Non-equilibrium systems and growth of complexity

Michał Mandrysz
Instytut Fizyki, Uniwersytet Jagielloński,
ul. Łojasiewicza 11, 30-348 Kraków, Polska

July 29, 2017

Contents

1 Introduction

The inspiration for writing this work and thus, for inquiry of the subject of non-equilibrium statistical mechanics has been the ineffable complexity of the world which slowly begins too become within reach of physics. In particular, recent developments in non-equilibrium statistical physics have convinced me that the time is ripe for a review of the vast subject concerning the small and the big, the fluctuation theorems on scales of single DNA strand and MaxEP on the planetary scales.

The investigations here presented start from a historical review of the subject (sections 2,3) which presents some less known points of view which aid in understanding; not only the chronological development of ideas, but also the various paradoxes and misconceptions (such as the source of irreversibility or the meaning of entropy) that arose on the way of intense research.

Indeed, the topic of far from equilibrium statistical mechanics developed mostly in the last 30-40 years, therefore a short review of the older, linear response near-equilibrium statistical mechanics of Onsager, Kubo and Prigogine is in place and presented in section 4. This is then followed in section 5, by a crude and simple model illustrating the mechanism of entropy lowering during non-equilibrium conditions, thus allowing for formation of complexity.

In section 6, a macroscopic "big-brother" of the fluctuation theorems and its distributions of entropy are presented. Building on introduced there measure of irreversibility we transition naturally to section 7, where we develop the topic of fluctuation theorems in one of their deterministic formulations, leading to derivation of Crooks fluctuation theorem and Jarzynski (in)equality.

Some interesting and innovative applications of the former, to self-replicators and traversals of energy landscapes are then reviewed in section 8. In fact the topic of applications is so rich and developing so fast that one could focus on it exclusively.

In the final section 9, a review of a attention-catching, but incomplete topic of extrema principles for non-equilibrium physics is performed. Subsequently, the most promising result, extending the MaxEnt approach is presented with an example application to planetary climates.

Finally, the topic of quantum mechanics and von Neumann entropy is only briefly scratched in section 3, leaving place for further work.

True logic of this world lies in
the calculus of probabilities

James Clerk Maxwell

2 Historical outline

2.1 The founding fathers of thermodynamics

The history of thermodynamics reaches back to the 1600s, when the first rudimentary thermoscopes (the ancestor of the thermometer) started to be constructed and a scientists, like Francis Bacon, began to formulate the right ideas about the nature of heat.

It took however until 1850s, after the experiments of James Joule, for the wide scientific community to finally accept heat as a form of energy. The relation between heat and energy was important for the development of steam engines and led to the description of idealized heat engines and their theoretical efficiency in 1824 by Sadi Carnot.

In the 1850s Rudolf Clausius and William Thomson (Lord Kelvin) stated both the First Law (the conservation of total energy), as well as the Second Law (heat does not spontaneously flow from colder to hotter objects). Other formulations followed quickly and the general implications of the laws were understood.

Vastness of important developments came after the recognition by Rudolf Clausius and James Clerk Maxwell in 1850s that gases consist of molecules at motion. This simple idea allowed Maxwell to derive and calculate many macroscopic properties of gases at equilibrium.

Shortly after that, Clausius introduced the notion of entropy, defined as the ratio of heat and temperature which led to the redefinition of the Second Law, now stating that for isolated systems this quantity (entropy) can only increase in time.

In 1870s Ludwig Boltzmann constructed an equation that he thought could describe the detailed time development of any gas and used it to derive the so-called H-theorem. The theorem stated that a quantity equal to entropy must always increase in time. Therefore it seemed that Boltzmann had successfully proved the Second Law. During his times however, a famous and simple objection was poised known as the Loschmidt paradox which stated, that due to time-reversal property of Newton laws the evolution could be run in reverse leading to decrease in entropy.

The resolution of this paradox was noted much later and should probably classified as hard to grasp or at least hard to get accustomed to, because even today one can find discussions and erroneous statements about the "arrow of time" in the literature. Indeed, those difficulties were noted already by Gibbs, in his words [1]:

> "Any method involving the notion of entropy, the very existence of which depends on the second law of thermodynamics, will doubtless seem to many

far-fetched, and may repel beginners as obscure and difficult of comprehension."

For this reason we will intentionally postpone the discussion of the Loschmidt paradox to the later part of this section in which we go through it thoroughly and highlight the more recent paths of developments in non-equilibrium thermodynamics and statistical physics.

In responding to some of the other objections Boltzmann realized that, generally speaking there are many more states that seem chaotic and random than seem orderly. This realization led him to argue that entropy must be proportional to the logarithm of the number of possible states of a system and the nature of the Second Law must be probabilistic.

Around 1900s, Williard Gibbs formulated statistical mechanics in more general context and introduced the notion of ensemble - a collection of many macroscopically similar copies of the system, upon which the notion of ergodicity was built. He argued that if a single particle visits every possible piece of the phase space, then when averaged over a sufficiently long time, then a property in question would have the same value if one would instead think of ensembles.

Gibbs also introduced another definition of entropy, which, as noted by him [1] would only increase in a closed system if it was measured in a "coarse-grained" way in which nearby states were not distinguished to perfect accuracy. In literature one can sometimes find statements [2] that this property of Gibbs entropy is problematic, but in fact the resolution of this paradox in very similar to the resolution of the Loschmidt paradox.

During the beginning of the XX century, just before the development of quantum theory, which largely overshadowed the development of thermodynamics, Albert Einstein, on par with Marian Smoluchowski, published his paper on Brownian motion [3]. This was the first work, where the idea of relating the amplitude of dissipation to that of the fluctuations was employed, which started a very successful line of investigations, followed by Paul Langevin and the use of stochastic methods.

Then during 1930s Lars Onsager published two very influential papers [4, 5], which may be considered as the first out-reach towards non-equilibrium phenomena. Besides refining the understanding of transport coefficients and their cross relations it hinted on the validity of the so-called Onsager hypothesis or regression hypothesis, a non-equilibrium parallel to the Boltzmann ergodic hypothesis. It stated that the relaxation of a macroscopic non-equilibrium perturbation follows the same laws which govern the dynamics of fluctuations in equilibrium systems, it's controversy stimulated a lot of new developments.

In the 1950s the theory guided by this principle was further extended by Herbert Callen's and Theodore Welton's fluctuation-dissipation theorem, followed by the work of Melville Green and Ryogo Kubo in what is now known as Green-Kubo relations [6] - a family of surprising relations connecting transport coefficients to the correlation functions

of observables - those will be presented in further sections.

Meanwhile, thanks to the development of numerical methods and early computers the properties of various, otherwise not easily solvable, mechanical models were investigated. This ultimately led to the discovery of chaos (chaotic behaviour of systems) and development of information theory.

2.2 The information era and Schrödinger's influence on physics

In the 1940s Claude Shannon introduced the notion of information quantity [7] and during the 1950s, it was recognized that entropy is simply the negative of Shannon's quantity. This way a fundamental link between information theory and thermodynamics was established. This almost coincided with the groundbreaking discovery of the structure of DNA by James Watson and Francis Crick which sparked enthusiasm and inspired generations of physicists to answer the alluring (though not easy) question of the role of physics in biological processes. One of the founding fathers of quantum theory, Erwin Schrödinger, wrote an influential book on the subject, entitled "What is life?" [8] which posed some important questions intensifying the curiosity. Those ideas we will now briefly review.

First, their operation (living organisms) as a macroscopic system resembles approximately, a purely mechanical system rather than a thermodynamical system. Even though their size is far from what is considered a thermodynamic limit, they tend to stay unaffected (in special environments) by random molecular motion and; at the same time, evade the decay towards equilibrium for an unusually long time. This can be seen as, essentially, the definition of a living organisms.

Secondly, he notices that the way an organism accomplishes the above is through the exchange of energy and matter with it's environment, that leaves its own internal state in low entropy. He withdraws from considerations of free energy, although he acknowledges that the exact physical understanding should be accomplished through it rather than through entropy. However, perhaps worth mentioning is his hypothesis of "life intensity", the term which ought to relate to the rate at which the system produces entropy or dissipates heat.

Thirdly, each cell depends on very small group of atoms, the genetic code, which determine its evolution, something unprecedented, beyond the description of ordinary statistical physics. He proposes, that perhaps, a partial explanation for this dynamical behaviour (rather than statistical) can be traced to rigidity and tightness of chemical bonds. However the very vital point Schrödinger tries to make is the hypothesis, that there must exist a yet unknown, new law of physics that would explain fully how order can be produced out of disorder.

Lastly, even though Schrödinger introduces some quantum mechanics principles, like the uniqueness of Heitler-London bond in order to defend the theory laid down by Delbrück, he assures that quantum indeterminacy should play only marginal role in the future

laws of dynamics of living systems[1].

As mentioned earlier, Schrödinger influence driven many researchers to focus on the topic of non-equilibrium phenomena, however their individual approaches diverged widely, due to, as we will see the resolution of the Loschmidt paradox.

2.3 The resolution of the Loschmidt paradox

The Loschmidt paradox confronts the fact that the fundamental equations of motion are time-reversible. How therefore the irreversibility enters the picture?

The answer of statistical physics lies in the time-asymmetric probabilistic way in which we make predictions about the world. Besides the pure probabilistic description we need the common sense, axiom of causality, in order to obtain the time-asymmetric description. We use it so frequently implicitly, that we often forget about it [2].

Indeed, Boltzmann himself didn't notice that the way in which he derived the H-Theorem from his equation, implicitly assumed that the particles are uncorrelated before the collisions (Stosszahlansatz), but become correlated after the collision [9, 10], thus causing the time-reversal asymmetry[2].

One can also see this most clearly in a generic example, reviewing the procedure in which we compute some future macroscopic state from an initial macrostate. The final state is obtained by taking the *sum* of the probabilities over the indistinguishable microstates, while on the other hand the initial state is obtained by taking the *avarage* over the initial microstates.

However, if we would consider a scenerio where we either know the initial configuration exactly or have the ability to study all degrees of freedom until the final state, the arrow of time would indeed disappear, as in the case with microscopic structureless objects[3].

2.4 Different approaches towards irreversibility and non-equilibrium

As one might tell from the large amount of literature on the subject [11, 12, 13, 14, 15] the just presented explanation of irreversibility, noticed long time ago by Clausius, Boltzmann [13], Kelvin, Maxwell [16] and also Einstein (in the polemic with Walter Ritz over time-reversal symmetry of Maxwell equations) still leaves dissatisfaction in many.

Different alternatives for explanations of irreversible processes have been proposed over the years, including time-asymmetric electromagnetism, randomness of the radiative process, quantum mechanics, CP violation and even gravity. We now will shortly discuss

[1]The possibility that remains is that the origins of life, not their evolution could be quantum mechanical.

[2]This topic is closely related to molecular chaos and the hypothesis of Onsager presented in section 4.3.

[3]There is a slight subtlety on the road towards the microscopic description connected with non-monotonic behaviour of entropy which will be discussed later.

each of those approaches and their weaknesses, starting our discussion from the most direct one, given Ilya Prigogine.

The motivation for such development was clearly articulated by Prigogine [17]:

> "I have always found it difficult to accept this conclusion [macroscopic irreversibility emerging from initial conditions] ... especially because of the constructive role of irreversible processes. Can dissipative structures be the result of mistakes?"

His personal dissatisfaction with lack of irreversibility at the microscopic level together with his belief in stochastic nature of the world led him a radical proposition [18]. The evolution of density matrix, should, in his view, be extended with additional terms guarantying irreversibility at the cost of unitarity [17]. The details of this approach, known in literature as Misra-Prigogine-Courbage theory [15], will not be presented here, as in so far no evidence for its validity has been found [19].

Somewhat related to it, is the Ritz argument about the irreversibility of Maxwell's laws of electromagnetism. In a polemic with Einstein he states that the retarded and advanced potentials should not be treated on equal footing, to which Einstein disagreed. According to John Fox, who was a later commentator of the debate, based on the long lifetimes of fast muons (which are taken as evidence for time dilation) and the speed-of-light gamma rays from rapidly moving sources, the evidence stands in favor of Einstein's time-reversable explanation [20].

Now to see why irreversibility cannot be driven by quantum mechanics one just needs to notice that all quantum phenomena are controlled by Planck's constant, while the manifestations of the irreversibility - such as friction - are clearly macroscopically large.

A different school of thought [21, 22], claims that gravity is the source of irreversibility. However, it is a well known fact that in the absence of gravitational fields friction and irreversibility occurs as well. Beside that, similarly to the quantum case, the Newton's constant is to small to have such an effect. It is demonstrable that the magnitude of friction is controlled by atomic physics and electromagnetism and therefore is a *local* effect.

One of other hypothesised causes of irreversibility is CP violation. The weak nuclear interactions violate the CP symmetry which is equivalent to saying that they violate the T symmetry, because to our best knowledge, the CPT symmetry is strictly conserved. However the case here, is again similar to the discussion of gravitational and quantum mechanical effects, namely the effect is again too small to explain the friction force. Friction would have to be proportional to the small angle from the Cabibbo-Kobayashi-Matrix matrix. This is clearly not the case because the friction force is much stronger and it is controlled by electromagnetic collisions - collisions caused by a force whose microscopic description is time-reversal symmetric.

To summarise, one can see that the Second Law and irreversibility are intimately connected with statistics, inference methods and our lack of full information about systems.

2.5 Jaynes formulation of statistical mechanics (MaxEnt)

In 1957 Edwin Jaynes published an illuminating article on the information theoretical basis of statistical mechanics which, with agreement to our earlier discussion, equated entropy to our lack of knowledge about the system. From this approach it follows that the maximum entropy state i.e. equilibrium state of statistical physics can be viewed through information theory as the least biased state given the available information (e.g. energy constraints). Statistical mechanics then becomes, in a strict sense, a form of statistical inference rather than a physical theory.

As an illustration, let's consider a situation with n potential outcomes and $m < n$ constraints in the form of known values F_k of some functions f_k ($1 \leq k \leq m$). One then searches for a probability distribution that maximizes Shannon Entropy,

$$S = -\sum_{i=1}^{n} p_i \log p_i \tag{1}$$

subject to the constraints:

$$\langle f_k \rangle \equiv \sum_{i=1}^{n} p_i f_k(i) = F_k, \quad 1 \leq k \leq m. \tag{2}$$

This procedure, familiar to obtaining Maxwell-Boltzmann distribution in statistical physics, gives the least-biased probability distribution consistent with the available information.

