''}(n_t(r))\ \mathrm{Id}$$

for the identity matrix Id and the local thermodynamic entropy s.
Such a quadratic form for $\mathcal{L}$ can be derived for (2.2) for a number of stochastic dynamics but it is believed to be very general for diffusion processes, see e.g. [3].
Clearly, (2.3) equals

$$S(\omega) = \int_0^\tau dt \int_V dr\, \big(\nabla^{-1}(\frac{\partial n_t(r)}{\partial t}), \chi(n_t(r))^{-1}D(n_t(r))\nabla n_t(r)\big)$$

or

$$S(\omega) = -\int_0^\tau dt \int_V dr\, \big(\nabla^{-1}(\frac{\partial n_t(r)}{\partial t}), \nabla s'(n_t(r))\big)$$

That is of the form, current $\nabla^{-1}(\partial_t n_t(r))$ times a gradient in the local chemical potential $-s'(n_t(r))$ as we are used to find in expressions of entropy production for local thermodynamic equilibrium. If there is no particle current in or out of the system, we can integrate by parts to find

$$S(\omega) = \int_V dr\,[s(n_\tau(r)) - s(n_0(r))] = S(\tau) - S(0)$$

the total change of entropy.
Apart for the nonlinearities in (2.11) that is fully in line with the Onsager-Machlup set-up of 50 years ago, [58].

2.2 Lagrangian statistical mechanics

In the above examples, the Lagrangian is constructed from a model dynamics. These dynamics need not be fully microscopic; they are valid, effective and useful in some regimes, under some approximations and for some purposes. The thought arises therefore whether one could not as well start the business from specifying the Lagrangian, leaving aside its detailed dynamical origin. After all, also in equilibrium statistical mechanics one constructs models and theories based on approximate Hamiltonians, not worrying all the time about the deeply hidden origins. In that Lagrangian statistical mechanics, far or close to equilibrium, the Gibbs formalism again applies but now for distributions of space-time histories. One advantage is that the step to quantum mechanical processes is much easier to make. Another advantage is that the stationary distribution, the projection to the temporal hyperplane, becomes available for perturbation theory as it is the exponential of Hamilton's principle function, see [3, 36].
At this moment, one requirement for the Lagrangian stands out: that the antisymmetric part under time-reversal be given by the entropy production. The Lagrangian is however more than just its time-antisymmetric part. For perturbation theory beyond linear order, also the symmetric part matters, see Section 10.2. As the fluctuation symmetries of Section 9 will just be reformulations of the main observation (2.1), they cannot be the starting point of a systematic perturbation expansion. For that we need the full Lagrangian, see e.g. [53].

3 The second law in thermodynamics

In the beginning of the 19th century Sadi Carnot followed his father Lazare in meditating over perfect machines and efficiency, see e.g. [23, 12, 34]. That abstraction led Sadi Carnot to think of a generalized heat engine which draws heat from a source and delivers useful work. To operate continuously, the engine requires also a cold reservoir to which heat can be disposed. The new idea of Carnot was to think of a reversible machine, one that can be run in the opposite direction thereby restoring its original thermodynamic state. He realized that no heat engine can be more efficient than a reversible one operating between the same temperatures. These pure thoughts (*Réflexions sur la puissance motrices du feu*, 1824) imply practically useful insights, such as that one need not worry about working substances other than water in the design of steam engines.
Still room was left for further theoretical reflections. For example, since the reversible efficiency is a universal function of the reservoir temperatures, by inverting the reasoning, we can define temperature on a universal scale independent of the particular substance. To that, Kelvin and Joule added their insights that constitute the first law of thermodynamics, the conservation of energy in thermal processes as elevated from the principle of conservation of mechanical energy. Heat and work, naturally appearing in expressions of efficiency, can now be written in the same units and Carnot's principle takes the form

$$\frac{J_1}{T_1} + \frac{J_2}{T_2} \geq 0$$

where $J_1(J_2)$ is the heat given by the engine to the first (second) reservoir at absolute temperature $T_1(T_2)$. Its extension to more reservoirs was given by Kelvin in 1854 but it was Clausius who came with integrals:

$$\oint \frac{dJ}{T} \geq 0 \tag{3.1}$$

where one now imagines a cyclic process through which the engine makes contact with reservoirs at temperature T. If the process is reversible, the equality applies in (3.1) and T is also the (instantaneous) temperature of the system. But since then

$$\oint \frac{dJ}{T} = 0$$

for every cycle, we get the path-independence of the line integral

$$\int_\alpha^{\alpha'} \frac{dJ}{T} = S(\alpha') - S(\alpha) \tag{3.2}$$

where we integrate over a reversible path (equilibrium states). Thus Clausius discovered in 1865 a new function of the thermodynamic state. He called S the entropy (from the Greek $\tau\rho o\pi\eta$, turn, a turning point). As a direct consequence of (3.1)–(3.2), by closing the cycle,

$$\int_{\alpha'}^{\alpha} \frac{dJ}{T} \geq S(\alpha') - S(\alpha) \tag{3.3}$$

for all paths between macrostate α' and α of the system and processes where the system is in contact with a reservoir at temperature T and gives it heat dJ. The left-hand side is the total heat dissipated into the reservoirs and hence, is the change of entropy in the rest of the universe. Putting everything at the same side, we thus see the familiar

$$S(\text{final}) - S(\text{initial}) \geq 0 \tag{3.4}$$

for the total entropy of the universe, or, as interpreted by Clausius, *Die Entropie der Welt strebt einem Maximum zu.*

The equality (3.2) is the operational definition of entropy. Entropy is only defined here for certain thermodynamic states, what are called equilibrium states under specified restraints on composition, energy, etc. They are interconnected via reversible processes.
One does not need to remain with energy exchanges only. If we have an equilibrium system with energy E and particle numbers N_i in a volume V, the entropy $S(E, N, V)$ is defined as in (3.2) from the exact differential

$$dS = \frac{1}{T}(dE + pdV - \sum \mu_i dN_i)$$

where T is the absolute temperature, p is the pressure and μ_i are the chemical potentials.
That definition of thermodynamic entropy can be extended to systems in local thermodynamic equilibrium. Thermodynamics teaches that entropy is additive and when surface effects can be neglected, we also get extensivity: $S(E, N, V) = V s(e, n)$. We then introduce local energy densities $e(r)$, particle densities $n(r)$ (for one component) and a velocity profile $u(r)$ for which $\int_V n(r)u(r)\,dr = 0$ to write

$$S(e, n, u) = \int_V s\big(e(r) - \frac{1}{2}mn(r)v^2(r)\,,\, n(r)\big)\,dr \tag{3.5}$$

One imagines here the system as composed of microscopically very large domains that are still much smaller than the thermodynamic size of the system.
The inequality (3.4) describes the net result, nothing intermediate, of a process that starts and ends in equilibrium. One can also apply it to (3.5) with the understanding that during the process the system remains in local thermodynamic equilibrium. One must then add the macroscopic evolution equations for the quantities $e_t(r), n_t(r)$ and $u_t(r)$ and introduce intensive quantities like local temperature to obtain the entropy balance equation. For example, if the system is isolated with local temperature $T(r)$ and constant $n(r)$ and $u(r) = 0$, the changes in energy satisfy the conservation law $de_t(r)/dt + \text{div}J(r) = 0$ with J the heat current and we have by partial integration

$$\frac{d}{dt}\int_V s(e(r))\,dr = \int_V J \cdot \nabla \frac{1}{T}\,dr$$

The linear transport law (here, Fourier's law) defines the coefficient λ for which $J = -\lambda/T^2\,\nabla T$ and the entropy production rate equals

$$\frac{d}{dt}\int_V s(e(r))\,dr = \int_V \lambda\,(\nabla \frac{1}{T})^2\,dr \geq 0$$

requiring $\lambda \geq 0$. These manipulations serve many different scenario's, for open or for closed systems, in a stationary or in a transient regime, and are typical for, or, what is worse, are restricted to irreversible thermodynamics close to equilibrium.

The above summary contains the standard propositions of macroscopic phenomenology and can be found in many textbooks, e.g. [2]. Yet, it is not particularly illuminating when asked what entropy really is and how to extend it to truly nonequilibrium situations. The microscopic understanding of entropy and the second law was first present in the works of Maxwell, Boltzmann and Gibbs. It is in fact the start of statistical mechanics.

4 Second law from microscopic considerations

The relation (3.2) can be studied in quite some detail for an ideal gas. A simple calculation shows that there the (equilibrium) entropy $S(E)$ can be related to the total phase volume W. More specifically, up to an additive constant,

$$S = k_B \log W \tag{4.1}$$

where, for an ideal gas of N atoms in volume V, for which the energy lies in $(E, E + dE)$,

$$W \equiv \int_{\sum p_i^2/2m \simeq E, q_i \in V} dq_1 \dots dq_N \, dp_1 \dots dp_N$$

The formula (4.1) appears on the tombstone of Boltzmann in Vienna but it was first written down in that form (and criticized) by Planck who introduced the W as thermodynamic *Wahrscheinlichkeit*. It guided the young Einstein to propose experimental verifications of the corpuscular nature of fluids and of radiation. Already by its simplicity we cannot but feel that the first mysteries regarding entropy should now evaporate. Boltzmann had truly realized that counting is essential for predicting macroscopic behavior and that entropy was just that.

4.1 Counting

The formula (4.1) can be read and generalized as follows. An immense number of microstates belong to one and the same macrostate (manifest condition). The entropy counts them.
Counting seems to refer to something discrete and introducing discreteness in classical phase space has something arbitrary. One can of course refer to quantum mechanics but that introduces again other problems (more about that later). I give more precise definitions later but in fact, not much is necessary. Most important however is to treat all microstates, that is every assignment of values of positions and momenta of the particles, as equivalent. That is the microcanonical distribution and its relevance is mostly derived from it being left invariant by the Hamiltonian equations of motion (Liouville's theorem).

Denote by Ω the phase space of a perfectly closed and isolated mechanical system. The elements $x \in \Omega$ represent the microstates, i.e., $x = (q_1, \dots, q_N, p_1, \dots, p_N)$ gives the canonical variables for a classical system of N particles. The equations of Hamilton produce the flow $x \mapsto \varphi_t(x)$ on Ω and preserve the phase space volume: $|d\varphi_t(x)/dx| = 1$ for each t; the Liouville measure dx is time-invariant. We think of a large volume in which the many particles are conserved and the dynamics also conserves the total energy. We introduce therefore the state space $\Omega_E \equiv \Gamma$, the energy shell, corresponding to energies within $(E, E + dE)$ and denote by $|B|$ the phase space volume of a region $B \subset \Gamma$ given by the projection of the Liouville measure into Γ.
Now we change scales. For comparison with experimental situations, we look at variables that summarize the manifest image of the system. We have in mind a collection of macroscopic variables $A_r(x)$, mostly approximately additive and locally conserved functions on phase space. We obtain a subdivision of Γ by cutting it up in phase cells defined by $\alpha_r < A_r(x) < \alpha_r + \Delta\alpha_r$ (with some tolerance $\Delta\alpha_r$). As a result, to every $x \in \Gamma$ is associated a region $M(x)$ in phase space consisting of all microstates that share with x the same manifest image. Boltzmann defines then the entropy as

$$S(x) = k_B \log |M(x)| \tag{4.2}$$

where we leave out some irrelevant constants that appear when the number of particles can change. That is the meaning of (4.1). Most important to remember in scrutinizing definition (4.2) is the great disparity in sizes of $W = |M(x)|$ and the smallness of Boltzmann's constant $k_B = 1.381 \times 10^{-23}$ J/K. An entropy difference $S' - S$ of about 0.1 millicalorie at room temperature corresponds to a phase volume ratio of

$$\frac{W}{W'} = \exp -(S' - S)/k_B = e^{-10^{20}}$$

Since entropy is extensive in the number of particles, any visible change in the entropy per particle (as measured in units of k_B) corresponds to a ratio of phase volumes that is exponential in the number of particles.

4.2 The hand-waiving part

Equilibrium is the state of maximal entropy; it has the overwhelmingly largest volume in phase space which makes it the typical endpoint for about every evolution. Equilibrium can thus be described via a variational principle, maximizing the entropy while keeping some constraints fixed. If a constraint is lifted, even in a time-dependent way, new macrostates can be explored and the entropy will increase until a new equilibrium is installed. In that sense, the very definition of equilibrium as solution of a variational principle is already incorporating the second law of thermodynamics.