The practice of MaxEnt approach can also make use of the formula for maximization of relative entropy (negative Kullback-Leibler divergence),

$$H(p \vee q) = -\sum_{i} p_i \ln \frac{p_i}{q_i} \tag{3}$$

with respect to p_i (new a posteriori distribution). $H(p||q)$ can be interpreted as the information gained by using p_i instead of q_i, which really conveys the essence of the theory.

After Shannon and Jaynes established the links between information theory, statistical mechanics and thermodynamics, there was a growing need to include the concept of computation as well. Following some preliminary statements by John von Neumann, it was thought that any computational process must necessarily increase entropy. However in 1960s Rolf Landauer and later in 1970s Charles Bennett pointed out that it is not the case [23, 24], instead, entropy is raised during the process of erasure of information. This interesting topic however will not be developed in this work. Instead we will focus on a more recent line of work driven by a renewed interest toward the theory of fluctuations, considering non-equilibrium systems with reversible Nosé-Hoover thermostats.

2.6 The discovery of fluctuation theorems

The first time-reversible, deterministic thermostats and ergostats (homogeneous thermostats) were invented in the early 1980s by Hoover, Ladd and Moran[25]. Prior to this development there was no satisfactory mathematical way of modelling thermostated non-equilibrium steady states.

In the early 1990s Evans, Cohen and Morriss [26] considered the fluctuations of the entropy production rate in a shearing fluid, and proposed the so-called Fluctuation Relation (FR), which represents a general result concerning systems arbitrarily far from equilibrium, which, in the near equilibrium, stays consistent with the Green-Kubo and Onsager relations. Nevertheless, its significance was not immediately clear. A few years later, in different conditions, Giovanni Gallavotti and Ezechiel Cohen [27] derived a similar relation. However, the work that drew the widest attention was the equality derived just two years later by Christopher Jarzynski [28]. His simple and very practical relation allowed one to obtain equilibrium free energy differences between two configuration of a system in terms of the average[4] work performed during parametric switching in between those two configurations. It took a few years, to see a more general picture, in which, the Jarzynski inequality was just a special case of the Fluctuation Theorems. Theorems which, by the works of Gavin Crooks [29] also became more practical.

Those 20+ years have produced a whole new theoretical framework which encompasses the previous linear response theory and goes beyond that, including the far from equilibrium phenomena and the behaviour in between micro and macro scales. The last 5 years of this continued research paved the way for Jeremy England and coworkers [30, 31] who began description of objects living in that scale e.g. proteins and DNA molecules, implementing what could be named, Schrödinger's plan.

3 Treatments of entropy in the standard contexts

3.1 Gibbs entropy

The Gibbs Entropy is defined with the use of the N particle probability distribution function ρ and the Boltzmann constant k_B:

$$S_G = k_B \int \rho \log \rho, \tag{4}$$

and generalizes both results of Boltzmann: Boltzmann H, which is useful only for description of systems of non-interacting molecules [32] and the Boltzmann entropy $S_B = k_B \log W$ where W represents the number of possible microscopic configuration of a macrostate (phase volume). From this second fact we see that $\log k_B^{-1} S_G$ is a measure of the phase

[4]To be precise, ensemble finite-time measurements of the work performed.

volume of microstates or measure of our degree of ignorance as to the true unknown microstate.

Moreover, it can also be demonstrated [32] that the change of Gibbs entropy over a reversible path is equal to Clausius entropy:

$$\Delta S_G = \int_1^2 \frac{d\langle K + B\rangle + \langle P\rangle d\Omega}{T} \quad = \frac{dQ}{T}, \tag{5}$$

where K is the total kinetic energy, V is the interparticle potential and P, T are pressure and temperature. Due, to the generality of Gibbs entropy we may therefore drop the G, and from now on speak of entropy S and it's properties.

In closed Hamiltonian systems Gibbs entropy stays constant. This feature of entropy was first noted by Gibbs himself and was solved by a coarse-graining procedure [1]. The alleged arbitrariness of this procedure was subject to critique [2].

However, staying for a moment with non-coarse-grained entropy one can propose a thought experiment, leading to paradox and immediately resolving it to understand the case more clearly.

Let's consider a case of a simple gas (e.g. ideal) closed in an isolated box container of size L which molecules are localized in an imaginary box-like area of size $L/2$ at time t_0 (Figure 1a). It is obvious that the gas will expand, but according to the constancy of Gibbs entropy for an isolated system the entropy will not change. Of course the answer to this apparent paradox is very simple - if we had the ability to wait the time necessary for particles to spontaneously localize in the volume $L/2$, we would probably never need the Second Law of thermodynamics. In every imaginable case, we would need to place the particles in the initial state by hand, that is, close them in an actual box of size $L/2$ and then release (Figure 1b). In this scenerio, the final phase space is of course larger than the initial one and Gibbs entropy increases, as expected.

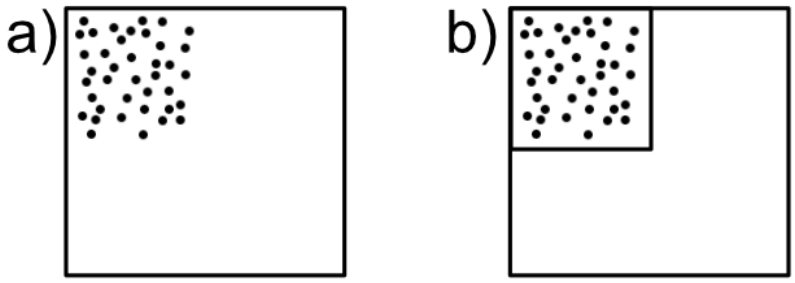

Figure 1: Gas enclosed in part of a container

The means of practical use of still non-coarse-grained entropy in closed systems subject to an adiabatic change was explained with the use of MaxEnt approach by Jaynes [32]. If we knew that on the beginning, at time t_0 the system is in *complete* thermodynamic equilibrium having entropy S, then we know that at the later time; after the external adiabatic influence ceased the new "test" or experimental distribution function will have

entropy $S_e \geq S$. By saying that we demand *complete* thermodynamic equilibrium at the beginning, we say that the *systems history* has to be followed by the experimenter in order to become confident of the obtained equilibrium, as some otherwise unexplainable exceptions exist, such as the Hahn experiment.

It is perhaps worth underlying, that the increase of entropy is linked to our knowledge about the system, rather than anything it is doing internally. This should not come up as particularly surprising as our division between work and heat is somewhat arbitrary. In fact, as just seen, Jaynes advocated an interpretation of entropy as a measure of reproducibility, rather than disorder [33]. Furthermore, even the exact parameters of entropy depend on the situation, so does their number. One can increase their number arbitrarily moving toward classical deterministic description where the notion of entropy collapses[5].

3.2 von Neumann entropy during measurement process

Although this work is meant to stay within the classical limit, it might be worth while to clear out the notion of entropy in quantum context.

The von Neumann entropy is defined as

$$S_{vN} = -\operatorname{Tr}(\hat{\rho}\log\hat{\rho}). \tag{6}$$

For which the general form of the density matrix operator is

$$\hat{\rho} = \sum_k p_k \left|\psi_k\right\rangle \left\langle\psi_k\right|. \tag{7}$$

For the case of pure state $|\psi\rangle$ the density matrix is simply

$$\hat{\rho} = \left|\psi\right\rangle \left\langle\psi\right|. \tag{8}$$

and it is easy to verify that the entropy of a pure state is equal to zero. The entropy of mixed state is always greater than zero. If the system is in a pure state, it will continue to be in a pure state as long as it stays isolated. For a mixed state, the degree of non-purity measured by the entropy will stay constant as long as it is isolated. This follows from the fact that the time evolution is unitary and the eigenvalues of the density operator do not change in time under such conditions.

An interesting question one might ask (and not really discussed in textbooks) is how the entropy changes after a measurement of a particle in many-body system which, initially, was in pure state.

Without loss of generality, let's consider an isolated system of two identical particles described solely by their momentum states. In the scenario of two particles of identical momentum we can write the initial pure state as

$$|2, 0, 0, ...\rangle \tag{9}$$

[5]This is not the case for von Neumann entropy.

which has, of course, zero entropy. Now we would like to perform a measurement on that state. In the second quantization formalism the measurement of a particle is realized by the field operator $\hat{\Psi}(x) = \sum_k \phi_k(x)\hat{a}_n$, which annihilates a single particle at position x. Therefore after the measurement of particle at some position x, one particle is "virtually" removed from the system under consideration, but the system stays in pure state

$$\hat{\Psi}(x)\,|2,0,0,...\rangle = \phi_1(x)\,|1,0,0,...\rangle\,, \tag{10}$$

with zero entropy. It is important to notice though, that a particle is lost from our considerations and therefore the entropies before and after the measurement describe two different systems. Of course, in reality the particle doesn't disappear. After determination of it's position by experiment ($\Delta x \to 0$), the uncertainty of it's momentum approaches infinity ($\Delta p \to \infty$), which means that we can reconstruct the real state using a linear combination of states with *any* value of momentum:

$$c_1\ |2,,0,...\rangle + c_2\ |1,1,0,...\rangle + c_3\ |1,0,1,...\rangle + ... \tag{11}$$

where the squared modulus of the coefficients has to sum up to one ($\sum_i |c_i|^2 = 1$).

Now depending on the precision of the measurement we can recalculate our entropy, now getting a value greater than zero. If we would perform the same analysis for a pure state of two particles in different states i.e. $|1,1,0,...\rangle$ then we would obtain an increase of entropy even without accounting for the lost particle. This crude example gives a clear illustration of the fact that after **any** measurement the von Neumann entropy has to increase and that it's change is ultimately related to lost information about the system in the act of the measurement.

There's another interesting feature of quantum entropy, namely inequalities that it fullfills. If we bipartite the system into subsystems A and B each containing it's own set of commuting observables, then in order to calculate the entropy S_A of a subsystem A we need to calculate the entropy with respect to density matrix traced over the other subsystem, namely

$$\hat{\rho}^A = \mathrm{Tr}_B\,\hat{\rho}, \tag{12}$$

then in general the following identities are satisfied

$$\begin{aligned} S(\rho) &\leq S_A + S_B \\ S(\rho) &\geq |S_A - S_B|\,. \end{aligned} \tag{13}$$

The interpretation of the first inequality is that the full information about the states of the subsystems A and B will in general not be sufficient to give full information about the state of the total system $A + B$. Or in other words, when there are correlations between the two subsystems, these are not seen in the description of A and B separately.

4 Near-equilibrium thermodynamics

4.1 Local equilibrium and entropy production

The term local equilibrium describes the situation in which the thermodynamic quantities of the system such as density, temperature, pressure, etc. can vary spatially and with time, but in each volume element the thermodynamic relations between the values which apply locally are obeyed. Such conditions are possible through the assumption of efficient dissipation of effects imposed by gradients and chemical affinities through molecular collisions. We therefore expect that this to hold for fluids, moderately dense gases and many solids. Under this assumptions, extension of equilibrium thermodynamics is possible[34].

An approached pioneered by Onsager, describing entropy in case of open systems, is an extension of Clausius entropy for isolated systems. It states that the variation of system's entropy dS is the sum of two contributions:

$$dS = dS_i + dS_e, \tag{14}$$

the entropy produced within the system dS_i (for example by change of microscopic configuration) and the entropy transferred into (or out of) the system through it's boundaries dS_e.

The Second Law then states that internal entropy production dS_i must be zero for reversible (equilibrium) transformations and positive for non-equilibrium transformations of the system [35]:

$$dS_i \geq 0, \tag{15}$$

or per unit time:

$$\frac{dS_i}{dt} = \int \sigma dV \geq 0 \tag{16}$$

where σ is the entropy production source per unit volume and dV denotes infinitesimal volume element.

The entropy supplied, dS_e, on the other hand may be positive, zero or negative, depending on the interaction of the system with it's surroundings and may be written in terms of entropy flows through the boundaries of the system:

$$\frac{dS_e}{dt} = \int J d\Omega \geq 0 \tag{17}$$

where J denotes the flux and $d\Omega$ the infinitesimal element of the boundary of the system. For a closed system we have the Clausius relation in terms of exchanged heat:

$$dS_e = \frac{dQ}{T}. \tag{18}$$

The standard formalism of linear irreversible dynamics develops then a more explicit expression for entropy production per unit time, assuming that even outside equilibrium

(but near) entropy depends only on the same variables as at equilibrium, expanding it as follows:

$$\sigma(x,t) = \sum_i J_i(x,t) X_i(x,t) \geq 0. \tag{19}$$

The sources of entropy production from the point of view of coarse-grained Gibbs entropy have been intensely studied in non-equilibrium systems. A notable result [36, 37] states that entropy production itself is independent of the level of coarse-graining applied to Gibbs entropy.

There exist however some conceptual problems with the assumption of non-negative entropy sources. For example in an electric circuit close to equilibrium, entropy production is equal to the product of the electric current times the voltage divided by the ambient temperature. If the circuit has a complex impedance, there will necessarily be a phase lag between the applied voltage and the current. Therefore there will exist an interval in which entropy production will be negative. This fact was noted by Landauer in his analysis [38] of Glansdorff's and Prigogine's book [34]. The modern far from equilibrium approach, involving fluctuation theorems and *dissipation function* solves this problem [2].

4.2 Linear response, regression and fluctuations

A very common approximation made in the treatment of near-equilibrium thermodynamics is the assumption of linear response. If an adiabatically insulated system is perturbed out of equilibrium (but still very near to it) by some time dependent force $f(t)$, then the response of mean zero observable $\delta X = X - \langle X \rangle_{eq}$ should satisfy the linearity property

$$\delta X(\lambda f(t), t) = \lambda \delta X(f(t), t) \tag{20}$$

Linear response of a system driven from equilibrium can be described in terms of the *time correlation (autocorrelation) function* of the observable X (from now on we will assume that X is mean zero observable, that is $X = \delta X$):

$$C(t) = \langle X(t)X(0) \rangle = \frac{\mathrm{Tr}\{X(t)X(0)\rho_{eq}\}}{\mathrm{Tr}\{\rho_{eq}\}}. \tag{21}$$

where ρ_{eq} is the equilibrium density function.

With the use of correlation functions, we now study the effect of relaxation towards equilibrium, assuming that the external influence ceased at time $t = 0$. Then a general property of such auto-correlation function for times $t \geq 0$ is called *regression* and follows directly from the Schwarz inequality and $X^2(t) < X^2(0)$ - the assumption of fading disturbance:

$$|C(t)| \leq C(0). \tag{22}$$

In fact in the long time limit we expect to obtain the equilibrium values of observables and

$$\lim_{t\to\infty} C(t) = 0. \tag{23}$$

Some further properties useful for further discussion can also be noted. On the microscopic level of enumerated, time dependent observables X_i, the equations of motion are time reversible and time translation invariant [39], thus leading to[6]:

$$\langle X_i(t+\tau)X_j(t)\rangle = \langle X_i(t-\tau)X_j(t)\rangle = \langle X_i(t)X_j(t+\tau)\rangle. \tag{24}$$

Dividing by τ and going with it to the limit $\tau \to 0$ we obtain:

$$\langle \dot{X}_i(t)X_j(t)\rangle = \langle X_i(t)\dot{X}_j(t)\rangle. \tag{25}$$

Now, one might perform an analysis from the macroscopic view. Assume that some system is described by a set of macroscopic variables $\{\bar{X}_i\}$ for $i = 1, ..., N$ of zero mean $E(\bar{X}_i) = 0$, such that a non-zero value of $\bar{X}_i$ corresponds to an average deviation from the equilibrium value due to an applied external force f. Again we'll assume the case in which the force ceases to exist for $t > 0$. From experience one then postulates a set of phenomenological coupled equations bringing the system back to equilibrium state:

$$\dot{\bar{X}}_i = -\sum_j \lambda_{ij}\bar{X}_j. \tag{26}$$

Such coupling between between macroscopic variables is the source of many old relations, such as thermoelectric Peltier and Seeback effects. The probability of such deviations is then proportional to the phase volume given by exponential of entropy change (see 3.1):

$$P \propto \exp\left(\frac{S(\bar{X}_1, ..., \bar{X}_N) - S_0}{k_B}\right), \tag{27}$$

where S_0 is the equilibrium value of entropy. Since we consider near-equilibrium, the linear term in the expansion disappears and we're left with

$$S - S_0 = -\sum_{ij} S_{ij}\bar{X}_i\bar{X}_j, \tag{28}$$

where $S_{ij} = -\frac{1}{2}\frac{\partial^2 S}{\partial \bar{X}_i \partial \bar{X}_j}$ is a positive definite, symmetric matrix.

One then defines so-called **generalized thermodynamic forces** as

$$F_i = -\frac{\partial S}{\partial \bar{X}_i} = \sum_j S_{ij}\bar{X}_j. \tag{29}$$

From which, by matrix inversion one can obtain again the macroscopic variables $\bar{X}_i$:

$$\bar{X}_j = \sum_i (S^{-1})_{ji} F_i. \tag{30}$$

[6]Some of the variables X_i can in fact be odd under time reversal, thus for those $\langle X_i(t+\tau)X_j(t)\rangle = \langle -X_i(t)X_j(t+\tau)\rangle$.

Inserting those back to equation (26) one gets

$$\dot{\bar{X}}_i = -\sum_j \lambda_{ij} \sum_k (S^{-1})_{jk} F_k = \sum_k \gamma_{ik} F_k. \tag{31}$$

4.3 Onsager relations and hypothesis

Lars Onsager [4] shown that γ_{ik} from the previous paragraph is in fact symmetric. We can now repeat his derivation just that by combining the equation (25) with equation (31), thus obtaining Onsager relations:

$$\gamma_{ij} = \gamma_{ji}. \tag{32}$$

In general the relaxation of small macroscopic non-equilibrium disturbances need not to be related to the regression of microscopic fluctuations in the corresponding equilibrium system. However, Onsager conjectured that in the linear approximation they should be equal. To see why this is the case we give a heuristic argument for mechanical forces. If we assume that the external force f couples to the observable X then the Hamiltonian will exhibit an additional[7] term $H' = -fX$. Let's now consider the expression for $\bar{X}$ for time $t < 0$:

$$\bar{X}(0) = \frac{\mathrm{Tr}\{X(0)e^{-\beta(H-fX)}\}}{\mathrm{Tr}\{e^{-\beta(H-fX)}\}} \approx \beta f \langle X(0)X(0)\rangle = \beta f C(0) \tag{33}$$

where in approximation each exponential was Taylor expanded to first order. For time $t > 0$:

$$\bar{X}(t) = \frac{\langle X(t)e^{-\beta(H-fX)}\rangle}{\langle e^{-\beta(H-fX)}\rangle} \approx \beta f \langle X(t)X(0)\rangle = \beta f C(t). \tag{34}$$

Onsager hypothesis can now be seen as simply:

$$\frac{\bar{X}(t)}{\bar{X}(0)} = \frac{C(t)}{C(0)}. \tag{35}$$

As a practical note on application of Onsager relations, we quote Charles Kittel [40]:

> "It is rarely a trivial problem to find the correct choice of (generalized) forces and fluxes applicable to the Onsager relation."

4.4 Green-Kubo relations

The Green-Kubo formulae relate the macroscopic, linear transport coefficients of a system to its microscopic equilibrium fluctuations. A foretaste of the Green-Kubo formalism was already given in the previous section where we considered a small perturbation term $H' = -fX$ to the Hamiltonian H. However to keep the presentation simple we will now

[7] This comes from small displacements approximation and $f = -\frac{\partial}{\partial X} H$.

turn our attention to isothermal case and static force f. The term for small macroscopic deviations of Y due to field f is given by

$$\bar{Y} = \frac{\mathrm{Tr}\{Ye^{-\beta(H-fX)}\}}{\mathrm{Tr}\{e^{-\beta(H-fX)}\}} = \mathrm{Tr}\{Ye^{-\beta(H-F-fX)}\} \tag{36}$$

where F denotes the free energy coming from the partition function. The linear response approximation defines the static isothermal (T) susceptibility χ^T_{BA} by[8]:

$$\bar{Y} = \chi^T_{YX} f. \tag{37}$$

With the use of the following identity [6]:

$$e^{\beta(a+b)} = e^{\beta a}(1 + \int_0^\beta d\lambda\ e^{-\lambda a} b e^{\lambda(a+b)}), \tag{38}$$

with $a = H - F$ and $b = -fX$. One can notice that the integral part corresponds to the change in density function under which the ensemble average takes part, thus

$$\begin{aligned} \bar{Y} &= \int_0^\beta d\lambda\ \mathrm{Tr}\{Ye^{-\lambda(H-F)}Xe^{\lambda(H-F-fX)}\}f \approx \int_0^\beta d\lambda\ \mathrm{Tr}\{Ye^{-\lambda(H-F)}Xe^{\lambda(H-F)}\}f \\ &= \langle YX\rangle f \end{aligned} \tag{39}$$

where by approximating fX to be small we obtained a special case of Green-Kubo relations defining the cross term susceptibility between observables X and Y in terms of correlation functions in the static force, isothermal conditions:

$$\chi^T_{YX} = \langle YX\rangle. \tag{40}$$

As a note, let's mention a famous objection to the Kubo relations poised by van Kampen. The argument of van Kampen concerned the plausibility of taking the linear terms first and then the ensemble average (in general one should do the opposite). An answer to this objection is that the microscopic trajectories of particles affected by the perturbing fields experience a large number of collisions in exceedingly small times ($\approx 10^{-9}s$ for low density gases) which makes the approximation possible [10].

4.5 Steady states and the definition of temperature

The steady state is loosely defined as an emergent state of a system subject to some constant[9] driving force F_e or constant fluxes vector $\boldsymbol{f}$, for which values of some observables denoted as A_i stabilize after a sufficiently long time, i.e.

$$\begin{aligned} \lim_{t\to\infty} \langle A_i(t)\rangle_0 &= const, \\ \frac{1}{\tau}\int_0^\tau F_e(t)dt &= const, \end{aligned} \tag{41}$$

[8]Note, that here again Y is assumed to have mean zero.

[9]Constant in the sense of some finite-time τ average.

where "const", in the first case, denotes some different (for each i) constant values and the ensemble average here is taken with respect to initial probability distribution function.

In some cases, one can impose a condition of vanishing expectation value of $\partial\rho/\partial t$ over the probability density function $p(\rho)$ and non-vanishing expectation value of fluxes vector $\boldsymbol{f}$ over the probability density function $p(\boldsymbol{f})$ [41], i.e.

$$\begin{aligned} \langle\frac{\partial\rho}{\partial t}\rangle_{p(\boldsymbol{\rho})} &= 0 \\ \langle\boldsymbol{f}\rangle_{p(\boldsymbol{f})} &\neq 0. \end{aligned} \tag{42}$$

The first definition (constancy of the density function) cannot however strictly hold. Indeed, in cases far from equilibrium considered in section 7.1.7 it is shown to be false. This case is used as approximation in the MaxEP approach of section 9.2.

When discussing steady states one often uses Clausius entropy to describe entropy changes related to heat transfer, however the temperature used in it's definition is not always well defined. In cases of near-equilibrium that one may call local equilibrium, the definition of temperature differences are "smooth" enough, i.e. locally there is a reasonable definition of temperature and the temperature gradient determines the heat flux. In the opposite case, it is the molecular kinetics which determines the energy transfer. In most cases under consideration it happens much faster and is considered a transient state after which local equilibrium gets established.

On the other hand, in far from equilibrium conditions there might be several definitions of temperature. One of the solutions provided by Evans et al. [2] is to define temperature of non-equilibrium state by the temperature of the underlying equilibrium state to which the system would otherwise relax. This definition gives consistent results.

4.6 MinEP principle

The most well known contribution of Ilya Prigogine to statistical physics, often called the Minimum Entropy Production (MinEP) principle, sprouts from the analysis of second order excess entropy around a steady state $(\delta^2 S)_{ss}$.

If we perturb the system around it's equilibrium state we obtain:

$$S = S_0 + \delta S + \frac{1}{2}\delta^2 S, \tag{43}$$

the linear term, δS, vanishes near equilibrium, the latter is a quantity later used as a Lyapunov function and has certain benefits over other (not necessarily all) Lyapunov functions one could define. Its macroscopic meaning is conserved independently of microscopic details of the system under consideration and is also independent of the nature of particular (possibly inhomogeneous) fluctuations. It is important however, that this result holds only for near-equilibrium steady-states. Only then the quantity $(\delta^2 S)_{ss}$ generates probability of fluctuations, as Prigogine himself, insisted in response to criticism [42].

The term "dissipative structures" was also coined by Prigogine. In Prigogine's view the fluctuations are the trigger for the instabilities (or rather bifurcations in the equations of motion), which in turn give rise to spacetime dissipative structures.

An often given example of instabilities leading to formation of structures are the Rayleigh-Bénard convection cells, which simplified non-equilibrium (not near-equilibrium nor MinEP) treatment we describe in the next section.

5 Thermodynamic lowering of entropy in non-equilibrium conditions

The Second Law of course holds for isolated systems as a whole and one can therefore imagine (on the basis of additivity of entropy) that out of equilibrium some subsystems may maintain lower entropy.

Let's therefore consider a simple model consisting of three elements: the cooler C, the heater H and a system under consideration S, staying out of equilibrium. We assume, that the temperatures of the cooler and the heater stay constant, and that heat Q_H flows into the system S and heat Q_C flows out. The situation is illustrated by the figure 2. Treating the heater and the cooler as the environment, we can think of our system S as

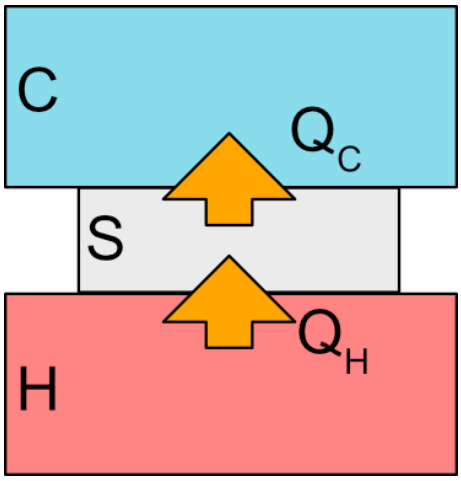

Figure 2: System (S) under the conditions of continuous heat transfer.

an open (but without particle transfer). Further on we will analyze the system S from the perspective of internal (i) entropy production and external (e) entropy flux, flowing *to* the system S. The change in entropy, as in equation (14), will be the sum of those two contributions:

$$dS_S = dS_i + dS_e. \tag{44}$$

In the current analysis let's consider a steady state situation in which the same amount of heat flows in as flows out, that is $dQ_C = -dQ_H$. Using this relation we get the following

term for the change of entropy:

$$dS_e = \frac{dQ_H}{T_H} + \frac{dQ_C}{T_C} = dQ_H \left(\frac{1}{T_H} - \frac{1}{T_C} \right) = dQ_H \left(\frac{T_C - T_H}{T_H T_C} \right) < 0. \tag{45}$$

From which it follows, that the heat flow takes the entropy out of our system. For the purpose of further discussion we introduce the concept of rate of entropy change connected with the heat flow:

$$j_e \equiv \frac{dS_e}{dt}. \tag{46}$$

In the considered scenerio, the j_e is held constant (steady-state) and we suspect a continuous fall in system's entropy. Yet, moving away from the equilibrium state we suspect, that the a balancing role will be played by dS_i moving the system back to equilibrium state. Similarly, as before we define the rate of internal entropy production:

$$j_i \equiv \frac{dS_i}{dt}. \tag{47}$$

When $T_H = T_C$, i.e. the system is in equilibrium with constant entropy S_{EQ}, therefore it follows that $j_i = 0$. We assume that the rate of internal entropy production j_i should be a function of system's entropy S_S, i.e. $j_i = j_i(S_S)$ with the boundary condition $j_i(S_S = S_{EQ}) = 0$. Around the equilibrium state $S_S = S_{EQ}$, we can Taylor expand the function $j_i(S_S)$ to it's linear term

$$j_i(S_S) = j_i\left(S_{EQ}\right) + \left(S_S - S_{EQ}\right) C_1 + \mathcal{O}\left(S_S^2\right), \tag{48}$$

fulfilling $j_i\left(S_{EQ}\right) = 0$. The dimensional and stability analysis tells us that C_1 has the dimension of inverse time and in the case of $j_e = 0$ should simply be equal to S_{EQ}, therefore we set $C_1 = -\frac{1}{\tau}$, where τ is a positive defined relaxation constant. Using the equation (44) we get

$$\frac{dS_S}{dt} = j_i\left(S_S\right) = \left(S_S - S_{EQ}\right) C_1. \tag{49}$$

The solution of the differential equation (49) is given by:

$$S_S(t) = S_{EQ} + (S_0 - S_{EQ}) e^{-t/\tau}, \tag{50}$$

where the initial condition was set $S_S(0) = S_0$. Now we include the term j_e into our considerations. In this case the equation (44) results in the following differential equation:

$$\frac{dS_S}{dt} = j_e + j_i\left(S_S\right) = j_e + \frac{S_{EQ} - S_S}{\tau}. \tag{51}$$

Given the boundary condition $S_S(0) = S_{EQ}$ it's solution is given by:

$$S_S(t) = S_{EQ} + j_e \tau \left(1 - e^{-t/\tau}\right), \tag{52}$$

where j_e is a negative constant. The graph of this function is presented on figure 3. One can see that in the limit $t \to \infty$ the entropy of the system falls to it's minimal value:

$$S_{min} = S(t \to \infty) = S_{EQ} + j_e \tau < S_{EQ}. \tag{53}$$

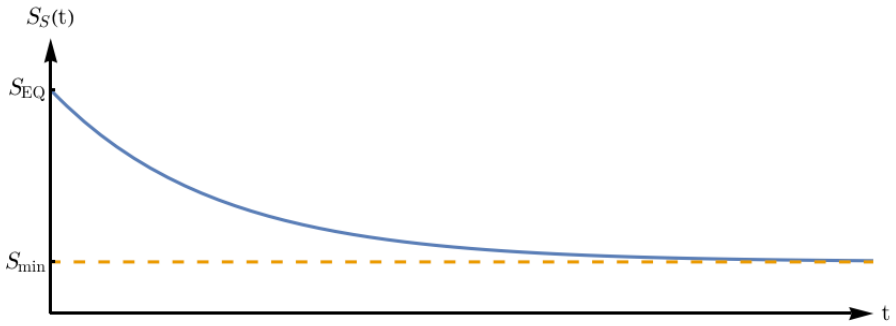

Figure 3: Lowering of entropy induced by heat flow as a function of time.