The microscopic definition (4.2) of entropy allows to move beyond equilibrium thermodynamics. We can now follow a microstate and its entropy as time runs. Accepting the counting procedure above and realizing the great difference in scales, that entropy will increase needs no further explanation. Aside from great conspiracies, it is expected that the dynamics takes the system into macrostates with a larger number of compatible microstates, [24, 4, 41]. If in a large collection of black and white balls the number of white balls is about a billion times larger than the number of black balls, then, when picking a ball by whatever which procedure that treats the balls equivalently, we would be surprised not to have picked a white one.

Violations of that microscopic version of the second law are now absolutely possible and in fact ought to be expected when the number of particles and/or the lapse of time over which the system is monitored gets small. It is only because of the huge amount of particles in one mole of a gas that we never witness Poincaré recurrencies. That is the answer to the Zermelo paradox.
The second law appears very rigid when applied in the laboratory situation or in the understanding of macroscopic behavior. After all that is why it is called a law. Nevertheless, its microscopic roots are statistical. In the end it is a probability statement, a *moral certainty* as Maxwell put it. Or, with Gibbs: *the impossibility of an uncompensated decrease of entropy seems to be reduced to an improbability.*
Even if we have a large number of particles, one can easily imagine initial conditions for which the evolution breaks the second law. The second law only applies to typical initial data, microstates randomly selected from within the phase volume that corresponds to the initial macrostate.

4.3 Time-(a)symmetry

It is in the last sentence of the previous section that appears the truly hard part about the second law. To state that more precisely let me first recomfort the reader that so far no arrow of time has appeared. Think indeed of a dynamically reversible time-evolution. By reversing all molecular velocities, the *same* equations of motion carry the system back along their very *same* trajectories to their initial positions where they end up with their velocities reversed. That is the reversibility that is present in the microscopic dynamics of a mechanical system. Everything we have said before to make the second law understandable applies to both directions of time. No *a priori* sign of the microscopic time needs to be selected and the change of entropy is then expected to be equal and positive in both time-directions. Furthermore, for every initial microstate from which the entropy increases, there is also a microstate from which the Boltzmann entropy decreases. Nevertheless,

that Loschmidt construction does not give rise to a paradox because the second law applies to a typical initial condition and the microstates that are obtained after evolution (with reversed molecular velocities) make only a very small fraction of the microstates that are actually corresponding to the final macrostate.
The deeper question is however why we can have nonequilibrium states in the first place. The second law can only come in action if the system is initially prepared in a macrostate with a much smaller phase volume than is reachable by the evolution. That has obviously been done by putting constraints of some sort but all the same it is again the endpoint of a previous evolution for a bigger system, perhaps including now ourselves. Iterating this argument an arbitrary number of times, we soon reach questions of physical cosmology. A healthy attitude, here and elsewhere, seems to be that instead of trying to (dis)prove the second law, one should use it. In fact, the second law now guides us to conclude within a particular theory as to the nature of the initial condition of the universe; it must allow for the huge entropy increases that have occurred ever since. Obviously, the theory of general relativity and hence gravity play there an all-important role. Within the standard theory of physical cosmology, that leads to precise mathematical estimates on various aspects of the initial singularity. We refer to Chapter 7 in [59] and to [35] for a first look at such derivations.

4.4 Shadowing the macrostate evolution

I am adding here some more mathematical relations to the discussion above. The advantage is not so much precision but generality and recognizing what is less and more essential.

Let Γ be the phase space of a dynamical system. We ask that Γ is equipped as a probability space with some probability measure ρ. Let φ_t be a map on Γ that leaves ρ invariant.
Since we plan to do statistical mechanics and have in mind systems composed of a huge number of particles, we specify a number of macroscopic variables

$$A_r(x) = \frac{1}{N}\sum_{k=1}^{N} a_r^k(x) \tag{4.3}$$

with a_r^k a function that only depends say on the state of the $k-$th particle. For example, we could divide the six-dimensional one-particle phase space in cells C_r and let a_r^k be one or zero depending on whether the state of the $k-$th particle $x_k \in C_r$ or not. Or, the index r can have a purely spatial interpretation in which case $A_r(x)$ gives the profile of some one-particle observable. We can of course go beyond one-particle variables and r can then indicate various other macrovariables but that is not essential here. In short, a macrostate $\alpha = (\alpha_r)$ is realized by the microstate x when $A_r(x) = \alpha_r$ plus/minus some tolerance. I write $\alpha(x)$ for it.

The macrostate α can be represented in different mostly equivalent ways. The most fundamental way appears to be the microcanonical formulation. We then think of Γ as the constant energy surface and ρ is the projection of the Liouville measure. We associate with α the phase volume M_α of all microstates x for which $\alpha(x) = \alpha$ and the entropy of such an x is then, following (4.2), $S(x) = \log |M_\alpha|$.
It is however often more convenient to imagine that the macrostate α is not given via a phase volume but via a distribution function ρ_α. I thus prefer to go to other Gibbs ensembles and we associate to α a probability measure ρ_α on Γ having a density with respect to ρ. The probability ρ is for example the Gibbs measure at some inverse temperature β with respect to some microscopic interaction and ρ_α has a density

$$\frac{d\rho_\alpha}{d\rho}(x) = \exp[-N\sum_{r=1}^{n} \lambda_r A_r(x) - \ln \frac{Z_\lambda}{Z}] \tag{4.4}$$

with respect to ρ. The Lagrange multipliers λ_r, conjugate to the A_r, are determined from requiring the ρ_α-expectations $\langle A_r\rangle_\alpha = \alpha_r$ and ρ_α is a constrained or, depending on the interpretation of the

index r, a local equilibrium measure in the canonical set-up; Z_λ and Z are the partition functions corresponding to ρ_α and ρ respectively.
It is important to select the pure phases for which $A_r(x) = \alpha_r$ with overwhelming $\rho_\alpha-$probability. These satisfy a law of large numbers. In all cases, the negative logarithm of the density

$$S(x) = -\log \frac{d\rho_{\alpha(x)}}{d\rho}(x) \tag{4.5}$$

corresponds to the variable entropy as in (4.2). Indeed, the negative logarithm of (4.4) corresponds exactly to the entropy governing the (equilibrium) fluctuations of ρ. Write $W_N(\alpha)$ as the $\rho-$probability to see $A_r(x) = \alpha_r$. Then, approximately for large N,

$$W_N(\alpha) = \exp[N \sum_{r=1}^{n} \lambda_r \alpha_r + \log \frac{Z_\lambda}{Z}], \quad W_N(\alpha(x)) = e^{S(x)} \tag{4.6}$$

The precise formulation (and proof) of (4.6) belongs to the theory of large deviations, see e.g. [71, 39, 22], but that is just an elaboration of the older Boltzmann-Planck-Einstein relation (4.1): the entropy governing the macroscopic fluctuations.

We have thus physical interest in considering (4.5) and we would hope that $S(\varphi_t(x)) \geq S(x)$ at least for most of the microstates x that are drawn out of a particular macrostate, represented by ρ_α. Expectations with respect to ρ_α are written as $\langle\cdot\rangle_\alpha$ and without subscript, $\langle\cdot\rangle$ is an expectation in ρ. I now claim that

$$\langle e^{-S(\varphi_t x)+S(x)}\rangle_\alpha = 1 \tag{4.7}$$

The reason is simple. For almost all x, under ρ_α, $\alpha(x) = \alpha$ and hence

$$\langle e^{-S(\varphi_t x)+S(x)}\rangle_\alpha = \langle e^{-S(\varphi_t x)}\rangle = \langle e^{-S(x)}\rangle = 1$$

by definition and by the invariance of ρ under φ_t. The equality (4.7) implies that

$$\langle S(\varphi_t x) - S(x)\rangle_\alpha \geq 0 \tag{4.8}$$

with strict inequality in all non-trivial cases (if $S(\varphi_t x) - S(x)$ is really variable, not constant equal to zero).

Observe that for microstates x,

$$\langle S\rangle_{\alpha(x)} = S(x)$$

On the other hand, if we define $F(\alpha) = \alpha'$ from $\alpha'_r \equiv \langle A_r \circ \varphi_t\rangle_\alpha$, i.e., the macrostate after evolution with φ_t, then

$$\langle S \circ \varphi_t\rangle_{\alpha(x)} = S(x')$$

for every x' with $\alpha(x') = F(\alpha(x))$. As a conclusion, the inequality (4.8) is for the change in entropies

$$S(x') - S(x) \geq 0$$

for all x' that realize the macrostate that comes out from evolving the macrostate to which x belongs. The word macrostate is here understood in the Gibbsian sense, i.e., in terms of the probability distributions ρ_α introduced before. That implies that we have not succeeded yet in deriving the microscopic picture and that the entropies in (4.8) remain equilibrium Gibbs entropies.

5 H-theorem

At the end of the previous paragraph, we got somehow an evolution from one macrostate to another one, both equilibrium, almost by definition and not because the dynamics is like that. Not surprisingly, we did not end up with the stronger Boltzmann formulation of the second law. Yet, we can obtain more than the second law by requiring that the evolution of macroscopic variables

is autonomous.

We think again about the microcanonical ensemble. In that case, ρ is the Liouville measure on a constant energy surface Γ. We make a partition of Γ according to the possible values of a collection of macroscopic variables $A_r(x)$ as in (4.3). The $M(x)$ of (4.2) is the phase volume containing all the $y \in \Gamma, A_r(y) = A_r(x)$. A macrostate α specifies one such phase volume, call it M_α and $M_{\alpha(x)} = M(x)$. If wished we can proceed as in the previous section and speak about the distribution ρ_α, now concentrated on one phase volume M_α and random within.

Suppose that at a certain moment, along the trajectory, x is our microstate with corresponding macrostate $\alpha = \alpha(x)$. If we now apply the dynamics φ_t, *a priori* $\varphi_t(x)$ can fall in various different macrostates even when we know, as we do, that $x \in M_\alpha$ initially. In other words, we get a non-trivial statistics on at least some macroscopic values and it seems we fall back in the framework of section 4.4, the canonical Gibbs formalism. One can proceed along these lines but let me however deviate here and instead make an additional assumption. I assume that the macroscopic partition is so well chosen for the dynamics φ_t that (at least approximately for systems composed of many particles) in fact the dynamics is autonomous on the macroscopic variables. That means that from knowing the macrostate α at some (arbitrary) time, we also know the macrostate at a later time: almost all $y \in M_\alpha$ satisfy $\varphi_t(y) \in M_{\alpha'}$ for the same macrostate α'. Then, clearly, the image of macrostate α under φ_t is concentrated within the phase volume $M_{\alpha'}, \varphi_t(M_\alpha) \subset M_{\alpha'}$ or better $|\varphi_t(M_\alpha)| \leq |M_{\alpha'}|$. As a consequence, by Liouville's theorem

$$|M_\alpha| = |\varphi_t(M_\alpha)| \leq |M_{\alpha'}|$$

and thus, the entropy (4.2) is typically non-decreasing. We have now obtained something even stronger than the second law, along a microscopic trajectory $S(\varphi_t(x)) \geq S(\varphi_s x), t \geq s$ from a typical initial microstate, i.e., one that realizes macrostates according to the autonomous equation.

Observe that the previous argument, first given in [33] and recently described in [25, 26] captures very well the intuitive meaning of a constant increase in entropy but it delegates the problem to establishing an autonomous equation for the macrovariables. That macroscopic reproducibility is well-documented in experience but few are the examples where we can actually achieve it starting from microscopic models. The most famous one is the Boltzmann equation where the $H-$theorem was first proven for dilute gases. In contrast to what is often claimed, one does not need an extra assumption of randomization or *Stoßzahlansatz*, see [40]. In general, it remains to be seen what are the essential conditions on the dynamics that produce an $H-$theorem. Certainly, the chaotic nature of the microscopic dynamics can help, as adding stochasticity helps to prove it. Yet that may not be necessary.
A formulation of what an $H-$theorem means in a quantum context is at the end of the next section.

6 Quantum entropies

The logic and usefulness of the ideas above are not at all restricted to classical statistical mechanics. In quantum mechanics, one can go quite a bit in the same direction. Literally *a bit*, because in some sense the quantum situation is more discrete (quantized) and therefore seems more amenable to counting. Yet there are problems. One has to do with the nature of the microstates for a quantum system, the phase space. We have to make sure that we know what exactly there is to count and what we mean by a quantum history. Secondly, there is the problem of macroscopic measurements. One can always say that macroscopic variables are classical variables because they commute and we can measure them simultaneously. Yet, before taking the thermodynamic limit, they do not commute. Thirdly, there is no well-established theory of large deviations. The basic relation between entropy and fluctuations of macroscopic quantities, even in equilibrium, has not been satisfactorily settled. I will not discuss these problems here, only mention them again in context, and I restrict myself to giving some definitions which appear to be no longer very standard.