This is of course consistent with the second law of thermodynamics as we are describing an open system. It is easy to notice that the total entropy change is equal to $dS = dS_i \geq 0$ (for simplicity it was assumed that the heater and cooler don't act as producers of entropy). We see that if the relaxation constant $\tau = 0$, then the system would stay in equilibrium the whole time, of course, $\tau > 0$ for most materials (if not all). The most important part due to which this low entropy state was obtained is of course the entropy out-flow j_e, without which the non-equilibrium condition would not form.

6 Measure of irreversibility and the Second Law

In the following section a general protoplast of the Fluctuation Theorem is presented. It is derived using nothing else than simple probability calculus and the hypothesis of equal a priori probabilities. The result that one obtains is an exact form of the entropy produced in terms of microscopic transition probabilities for macroscopic objects.

Let's consider a generic statistical mechanical system at two times, at initial time t_0 and at final time t_1, each described by a complete set of possible macrostates $\{A_i\}$ for $i = 1, ..., N_A$ and $\{B_j\}$ for $j = 1, ..., N_B$ respectively. Each initial macrostate consists of some number of corresponding microstates denoted by M_i, and similarly, each final macrostate consists of some number of final microstates denoted by N_j. The deterministic, microscopic equations of motion then evolve a certain number of the microstates K_{ij} from an initial macrostate A_i to some final macrostate B_j, as illustrated on figure 4. The probability of the forward transition $P(B_j|A_i)$ is then equal to

$$P(B_j|A_i) = \frac{K_{ij}}{M_i}. \tag{54}$$

Now for the time reversed case, that is to obtain $P(A_i|B_j)$, we use the Bayes theorem

$$P(A_i|B_j) = \frac{P(B_j|A_i)P(A_i)}{P(B_j)}, \tag{55}$$

but on the way of doing so, we note that A_i is still our "hypothesis" and B_j is our evidence. Now, since we have no a priori knowledge about the initial macrostates, each of them is

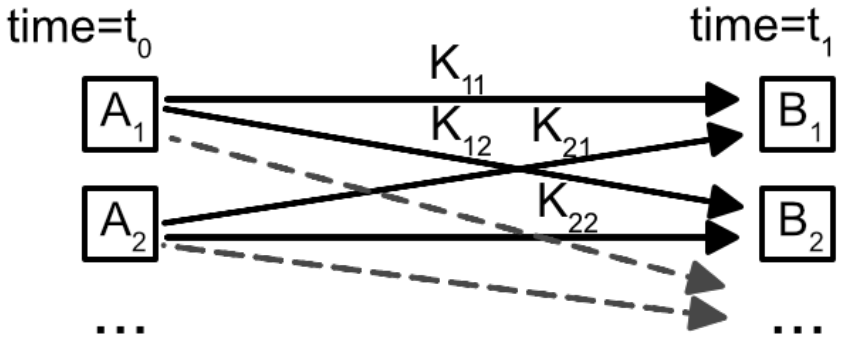

Figure 4: Transitions between macrostates A_i and macrostate B_i dictated by the 'transition' matrix K_{ij}.

equally probable $P(A_i) = 1/N_A$. $P(B_j)$ is then the marginal probability of evidence in all contradicting hypotheses or in other words a normalization factor obtained using the relation:

$$\sum_i P(A_i|B_j) = 1, \tag{56}$$

which leads to $P(B_j) = \sum_i P(B_j|A_i)P(A_i)$. Using this we obtain the following expression for the post-diction, sometimes called the *extended Bayes theorem*:

$$\begin{aligned} P(A_i|B_j) &= \frac{K_{ij}}{M_i} P(A_i) \left(\sum_k \frac{K_{kj}}{M_k} P(A_k) \right)^{-1} \\ &= \frac{K_{ij}}{M_i} \left(\sum_{k,m} \frac{K_{kj}}{K_{km}} \right)^{-1}, \end{aligned} \tag{57}$$

where in the last equation we made use of the fact that $\sum_m K_{km} = M_k$. This form is especially useful, because the matrix elements can be normalized and effectively we obtain a form of "stochastic" description. It is important to emphasize that the conditional probabilities $P(B_j|A_i)$ and $P(A_i|B_j)$ are entirely different in nature - the first represents a prediction, but the second is a post-diction. There is no symmetry between assumptions and assertions in conditional probability calculus.

Now, comparing the probability of forward macroscopic evolution to backward evolution probability one obtains:

$$\frac{P(B_j|A_i)}{P(A_i|B_j)} = e^{\ln \sum_{k,m} \frac{K_{kj}}{K_{km}}}. \tag{58}$$

The interpretation of the term in the exponent, can be done by noticing that there's only one known macroscopic quantity[10] that is strictly positive and can change signs

[10]Up to a numerical factor of k_B^{-1}.

after reversing the left side, namely the standard measure of irreversibility - change in entropy or so-called produced entropy[11]. We have thus described the entropy produced by a macroscopic system solely in terms of microscopic transition probabilities between two macroscopic states, without any assumptions about the dynamics and external forces influencing the system. Moreover, this is perhaps the most straight-forward argument against Loschmidt, which has the benefit of being easily visualizable.

With the use of real, positive random matrices, satisfying the conditions $\sum_j K_{kj} = 1$ for any k, one can obtain the probability distributions for the possible values of $\Delta S = \ln \sum_{k,m} \frac{K_{kj}}{K_{km}}$, entropy produced in the transitions. The results are presented on figure 5. A remarkable feature of this results is that the obtained distributions are not gaussian and

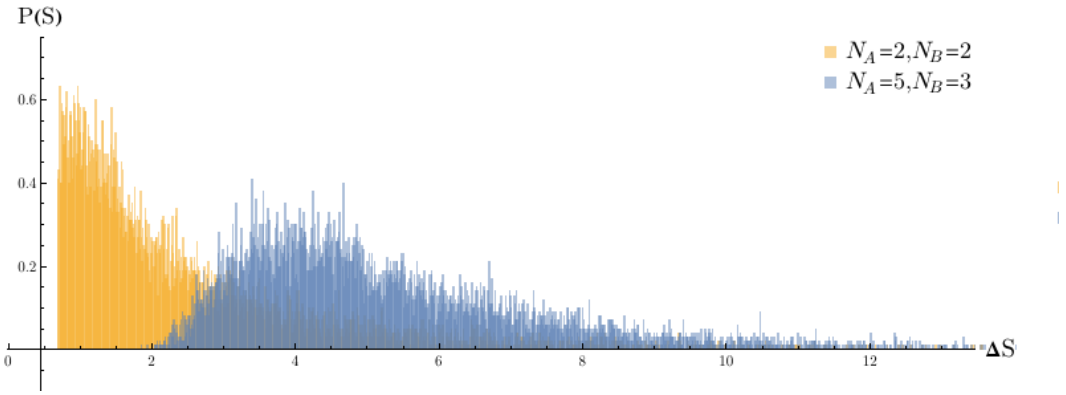

Figure 5: Probability distributions of entropy produced in transitions between N_A initial and N_B final macrostates, calculations performed for $N = 10000$ random matrices.

the most likely value is proportional to the number of macrostates. Note, that the steep beginning of the first graph is not an artefact.

Despite its generality one has to remember that this result holds for macroscopic systems with multiplicity of macrostates. Have we had just one initial or just one final macrostate the analysis would collapsed to a time-reversable case. A tempting question one could ask is, what would happen if we had an ability to partially follow the micro-trajectories? As we will see shortly, this is exactly the case covered by the Fluctuation Theorems.

7 Fluctuation theorems

In the previous section we have seen how irreversibility arrises in macroscopic systems, now we wish to extend it to small systems living on the boundary between micro- and

[11]Later we will see that in the case of fluctuation theorems this quantity can be also negative. Then our association to Shannon entropy breaks down and in this case a secondary notion of *dissipation* is sometimes introduced.

macroscopic descriptions. The crucial element of this extension is the same, natural measure of irreversibility introduced before, however now we will be taking into the account the possible microscopic trajectories of the systems. An early indicator that this direction of extension might indeed be needed was given in case of the mentioned in 3.1, Hahn spin echoes experiment.

The fluctuation theorems and their derivatives including Jarzynski equality have been demonstrated for a wide range of systems and in a number of experiments including DNA stretching [43], optically trapped colloids [44], shearing systems [26], pendulums [45], molecular motors [46] and various quantum mechanical systems [47].

Since the development of the first fluctuation theorem of Evans, Cohen and Morriss the conditions of applicability were heavily researched and in the result there exist many approaches for deriving fluctuation theorems. Those can be roughly described as deterministic [48, 2] or stochastic [49, 50].

In the deterministic framework of Evans, irreversibility finds its origins in non-linear terms which provide the contraction of phase space, this is in contrast to the stochastic descriptions in which irreversibility is given a priori - it can be perceived as merit, for one finds fewer technical difficulties [29]. In fact, the Gallavotti-Cohen theorem, which is a stationary fluctuation theorem for systems in contact with a deterministic Gaussian thermostat, breaks down in some systems and in most cases when the forcing is very strong. Those technical difficulties have their source in ergodicity. The Fluctuation Relation in the deterministic case holds only if the system has certain 'ergodic' properties [51]. Putting it simply, if the aim of ergodic theory is to understand how randomness arises from deterministic constituents, once stochasticity is added "by hand" the question is artificially bypassed.

However, the assumption that stochastic processes retains it's stochastic character at arbitrary small temporal and spatial scales leads to a conclusion that the rate of production of entropy is infinite at the limit of infinite resolution [10].

Despite of the acknowledged difficulties, the deterministic approach of Searles and Evans will be described for reasons of generality, aesthetics and also because it distances itself from the notion of entropy.

7.1 The deterministic approach

The main objective of this section will be the derivation of the dissipation function and Evans-Searles fluctuation relation (theorem), while introducing the bare minimum amount of necessary concepts. On basis of this results the Crooks fluctuation theorem and Jarzyński equality will then be derived. In the following considerations we will assume that classical mechanics gives an adequate description of the dynamics. We will also assume that quantum and relativistic effects can be safely ignored.

7.1.1 Thermostated, but time reversible systems

The construction of thermostated time reversible systems is usually done inserting some time-reversible, but non-Hamiltonian terms into the equations of motion defined as the surroundings. Surroundings in first approximation are assumed to stay far from the system of interest and should not affect the system under consideration. The work done on a system, should be on average, converted into heat, which is conducted through the system of interest and eventually removed by the non-Hamiltonian terms residing usually, in the remote boundaries.

For Gaussian isokinetic thermostat the mentioned time-reversible term is usually of the form $-S_i \alpha p_i$ where α is determined by the thermostatting condition[12] and $\mathbf{S}$ is a diagonal matrix. The term serves as a mean by which we can add or remove heat from the particles in the reservoir region ($S_i = 1$ for reservoir region and $S_i = 0$ outside this region).

In case of isolated hyperbolic systems, interacting with a thermostat the phase space is usually contracted to an attractor and the sum of Lyapunov exponents is less then zero. This is in contrast with time-reversibility of the thermostated equations of motion. The solution to this apparent paradox is related to the fractal nature of the attractor. The attractor can be described by a smooth measure in the unstable directions and a fractal measure in the stable directions. This kind of measures are known as SRB (Sinai-Ruelle-Bowen) measures [10].

7.1.2 Time reversible setup

Let's consider a closed and adiabatic Hamiltonian system of interacting particles which can exchange energy with its environment in the form of work. For example, it might be desirable to change the mean internal energy, $U = \langle H \rangle$ of the system, by externally manipulating some parameter $\lambda(s)$ (in general dependent on time s) in the potential energy function. For an externally driven adiabatic system, the rate of increase of H must be identically equal to the rate of work $\dot{W}$ done on the system by the environment, thus

$$\dot{W}^{ad} = \dot{H}^{ad} = \dot{\lambda}\frac{\partial H}{\partial \lambda} + \dot{\boldsymbol{q}}\frac{\partial H}{\partial \boldsymbol{q}} + \dot{\boldsymbol{p}}\frac{\partial H}{\partial \boldsymbol{p}}, \tag{59}$$

where the superscript *ad* emphasises adiabatic conditions.

When an external agent (not described by the potential) does work on a system without changing the underlying equilibrium state of mean energy U, we refer to that field as a purely *dissipative field*, denoting it generally by vector $\boldsymbol{F}_e$.

In the microscopic picture, the systems phase space $\{\boldsymbol{q_1}, ..., \boldsymbol{q_N}, \boldsymbol{p_1}, ..., \boldsymbol{p_N}\} \equiv (\boldsymbol{q}, \boldsymbol{p}) \equiv \boldsymbol{\Gamma}$ (where $\boldsymbol{q_i}, \boldsymbol{p_i}$ are denoting position and conjugate momenta of the particle i) evolves according to Hamiltonian equations of motion with additional terms connected with the dissipative field $\boldsymbol{F}_e$ and the time-reversible thermostat:

[12] By such construction, the system will relax to an equilibrium state $f_{eq}(\boldsymbol{\Gamma}) = Z^{-1} e^{-\beta H} \delta(\boldsymbol{p} \cdot \boldsymbol{S} \cdot \boldsymbol{p} - 2mK)$ where K denotes the kinetic energy of the system and Z the partition function.

$$\begin{aligned}\dot{\boldsymbol{q}} &= \frac{\partial H(\boldsymbol{\Gamma}, s)}{\partial \boldsymbol{p}} + \boldsymbol{C}(\boldsymbol{F}) \cdot \boldsymbol{F}_e(s), \\ \dot{\boldsymbol{p}} &= -\frac{\partial H(\boldsymbol{\Gamma}, s)}{\partial \boldsymbol{q}} + \boldsymbol{D}(\boldsymbol{F}) \cdot \boldsymbol{F}_e(s) - \alpha(\boldsymbol{\Gamma}) \boldsymbol{S} \cdot \boldsymbol{p}.\end{aligned} \tag{60}$$

Note that the thermostatting term $(-\alpha(\boldsymbol{\Gamma})\boldsymbol{S} \cdot \boldsymbol{p})$ was added to our otherwise adiabatic system in an ad-hoc manner and should not influence the adiabatic work relation of equation (59), which after the addition of the dissipative field $\boldsymbol{F}_e$ reduces to:

$$\dot{W}^{ad}(\boldsymbol{\Gamma}, s) = \dot{\lambda} \frac{\partial H(\boldsymbol{\Gamma}, s)}{\partial \lambda} - V \boldsymbol{J}(\boldsymbol{\Gamma}) \cdot \boldsymbol{F}_e(s) \tag{61}$$

where V is the volume of the system, and $\boldsymbol{J}(\boldsymbol{\Gamma})$ is the dissipative flux due to field $\boldsymbol{F}_e(\lambda)$, which one (ignoring the thermostat term) can easily obtain:

$$V \boldsymbol{J}(\boldsymbol{\Gamma}) = -\Big(\frac{\partial H}{\partial \boldsymbol{q}} \cdot \boldsymbol{C} + \frac{\partial H}{\partial \boldsymbol{p}} \cdot \boldsymbol{D} \Big). \tag{62}$$

Now, using the First Law of Thermodynamics we can include the thermostat into our considerations, concluding that the rate of heat exchange with the thermostat follows from $\dot{Q} = \dot{H} - \dot{W}$ and is therefore given by

$$\dot{Q}(\boldsymbol{\Gamma}, s) = -\alpha(\boldsymbol{\Gamma}) \frac{\partial H}{\partial \boldsymbol{p}} \cdot \boldsymbol{S} \cdot \boldsymbol{p}. \tag{63}$$

The total work done and heat added to the system then depends only on the initial phase space point $\boldsymbol{\Gamma}_0$ and time duration t. That is, $W(\boldsymbol{\Gamma}_0, t) = \int_0^t \dot{W} dt$ and $Q(\boldsymbol{\Gamma}_0, t) = -\int_0^t \dot{Q} ds$.

Introducing the basics of thermostats, we now wish to move back to some more general considerations. We define a time reversal mapping M^T (written for short as superscript star) to be an operator acting on phase space (the brackets indicate that the operator acts here on the phase exclusively):

$$\boldsymbol{\Gamma}^* = M^T[\boldsymbol{\Gamma}] \equiv (\boldsymbol{q}, -\boldsymbol{p}) \tag{64}$$

and a p-Liouvillean operator iL defined by the solution of differential equation

$$\dot{\boldsymbol{\Gamma}} \equiv iL(\boldsymbol{\Gamma})\boldsymbol{\Gamma}, \tag{65}$$

which is given by

$$\boldsymbol{\Gamma}_t = S^t \boldsymbol{\Gamma} \equiv \exp[iL(\boldsymbol{\Gamma})t]\boldsymbol{\Gamma}. \tag{66}$$

where the $\boldsymbol{\Gamma}_t$ denotes phase evolved to time t and the $exp[iL(\boldsymbol{\Gamma})t]$ is known as the p-propagator or the phase space propagator. A property useful in later part of this work is:

$$\frac{d}{dt}(S^t \boldsymbol{\Gamma}) = iL(\boldsymbol{\Gamma}) \exp[iL(\boldsymbol{\Gamma})t]\boldsymbol{\Gamma} = S^t \dot{\boldsymbol{\Gamma}}. \tag{67}$$

The time reversal dynamics satisfies an easy to check equation:

$$M^T S^t M^T S^t \ \mathbf{\Gamma} = \mathbf{\Gamma}, \tag{68}$$

where the action of operators M^T and S^t is evaluated from the right side to the left side.

The introduced equations indicate to an the existence of time reversed trajectories (anti-trajectories), i.e. if we generate a trajectory starting at $\mathbf{\Gamma}_0$ and terminating at $\mathbf{\Gamma}_t$, then under the same dynamics, we start at $\mathbf{\Gamma}_t^*$ and arrive back to $\mathbf{\Gamma}_0^*$. In most cases presented here though, we will be interested in *bundles* of trajectories $d\mathbf{\Gamma}$ and *bundles* of anti-trajectories $d\mathbf{\Gamma}^*$ passing through a volume element centered around the point $\mathbf{\Gamma}$.

We now proceed to introduce a handful of important concepts.

Phase space distribution function: The phase space distribution function $f(\mathbf{\Gamma}; t)$ gives the probability per unit phase space volume of finding phase members near the phase vector $\mathbf{\Gamma}$ at time t.

Probabilities of phase space trajectories: The probability $p(dV_{\mathbf{\Gamma}_t}, t)$, that a phase $\mathbf{\Gamma}_t$, will be observed within an infinitesimal phase space volume of size $dV_{\mathbf{\Gamma}_t}$ about $\mathbf{\Gamma}_t$ at time t, is given by,

$$p(dV_{\mathbf{\Gamma}_t}, t) = f(\mathbf{\Gamma}_t; t) dV_{\mathbf{\Gamma}_t}. \tag{69}$$

Ensemble averages: Value of any phase function $A(\mathbf{\Gamma})$ can be obtained with the use of ensemble averages by taking $N_{\mathbf{\Gamma}}$ time evolved initial phases $\mathbf{\Gamma}$ consistent with macroscopic constraints

$$\langle A(t) \rangle = \lim_{N_{\mathbf{\Gamma}} \to \infty} \sum_{j=1}^{N_{\mathbf{\Gamma}}} A(S^t \mathbf{\Gamma}_j)/N_{\mathbf{\Gamma}}, \tag{70}$$

or in the continuous limit by specifying the initial phase space probability density $f(\mathbf{\Gamma}; 0)$ or time-dependent evolution of this density $f(\mathbf{\Gamma}; t)$:

$$\langle A(t) \rangle = \int d\mathbf{\Gamma} A(\mathbf{\Gamma}) f(\mathbf{\Gamma}; t) = \int d\mathbf{\Gamma} A(S^t \mathbf{\Gamma}) f(\mathbf{\Gamma}; 0). \tag{71}$$

This equation can be seen as an application of equivalence of Heisenberg and Schrödinger representations - either the observable or the state is evolved. Time stationarity of an ensemble average is then defined simply by

$$\langle A(t) \rangle = \langle A(t + \Delta) \rangle, \tag{72}$$

for any time $\Delta > 0$.

Ergodicity: Stationary system is said to be physically ergodic if the time average of the phase function representing a physical observable, along a trajectory that starts *almost anywhere*[2] in the ostensible phase space, is equal to the ensemble average taken over an ensemble of systems consistent with the small number of macroscopic constraints on the system:

$$\lim_{t\to\infty}\langle A(t)\rangle = \lim_{t\to\infty}\frac{1}{t}\int_0^t ds\ A(S^s\mathbf{\Gamma}). \tag{73}$$

One may also talk about, so-called *ergodic consistency condition* in many context, the reason for this requirement stems from the requirement of existence of distributions sitting in the denominator of various theorems for example in the definition of (later introduced) dissipation function.

7.1.3 Phase Continuity equation

The motion of the phase space distribution function is governed by a Lagrangian form of the phase continuity equation (also known as *streaming*):

$$\frac{df(\mathbf{\Gamma};t)}{dt} = -f(\mathbf{\Gamma};t)\frac{\partial}{\partial\mathbf{\Gamma}}\cdot\dot{\mathbf{\Gamma}}(\mathbf{\Gamma}) = -f(\mathbf{\Gamma};t)\Lambda(\mathbf{\Gamma}) \tag{74}$$

this equation follows directly from a well known form of Liouville equation

$$\frac{\partial f(\mathbf{\Gamma};t)}{\partial t} = -\frac{\partial}{\partial\mathbf{\Gamma}}\cdot[\dot{\mathbf{\Gamma}}(\mathbf{\Gamma})f(\mathbf{\Gamma};t)] = -(\frac{\partial}{\partial\mathbf{\Gamma}}\cdot\dot{\mathbf{\Gamma}} + \dot{\mathbf{\Gamma}}\cdot\frac{\partial}{\partial\mathbf{\Gamma}})f(\mathbf{\Gamma};t) \tag{75}$$

where by moving the last term to the left side we get back equation (74).

We say that a system fulfills the *adiabatic incompressibility of phase space* ($AI\mathbf{\Gamma}$) if in the **absence** of the thermostatting terms the equations of motion preserve the phase space volume, that is

$$\Lambda \equiv (\partial/\partial\mathbf{\Gamma})\cdot\dot{\mathbf{\Gamma}} = 0 \tag{76}$$

This condition gives a restriction for equation (60) on the coupling tensors $\mathbf{C}$ and $\mathbf{D}$

$$\frac{\partial}{\partial\boldsymbol{q}}\cdot\boldsymbol{C}\cdot\boldsymbol{F}_e = \frac{\partial}{\partial\boldsymbol{p}}\cdot\boldsymbol{D}\cdot\boldsymbol{F}_e = 0. \tag{77}$$

For thermostated systems in a driven steady state, a contraction of phase space occurs continually, as the initial phase volume shrinks to a fractal attractor of lower dimension. It can be shown[2] that for isokinetic or isoenergetic systems with fixed total momentum and satisfying $AI\mathbf{\Gamma}$, the phase space expansion factor is exactly $\Lambda = -(3N_{th}-4)\boldsymbol{\alpha}$, where N_{th} denotes the number of thermostated particles.

Moreover, for appropriate selection of thermostats (including Gaussian isokinetic and Nosé-Hoover thermostat) the phase-space contraction factor is proportional to the rate of heat exchange with the thermostat:

$$\dot{Q}(\mathbf{\Gamma}) = k_B T \Lambda(\mathbf{\Gamma}). \tag{78}$$

If we make the following substitution $\mathbf{\Gamma} \to S^t\mathbf{\Gamma}$ in equation (74) this first-order ordinary differential equation is solved by

$$f(\mathbf{\Gamma}_t; t) = \exp\Big[-\int_0^t ds\ \Lambda(\mathbf{\Gamma}_s)\Big] f(\mathbf{\Gamma}; 0). \tag{79}$$

The measure of an infinitesimal phase space volume $dV_{\mathbf{\Gamma}_s}$ centered on the streamed position $S^s\mathbf{\Gamma} : 0 \leq s \leq t$ along the phase space trajectory also changes, but in the opposite direction (in order to keep the probability constant):

$$dV_{\mathbf{\Gamma}_s} = \exp\Big[\int_0^t ds\ \Lambda(\mathbf{\Gamma}_s)\Big] dV_{\mathbf{\Gamma}_0}. \tag{80}$$

In a lot of cases the phase space volume goes to zero, while the density approaches infinity [2].

7.1.4 Dissipation and the macroscopic irreversibility

The dissipation function serves as mathematical replacement for entropy production. When entropy production can be defined, it is equal, on average, to the dissipation function. The main advantage though, is that, unlike the entropy production, the dissipation function can, for ergodically consistent systems be always well defined [2]. Another justification for introducing a new quantity is that in some cases dissipation can be negative, but strictly speaking entropy, as interpreted in the light of information theory, should be positive. This should not come up as a surprise since we're getting close to scales at which the notion of thermodynamic entropy emerges.

The dissipation function was first properly (not implicitly) defined in 2000 by Searles and Evans [52]. It is similar to the entropy production, and although it's not a state function, it provides description of non-equilibrium systems through various fluctuation theorems. The most straight forward definition of the dissipation function is derived from the ratio of the probabilities p at time zero, of observing sets of phase space trajectories originating inside infinitesimal volumes of phase space $dV_{\mathbf{\Gamma}_0}$ and $dV_{\mathbf{\Gamma}_t^*} \equiv dV(M^T S^t \mathbf{\Gamma}_0)$:

$$\frac{p(dV_{\mathbf{\Gamma}_0}, 0)}{p(dV_{\mathbf{\Gamma}_t^*}, 0)} = \frac{f(\mathbf{\Gamma}_0; 0) dV_{\mathbf{\Gamma}_0}}{f(\mathbf{\Gamma}_t^*; 0) dV_{\mathbf{\Gamma}_t^*}}. \tag{81}$$

Now by noting that the Jacobian for the time reversal map is unity, $dV_{\mathbf{\Gamma}^*}/dV_{\mathbf{\Gamma}} = 1$, together with equation (80) we get

$$\frac{p(dV_{\mathbf{\Gamma}_0}, 0)}{p(dV_{\mathbf{\Gamma}_t^*}, 0)} = \frac{f(\mathbf{\Gamma}_0; 0)}{f(\mathbf{\Gamma}_t^*; 0)} \exp\Big[-\int_0^t ds\ \Lambda(\mathbf{\Gamma}_s)\Big]. \tag{82}$$

The logarithm of this equation will now be used as the definition of the time integral of dissipation Ω:

$$\int_0^t ds\ \Omega(\mathbf{\Gamma}_s) \equiv \ln\left(\frac{f(\mathbf{\Gamma}_0;0)}{f(\mathbf{\Gamma}_t^*;0)}\right) - \int_0^t ds\ \Lambda(\mathbf{\Gamma}_s) \equiv \Omega_t(\mathbf{\Gamma}_0). \tag{83}$$

The dissipation function Ω_t is completely determined for a deterministic trajectory by the initial coordinate, $\mathbf{\Gamma}_0$, and the duration of the trajectory, t. Moreover, the time-reversibility of the dynamics dictates that conjugate pairs of trajectories have the same value of dissipation Ω_t, but with opposite sign:

$$\Omega_t(\mathbf{\Gamma}_t^*) = -\Omega_t(\mathbf{\Gamma}_0). \tag{84}$$

One should perhaps underline that this is the place in which we postulate that some time-asymmetry takes place, otherwise the equation (81) would be equal to one. A possible interpretation of this equation states that dissipation function is a measure of the temporal asymmetry inherent in sets of trajectories originating from an initial distribution of states. To underline the extensive (in time) property of dissipation, an auxiliary quantity $\bar{\Omega}_t$ called *time-averaged dissipation* is sometimes defined through the relation $\Omega_t(\mathbf{\Gamma}) \equiv \bar{\Omega}_t(\mathbf{\Gamma})t$.