One would like to say that quantum entropy is the logarithm of the dimension of the subspace in the Hilbert space $\mathcal{H}$ of the system, that corresponds to the manifest condition. I suppose that $\mathcal{H}$ is finite dimensional (but very large). Given a microstate $\psi \in \mathcal{H}$ one then follows von Neumann [57] in writing

$$S(\psi) \equiv \sum_\alpha (\psi, P_\alpha \psi) \log d_\alpha - \sum_\alpha (\psi, P_\alpha \psi) \log (\psi, P_\alpha \psi) \tag{6.1}$$

corresponding to the decomposition

$$\mathcal{H} = \bigoplus \mathcal{H}_\alpha \tag{6.2}$$

into linear subspaces. The manifest condition (or macrostate) is represented by the projections P_α on $\mathcal{H}_\alpha, P_\alpha P_\beta = \delta_{\alpha,\beta} P_\alpha, \sum_\alpha P_\alpha = \mathrm{id}$. We write d_α for the dimension of $\mathcal{H}_\alpha$; it is the analogue of the phase space volume in the classical case.
A problem here is that if we have more than one macroscopic observable, say the magnetization in the $z-$ and in the $x-$direction for a collection of spin 1/2-particles, in general not commuting before the thermodynamic limit is taken, then they do not have a joint eigenspace decomposition and we are in the dark as to what to write for P_α.

The sum in (6.1) reflects that the wavefunction can still correspond to different (mutually exclusive) macrostates (here labeled by the running index α). That feels like the description starting from the wavefunction as microstate is not complete. That discussion would take me way too far. Part of the problem is what we mean by statistical ensembles. As an example, the identification of probabilities on wavefunctions and density matrices does not seem one-to-one.

One can move definition (6.1) to the level of density matrices and call

$$S(\rho) \equiv \sum_\alpha \mathrm{Tr}[P_\alpha \rho] \log d_\alpha - \sum_\alpha \mathrm{Tr}[P_\alpha \rho] \log \mathrm{Tr}[P_\alpha \rho] \tag{6.3}$$

the entropy of the state represented by ρ. Note that (6.3) (just like (6.1)) only depends on ρ through its projection $p(\rho)$ (a probability measure) on the macroscopic states:

$$p(\rho)(\alpha) \equiv \mathrm{Tr}[P_\alpha \rho] \tag{6.4}$$

Underlying is the variational principle

$$S(\rho) = \sup_{p(\rho')=p(\rho)} - \mathrm{Tr}[\rho' \log \rho'] \tag{6.5}$$

with the supremum reached at

$$\rho' = \sum_\alpha \frac{p(\rho)(\alpha)}{d_\alpha} P_\alpha \tag{6.6}$$

The above mimics rather well the situation in classical statistical mechanics and the relation there between Boltzmann and Gibbs entropies and how (constrained) equilibrium is characterized by a variational principle. So we can move beyond (6.5) and construct the quantum equilibrium entropy just as one does in the Gibbs formalism, see section 4.4. The constraints are then in the form of specifying expectations like $\mathrm{Tr}[\rho' A_r]$, for a class of macroscopic observables A_r, and we obtain the quantum equilibrium states, generalizing (6.6) to the standard Gibbs form. The problem is however that we do not have the fluctuation formula (4.6) so important for (4.1). In fact, I even believe it is not true in general. For example, suppose we have a collection of N spin 1/2-particles with a local Hamiltonian H at inverse temperature β. The equilibrium reference state has thus a density matrix $\exp[-\beta H]/Z$. Suppose now the additional constraint that

$$\mathrm{Tr}[M_z] = \alpha_z$$

where $M_z = \sum_{i=1}^N \sigma_i^z / N$ is the magnetization in the $z-$direction. The constrained state has density $\exp[-\beta H - \lambda N M_z]/Z_\lambda$ for suitable Lagrange multiplier $\lambda = \lambda(\alpha_z)$. Its free energy is

$$\frac{1}{N} \log \mathrm{Tr}[e^{-\beta H - \lambda N M_z}] \tag{6.7}$$

On the other hand, the probability in equilibrium that the magnetization is about α_z is

$$\frac{1}{Z}\mathrm{Tr}[e^{-\beta H}\, P_{\alpha_z}] \tag{6.8}$$

where P_{α_z} is the projection on the space $\mathcal{H}_{\alpha_z}$ with eigenvalues of M_z around α_z. Now, in contrast with (4.6), the logarithm of (6.8) will not be given by the Legendre transform of the free energy (6.7) but rather by Legendre transforming

$$\frac{1}{N}\log \mathrm{Tr}[e^{-\beta H}\, e^{-\lambda N M_z}] \tag{6.9}$$

and these might give different results even in the limit $N \uparrow +\infty$ when the commutator $[H, M_z]$ is of order one.

Let me finally present the logic of any $H-$theorem that also works in a quantum set-up, [10]. To a density matrix ρ, we associate a macrostate $\alpha = \alpha(\rho)$ that collects the values of expectations under ρ for a selection of macroscopic observables. To a macrostate α is associated a constrained equilibrium state ρ_α which maximizes

$$-\mathrm{Tr}\ [\rho' \log \rho'], \quad \alpha(\rho') = \alpha$$

The maximum is the entropy $S(\alpha) = \text{-}\mathrm{Tr}[\rho_\alpha \log \rho_\alpha]$ of α. I write $S(\rho) = S(\alpha(\rho))$ for the entropy corresponding to the state represented by ρ.
Now dynamics comes; let $\rho(s)$ denote the density matrix at time s under a unitary evolution starting from ρ. We have, by Liouville-von Neumann,

$$\begin{aligned} S(\rho(s)) &= S(\alpha(\rho(s))) = S(\rho_{\alpha(\rho(s))}) \\ &= S(\rho_{\alpha(\rho(s))}(t-s)) \end{aligned} \tag{6.10}$$

The macroscopic autonomy-assumption of Section 5 is written as follows:

$$\text{if } \alpha(\rho(s)) = \alpha(\rho'(s)), \text{ then } \alpha(\rho(t)) = \alpha(\rho'(t)), \quad t \geq s$$

In (6.10), under that autonomy-assumption,

$$\alpha(\rho_{\alpha(\rho(s))}(t-s)) = \alpha(\rho(t))$$

and hence, by the variational characterization of the entropy

$$S(\rho_{\alpha(\rho(s))}(t-s)) \leq S(\rho(t))$$

That gives the quantum $H-$theorem

$$S(\rho(s)) \leq S(\rho(t)), \quad s \leq t$$

when combined with (6.10). As for classical dynamics, the main problem remains to understand when the autonomy is established.

In the rest of the paper, I will not come back to quantum aspects. It does not mean that there are no quantum versions of what follows. The relation between time-reversal and quantum entropy is discussed in [5]. The simplest example of relaxation to equilibrium for a unitary evolution on quantum spins, with an associated $H-$theorem can be found in [11]. An extension to some quantum versions of a fluctuation relation that connects irreversible work with free energy or to so called transient fluctuation theorems is in [38, 9, 70, 54, 55, 69, 74].

7 Time-reversal and entropy

In the present section, I give a general argument for believing in the main observation under (2.1). It is a compact upgrade of what was written in [46] in collaboration with K. Netočný. Those that are happy with examples are referred to Section 2.1 and to [48] for even more examples.

7.1 Mathematical set-up

I start with formalities that generate a quite broad class of possible scenarios. Illustrations come in the next subsection.
As we have seen before in Sections 4.3–4.4, it is not necessary for the second law that the microscopic dynamics be time-reversal invariant. From now on however, I will only consider the case that the microscopic dynamics is reversible; an extension to microscopically irreversible dynamics is in [45].

Let Γ be the phase space. There are dynamics on Γ, invertible transformations f_t, possibly depending on t, and I write

$$\varphi_t = f_t \circ f_{t-1} \circ \dots f_1, \quad t = 1, \dots, \tau$$

for the time-inhomogeneous updating after τ steps. If we reverse the order (for fixed τ), we get the reversed dynamics

$$\widetilde{\varphi_t} = f_{\tau-t+1} \circ \dots \circ f_{\tau-1} \circ f_\tau$$

I skip the continuous time version.
The phase space Γ supports a measure ρ that is left invariant by φ_τ: $\rho(\varphi_\tau^{-1}B) = \rho(B)$. Γ is further equipped with an involution π that also leaves ρ invariant. I assume dynamical reversibility in the sense that for all t,

$$f_t \circ \pi = \pi \circ f_t^{-1}$$

As a consequence, $\pi\tilde{\varphi_t}^{-1}\pi = f_\tau \circ \dots f_{\tau-t+1}$.

The statistical physics begins from considering a collection $\mathcal{A} = \{A_r\}$ of functions $A_r : \Gamma \to \boldsymbol{R}$ indexed via subscript r. It divides Γ in volumes M_α containing all $x \in \Gamma$ with $A_r(x) = \alpha_r$, plus/minus some tolerance that I ignore here. The label α runs the contracted description.
A statistics μ on $\mathcal{A}, \mu : \mathcal{A} \to \boldsymbol{R}$ assigns a real number to every function A_r.
It is also understood that the collection $\mathcal{A}$ is globally invariant under π so that we can define $\pi(A_r(x)) = A_r(\pi x) = A_{\pi r}(x)$ and $\pi\alpha$. The time-reversal of a statistics μ is written $\tilde{\mu}$ with $\tilde{\mu}(A_r) = \mu(\pi A_r)$.

To a statistics μ I associate a probability measure on Γ. That is done via a variational principle (which I assume is well-posed).
The entropy of a statistics μ is the supremum

$$S(\mu) = \sup - \int_\Gamma g(x)\, \log g(x)\, \rho(dx) \tag{7.1}$$

over all probability densities $g \geq 0$ with

$$\int g(x)\, A_r(x)\, \rho(dx) = \mu(A_r), \quad \int g(x)\, \rho(dx) = 1$$

That entropy governs the fluctuations of ρ.
I denote by g_μ the density that reaches the supremum in (7.1) and ρ_μ is the probability measure on Γ with density g_μ with respect to ρ. The variable and μ-dependent entropy of x is

$$S_\mu(x) = -\log g_\mu(x) \tag{7.2}$$

The probability ρ_μ evolves under the dynamics and gives rise to a new statistics

$$\mu_\tau(A_r) = \int A_r(\varphi_\tau(x))\, g_\mu(x)\, \rho(dx)$$

at later time τ.

Consider now a trajectory $\omega = (\alpha(0), \alpha(1), \dots, \alpha(\tau))$ collecting all $x \in \Gamma$ for which $A_r(x) = \alpha_r(0), \dots, A_r(\varphi_\tau(x)) = \alpha_r(\tau)$ for all r. Its time-reversal is $\Theta\omega = (\pi\alpha(\tau), \pi\alpha(\tau-1), \dots, \pi\alpha(0))$. The

probability to see ω under the dynamics φ_t started from ρ_μ is denoted by $P_\mu(\omega)$. The probability to see $\Theta\omega$ under the reversed dynamics $\tilde\varphi_t$ started from $\rho_{\widetilde{\mu_\tau}}$ is denoted by $\widetilde{P}_{\widetilde{\mu_\tau}}(\Theta\omega)$. Observe that it is natural to run the reversed dynamics because if for some $x \in \Gamma, y_0 = x, \ldots, y_n = \varphi_\tau x$, then $(\pi y_\tau, \pi y_{\tau-1}, \ldots, \pi y_0)$ is an orbit for the reversed dynamics.

Now comes the major claim that constitutes a part of the first observation in Section 2. Take $x \in M_0$, i.e., with the $A_r(x) = \alpha_r(0)$ and suppose that $\varphi_\tau(x) \in M_\tau$, i.e., with the $A_r(\varphi_\tau x) = \alpha_r(\tau)$. Then,

$$S_{\mu_\tau}(\varphi_\tau x) - S_\mu(x) = \log \frac{P_\mu(\omega)}{\widetilde{P}_{\widetilde{\mu_\tau}}(\Theta\omega)} \tag{7.3}$$

It is only a part of the first claim in 2 because (7.3) really talks about a transient behavior and not a steady state situation of a driven system maintained in a nonequilibrium state. Nevertheless, (7.3) is the mother of all further relations, see below in Section 7.3, but let us first see why (7.3) is true. Observe that

$$P_\mu(\omega) = e^{-S_\mu(x)}\, \rho[\cap_{t=0}^{\tau} \varphi_t^{-1} M_t]$$

while

$$\widetilde{P}_{\widetilde{\mu_\tau}}(\Theta\omega) = e^{-S_{\mu_\tau}(\varphi_\tau x)}\, \rho[\cap_{t=0}^{\tau} \tilde\varphi_t{}^{-1} \pi M_{\tau-t}]$$

But the last factor can be rewritten by using that $\rho(B) = \rho(\varphi_\tau^{-1}\pi B)$,

$$\rho[\cap_{t=0}^{\tau} \tilde\varphi_t{}^{-1} \pi M_{\tau-t}] = \rho[\cap_{t=0}^{\tau} \varphi_\tau^{-1} \pi \tilde\varphi_t{}^{-1} \pi M_{\tau-t}]$$

Finally use that $\varphi_\tau^{-1}\pi\tilde\varphi_t{}^{-1}\pi = \varphi_{\tau-t}^{-1}$ to conclude (7.3).

7.2 Illustrations

A simple illustration of the above notation is to take for Γ the phase space of constant energy for N particles in a volume V. The dynamics $f_1 = \ldots = f_\tau, \varphi_\tau = f^\tau$ is a discretization of a conservative Hamiltonian dynamics with ρ the projection of the Liouville measure on $\Gamma : \rho(B) = |B|$. One can imagine a finite partition of Γ corresponding to values of some set of macroscopic variables and each phase volume M_α collects the microstates that show the same manifest condition or macrostate α. A statistics μ gives an initial probability measure on the macrostates, $\mu(\alpha)$ and the constrained equilibrium ρ_μ, that solves the variational principle (7.1) is

$$\rho_\mu(B) = \sum_\alpha \mu(\alpha) \frac{\rho(B \cap M_\alpha)}{|M_\alpha|}$$

Its density with respect to ρ is thus $g_\mu(x) = \mu(\alpha)/|M_\alpha|$ when $x \in M_\alpha$. The entropy (7.2) then equals $S_\mu(x) = \log |M_{\alpha(x)}| - \log \mu(\alpha(x))$. Its expectation under ρ_μ is

$$\langle S_\mu \rangle_\mu = \sum_\alpha \mu(\alpha)\, \log |M_\alpha| - \mu(\alpha)\, \log \mu(\alpha)$$

If μ is concentrated on just one macrostate α, then $\langle S_\mu \rangle_\mu = \log |M_\alpha|$ which corresponds to the Boltzmann entropy if that macrostate is selected from a microstate x as $\alpha_r = A_r(x)$.

A second illustration is obtained if we write ρ_μ in the Gibbs form. Say,

$$g_\mu(x) = \frac{1}{Z} e^{-\sum_r \lambda_r A_r(x)}$$

which solves the variational principle (7.1) for a suitable choice of λ_r made from the $\mu(A_r)$. The entropy $S(\mu)$ is the usual equilibrium Gibbs entropy in the ensemble determined by the A_r. The variable entropy $S_\mu(x)$ has $\rho_\mu-$expectation exactly equal to $S(\mu)$. That is generally true; taking expectations of (7.3) under ρ_μ gives the change of the Gibbs entropy in terms of the expected breaking of time-reversal symmetry.

7.3 Open and driven

Just like in equilibrium statistical mechanics, it is useful to consider reservoirs that interact with the system and to derive a thermodynamics in terms of quantities that depend solely on the system's state. In other words, we want to integrate out the degrees of freedom of reservoirs, heat or particle baths, and introduce variables that describe what is going on in terms of the evolution of the system. In nonequilibrium statistical mechanics, there are plenty of effective models that describe just that. We have seen two examples of that, one a Markov diffusion process and the other a Markov jump process in Section 2.1. It is however interesting to see whether the extensions of (7.3) that were obtained there, have a more general validity. The goal is therefore to integrate out the reservoirs from the right-hand side in (7.3) and to obtain an expression which is again of the same form (as the right-hand side of (7.3)) but now in terms of probabilities for a history of the system only. That question was discussed in [46, 45] and I give here a summary of the positive result.

I consider now a system in contact with k reservoirs. The microstate x is decomposed as $x = (x_S, x_1, \ldots, x_k)$ with x_S representing the system variables and x_v stands for the microstate of the $v-$th reservoir. The macrostates α will now only be macroscopic in so far as the reservoirs are concerned; the system remains described by x_S. So we have macroscopic observables $A_r^v(x_S, x_v)$ whose values are determined from knowing the state x_S of the system and the state x_v of the $v-$th reservoir. An α is given if we know x_S and also $A_r^v(x_S, x_v)$ for all r and v. In the course of time x_S and x_v can change and hence the $A_r^v(x_S, x_v)$ are variable in time. That is how we get non-constant trajectories ω.
Yet, the intensive variables that characterize each reservoir are kept steady. Otherwise, I would not call them reservoirs. That is the steady approximation: the temperature, pressure or the electro/chemical potentials of each reservoir remain fixed and unchanged over the time. It does not mean that there can be no time-dependent force acting on the system making the dynamics inhomogeneous as we had it before. For example, the Hamiltonian of the system can change while it is in contact with a heat bath at fixed temperature, see Section 8.

A second specification, besides the fact that the intensive variables of the reservoir remain fixed during the evolution, is that the reservoirs are spatially separated and locally coupled to the system. What I want is that from knowing the trajectory $\gamma = (\eta(0), \eta(1), \ldots, \eta(\tau))$ of the system, I know exactly what currents $J_r^v(\gamma)$ have been flowing in what reservoir. So, to the change $\eta(t) \to \eta(t+1)$ at time t corresponds a displacement $A_r^v(\eta(t+1), x_v(t+1)) - A_r^v(\eta(t), x_v(t))$ that can be calculated from the couple $(\eta(t), \eta(t+1))$ alone. In physical realizations, that is obtained from local conservation laws.

As a consequence of the previous hypothesis, trajectories ω as we had them before in (7.3) for the total system, are of the form $\omega = ((\eta(0), U(\eta(0)), \ldots, (\eta(\tau), U(\eta(0)) + J(\gamma))$ where $U(\eta(0))$ specifies the values $A_r^v(\eta(0), x_v(0)) = U_r^v(\eta(0))$ depending on $\eta(0)$ and from then on, all the $A_r^v(\eta(t), x_v(t)), t \geq 1$ are completely determined by the trajectory $\gamma = (\eta(0), \eta(1), \ldots, \eta(\tau))$ of the system only. In particular, we know

$$A_r^v(\eta(\tau), x_v(\tau)) = U_r^v(\eta(0)) + J_r^v(\gamma)$$

where $J_r^v(\gamma)$ is the time-integrated flux of type r flowing into the $v-$th reservoir.

For the reference measure ρ I take the product $dx = dx_S\, dx_1 \ldots dx_k$ over all momenta and positions of the particles. The initial statistics μ determines the intensive variables of the reservoirs and a probability distribution for the microstate x_S of the system. Denote by $h(x_S)$ the probability density of the system. That means that the initial density $g_0 = g_\mu$ can be written as

$$g_0(x) = g_\mu(x) = h(x_S) \prod_{v=1}^{k} \frac{e^{-\sum_r \lambda_r^v A_r^v(\eta, x_v)}}{Z_v(\eta)} \tag{7.4}$$

The h and the Lagrange multipliers λ_r^v sit in μ. g_μ corresponds to an equilibrium state of the

environment conditioned on the state of the system, as distributed via h.

The dynamics defines the statistics $\widetilde{\mu_\tau}$ at time τ as we had it before. The only thing that changes with respect to (7.4) for the new $g_{\widetilde{\mu_\tau}}$ is that we now get a distribution h_τ instead of h for the system and, in general, new λ_r^v. As I have mentioned before however, we want to model reservoirs with fixed intensive quantities. Therefore, introduce

$$g_\tau(x) = h_\tau(x_S) \prod_{v=1}^{k} \frac{e^{-\sum_r \lambda_r^v A_r^v(\eta, x_v)}}{Z_v(\eta)} \tag{7.5}$$

where, compared with (7.4), I have changed the distribution of the system and left the λ_r^v untouched. It is that density (7.5) that I will use to start the reversed dynamics.

Let us first look at $P_\mu(\omega)$ of (7.3). According to our set-up,

$$P_\mu(\omega) = \int dx\, g_0(x) \prod_{t=0}^{\tau} \delta(x_S(t) - \eta(t)) \prod_{r,v} \delta(A_r^v(x_S, x_v) - U_r^v(\eta(0)))$$

or

$$P_\mu(\omega) = h(\eta(0)) \frac{e^{-\sum_{r,v} \lambda_r^v U_r^v(\eta(0))}}{Z_v(\eta(0))} \times$$
$$\int dx \prod_{t=0}^{\tau} \delta(x_S(t) - \eta(t)) \prod_{r,v} \delta(A_r^v(x_S, x_v) - U_r^v(\eta(0)))$$

depends only on the trajectory $\gamma = (\eta(0), \eta(1), \ldots, \eta(\tau))$ of the system. In the same way, for $\Theta\omega$ under the reversed dynamics (indicated with a tilde),

$$\tilde{P}_{\widetilde{\mu_\tau}}(\Theta\omega) = h_\tau(\eta(\tau)) \frac{e^{-\sum_{r,v} \lambda_r^v [U_r^v(\eta(0)) + J_r^v(\gamma)]}}{Z_v(\eta(\tau))} \times$$
$$\int dx \prod_{t=0}^{\tau} \delta(\tilde{x_S}(t) - \pi\eta(\tau - t)) \prod_{r,v} \delta(A_r^v(x_S, x_v) - U_r^v(\eta(0) - J_r^v(\gamma)))$$

By the same reasons that led us to (7.3), upon dividing, we get that the remaining integrals in $P_\mu(\omega)$ and $\tilde{P}_{\widetilde{\mu_\tau}}(\Theta\omega)$ cancel each other to yield

$$\log \frac{P_\mu(\omega)}{\tilde{P}_{\widetilde{\mu_\tau}}(\Theta\omega)} = \sum_v \log \frac{h(\eta(0))\, Z_v(\eta(n))}{h_\tau(\eta(\tau))\, Z_v(\eta(0))} + \sum_{r,v} \lambda_r^v J_r^v(\gamma) \tag{7.6}$$

The last term is the time-integrated dissipation into the reservoirs (the change of entropy in the environment), which is a function of the trajectory γ of the system. Only that term can be extensive in time, the other terms are temporal boundary terms. Of course, when the system does not get frustrated by the presence of more reservoirs that tell it opposite things, that term becomes also a total time-derivative. For example, suppose there is just one heat bath coupled to the system and no work is done: conservation of energy requires $J(\gamma) = H(\eta(0)) - H(\eta(\tau))$ where H is the Hamiltonian of the system.
The term $\log Z_v(\eta(0))/Z_v(\eta(\tau))$ is the difference in equilibrium free energy of the v-th reservoir for boundary conditions $\eta(\tau)$ and $\eta(0)$ as imposed by the system. These terms in (7.6) are not only boundary terms in time but also in the boundary of the system. One can then expect that they are typically vanishing under weak coupling assumptions, see also in Sections 8.1 and 8.2.

The identity (7.6) complements (7.3) and fulfills the first observation in Section 2. If we allow the steady state condition $h_\tau = h$ and we take the expectation of (7.6), we get the mean steady state entropy production in the world. We can however now also study its fluctuations. That brings us to the second main observation alluded at in Section 2.

8 Jarzynski relation

Thermodynamic potentials are everywhere in applications of thermodynamics. They are tabulated and predict what processes are workable, under what conditions. For example, for a system that can extract heat from an environment at constant temperature T, the energy that is available to do work is exactly the free energy $F \equiv U - TS$, that is its energy U minus the heat term TS where S is the entropy of the system. Turning it around, it suffices to measure the work done under isothermal conditions in changing the parameters of the system and it will be equal to the free energy difference. That however is only valid if the thermodynamic process involved is sufficiently slow, quasi-static, a scenario that cannot be hoped for in many cases. It was therefore very welcome that an extended relation between free energy and work was proposed and exploited in a series of papers since the pioneering work of Jarzynski in 1997, [28]. That identity reads

$$e^{-\beta \Delta F} = \langle e^{-\beta W} \rangle \tag{8.1}$$

In the left-hand side ΔF is what we want to measure, the difference in free energies between two equilibria, say with parameter values κ_τ and κ_0 in the system Hamiltonian. That parameter could for example correspond to a spring constant. The right-hand side is an average over all possible paths that take the system in equilibrium for parameter value κ_i in its initial Hamiltonian to a state where that parameter is changed into κ_f. The work done W depends on the path if the process is not adiabatic (i.e., without heat transfer) or if it is not quasi-static. The protocol, i.e., the sequence of forcing in the time-dependent Hamiltonian, is always kept fixed.