7.1.5 Evans-Searles Fluctuation Theorem

If we now choose our initial volume elements $\mathbf{\Gamma}_0$ in such a way that all the trajectories originating at time zero have the time-averaged dissipation function $\bar{\Omega}_t(\mathbf{\Gamma}) = (A \pm \delta A)$, then the probability that time-average dissipation takes the value A between $A \pm \delta A$ is given by

$$p(\bar{\Omega}_t = A) = \int d\mathbf{\Gamma}_0\ \delta[\bar{\Omega}_t(\mathbf{\Gamma}_0) - A] f(\mathbf{\Gamma}_0, 0). \tag{85}$$

Similarly, by changing the dummy variable of integration $\mathbf{\Gamma}_0$, the probability of time average dissipation holding the value $-A$ between $-A \pm \delta A$ is equal to:

$$p(\bar{\Omega}_t = -A) = \int d\mathbf{\Gamma}_t^*\ \delta[\bar{\Omega}_t(\mathbf{\Gamma}_t^*) + A] f(\mathbf{\Gamma}_t^*, 0). \tag{86}$$

Combining those two equations with equation (83), we get the Evans-Searles Fluctuation Theorem (ESFT):

$$\frac{p(\bar{\Omega}_t = A)}{p(\bar{\Omega}_t = -A)} = \exp[A\ t]. \tag{87}$$

This fluctuation relation is valid for arbitrary system size (the thermodynamic limit was not required) and can be applied to small systems observed for short periods of time. The conditions of *ergodic consistency* and microsopic time reversibility are all that is required. The relation has been verified experimentally by Wang(2002) [53], Carberry(2007) [44].

One should note that this approach considers probabilities of infinitesimal sets of trajectories fulfilling given requirements, instead of individual trajectories and only at equilibrium all individual trajectories cancel out. The underlying cause of irreversibility in

this case is not exactly obvious, one can however argue [2] that *causality* is the source of irreversibility, as we have started our considerations with the assumptions of known initial distribution function, which then undergoes the influence of the dissipative field F_e and adiabatic variation of λ. If we have instead assumed that the end state is the known state we would get an opposite and the ensemble-averaged dissipation $\langle \Omega_t \rangle$ (presented below) would be negative.

The Second Law of thermodynamics can be derived from ESFT in a trivial manner, by showing that time averages of the ensemble-averaged dissipation are non negative, i.e.

$$\langle \Omega_t \rangle \geq 0, \ \forall t > 0. \tag{88}$$

The proof follows from simple integration of equation (87):

$$\begin{aligned} \langle \Omega_t \rangle &= \int_{-\infty}^{\infty} dB \ p(\Omega_t = B) B \\ &= \int_{0}^{\infty} dB \ p(\Omega_t = B) B + \int_{\infty}^{0} dB \ p(\Omega_t = B) B \\ &= \int_{0}^{\infty} dB \ p(\Omega_t = B) B - \int_{0}^{\infty} dB \ p(\Omega_t = -B) B \\ &= \int_{0}^{\infty} dB \ p(\Omega_t = B) B (1 - \exp[-B]) \geq 0 \end{aligned} \tag{89}$$

At this point it's useful to define equilibrium system as the system for which, over the phase space domain D, the time-integrated dissipation function is identically zero:

$$\bar{\Omega}_{eq,t}(\mathbf{\Gamma}) = 0, \forall \mathbf{\Gamma} \in D, \forall t > 0 \Rightarrow \langle \Omega_t \rangle = 0, \forall t > 0. \tag{90}$$

Another quantity derivable from ESFT is the *Kawasaki identity* also known as *Non-equilibrium Partition Identity* (NPI) which was first implied for Hamiltonian systems by Yamada and Kawasaki (1967) [54] who stated:

$$\langle \exp\left[-\bar{\Omega}_t t\right] \rangle = 1. \tag{91}$$

One derives this result from ESFT given by equation (87) as follows:

$$\begin{aligned} \langle \exp\left[-\bar{\Omega}_t t\right] \rangle &= \int_{-\infty}^{\infty} dA \ p(\bar{\Omega}_t = A) \exp[-At] \\ &= \int_{-\infty}^{\infty} dA \ p(\bar{\Omega}_t = -A) \\ &= \int_{-\infty}^{\infty} dA' \ p(\bar{\Omega}_t = A') = 1. \end{aligned} \tag{92}$$

7.1.6 Instantaneous dissipation function

The dissipation function is a functional of both the dynamical equations that evolve the phase $S^t\mathbf{\Gamma} = \exp[iL(\mathbf{\Gamma})t]\mathbf{\Gamma}$ and also the initial distribution $f(\mathbf{\Gamma}; 0)$. One could get an

equation independent of duration of evolution t, by differentiation of equation (83):

$$\begin{aligned}\frac{\partial}{\partial t}\int_0^t ds\ \Omega(\mathbf{\Gamma}_s) &= \Omega(\mathbf{\Gamma}_t)\\ &= \frac{\partial}{\partial t}[\ln f(\mathbf{\Gamma};0) - \ln f(\mathbf{\Gamma}_t;0) - \int_0^t ds\ \Lambda(\mathbf{\Gamma}_s)]\\ &= -\frac{1}{f(\mathbf{\Gamma}_t;0)}\frac{\partial f(\mathbf{\Gamma}_t;0)}{\partial t} - \Lambda(\mathbf{\Gamma}_t)\\ &= -\frac{1}{f(\mathbf{\Gamma}_t;0)}\frac{\partial(\mathbf{\Gamma}_t)}{\partial t}\frac{\partial f(\mathbf{\Gamma}_t;0)}{\partial(\mathbf{\Gamma}_t)} - \Lambda(\mathbf{\Gamma}_t)\\ &= -\frac{1}{f(\mathbf{\Gamma}_t;0)}S^t\dot{\mathbf{\Gamma}}\frac{\partial f(\mathbf{\Gamma}_t;0)}{\partial(S^t\mathbf{\Gamma})} - \Lambda(\mathbf{\Gamma}_t)\end{aligned} \tag{93}$$

where the last line was obtained using equation (67). If we now set $t = 0$ we obtain the expression for the *instantaneous dissipation function*:

$$\Omega(\mathbf{\Gamma}) = -\frac{1}{f(\mathbf{\Gamma};0)}\dot{\mathbf{\Gamma}}(\mathbf{\Gamma})\frac{\partial f(\mathbf{\Gamma};0)}{\partial\mathbf{\Gamma}} - \Lambda(\mathbf{\Gamma}). \tag{94}$$

We can immediately read off this equation that dissipation is lessen by the phase space expansion.

7.1.7 Dissipation Theorem

As we have seen the dissipation function takes a central role in the fluctuation theorem and the second law inequality. However dissipation is also important in quantifying a range of non-equilibrium behaviours, including nonlinear response and relaxation towards equilibrium. Starting from the solution of the Lagrangian form of the Liouville equation (79) we can use dissipation function equation (83) to derive

$$\begin{aligned}f(S^t\mathbf{\Gamma};t) &= \exp\left[-\int_0^t ds\ \Lambda(S^s\mathbf{\Gamma})\right]f(\mathbf{\Gamma};0)\\ &= \exp\left[-\int_0^t ds\ \Lambda(S^s\mathbf{\Gamma})\right]f(S^t\mathbf{\Gamma};0)\exp\left[\int_0^t ds\ \Omega(S^s\mathbf{\Gamma}) + \int_0^t ds\ \Lambda(S^s\mathbf{\Gamma})\right]\\ &= f(S^t\mathbf{\Gamma};0)\exp\left[\int_0^t ds\ \mathbf{\Omega}(S^s\mathbf{\Gamma})\right],\end{aligned} \tag{95}$$

after substitution $\mathbf{\Gamma} \to S^{-t}\mathbf{\Gamma}$ and change of variables we get

$$f(\mathbf{\Gamma};t) = f(\mathbf{\Gamma};0)\exp\left[\int_0^t ds\ \mathbf{\Omega}(S^{-s}\mathbf{\Gamma})\right], \tag{96}$$

which states that the forward in time propagator for the N-particle distribution function is given by the exponential (backward) time integral of the dissipative function. An immediate conclusion one can draw from this is that for all non-equilibrium deterministic systems the N-particle distribution function has explicit time dependence and cannot be

written in a closed, time-stationary form. As with ESFT, this result can be applied to any initial ensemble and time-reversible dynamics satisfying the $AI\mathbf{\Gamma}$ condition, a more detailed analysis of the time-dependant case can be found in [55].

From equation (96) one can calculate non-equilibrium ensemble averages of any physical phase function $B(t)$ using the Schrödinger representation:

$$\begin{aligned}\langle B(t)\rangle &= \int_D d\mathbf{\Gamma}\ B(\mathbf{\Gamma}) \exp\left[\int_0^t ds\ \Omega(S^{-s}\mathbf{\Gamma})\right] f(\mathbf{\Gamma};0) \\ &= \langle B(0) \exp\left[\int_0^t ds\ \Omega(S^{-s}\mathbf{\Gamma})\right]\rangle_{f(\mathbf{\Gamma};0),\boldsymbol{F}_e}.\end{aligned} \tag{97}$$

By differentiating the last equation with respect to time and switching to the Heisenberg representation we get:

$$\begin{aligned}\frac{d\langle B(t)\rangle}{dt} &= \int_D d\mathbf{\Gamma}\ B(\mathbf{\Gamma})\Omega(S^{-t}\mathbf{\Gamma}) f(\mathbf{\Gamma};t) \\ &= \int_D d\mathbf{\Gamma}\ B(S^t\mathbf{\Gamma})\Omega(\mathbf{\Gamma}) f(\mathbf{\Gamma};0) \\ &= \langle B(t)\Omega(0)\rangle_{f(\mathbf{\Gamma};0),\boldsymbol{F}_e}.\end{aligned} \tag{98}$$

If we now integrate it in time, we can write the averages of physical phase functions as:

$$\langle B(t)\rangle_{f(\mathbf{\Gamma};0)} = \langle B(0)\rangle_{f(\mathbf{\Gamma};0)} + \int_0^t ds \langle B(s)\Omega(0)\rangle_{f(\mathbf{\Gamma};0),\boldsymbol{F}_e}, \tag{99}$$

getting the Dissipation Theorem, which states that the nonlinear response of an arbitrary phase variable can be calculated from the time integral of the non-equilibrium transient time correlation function (TTCF) of the phase variable with the dissipation function. Two simple limits of this theorem can be read of immediately, first one being the equilibrium case - in which we have no dissipation, so the ensemble averages stay constant. Second is the case in which the external field drives the system out of equilibrium in a linear manner (weak field), equation (99) reduces then to the Green-Kubo linear response relation.

7.1.8 Mixing properties and their relations

Let's consider a system with at least two zero-mean phase variables $A(\mathbf{\Gamma})$ and $B(\mathbf{\Gamma})$.

Mixing A system is said to be mixing if for integrable, reasonably smooth physical phase functions, time correlation functions $\langle A(0)B(t)\rangle_\mu$ taken over a stationary distribution μ factorize in the long time limit:

$$\lim_{t\to\infty} \langle A(0)B(t)\rangle_\mu = \langle A\rangle_\mu \langle B\rangle_\mu \tag{100}$$

Weak T-mixing Weak T-mixing is a direct generalization of mixing for transient rather stationary distributions. Mixing is for correlation functions in systems that have stationary averages of physical phase functions such as equilibrium or steady-state distributions. If in a system either $\langle A(0)\rangle$ or $\langle B(t)\rangle = 0, \forall t$, then such a system is called weakly T-mixing if:

$$\lim_{t\to\infty} \langle A(0)B(t)\rangle = 0 \tag{101}$$

T-mixing If a system is weakly T-mixing and the decay of transient correlation takes place at a rate faster than $1/t$ then we say that the system is T-mixing and will be stationary at long times. In other words its TTCFs must converge to finite values:

$$\left| \int_0^\infty ds \langle A(0)B(s)\rangle \right| = const < \infty \tag{102}$$

ΩT-mixing We say that a system possesses the property of ΩT mixing if the integral

$$\left| \int_0^\infty ds \langle B(s)\Omega(0)\rangle \right| = const < \infty \tag{103}$$

is bounded from above. This requirement let's us predict that the system will relax either to a non-equilibrium steady state or toward an equilibrium. In other words, it is a *necessary* condition for ensemble averages to be time-independent or stationary at long times. T-mixing systems are ΩT-mixing, but not all ΩT-mixing are T-mixing. All T-mixing systems must relax to time stationary states in the long time limit.

7.1.9 Relaxation

Non-equilibrium system can relax to equilibrium in two ways: conformally and non-conformally. A conformal system relaxes in such manner that the non-equilibrium distribution is of the form

$$f(\mathbf{\Gamma}; t) = \exp[-\beta H(\mathbf{\Gamma}) + \lambda(t) g(\mathbf{\Gamma})] Z^{-1} \tag{104}$$

for all times t and the deviation function, g, is a constant over the relaxation. As one might suspect conformal relaxation is an exception rather than the norm in natural relaxation processes.