Derivations of the Jarzynski relation (8.1) have been made in various ways and in various approximations, see [8, 28, 29, 30, 31, 32, 48, 46]. From such a relation free energy differences can be measured even in situations where the process of changing the parameters is not so well-controlled. That has already been experimentally realized in e.g. molecular systems [27, 43, 65]. An important statistical problem there is to estimate how many runs one needs to estimate the right-hand side in (8.1). After all, one would like to see trajectories for which the dissipated work is negative. Assuming a Gaussian shape for the (so called second law breaking) tail of the distribution of the dissipated work (as for example done in [65]) is practically useful but theoretically, remains unmotivated.
I will not discuss the various applications and restrict myself to showing how relation (8.1) can directly be obtained from (2.1) or from (7.6).

8.1 In Markov approximation

I start by giving the derivation in the context of a Markov jump dynamics. It is originally due to Crooks, in [8], but can be directly transformed into an illustration of (2.1).

One imagines a time-dependent Hamiltonian H_t for a system in contact with a heat bath at inverse temperature β. An effective dynamics can for example be obtained from a weak coupling limit, where then the driving protocol has to vary on the same time scale as the dissipation processes. At any rate, we start here from a discrete time-inhomogeneous Markov chain on a finite state space with transition probabilities $p_t(\eta, \eta')$ for $\eta \to \eta'$. That governs the thermal transitions in exchanging heat with the reservoir. There is detailed balance with respect to H_t: for any pair of states η, η',

$$\frac{p_t(\eta, \eta')}{p_t(\eta', \eta)} = e^{-\beta[H_t(\eta') - H_t(\eta)]} \tag{8.2}$$

If we start the system in equilibrium for H_0, the probability to see a trajectory $\gamma = (\eta(0), \eta(1), \ldots, \eta(\tau))$ is

$$P_\beta(\gamma) = \frac{e^{-\beta H_0(\eta(0))}}{Z_0}\, p_1(\eta(0), \eta(1)) \ldots\, p_\tau(\eta(\tau - 1), \eta(\tau))$$

Expectations, needed for the right-hand side of (8.1), are denoted by $\langle G \rangle_\beta = \sum_\gamma G(\gamma) P_\beta(\gamma)$. Obviously, that process does not instruct us about the dynamics of changing the parameters in the

Hamiltonian. One can think of an instantaneous change in which H_t is modified into H_{t+1} after every thermal transition.

The total change in energy is $\Delta U = H_\tau(\eta(\tau) - H_0(\eta(0))$ and the total heat that flows in the heat bath in the thermal transitions (8.2) is

$$J(\gamma) = -\sum_{t=1}^{\tau}[\, H_t(\eta(t)) - H_t(\eta(t-1))\,] \tag{8.3}$$

The total work is therefore defined as

$$W(\gamma) = J(\gamma) + \Delta U = \sum_{t=0}^{\tau-1}[H_{t+1}(\eta(t)) - H_t(\eta(t))] \tag{8.4}$$

The claim is now that

$$\langle e^{-\beta W}\rangle_\beta = e^{-\beta\Delta F} \tag{8.5}$$

where $\Delta F = -1/\beta \log Z_\tau/Z_0$ is the difference in free energies corresponding to H_τ and H_0 respectively.

The simplest way to prove (8.5) is to use the relations (7.6) or (2.1) between entropy production and time-reversal. The reversed dynamics just reverses the protocol and starts from the equilibrium distribution for H_τ. So we let

$$\widetilde{P_\beta}(\gamma) = \frac{e^{-\beta H_\tau(\eta(0))}}{Z_\tau}\, p_\tau(\eta(0),\eta(1))\dots p_1(\eta(\tau-1),\eta(\tau)) \tag{8.6}$$

and compute the promising

$$R(\gamma) = \log\frac{P_\beta(\gamma)}{\widetilde{P_\beta}(\Theta\gamma)} \tag{8.7}$$

A simple computation that uses (8.2) gives

$$R(\gamma) = \beta[H_\tau(\eta(\tau)) - H_0(\eta(0))] + \log\frac{Z_\tau}{Z_0} - \beta\sum_{t=1}^{\tau}[H_t(\eta(t)) - H_t(\eta(t-1))]$$

Hence, from the definitions (8.3)–(8.4), one arrives at

$$R = \beta\Delta U - \beta\Delta F + \beta J = \beta W - \beta\Delta F \tag{8.8}$$

Now comes the first time to profit from the form of R as in (2.1) or here in (8.7). The point is that we have the normalization condition $\langle\exp(-R)\rangle_\beta = 1$ or, explicitly,

$$\sum_\gamma P_\beta(\gamma)\frac{\widetilde{P_\beta}(\Theta\gamma)}{P_\beta(\gamma)} = 1$$

Inspection learns however that upon substituting (8.8) for (8.7), that normalization is exactly equivalent with (8.5) and we are done.

8.2 As application of Section 7.3

As I have illustrated in the previous lines, in our scheme, the Jarzynski relation follows essentially from a normalization condition. Let us see what that means in a more general context.

The main point is to arrive at (8.8) for the source term of time-reversal breaking. But simply look back at (7.6) and apply it for one heat bath:

$$R(\gamma) = \log\frac{h(\eta(0)\,Z(\eta(\tau))}{h_\tau(\eta(\tau))\,Z(\eta(0))} + \beta J(\gamma) \tag{8.9}$$

We choose $h(\eta) = \exp[-\beta H_0(\eta)]/Z_0$ and we choose $h_\tau(\eta) = \exp[-\beta H_\tau(\eta)]/Z_\tau$ (we are absolutely free to do that) and by the very construction we have

$$\langle e^{-R}\rangle_\beta = 1 \tag{8.10}$$

Hence, when we suppose that $Z(\eta(\tau))/Z(\eta(0)) \simeq 1$ (which is a weak coupling condition), then

$$\langle e^{-\beta\Delta U+\beta\Delta F-\beta J}\rangle_\beta = 1$$

and (8.1) follows from the first law, $\Delta U + J = W$.

There are other and mathematically even simpler derivations of (8.1). In fact, we already had it in (4.7). In the present section I wanted to relate it more closely to the relation between time-reversal and entropy production. What really happens is in (2.1). First use the observation that R is the total entropy production when the system starts and ends in equilibrium at the same temperature but for a different Hamiltonian. Since that total entropy production is the change of entropy in the system $= \beta\Delta U - \beta F$ plus the entropy production in the environment $= \beta J$, we get $R = \beta W - \beta\Delta F$. Next use that R is the logarithm of a probability density over trajectory-space and apply the normalization condition (8.10). Observe however that in the experimental realizations or verifications of the Jarzynski relation $\beta W - \beta\Delta F$ is not equal to the total entropy production since one does not wait to equilibrate the system. Therefore, the observation that for some trajectories $\gamma, \beta W(\gamma) - \beta\Delta F < 0$ does not imply transient violations of the second law.

9 Fluctuation relations

One of the ever returning themes in statistical mechanics is to find the right balance between dynamical and statistical considerations. Since the start of kinetic gas theory, there was a fruitful exchange of ideas between the theory of heat and the theory of dynamical systems. The thermodynamic formalism has become a standard chapter for studies in dynamical systems and, ever since Clausius, heat is understood as motion.

More recently, there has been a fruitful revival of connecting the two theories. In particular, programs are running for understanding the effect of nonlinearities on transport coefficients and for defining nonequilibrium ensembles in terms of Sinai-Ruelle-Bowen measures, [13, 21, 66]. A decade ago, [14] found numerically a remarkable symmetry in the fluctuations of the phase space contraction of a dynamical system. That phase space contraction played the role of entropy production within the effective set-up. The context is that of simulation via molecular dynamics and of thermostated dynamics, see [67]. Of course, even though they resemble Newtonian equations, these models are not microscopic and it is not clear how to derive them in some effective regime. They are mostly numerically interesting. There is also the inconvenience that different thermostats may give rise to different rates of phase space contraction under the same macroscopic conditions so that one needs very specific thermostats to get the phase space contraction coincide with the physical entropy production, see e.g. [72]. Nevertheless, Gallavotti and Cohen went on to prove a fluctuation symmetry for the steady distribution of the time-averages of the phase space contraction rate in some strongly chaotic dynamics and they hypothesized that this symmetry is much more general and relevant also for the construction of nonequilibrium statistical mechanics and the fluctuations of the entropy production in particular, [18]. That was confirmed by the results in [37, 42] for physically inspired stochastic dynamics but, ideologically, it could add to the feeling that intrinsic randomness, the stochastic or chaotic character of the dynamics, is essential for obtaining a universal fluctuation behavior around a strictly positive entropy production. Moreover, the suggested identification of phase space contraction with entropy production, or in other contexts, between physical entropy and various dynamical entropies, seems to reduce the concept of entropy and of the second law to a purely dynamical context. One is reminded of the words of Maxwell where he criticizes Clausius and Boltzmann for trying at one moment to derive the second law from dynamics, *as if any pure dynamical statement would submit to such an indignity.*

In [44] it was emphasized that the fluctuation symmetry of Gallavotti-Cohen results from the Gibbs formalism and that it is not so much the chaotic nature of the dynamics that should be held responsible but rather the Gibbsian nature of the space-time distribution. That is most easy to see in stochastic models where the method of [44] as applied in [48], not only simplifies the treatments in [42, 37] but also extends the study to non-Markovian dynamics and to the much more physical local fluctuation theorems, see [51].
Also for deterministic chaotic dynamics, the Gibbsian idea extends the results of [18] to the larger class of expansive homeomorphisms with the specification property, where the technology of Markov partitioning and symbolic dynamics are not available, see [52]. It was found that phase space contraction obtains its formal analogy with the physical entropy production as source term in the potential for time-reversal breaking, as already explained under example 2.1.3.

The subject of the present section is to show how the observation in Section 2, in particular equations (2.1), and how its confirmations in (7.3) and (7.6) in a statistical mechanical context, are related to and extend previously obtained fluctuation symmetries.

9.1 General idea

I explain here what a fluctuation symmetry is and how it formally arises in great generality.

Let ω be a variable distributed according to some probability measure $P(d\omega)$. Have in mind that ω is a trajectory for reduced variables and that P is the steady state distribution. Suppose that $R(\omega)$ is a real function of ω that satisfies, formally,

$$\frac{\text{Prob}[R(\omega) = Q]}{\text{Prob}[R(\omega) = -Q]} = e^Q \tag{9.1}$$

at least for a range of Q's. The symmetry expressed by (9.1) is the so called fluctuation symmetry. It is easy to find a distribution for which (9.1) holds exactly true. As an example take R a real variable with distribution

$$P(dR) = p(R)\,dR = g(R)\,\exp[-(R-1)^2/4]\,dR$$

where $g(R) = g(-R)$ is symmetric. Then, the probability density $p(R)$ satisfies

$$\frac{p(Q)}{p(-Q)} = e^Q$$

which means that (9.1) is satisfied for all Q. In particular, a proper rescaling of an arbitrary Gaussian random variable satisfies (9.1). The opposite is of course not true: the fluctuation symmetry does not at all imply that the (large) fluctuations of the entropy production are Gaussian.
There are various modifications of (9.1). Most interesting is to consider in (9.1) not one R but a sequence S_τ indexed by τ and to require that

$$\lim_{\tau\uparrow+\infty} \frac{1}{\tau} \log \frac{\text{Prob}[S_\tau(\omega) = \tau q]}{\text{Prob}[S_\tau(\omega) = -\tau q]} = q \tag{9.2}$$

Suppose for example that $S_\tau = R + b_\tau$ where $b_\tau(\omega)/\tau$ goes to zero (uniformly) with τ and R satisfies (9.1) exactly. Then, S_τ will satisfy (9.2).
One can also make versions where the probabilities are $\tau-$dependent or where the probabilities in the numerator and in the denominator of (9.1) and (9.2) are not quite the same. These have obtained names in the literature as detailed or transient versus steady, global versus local fluctuation symmetries. Let me however leave that zoology for a moment and discuss how fluctuation symmetries might arise.

I start with the exact symmetry (9.1). Suppose indeed that

$$R(\omega) = \log \frac{p(\omega)}{p(\Theta\omega)} \tag{9.3}$$

where $P(d\omega) = p(\omega)d\omega$ and where Θ is some involution, $\Theta^2 = \mathrm{id}$. That is close to what we had before, e.g. in (2.1) or in (7.3) and (7.6). Observe that $R(\Theta\omega) = -R(\omega)$. As a consequence

$$\int G(\omega)\, dP(\omega) = \int G(\Theta\omega) e^{-R(\omega)}\, dP(\omega) \tag{9.4}$$

for all functions G. In particular,

$$\int G(R(\omega))\, dP(\omega) = \int G(-R(\omega)) e^{-R(\omega)}\, dP(\omega) \tag{9.5}$$

Since we can take here G to be the indicator function of the event that $R(\omega) = Q$, we recover (9.1) as a special case of (9.5). Here is another special case: take $G(R) = \exp[-zR]$ for some complex number z,

$$\int e^{-zR(\omega)}\, dP(\omega) = \int e^{-(1-z)R(\omega)}\, dP(\omega) \tag{9.6}$$

expressing a symmetry in the generating function for the distribution of R. In fact, in the sense of Legendre transforms, (9.1) is dual to (9.6).
Clearly, the asymptotic symmetries like in (9.2) can also be obtained from modifications of (9.3) yielding an asymptotic version of e.g. (9.6).

The translation of the above formalities to the framework of the rest of the paper is easy. The R has already appeared in Section 2 and it has reappeared in Section 7. The physical entropy production S_τ over a time-span τ equals R, essentially, and hence, a fluctuation symmetry is immediate. In conclusion, we see that once we have understood that the entropy production is the time-reversal breaking part of the Lagrangian, then we understand that its fluctuations are governed as in (9.5).

9.2 Examples of fluctuation symmetries

9.2.1 Steady state fluctuation theorem

The story began with the numerical work of [14] and was firmly brought into the context of dynamical systems by the fluctuation theorem of Gallavotti and Cohen, see [18, 66]. It asserts that for a class of dynamical systems the fluctuations in time of the phase space contraction rate obey a general law. That means the following:

One is asked to consider a reversible smooth dynamical system $x \mapsto \varphi(x), x \in \Gamma$. The phase space Γ is in some sense bounded carrying only a finite number of degrees of freedom (a compact and connected manifold). The transformation φ is a diffeomorphism of Γ. Reversibility means that there is a diffeomorphism π on Γ with $\pi^2 = 1$ and $\pi\varphi\pi = \varphi^{-1}$. It is assumed that the dynamical system satisfies some chaotic (uniformly hyperbolic) condition: it is a transitive Anosov system. It ensures a Markov partition (and the representation via some symbolic dynamics) and the existence of a natural stationary probability measure ρ, the so called Sinai-Ruelle-Bowen measure of the dynamics, with expectations

$$\rho(G) = \lim_\tau \frac{1}{\tau} \sum_0^\tau G(\varphi_t x) \tag{9.7}$$

corresponding to time-averages for almost every randomly chosen initial point $x \in \Gamma$.
Consider now minus the logarithm of the Jacobian determinant D which arises from the change of variables implied by the dynamics; write $J = -\log D$, $J(x)$ is the phase space contraction rate and is suggested to play here the role of entropy production, see [1]. One assumes (and sometimes proves) dissipativity: the expected contraction

$$\rho(J) > 0 \tag{9.8}$$

is strictly positive.
One is interested in the fluctuations of

$$w_\tau(x) = \frac{1}{\rho(J)\tau} \sum_0^\tau J(\varphi_t(x)), \tag{9.9}$$

for large time τ. The fluctuation theorem then states that $w_\tau(x)$ has a distribution $P_\tau(w)$ with respect to the stationary state ρ such that

$$\lim_\tau \frac{1}{\tau\rho(J)w} \log \frac{P_\tau(w)}{P_\tau(-w)} = 1 \tag{9.10}$$

always. Less precise and more clear,

$$\frac{\text{Prob}[\sum_0^\tau J(\varphi_t x) = q\tau]}{\text{Prob}[\sum_0^\tau J(\varphi_t x) = -q\tau]} = \exp \tau q$$

for large τ. In other words, the distribution of entropy production (read: phase space contraction) over long time intervals satisfies the symmetry (9.2).
I have already explained in Section 2.1.3 how that fluctuation symmetry can be explained from our general perspective: the phase space contraction rate is the source-term for time-reversal breaking in the action. Detailed proofs are to be found in [18, 66, 52] and [20] contains the continuous time version.

Kurchan pointed out that this fluctuation theorem also holds for certain diffusion process, finite systems undergoing Langevin dynamics, [37]. That was extended by Lebowitz and Spohn in [42] to quite general finite Markov processes. The method based on (2.1) is however simpler and more powerful. I have illustrated that already with example 2.1.2. There is a family of stationary distributions ρ_u with u specifying the density. They are Bernoulli measures with density u. The R can be easily calculated along the lines of (2.9). It is exactly equal to the physical entropy production, the variable Joule heating, and we get an exact steady fluctuation symmetry (9.1). There are many more illustrations of that scheme, see e.g. [48]. The first analytically hard steady state fluctuation theorem was proven in a model of heat conduction by Rey-Bellet and Thomas, [63, 62] whereas the physical heuristics follows exactly what was written under example 2.1.1.

The fluctuation theorem wants to speak about the fluctuations of the entropy production. To make it observationally accessible, in particular for bulk driven systems, we better not consider the global fluctuations as they will be damped exponentially in the size of the system. We are therefore interested in spatially localized fluctuations. To explain, I give an example which is not immediately related to a dynamics.
Consider the standard two-dimensional Ising model in a lattice square V with say periodic boundary conditions. Its Hamiltonian is

$$H_V(\sigma) = -\sum_{|i-j|=1} \sigma_i\sigma_j - h\sum_i \sigma_i$$

The last term (with bulk magnetic field $h \neq 0$) breaks the spin flip symmetry $\sigma \to -\sigma$. We want to find out whether the magnetization in some subsquare Λ which is much smaller than the volume V satisfies a fluctuation symmetry. The object of study is thus

$$M_\Lambda = \sum_{i\in\Lambda} \sigma_i$$

with distribution obtained from the Gibbs measure $P_V \sim \exp[-\beta H_V]$. The idea is that we first let V be very very large and only afterwards consider growing Λ. Thus, we want to show a symmetry $q \to -q$ in the behavior of

$$p_\Lambda(q) = \lim_V P_V[M_\Lambda \simeq q|\Lambda|]$$

as Λ gets larger.
Here is how to get it in complete analogy with the suggestions of Section 2 but with spin configurations replacing space-time trajectories, spin flip replacing time-reversal, the Ising Hamiltonian replacing the space-time Lagrangian and two dimensions being thought of as 1 spatial and 1 temporal dimension. Let Θ_Λ apply a local spin flip: $(\Theta_\Lambda\sigma)_i = -\sigma_i$ if $i \in \Lambda$ and $(\Theta_\Lambda\sigma)_i = \sigma_i$ otherwise. It is immediate that

$$\begin{aligned} R_\Lambda(\sigma) &= \log \frac{P_V(\sigma)}{P_V(\Theta_\Lambda\sigma)} \\ &= \beta H_V(\Theta_\Lambda\sigma) - \beta H_V(\sigma) = 2h\beta \sum_{i\in\Lambda} \sigma_i + O(|\partial\Lambda|) \end{aligned} \tag{9.11}$$

where the last term is of the order of the boundary of Λ. Observe that the volume V has disappeared. As always, R satisfies an exact fluctuation symmetry (9.1), for every V, hence also in the limit:

$$\langle e^{-zR_\Lambda}\rangle_{\beta,h} = \langle e^{-(1-z)R_\Lambda}\rangle_{\beta,h}$$

From here, we substitute (9.11), and see, by easy manipulations that also M_Λ satisfies a fluctuation symmetry up to corrections of order $|\partial\Lambda|/|\Lambda|$:

$$\lim_\Lambda \frac{1}{|\Lambda|} \log \frac{p_\Lambda(q)}{p_\Lambda(-q)} = 2h\beta\, q$$

(The prefactor $2h\beta$ can of course be scaled away by a proper normalization of M_Λ.) That is called a local fluctuation theorem.
In a dynamical context it was obtained in [15, 19] for coupled chaotic maps and in [51] for reaction-diffusion processes. Instead of considering the full spatial volume, one looks at a finite space-time window which only afterwards is growing larger. The underlying idea remains the fact that on space-time the distribution of trajectories is Gibbsian-like and we can apply exactly the same ideology as above for the Ising model. Even though these local fluctuation relations are more physical, we are still waiting for convincing experimental realizations, see however [6].

9.2.2 Transient fluctuation theorem

The transient case is exactly similar to the steady case except that the dynamics is started from a distribution that changes with time. Within the set-up of dynamical systems, that can make a great difference. Singularities in the stationary measure present extra mathematical difficulties that are not present in the transient case, see e.g. [7]. For statistical mechanical purposes, it makes no great difference. The idea remains the same, is easier to prove but rests again on the formula (2.1). In a way, a transient fluctuation theorem is an extension of the Jarzynski relation. Just look at (9.6) for $z = 1$ and take R not starting from the steady state but from a transient state: one sees the relation (8.10). Instead of spelling out the mathematical details, let me instead turn to a recent experiment.

9.2.3 Experimental verification

An experimental test and to some extent, verification of a transient fluctuation symmetry is contained in [73]. The authors consider a colloidal particle captured in an optical trap that is translated relative to the surrounding water. The particle is micron-sized, the force is of order of pico-Newton and about 500 particle trajectories were recorded for times up to 2 seconds after initiation. The particle Hamiltonian is time-dependent

$$H_t(p,q) = \frac{p^2}{2m} + \frac{\kappa}{2}\,(q - a(t))^2 \tag{9.12}$$

with $a(t)$ the time-dependent position of the trap, approximated as the position of the minimum in a harmonic potential with spring constant κ. The force exerted on the particle is $F_t(q) = -\kappa(q - a_t)$.

The motion of $a(t)$ is rectilinear. The total work W done on the system over a time τ is

$$W(\gamma) = \int_0^\tau dt\, \dot{a}_t\, F_t(q(t)) \tag{9.13}$$

depending on the trajectory $\gamma = (q(t))$ of the particle and the protocol (a_t). It is useful to check that (9.13) satisfies the decomposition in (8.4), here in continuous time,

$$W(\gamma) = \kappa\, \frac{(a_\tau - q(\tau))^2 - (a_0 - q(0))^2}{2} - \kappa \int_0^\tau dt\, v(t)\, (q(t) - a_t) \tag{9.14}$$

with $v(t) = \dot{q}(t)$. The total work need of course not equal the entropy production. As can be seen in (8.8) or as I have written already under Section 8.2, the entropy produced in the world is (β times) the total work minus the mechanical work.