7.1.10 Relaxation Theorem

The Relaxation Theorem states that if an arbitrary initial ensemble of ergodic Hamiltonian systems is in contact with a heat bath and there is a decay of temporal correlations, then the system will at long times, relax to the Maxwell-Boltzmann distribution. Further, this

distribution has zero dissipation everywhere in phase space. For such systems no other distribution has zero dissipation everywhere.

$$\lim_{t\to\infty} \Omega(\mathbf{\Gamma}; f(\mathbf{\Gamma}, t)) = 0, \forall\ \Gamma$$

This result is exact arbitrarily far from equilibrium and independent of system size, dervi-ation can be found in [2].

7.1.11 Driven systems

Driven systems are a subcategory of non-equilibrium systems which are subject to an external dissipative field $\boldsymbol{F}_e$. For such systems systems the dissipation function has the form:

$$\Omega(\mathbf{\Gamma}) \equiv -\beta V \boldsymbol{J}(\mathbf{\Gamma}) \cdot \boldsymbol{F}_e(s) \tag{105}$$

where V is the volume of the system and $\boldsymbol{J}$ is the previously defined dissipative flux[13]. If the system that is driven was initially at equilibrium, then equation (99) can be rewritten as:

$$\langle B(t)\rangle_{f(\mathbf{\Gamma};0)} = \langle B(0)\rangle_{f(\mathbf{\Gamma};0)} - V\int_0^t ds\ \langle \beta \boldsymbol{J}(0) B(s)\rangle_{f(\mathbf{\Gamma};0),\boldsymbol{F}_e} \cdot \boldsymbol{F}_e \tag{106}$$

and at the approximation of zero field in the correlation function it reduces to Green-Kubo expression for the linear response.

If we consider a simple (fields and dissipation flux taken as scalars), nonequilibrium, thermostated system of volume V, consisting of charged particles driven by an external field F_e and time-average of the current density along a trajectory taken as $J_{c,t} = \frac{1}{t}\int_0^t J_c(s)\ ds$, then the fluctuation relation of equation (87) can be stated as:

$$\frac{p(J_{c,t} = A \pm dA)}{p(J_{c,t} = -A \pm dA)} = \exp[A\beta F_e V t] \tag{107}$$

From this equation, one can see that as the system size or time of observation is increased, the relative probability of observing positive to negative current density increases exponentially so the current density has a definite sign and the second law of thermodynamics is retrieved. In obtaining there results, nothing was assumed about the form of the distribution of current density (it does not have to be Gaussian). Moreover, in the weak field limit, the rate of entropy production, $\dot{S}$, is given by linear irreversible thermodynamics: $\dot{S} \equiv \sum \langle J_i\rangle V X_i/T$, where the sum is over the product of all conjugate thermodynamic fluxes, J_i, and thermodynamics forces, X_i, divided by the temperature of

[13]Even though in this definition dissipation is a linear functional of the field, we can always hide higher order dependence under F_e

the system, T. From those considerations the relation between dissipation function and entropy production stands out simply as:

$$\lim_{F_e \to 0} \dot{S}(t) = k_B \langle \Omega(t) \rangle. \tag{108}$$

The difference at high fields is because the temperature that appears in the dissipation function is that which the system would relax to if the fields were removed, rather than any non-equilibrium system temperature observed with the field on [2]. The change in entropy for a process will be similarly related to the time integral of the dissipation

$$\lim_{F_e \to 0} \Delta S = k_B \langle \Omega_t \rangle. \tag{109}$$

7.1.12 Non-equilibrium Steady States

From equation (72) we see that stationarity of a system implies that its physical properties do not vary in time. This can be understood in the sense of all times or sufficiently late times, however stationarity does not imply that the distribution function is stationary, as was already shown by equation (96). The time independent values of physical properties, on the other hand, can be dependent on the initial phase $\mathbf{\Gamma}$, if they are not; we call it a *physically ergodic non-equilibrium steady state* (peNESS):

$$\lim_{t \to \infty} \langle A(t) \rangle_0 = \lim_{t \to \infty} \frac{1}{t} \int_0^t ds \; A(S^s \mathbf{\Gamma}) \tag{110}$$

where $\langle ... \rangle_0$ denotes an ensemble over the initial time $t = 0$ and probability distribution $f(\mathbf{\Gamma}; 0)$. Contrary to intuition, not all NESSs are physically ergodic. An example is the Rayleigh-Bénard instability which occurs in a system with fixed boundary conditions and fixed geometry where a system might develop to a fixed number of rolls (two, four etc.) and persist in it indefinitely [2].

7.2 Generalized Crooks fluctuation theorem

Crooks fluctuation theorem together with Jarzynski equality were originally developed for determining the difference in free energy of canonical equilibrium states from experimental information taken from non-equilibrium paths that connects two equilibrium states. In order to establish a connection with Crooks fluctuation theorem, we will need another definition of so-called *generalized dimensionless "work"* $\Delta X_\tau(\mathbf{\Gamma})$ for a trajectory of duration τ originating from the phase point $\mathbf{\Gamma}$ as

$$\begin{aligned} \exp[\Delta X_\tau(\mathbf{\Gamma})] &\equiv \lim_{dV_{\mathbf{\Gamma}} \to 0} \frac{p_{eq,1}(dV_{\mathbf{\Gamma}}; 0) Z(\lambda_1)}{p_{eq,2}(dV_{\mathbf{\Gamma}_\tau}; 0) Z(\lambda_2)} \\ &= \frac{f_{eq,1}(\mathbf{\Gamma}) d\mathbf{\Gamma} Z(\lambda_1)}{f_{eq,2}(\mathbf{\Gamma}_\tau) d\mathbf{\Gamma}_\tau Z(\lambda_2)} \end{aligned} \tag{111}$$

where $Z(\lambda_i)$, the partition function for the system is just a normalization factor for the equilibrium function $f_{eq}(\mathbf{\Gamma}) = \exp[F(\mathbf{\Gamma})]/Z$, where $F(\mathbf{\Gamma})$ is some single-valued phase function. After time τ the system ends its parametric change in λ, however the system is *not* in equilibrium. That is $f(\mathbf{\Gamma};0) = f_{eq,1}(\mathbf{\Gamma})$, but $f(\mathbf{\Gamma};\tau) \neq f_{eq,2}(\mathbf{\Gamma})$ in general, since relaxation to complete thermal equilibrium cannot take place in finite time. It can be shown that generalized work defined this way is in fact a state-function when evaluated along quasi-static paths.

The Generalized Crooks Fluctuation Theorem (GCFT) considers probability $p_{eq,f}(\Delta X_t = B \pm dB)$ of observing values of ΔX_t in the range $B \pm dB$ for forward trajectories starting from the initial equilibrium distribution 1, $f_1(\mathbf{\Gamma};0) = f_{eq,1}(\mathbf{\Gamma})$, and the probability $p_{eq,r}(\Delta X_t = -B \mp dB)$ of observing ΔX_t in the range $-B \mp dB$ for reverse trajectories by starting from the equilibrium distribution given by $f_{eq,2}(\mathbf{\Gamma})$ of state 2. The probability that the phase variable ΔX_τ takes the value B for a forward evolved trajectories is given by

$$p_{eq,1}(\Delta X_{\tau,f} = B \pm dB) = \int_{\Delta X_{\tau,f}=B\pm dB} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma}). \tag{112}$$

Analogously, the probability of particular values for backward evolved trajectories starting from $f_{eq,2}(\mathbf{\Gamma})$ is given by

$$p_{eq,2}(\Delta X_{\tau,r} = -B \mp dB) = \int_{\Delta X_{\tau,r}=-B\mp dB} d\mathbf{\Gamma} f_{eq,2}(\mathbf{\Gamma}). \tag{113}$$

Looking at the ratio of those probabilities we get (to simplify the notion we will suppress $\pm B$ and instead use $+/-$ in the superscript of ΔX):

$$\frac{p_{eq,1}(\Delta X^+_{\tau,f})}{p_{eq,2}(\Delta X^-_{\tau,r})} = \frac{\int_{\Delta X^+_{\tau,f}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma})}{\int_{\Delta X^-_{\tau,r}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,2}(\mathbf{\Gamma})}. \tag{114}$$

Now, using the definition of generalized work, equation (111), two times first get $f_{eq,2}(\mathbf{\Gamma})d\mathbf{\Gamma} = \exp[\Delta X_{r,\tau}(\mathbf{\Gamma})] f_{eq,1}(S^T\mathbf{\Gamma}) d(S^T\mathbf{\Gamma}) Z(\lambda_1)/Z(\lambda_2)$, then by inserting $\mathbf{\Gamma} \to M^T S^\tau \mathbf{\Gamma}$, we see that $\Delta X_{\tau,r}(\mathbf{\Gamma}) = -\Delta X_{\tau,f}(M^T S^\tau \mathbf{\Gamma})$ or $\Delta X^-_{\tau,r}(\mathbf{\Gamma}) = \Delta X^+_{\tau,f}(M^T S^\tau \mathbf{\Gamma})$ in the simplified notion. Using those two results we perform the transformations:

$$\begin{aligned}
\frac{p_{eq,1}(\Delta X^+_{\tau,f})}{p_{eq,2}(\Delta X^-_{\tau,r})} &= \frac{\int_{\Delta X^+_{\tau,f}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma})}{\int_{\Delta X^-_{\tau,r}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,2}(\mathbf{\Gamma})} \\
&= \frac{\int_{\Delta X^+_{\tau,f}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma}) Z(\lambda_2)/Z(\lambda_1)}{\int_{\Delta X^+_{\tau,f}(M^T S^\tau \mathbf{\Gamma})} \exp\big[-\Delta X^+_{\tau,f}(M^T S^\tau \mathbf{\Gamma})\big] d(M^T S^\tau \mathbf{\Gamma}) f_{eq,1}(M^T S^\tau \mathbf{\Gamma})} \\
&= \frac{\int_{\Delta X^+_{\tau,f}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma}) Z(\lambda_2)/Z(\lambda_1)}{\int_{\Delta X^+_{\tau,f}(\mathbf{\Gamma}')} d\mathbf{\Gamma}' \exp\big[-\Delta X^+_{\tau,f}(\mathbf{\Gamma}')\big] f_{eq,1}(\mathbf{\Gamma}')} \\
&= \exp[B] \frac{Z(\lambda_2)}{Z(\lambda_1)}.
\end{aligned} \tag{115}$$

Rewriting this again in full notion:

$$\frac{p_{eq,1}(\Delta X_{\tau,f} = B \pm dB)}{p_{eq,2}(\Delta X_{\tau,r} = -B \mp dB)} = \exp[B]\frac{Z(\lambda_2)}{Z(\lambda_1)} \tag{116}$$

we obtained the generalized Crooks fluctuation relation (GCFR).

In order to use GCFT we specialize the obtained result to an actual statistical mechanical ensemble and system of dynamics, obtaining the canonical form of CFT between initial and final equilibrium states with the same values of temperature, volume and number of particles (T, V, N). Referring to equation (60), the equilibrium distribution function $f(\mathbf{\Gamma}; 0)_{eq}$, the related free energy $F(\lambda)$ and partition function Z are given by

$$\begin{aligned} f(\mathbf{\Gamma}; 0)_{eq} &= Z^{-1} \exp[-\beta H(\mathbf{\Gamma}, 0)], \\ F(\lambda) \equiv -k_B T \ln Z(\lambda) &= -k_B T \ln[\int d\mathbf{\Gamma} \exp[-\beta H(\mathbf{\Gamma}, \lambda)]]. \end{aligned} \tag{117}$$

The Hamiltonian is varied parametrically from $\lambda_1 = \lambda(0)$ to the final, unique equilibrium state[14] $\lambda_2 = \lambda(\tau)$. In coupled system of equations (60), the phase space volume changes together with the Hamiltonian:

$$\begin{aligned} \Delta X_\tau &= \beta[H(S^\tau \mathbf{\Gamma}, \lambda(\tau)) - H(\mathbf{\Gamma}, \lambda(0))] + \ln\left[\frac{d\mathbf{\Gamma}}{d(S^\tau \mathbf{\Gamma})}\right] \\ &= \beta[H(S^\tau \mathbf{\Gamma}, \lambda(\tau)) - H(\mathbf{\Gamma}, \lambda(0))] + \int_0^\tau ds\ \Lambda(S^s \mathbf{\Gamma}) \\ &= \beta[H(S^\tau \mathbf{\Gamma}, \lambda(\tau)) - H(\mathbf{\Gamma}, \lambda(0)) + \Delta Q_\tau] \\ &= \beta \Delta W_\tau \end{aligned} \tag{118}$$

The generalized dimensionless "work" became identifiable as β times the work performed over a period of time τ:

$$\frac{p_1(\Delta W_\tau = W)}{p_2(\Delta W_\tau = -W)} = exp[\beta(W - \Delta F)], \tag{119}$$

thus obtaining the standard Crooks fluctuation theorem.

7.3 Jarzynski Equality

In ordinary statistical physics when transitions between two equilibrium states, say A and B, are performed infinitely slowly along some path then the total work W performed on such a system is equal to the Helmholtz free energy difference ΔF between the initial and final configurations. However, this is not the case when non-equilibrium transitions are considered. In fact, on average the work performed on the system will exceed Helmholtz free energy $\langle W \rangle \geq \Delta F$ and the difference will be equal to the dissipated energy, associated with increase of entropy during an irreversible process. The exact relation stating

$$\langle \exp(-\beta W) \rangle = \exp(-\beta \Delta F), \tag{120}$$

[14] To which it will relax thanks to the property of T-mixing

was found by Jarzynski [28]. Having already derived GCFR we will will use it to derive Jarzynski Equality also in it's generalized form, which expresses the free energy difference between two equilibrium states in terms of an average over irreversible paths. Subsequently, the generalized Jarzynski equality (GJE) follows from:

$$\begin{aligned}\langle \exp[-\Delta X_\tau(\mathbf{\Gamma})]\rangle_{eq,1} &= \int_{-\infty}^{\infty} dB\ p_f(\Delta X_\tau = B)\exp[-B] \\ &= \int_{-\infty}^{\infty} dB\ p_r(\Delta X_\tau = -B)\frac{Z(\lambda_2)}{Z(\lambda_1)} \\ &= \frac{Z(\lambda_2)}{Z(\lambda_1)},\end{aligned} \tag{121}$$

where the brackets $\langle ...\rangle_{eq,1}$ denote an equilibrium ensemble average over the initial equilibrium distribution. The usual practice is to use the inequality $e^x \geq 1 + x$ to rewrite it in a form of an inequality:

$$\begin{aligned}\frac{Z(\lambda_2)}{Z(\lambda_1)} &= \langle \exp[-\Delta X_\tau]\rangle_1 \\ &= exp[-\langle \Delta X_\tau\rangle_1]\langle \exp[-\Delta X_\tau + \langle \Delta X_\tau\rangle_1]\rangle \\ &\geq exp[-\langle \Delta X_\tau\rangle_1]\langle 1 - \Delta X_\tau + \langle \Delta X_\tau\rangle_1\rangle \\ &= exp[-\langle \Delta X_\tau\rangle_1]\end{aligned} \tag{122}$$

or by taking the logarithm:

$$\langle \Delta X_\tau\rangle \geq \ln\left[\frac{Z(\lambda_1)}{Z(\lambda_2)}\right] = \beta \Delta F_{21}, \tag{123}$$

where the right side is the free energy difference. Now, it's time to specialize our generalized work with a specific definition:

$$\Delta X = \beta \int_0^\tau ds\ W(s), \tag{124}$$

where W denotes the work. The inequality (123) implies $\Delta W_{21} \geq \Delta F_{21}$, so the minimum work is expended if the path is reversible or quasi-static, then the work is, in fact, the difference between the free energies.