Let us see what fluctuation symmetry we can expect from our main observation (2.1), or more specifically for our purpose here, from (8.7):

$$R(\gamma) = \log \frac{P_\beta(\gamma)}{\widetilde{P}_\beta(\Theta\gamma)} \tag{9.15}$$

where P_β gives the probability of the trajectory γ over a time τ, when starting from the particle in equilibrium with H_0 and for the protocol (a_t); $\widetilde{P}_\beta$ gives the probabilities when starting from the particle in equilibrium with H_τ and for the reversed protocol $(a_{\tau-t})$; Θ reverses the particle-trajectory. For the reversed protocol I write

$$\widetilde{R}(\gamma) = \log \frac{\widetilde{P}_\beta(\gamma)}{P_\beta(\Theta\gamma)}$$

so that

$$R(\Theta\gamma) = -\widetilde{R}$$

I repeat the standard computation in much detail: for a function G of $R(\gamma)$,

$$\begin{aligned}
\langle G(R)\rangle_\beta &= \sum_\gamma G(R(\gamma))\, \frac{P_\beta(\gamma)}{\widetilde{P}_\beta(\Theta\gamma)}\, \widetilde{P}_\beta(\Theta\gamma) \\
&= \sum_\gamma G(R(\gamma))\, e^{R(\gamma)}\, \widetilde{P}_\beta(\Theta\gamma) \\
&= \sum_\gamma G(R(\Theta\gamma))\, e^{R(\Theta\gamma)}\, \widetilde{P}_\beta(\gamma) \\
&= \sum_\gamma G(-\widetilde{R}(\gamma))\, e^{-\widetilde{R}(\gamma)}\, \widetilde{P}_\beta(\gamma)
\end{aligned} \tag{9.16}$$

Taking in (9.16) the function $G(R)$ as 1 or 0 depending on whether $R = Q$ or $R \neq Q$, we get

$$\frac{P_\beta[R = Q]}{\widetilde{P}_\beta[\widetilde{R} = -Q]} = e^Q$$

Obviously, the $\widetilde{R}$ for the reversed protocol has under $\widetilde{P}_\beta$ the same distribution as has our original R under P_β:

$$\widetilde{P}_\beta[\widetilde{R} = -Q] = P_\beta[R = -Q]$$

and hence,

$$\frac{P_\beta[R = Q]}{P_\beta[R = -Q]} = e^Q \tag{9.17}$$

Observe that that is an exact fluctuation symmetry valid for all times τ no matter how small.
On the other hand, we know that $R = \beta(W - \Delta F)$, see e.g. (8.8). For the time-dependence in the Hamiltonian (9.12), the difference in Helmholtz free energies $\Delta F = 0$ and therefore $R = \beta(W - \Delta F) = \beta W$. We conclude

$$\frac{P_\beta[\beta W = Q]}{P_\beta[\beta W = -Q]} = e^Q \tag{9.18}$$

again, valid for all times τ, which is what Wang *et al* have verified experimentally in [73].
The problem remains that R of (9.15) or, what is the same, βW of (9.13) are not exactly equal to the physical entropy production. It would be that when τ is so large that after starting the particle in equilibrium for H_0, the particle gets in equilibrium with H_τ, after pulling it through the medium with the optical trap. Then indeed, the entropy production is $\beta(W - \Delta F) = \beta W$. That would however require a relaxation of the particle in the trap after it has stopped moving and that takes time.
The real entropy production over a trajectory γ is given by the second term in (9.14). Indeed, as nearly always, it differs from R and hence, here, from βW by a temporal boundary term. It remains true therefore that the (true) entropy production, the dissipated work

$$S(\gamma) = -\beta\kappa \int_0^\tau dt\, v(t)\,(q(t) - a_t)$$

satisfies a fluctuation symmetry but only its asymptotic version (9.2). The fluctuation symmetry of $S(\gamma)$ is therefore only valid for large enough τ but, here as in other cases, remains unobservable since the fluctuation symmetry is exactly implying that the occurrence of negative entropy production over time τ is exponentially damped in τ! Quite to the contrary of what Wang *et al* claim, they do not observe second law violations through the fluctuation symmetry; for small times τ there is no fluctuation symmetry for the entropy production and for large times τ, it is precisely the fluctuation symmetry that tells us how unlikely (and unobservable) are second law violations.

9.2.4 Integrated fluctuation symmetries

For completeness I introduce here yet some other forms of the basic fluctuation symmetry (9.1) or of its asymptotic forms. Most interesting is perhaps to see how conditioning on a negative entropy production makes the time run backward. I start from the exact expression (9.4) to make it simpler and take $G(\omega) = F(\omega)\delta(R(\omega) - Q)$:

$$\int_{R=Q} F(\omega)\, dP(\omega) = e^Q \int_{R=-Q} F(\Theta\omega)\, dP(\omega) \tag{9.19}$$

Divide that by the exact fluctuation symmetry for $F = 1$ to get conditional expectations:

$$\langle F | R = Q\rangle = \langle F \circ \Theta | R = -Q\rangle$$

Conditioning on the opposite entropy production is cancelled by applying the time-reversal Θ.
We can also start from (9.5): take $G(R) = F(R)$ if $R > 0$ and $G = 0$ otherwise:

$$\int_{R>0} F(R(\omega))\, dP(\omega) = \int_{R<0} F(-R(\omega)) e^{-R(\omega)}\, dP(\omega)$$

Combining the choice $F(R) = 1$ with the choice $F(R) = \exp -R$ leads directly to the integrated fluctuation symmetry

$$\frac{\mathrm{Prob}[R < 0]}{\mathrm{Prob}[R > 0]} = \langle e^{-R} | R > 0\rangle$$

which is sometimes easier to examine. Of course, the asymptotic fluctuation symmetries (only valid in some limit $\tau \uparrow +\infty$, as in (9.2)) have similar integrated versions.

Less known but useful are a number of inequalities for R and hence (be it in asymptotic form) for the entropy production. We can divide (9.4) by $\langle G\rangle$ for some $G > 0$ and obtain, via the Jensen inequality,

$$\langle R\, G\rangle \geq \langle G\rangle \, \log \frac{\langle G\rangle}{\langle G \circ \Theta\rangle}$$

which shows that $\langle R^n\rangle \geq 0$ for all powers n. In the same direction, since $\langle \exp -R\rangle = 1$, by applying a Chebyshev inequality,

$$\langle R\rangle \geq (e^{-\delta} - 1 + \delta)\mathrm{Prob}[|R| \geq \delta]$$

for all $\delta \geq 0$. If we take here $\delta = q\tau > 0$ with $\tau \uparrow +\infty$, we get

$$\lim_\tau \mathrm{Prob}[|R|/\tau \geq q] \leq \frac{1}{q} \lim_\tau \langle \frac{R}{\tau}\rangle$$

where the right-hand side gives the mean entropy production rate.
The above two inequalities imply that breaking of the detailed balance condition (of time-reversal invariance) implies a strictly positive entropy production. In most physically relevant cases that will imply that currents will flow in the direction as expected from the second law of thermodynamics. The opposite statement, whether one can have a maintained current even when time-reversal invariance is not broken, is referred to as spontaneous breaking of time-reversal invariance. An example of that is heat conduction in harmonic chains where the heat current does not vanish when taking the thermodynamic limit (no Fourier law), see [64, 56]. Such superconductors are in general excluded in classical particle systems with realistic interactions, see e.g. [49, 50], but in quantum mechanics, where the quantum statistics introduces a global and hence nonlocal effective interaction, they make some of the most interesting phenomena in nonequilibrium statistical mechanics.

10 Response relations

The relations (9.4) are a form of Ward identities (but for a discrete symmetry) since they generate correlation function identities by differentiation with respect to z and with respect to parameters present in the distribution $P(d\omega)$. In that sense, (9.1) or its dual (9.5) are speaking about fluctuation-dissipation relations.

10.1 In linear order

Let us take again (9.4) for an antisymmetric $G, G(\Theta\omega) = -G(\omega)$. An important example is $G(\omega) = J_r(\omega)$ a current of type r. By its antisymmetry under time-reversal

$$\int G(\omega)\, dP(\omega) = \frac{1}{2}\int G(\omega)(1 - e^{-R(\omega)})\, dP(\omega) \tag{10.1}$$

The integral over dP can be realized as a steady state expectation in a nonequilibrium state driven by thermodynamic fields $E = (E_r)$. Equilibrium corresponds to $E = 0 = R$. Close to equilibrium, when the E_r are very small, R is also small and we can expand to first order in the E_r:

$$\frac{\partial}{\partial E_{r'}}[\int G(\omega)\, dP(\omega)]_{E=0} = \frac{1}{2}\int G(\omega)[\frac{\partial R(\omega)}{\partial E_{r'}}]_{E=0}\, dP_{E=0}(\omega) \tag{10.2}$$

That is a generalized Green-Kubo relation. If we have

$$R = \sum_{r'} E_{r'} J_{r'}$$

and take $G = J_r$, then (10.2) becomes

$$\frac{\partial}{\partial E_{r'}}[\int J_r(\omega)\, dP(\omega)]_{E=0} = \frac{1}{2}\int J_r(\omega) J_{r'}(\omega)\, dP_{E=0}(\omega) \tag{10.3}$$

giving the usual expression for the linear transport coefficients with the Onsager reciprocity.
That structure is not rigorously valid for R equal to the entropy production because of the temporal boundary terms but asymptotic forms are obtained when dividing by τ and letting $\tau \uparrow +\infty$. I refer to [44] for a rigorous version of the argument in a context where everything is under control. Whether the linear response theory of irreversible thermodynamics can be rigorously derived from statistical mechanical models along the general formalism above is a technical question. The heuristics is clear and Green-Kubo relations follow from expanding to first order the fluctuation symmetry, see also in [16, 17, 42]. The question whether the space-time correlation functions present in the linear transport coefficients (10.3) are really integrable and under what conditions is an important physical question (about equilibrium dynamics) but the fluctuation symmetry is silent about that. Before worrying about rigorously deriving linear response theory, let us see what message the fluctuation symmetry perhaps holds for higher order response.

10.2 Second order

For that we look again at (9.4) but now we need it for a symmetric $G, G(\Theta\omega) = G(\omega)$, in particular for $G(\omega) = J_r(\omega)J_{r'}(\omega)$. Inspection of the resulting formula is disappointing: we cannot move beyond linear order. The fluctuation symmetry for the entropy production in all its versions is just a way of rewriting the basic observation (2.1) and it tells us nothing about the symmetric part under time-reversal of the Lagrangian $\mathcal{L}$. That part also is possibly non-trivially modified by the nonequilibrium driving. Let us take the example of heat conduction, see Section 2.1.1.

Consider the reversible reference process P^κ_β corresponding to the dynamics

$$\begin{aligned} dq_i &= p_i\,dt, \quad i \in V \\ dp_i &= -\frac{\partial U}{\partial q_i}(q)dt, \quad i \in V \setminus \partial V \\ dp_i &= -\frac{\partial U}{\partial q_i}(q)dt - \gamma\kappa_i p_i dt + \sqrt{\frac{2\gamma}{\beta_i}}dW_i(t), \quad i \in \partial V \end{aligned} \tag{10.4}$$

Taking $\kappa_i\beta_i = \beta$, $\forall i \in \partial V$, makes the process (10.4) reversible, as may be easily checked. As a consequence, for that choice, $P^\kappa_\beta = P^\kappa_\beta\Theta$ where $(\Theta\omega)_t = \pi\omega_{\tau-t}$ with $\omega = ((p_t, q_t), t \in [0,\tau])$ a trajectory and π reverses the sign of the momenta. The stationary density is the Gibbs measure $\rho^\beta \sim \exp -\beta H$ with respect to (2.4).
Let P_ρ denote the steady state path-space measure obtained from the dynamics (2.6), with stationary measure ρ. We compute the density of the process P_ρ with respect to P^κ_β. This makes (2.2) more precise. Writing the Radon-Nikodym derivative in the form

$$dP_\rho(\omega) = e^{-A_\rho(\omega)}\,dP^\kappa_\beta(\omega)$$

the action functional A_ρ is simply found by application of a Girsanov formula, see [47]:

$$\begin{aligned} -A_\rho(\omega) &= \sum_{i\in\partial V}\frac{1}{2}\Big[\int_0^\tau(\beta-\beta_i)p_i(t)dp_i(t) \\ &+ \int_0^\tau(\beta-\beta_i)\frac{\partial U}{\partial q_i}(q(t))p_i(t)dt \\ &+ \int_0^\tau\gamma(\beta\kappa_i-\beta_i)p_i^2(t)dt\Big] \\ &+ \log\rho(\omega_0) - \log\rho^\beta(\omega_0) \end{aligned}$$

The source of time-reversal breaking is $R_\rho(\omega) = A_{\rho\pi}(\Theta\omega) - A_\rho(\omega)$ and equals (2.7) (with here in the steady state, $\rho_0 = \rho_\tau = \rho$). That is the entropy production. The symmetric part is $Y_\rho(\omega) = A_{\rho\pi}(\Theta\omega) + A_\rho(\omega)$, here equal to

$$Y_\rho(\omega) = \log\frac{\rho^\beta(\omega_\tau)\,\rho^\beta(\omega_0)}{\rho(\omega_\tau)\,\rho(\omega_0)} - \gamma\beta\sum_{i\in\partial V}\frac{\varepsilon_i(\varepsilon_i+2)}{\varepsilon_i+1}\int_0^\tau p_i^2(t)dt$$

where I have written $\beta/\beta_i = \kappa_i = 1 + \varepsilon_i$. We conclude that there is a term, extensive in time, which also depends on the different driving $\varepsilon_i \neq 0$. That will need to be taken into account when computing higher order response functions.

Acknowledgment. I am very grateful to Karel Netočný for many useful discussions and correspondence, in particular in connection with Section 7. Part of this text was written while visiting the Isaac Newton Institute (Cambridge) in the programme *Interaction and Growth in Complex Stochastic Systems*. Its hospitality and unique atmosphere are gratefully acknowledged.

References

[1] L. Andrey, The rate of entropy change in non-Hamiltonian systems, *Phys. Lett.* **11A**, 45–46 (1985).

[2] R. Balian, From Microphysics to Macrophysics: methods and applications of statistical physics, Vol. II. Springer-Verlag (Berlin Heidelberg) 1991.

[3] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio and C. Landim, Macroscopic Fluctuation Theory for Stationary Non-Equilibrium States, *J. Stat. Phys.* **107**, 635–675 (2002).

[4] J. Bricmont, *Bayes, Boltzmann and Bohm: Probability in Physics*. In: Chance in Physics, Foundations and Perspectives. Eds. J. Bricmont, D. Dürr, M.C. Galavotti, G. Ghirardi, F. Petruccione, and N. Zanghi, (Springer-Verlag, 2002).

[5] I. Callens, W. De Roeck, T. Jacobs, C.Maes and K. Netočný, Quantum entropy production as a measure for irreversibility, *Physica D: Nonlinear phenomena* **187**, 383–391 (2004).

[6] S. Ciliberto and C. Laroche, An experimental verification of the Gallavotti-Cohen fluctuation theorem, *J. Phys. IV (France)* **8**, 215–222 (1998).

[7] E.G.D. Cohen and G. Gallavotti, Note on Two Theorems in Nonequilibirum Statistical Mechanics, *J. Stat. Phys.* **96**, 1343–1349 (1999).

[8] G.E. Crooks, Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems, *J. Stat. Phys.* **90**, 1481 (1998).

[9] W. De Roeck and C. Maes, A quantum version of free energy - irreversible work relations, to appear in *Phys. Rev. E* (2004).

[10] W. De Roeck, C. Maes and K. Netočný, *Mathematical formulation and example of the conditions that lead to a quantum H-theorem*, private communication.

[11] W. De Roeck, T. Jacobs, C. Maes and K. Netočný, An Extension of the Kac Ring Model, *J. Phys. A: Math. Gen.* **36**, 11547–11559 (2003).

[12] Jean et Nicole Dhombres: Lazare Carnot. Fayard (Paris) 1997.

[13] J.R. Dorfman, An Introduction to chaos in nonequilibrium statistical mechanics. Cambridge University Press (Cambridge) 1999.

[14] D.J. Evans, E.G.D. Cohen and G.P. Morriss, Probability of second law violations in steady flows, *Phys. Rev. Lett.* **71**, 2401–2404 (1993).

[15] G. Gallavotti, A local fluctuation theorem, *Physica A* **263**, 39–50 (1999).

[16] G. Gallavotti, Chaotic hypothesis: Onsager reciprocity and the fluctuation dissipation theorem, *J. Stat. Phys.* **84**, 899–926 (1996).

[17] G. Gallavotti, Extension of Onsager's reciprocity to large fields and the chaotic hypothesis, *Phys. Rev. Lett.* **77**, 4334–4337 (1996).

[18] G. Gallavotti and E.G.D. Cohen, Dynamical ensembles in nonequilibrium Statistical Mechanics, *Phys. Rev. Letters* **74**, 2694–2697 (1995). Dynamical ensembles in stationary states, *J. Stat. Phys.* **80**, 931–970 (1995).

[19] G. Gallavotti and F. Perroni, *An experimental test of the local fluctuation theorem in chains of weakly interacting Anosov systems*, `mp_arc` # 99–320, `chao-dyn`/9909007.

[20] G. Gentile, Large deviation rule for Anosov flows, *Forum Math.* **10**, 89–118 (1998).

[21] P. Gaspard, Chaos, Scattering and Statistical Mechanics. Cambridge University Press (Cambridge) 1998.

[22] H.-O. Georgii, *Probabilistic Aspects of Entropy*, In: Entropy, Princeton University Press, Princeton and Oxford. Eds. A. Greven, G. Keller and G. Warnecke, 2003.

[23] C.C. Gillispie, Lazare Carnot, Savant. Princeton University Press (Princeton) 1971.

[24] S. Goldstein, *Boltzmann's Approach to Statistical Mechanics*. In: Chance in Physics, Foundations and Perspectives. Eds. J. Bricmont, D. Dürr, M.C. Galavotti, G. Ghirardi, F. Petruccione, and N. Zanghi, (Springer-Verlag, 2002).

[25] S. Goldstein and J.L. Lebowitz: On the (Boltzmann) Entropy of Nonequilibrium Systems, to appear in *Physica D* `cond-mat/0304251`.

[26] P.L. Garrido, S. Goldstein and J.L. Lebowitz, The Boltzmann Entropy for dense fluids not in local equilibirum, `cond-mat/0310575`.

[27] G. Hummer and A. Szabo, Free energy reconstruction from nonequilibrium single-molecule pulling experiments, *PNAS* **98**, 3658–3661 (2001).

[28] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, *Phys. Rev. Lett.* **78**, 2690–2693 (1997).

[29] C. Jarzynski, Equilibrium free-energy from nonequilibrium measurements: a Master-equation approach, *Phys. Rev. E* **56**, 5018–5035 (1997).

[30] C. Jarzynski, Equilibrium free energies from nonequilibrium processes, *Act. Phys. Pol. B* **6**, 1609–1622 (1998).

[31] C. Jarzynski, Microscopic analysis of Clausius-Duhem processes, *J. Stat. Phys.* **96**, 415–427 (1999).

[32] C. Jarzynski, Hamiltonian derivation of a detailed fluctuation theorem, *J. Stat. Phys.* **98**, 77 (2000).

[33] E.T. Jaynes, Gibbs vs Boltzmann Entropies, *Am. J. Phys.* **33**, 391–398 (1965). *Papers on Probability, Statistics and Statistical Physics*, Ed. R. D. Rosencrantz (Reidel, Dordrecht 1983).

[34] E.T. Jaynes, *The Evolution of Carnot's Principle*, In: Maximum-Entropy and Bayesian Methods in Science and Engineering (Vol.I), Eds. G.J. Erickson and C.R. Smith, Kluwer, 267–281 (1988).

[35] M. K-H. Kiessling, *How to implement Boltzmann's probabilistic ideas in a relativistic world?*, In: Chance in Physics, Foundations and Perspectives. Eds. J. Bricmont, D. Dürr, M.C. Galavotti, G. Ghirardi, F. Petruccione, and N. Zanghi (Springer-Verlag, 2002).

[36] R. Kubo, K. Matsuo and K. Kitahara, Fluctuation and Relaxation of Macrovariables, *J. Stat. Phys.* **9**, 51–95 (1973).

[37] J. Kurchan, Fluctuation theorem for stochastic dynamics, *J. Phys. A: Math. Gen.* **31**, 3719–3729 (1998).

[38] J. Kurchan, *A Quantum Fluctuation Theorem*, `cond-mat/0007360`.

[39] O.E. Lanford III, *Entropy and equilibrium states in classical statistical mechanics*, In *Statistical Mechanics and Mathematical Problems (Batelle Seattle Rencontres 1971)*, Lecture Notes in Physics No. 20 (Springer-Verlag, Berlin), 1–113 (1973).

[40] O. Lanford, *Time Evolution of Large Classical Systems*, In: LND **38**, Ed. J. Moder, Springer (1975); *Physica A* **106**, 70 (1981).

[41] J.L. Lebowitz, Microscopic Origins of Irreversible Macroscopic Behavior, *Physica A* **263**, 516–527 (1999). Round Table on Irreversibility at STATPHYS20, Paris, July 22, 1998.

[42] J.L. Lebowitz and H. Spohn, A Gallavotti-Cohen type symmetry in the large deviations functional of stochastic dynamics, *J. Stat. Phys.* **95**, 333–365 (1999).

[43] J. Liphardt, S. Dumont, S.B. Smith, I. Tinoco and C. Bustamante, Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality, *Science* **296**, 1832–1835 (2002).

[44] C. Maes, The Fluctuation Theorem as a Gibbs Property, *J. Stat. Phys.* **95**, 367–392 (1999).

[45] C. Maes, Fluctuation relations and positivity of the entropy production in irreversible dynamical systems, preprint (2003).

[46] C. Maes and K. Netočný, Time-reversal and Entropy, *J. Stat. Phys.* **110**, 269–310 (2003).

[47] C. Maes, K. Netočný and M. Verschuere, Heat Conduction Networks, *J. Stat. Phys.* **111**, 1219–1244 (2003).

[48] C. Maes, F. Redig and A. Van Moffaert, On the definition of entropy production via examples, *J. Math. Phys.* **41**, 1528–1554 (2000).

[49] C. Maes, F. Redig and M. Verschuere, Entropy Production for Interacting Particle Systems, *Markov Proc. Rel. Fields* **7**, 119–134 (2001).

[50] C. Maes, F. Redig and M. Verschuere, No current without heat, *J. Stat. Phys.* **106**, 569–587 (2002).

[51] C. Maes, F. Redig and M. Verschuere, From Global to Local Fluctuation Theorems, *Moscow Mathematical Journal* **1**, 421–438 (2001).

[52] C. Maes and E. Verbitskiy, Large Deviations and a Fluctuation Symmetry for Chaotic Homeomorphisms, *Commun. Math. Phys.* **233**, 137–151 (2003).

[53] R.A. Minlos, S. Roelly and H. Zessin, Gibbs states on spacetime, *J. Potential Analysis* **13**, Issue 4 (2001).

[54] T. Monnai and S. Tasaki, *Quantum Correction of Fluctuation Theorem*, `cond-mat/0308337`.

[55] S. Mukamel, Quantum Extension of the Jarzynski Relation; Analogy with Stochastic Dephasing, *Phys. Rev. Lett.* **90**, 170604 (2003).

[56] H. Nakazawa, On the lattice thermal conduction, *Suppl. Prog. Theor. Phys.* **45**, 231–262 (1970).

[57] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*. Princeton University Press (Princeton) 1955.

[58] L. Onsager and S. Machlup, Fluctuations and Irreversible Processes, *Phys. Rev.* **91**, 1505–1512 (1953).

[59] R. Penrose, *The Emperor's New Mind*, Oxford University Press (Oxford), 1989.

[60] M. P. Qian, M. Qian and C. Qian, Circulations of markov chains with continuous time and probability interpretation of some determinants, *Sci. Sinica* **27**, 470–481 (1984).

[61] M. P. Qian and M. Qian, *The entropy production and reversibility of Markov processes* Proceedings of the first world congress Bernoulli soc. 1988, 307–316.

[62] L. Rey-Bellet, Statistical mechanics of anharmonic lattices, `mp_arc 03/97`.

[63] L. Rey-Bellet and L.E. Thomas, Fluctuations of the Entropy Production in Anharmonic Chains, *Ann. Henri Poincaré* **3**, 483–502 (2002).

[64] Z. Rieder, J.L. Lebowitz and E. Lieb, Properties of a harmonic crystal in a stationary nonequilibrium state, *J. Math. Phys.* **8**, 1073–1078 (1967).

[65] F. Ritort, C. Bustamante and I. Tinoco,Jr., A two-state kinetic model for the unfolding of single molecules by mechanical force, *PNAS* **99**, 13544–13538 (2002).

[66] D. Ruelle, Smooth Dynamics and New Theoretical Ideas in Nonequilibrium Statistical Mechanics, *J. Stat. Phys.* **95**, 393–468 (1999).

[67] S. Sarman, D.J. Evans and P.T. Cummings, Recent developments in non-Newtonian molecular dynamics, *Physics Reports*, Elsevier (Ed. M.J. Klein), **305**, 1–92 (1998).

[68] J. Schnakenberg, Network theory of behavior of master equation systems, *Rev. Mod. Phys.* **48**, 4, 571–585 (1976).

[69] H. Tasaki, *Jarzynski Relations for Quantum Systems and Some Applications*, technical note, `cond-mat/0009244`.

[70] S. Tasaki and T. Matsui, *Fluctuation Theorem, Nonequilibrium Steady States and MacLennan-Zubarev Ensembles of L^1-Asymptotic Abelian C^* Dynamical Systems*, `mp_arc 02-533`.

[71] S.R.S. Varadhan, *Large Deviations and Entropy*, In: Entropy, Princeton University Press, Princeton and Oxford. Eds. A. Greven, G. Keller and G. Warnecke, 2003.

[72] C. Wagner, R. Klages and G. Nicolis, Thermostating by deterministic scattering: Heat and shear flow, *Phys. Rev. E* **60**, 1401–1411 (1999).

[73] G.M. Wang, E.M. Sevick, E. Mittag, D.J. Searles and D.J. Evans, Experimental Demonstration of Violations of the Second Law of Thermodynamics for Small Systems and Short Time Scales, *Phys. Rev. Lett.* **89**, 050601 (2002).

[74] S. Yukawa, A Quantum Analogue of the Jarzynski Equality, *J. Phys. Soc. Jpn.* **69**, 2367 (2000).