We can also use the derived generalized relations directly to get another interesting result. If we choose the second equilibrium to be in fact our first equilibrium ($Z_1/Z_2 = 1$), therefore inducing a closed cycle, then the inequality (123) implies:

$$\oint ds \langle X(s)\rangle \geq 0, \tag{125}$$

i.e. the ensemble average of the cyclic integral of the generalized work is nonnegative. Although it's appearance is similar to Clausius inequality for the heat the derivation of the Clausius inequality is more demanding, but presented in [2]. The reason for this state of affairs is that, we have to complete many cycles until the system settles into a periodic response of the cyclic protocol before we can apply the cyclic integral for the heat. Moreover, not all systems do settle into a cyclic response.

8 Application to self-replication and adaptation

The main "new truth" obtained from fluctuation theorems is that, we can partially follow what we consider "microscopic trajectories", but those are not the real microscopic trajectories of fundamental particles, thus dissipation of heat occurs along the way.

Now we wish to "zoom out" from our theoretical considerations about the foundations of fluctuation theorems and take them for granted in order to explore some recent and interesting applications [56, 57]. Since we are no longer interested in the details, we proceed by switching to a stochastic description in which the Crooks equation (119) takes the form

$$\frac{\pi(j \to i; \tau)}{\pi(i \to j; \tau)} = \langle \exp\left[-\beta \Delta Q^{\tau}_{i \to j}\right] \rangle_{i \to j}, \tag{126}$$

where π's are the probability distributions of transitions over trajectories, during time τ either from $j \to i$ or $i \to j$ and Q is the dissipated heat. Let's define two macroscopic states denoted by I and II. We can then define the probabilities of transitions from macrostate I to macrostate II with the use of conditional probabilities[15]:

$$\pi(I \to II) = \int_{II} dj \int_{I} di \; \pi(i \to j) p(i|I) \tag{127}$$

and similarly,

$$\pi(II \to I) = \int_{I} di \int_{II} dj \; \pi(j \to i) p(j|II). \tag{128}$$

Their ratio is then given by:

$$\begin{aligned} \frac{\pi(II \to I)}{\pi(I \to II)} &= \frac{\int_I di \int_{II} dj \; \pi(j \to i) \frac{p(j|II)}{p(i|I)} p(i|I)}{\int_{II} dj \int_I di \; \pi(i \to j) p(i|I)} \\ &= \frac{\int_I di \int_{II} dj \; \pi(i \to j) \langle \exp\left[-\beta \Delta Q^{\tau}_{i \to j}\right] \rangle_{i \to j} \frac{p(j|II)}{p(i|I)} p(i|I)}{\int_{II} dj \int_I di \; \pi(i \to j) p(i|I)} \\ &= \langle \langle e^{-\beta \Delta Q^{\tau}_{i \to j}} \rangle_{i \to j} e^{\ln[\frac{p(j|II)}{p(i|I)}]} \rangle_{I \to II}, \end{aligned} \tag{129}$$

where in the last step we made use of equation (126) and $\langle ... \rangle_{I \to II}$ denotes an average over all paths from some microstate i in the initial ensemble I to some microstate j in the final ensemble II, with each path weighted by its likelihood. For clarity we rewrite the equation (129) as

$$\frac{\pi(II \to I)}{\pi(I \to II)} = \langle e^{-\beta \Delta Q^{\tau}_{i \to j} + \ln[\frac{p(j|II)}{p(i|I)}]} \rangle_{I \to II}, \tag{130}$$

remembering that $\Delta Q_{i \to j}$ contains a path ensemble average. One can now compare it with the equation (58) to see the essential difference between those two equations, namely the partial knowledge about microscopic trajectories and their dissipated heat. Proceeding, by moving the left side of (130) to the right side and under the ensemble average one gets

$$\langle e^{-\beta \Delta Q^{\tau}_{i \to j}} e^{\ln[\frac{p(j|II)}{p(i|I)}]} e^{-\ln[\frac{\pi(II \to I)}{\pi(I \to II)}]} \rangle_{I \to II} = 1, \tag{131}$$

[15] Note that we can do that, because we take into account what happens at the microscopic level

which by making use of inequality $e^x \geq 1 + x$ reduces to:

$$\beta\langle\Delta Q^{\tau}_{i\to j}\rangle_{I\to II} + \langle\ln[\frac{p(i|I)}{p(j|II)}]\rangle_{I\to II} + \ln\frac{\pi(II\to I)}{\pi(I\to II)} \geq 0. \tag{132}$$

The second term can now be identified with Shannon entropy between two macroscopic states $\Delta S_{int} = S_{II} - S_I$, obtaining:

$$\beta\langle\Delta Q^{\tau}_{i\to j}\rangle_{I\to II} + \ln\frac{\pi(II\to I)}{\pi(I\to II)} + \Delta S_{int} \geq 0. \tag{133}$$

This general result holds for wide range of transitions between the coarse-grained starting and ending states and has relevance to the known Landauer bound for heat generated by the erasure of a bit of information [57]. Here, however, we will proceed in applying it to a simple model of self replication.

8.1 Self-replication

Let's suppose we have a master equation for $n \gg 1$ governing the population

$$\dot{p}_n(t) = g\ n\left(p_{n-1}(t) - p_n(t)\right) - \delta\ n\left(p_n(t) - p_{n+1}(t)\right). \tag{134}$$

where $p_n(t)$ is the probability of having a population of n at time t with grow rate g and decay rate δ. If we now connect the state of 'living' with macrostate II and the state 'death' with macrostate I, then naturally we can assign $\pi(I \to II) = g\ \Delta t$ and $\pi(II \to I) = \delta\ \Delta t$ for some time Δt. The equation (133) then dictates

$$\Delta S_{int} + \beta\langle\Delta Q^{\tau}_{i\to j}\rangle_{I\to II} \geq \ln\frac{g}{\delta}. \tag{135}$$

which can interpreted as a general bound on self replication. An important thing to notice here is that ΔS_{int} is expected to be negative, because the self-replicator exists in non-equilibrium, 'living' state. By fixing all the terms other than the growth rate, one can gets the bound on the growth rate:

$$g \leq g_{max} = \delta\exp\left[\Delta S_{int} + \beta\langle\Delta Q^{\tau}_{i\to j}\rangle_{I\to II}\right]. \tag{136}$$

The most general observation that can be made from this equation, is that in order for the growth rate g to exceed the die rate δ the negative internal entropy change must be paid by the (strictly larger) dissipated heat. This dissipated energy, in case of a self-replicator, can have two sources: it is either stored in the reactants out of which the replicator gets built or comes from the work done on the system by some external driving field e.g. through the absorption of light.

Another comment can be made by considering two self-replicators with the same entropy change ΔS_{int} and die rate δ. In this scenario we see, that the one with larger heat dissipation will replicate faster. On the other hand an alternative route is also available by increasing the rate at which the self-replicator degrades δ and keeping the inner complexity, ΔS_{int}, low[16].

[16]Close to zero from the left side of the number axis.

8.2 Traversal of energy landscape

Let's consider a case of a driven thermostated system with two possible target macrostates II, III. We will be now interested in the probability ratio of transitions from state I to those two states. From equation (130) we get

$$\ln \frac{\pi(I \to II)}{\pi(I \to III)} = \ln \frac{\pi(II \to I)}{\pi(III \to I)} - ln \frac{\langle \exp\Big[-\beta \Delta Q_{i\to j} + \ln \frac{p_f^{II}}{p_i^I}\Big] \rangle_{I\to II}}{\langle \exp\Big[-\beta \Delta Q_{i\to k} + \ln \frac{p_f^{III}}{p_i^I}\Big] \rangle_{I\to III}}, \tag{137}$$

where the initial and final microstates where denoted by p_s and p_f. Because the system is driven and energy conserved, the work done on the system must go either to the heat or the systems Hamiltonian:

$$W_{i\to j} = \Delta Q_{i\to j} + H_* - H_I, \tag{138}$$

where the $*$ denotes a slot for the final state, here either II or III. If we now assume that the system is driven for a long time, we might neglect the correlations between the initial and final states and the work, giving us:

$$\ln \frac{\pi(I \to II)}{\pi(I \to III)} = \ln \frac{\pi(II \to I)}{\pi(III \to I)} - ln \frac{\langle \exp\Big[\beta(H_{II} - H_I) + \ln \frac{p_f^{II}}{p_i^I}\Big] \rangle_{I\to II}}{\langle \exp\Big[\beta(H_{III} - H_I) + \ln \frac{p_f^{III}}{p_i^I}\Big] \rangle_{I\to III}} - ln \frac{\langle e^{-\beta W} \rangle_{I\to II}}{\langle e^{-\beta W} \rangle_{I\to III}}. \tag{139}$$

The Hamiltonians can be obtained from the underlying equilibrium distributions $p_{eq} = e^{-\beta H_*}/Z^*_{eq}$ as follows:

$$H_* - H_I = -\beta^{-1}(\ln p^*_{eq} Z^*_{eq} - \ln p^I_{eq} Z^I_{eq}) = \beta^{-1}(\ln \frac{p^I_{eq}}{p^*_{eq}} + \ln \frac{Z^I_{eq}}{Z^*_{eq}}), \tag{140}$$

which after assuming that the initial distribution was at equilibrium, leaves us with:

$$\ln \frac{\pi(I \to II)}{\pi(I \to III)} = \ln \frac{\pi(II \to I)}{\pi(III \to I)} - ln \frac{\langle \frac{p_f^{II}}{p_{eq}^{II}} \rangle_{II}}{\langle \frac{p_f^{III}}{p_{eq}^{III}} \rangle_{III}} - ln \frac{\langle e^{-\beta W} \rangle_{I\to II}}{\langle e^{-\beta W} \rangle_{I\to III}} + \ln \frac{Z^{II}_{eq}}{Z^{III}_{eq}}. \tag{141}$$

One can notice that the second term will be zero if the final distributions are equilibrium distributions. Since, the free energy is defined as $F^* = -\beta \ln Z^*_{eq}$, we can combine the last two terms, introducing a term called dissipated work, defined by:

$$W_d = W - Z^*_{eq} + Z^I_{eq}, \tag{142}$$

thus obtaining:

$$\ln \frac{\pi(I \to II)}{\pi(I \to III)} = \ln \frac{\pi(II \to I)}{\pi(III \to I)} - ln \frac{\langle \frac{p_f^{II}}{p_{eq}^{II}} \rangle_{II}}{\langle \frac{p_f^{III}}{p_{eq}^{III}} \rangle_{III}} - ln \frac{\langle e^{-\beta W_d} \rangle_{I\to II}}{\langle e^{-\beta W_d} \rangle_{I\to III}}. \tag{143}$$

We might now try to interpret each of those terms. The intuitive meaning of the first one, is the fact that more likely are the states from which one can likely come back. The last term on the other hand, might be expanded with the use of cumulant expansion:

$$-\ln\langle\exp(-\beta W_d)\rangle = \beta\langle W_d\rangle - \Phi, \tag{144}$$

where Φ holds all higher order terms of the cumulant expansion. From the convexity of the exponential function and Jensen's inequality one might check[17] that we must have $\Phi \geq 0$. Thus, Φ can be thought of as a correction; due to the dispersion of the dissipated work distribution about the average $\beta\langle W_d\rangle$ that gives the heaviest weight to the leftward tail of the work distribution [58].

One might now argue that averaged dissipated heat $\langle W_d\rangle$ can be connected with particle drift through oscillatory energy barriers (details depending on the model), this was in fact recently demonstrated by Nikolay Perunov et al [56].

SHOW THIS, on own example

We have now seen how recent progress in non-equilibrium statistical mechanics and especially the fluctuation theorem allowed us to make plausible predictions about evolution of inanimate and possibly animate matter. Taking into account the information processing/storing side of the latter, one can ask a natural question - could there be a deeper relation behind the fluctuation theorem, perhaps involving information theory? Could fluctuation theorem be derived from it? The pursuit for the answer to this questions will be the topic of next paragraph.

9 Search for a unifying principle

The search for variational or extremization principles in physics has a long history of success. In classical mechanics, one finds the equations of motion from Lagrangian formalisms with the principle of least action. In thermodynamics and statistical physics of equilibrium state the principle of maximum entropy yields the true equilibrium states of a given system.

In 1912, Ehrenfest was the first who asked whether such a principle for a yet unknown function could exist for non-equilibrium steady states [41]. The search, however, could be dated to the works of Kirchoff and Rayleigh and their least dissipation theorems, which development ceased immediately after the establishment of equilibrium statistical mechanics. Interest in least dissipation principles was reborn in the 1950s due to joint paper of Onsager and Machlup [59] which modeled small microscopic fluctuations (away from criticality) around exponential relaxation to equilibrium and which provided a "microscopic" basis for the least dissipation principle. The validity of the least dissipation principles, including Prigogine's MinEP, is however restricted to near equilibrium states. In fact, in an example presented by Landauer [38], the steady state was generating the maximum amount of entropy.

[17] $e^{-\Psi+\Phi} = \langle\exp(-\beta W_d)\rangle \geq \exp(-\beta\langle W_d\rangle) = e^{-\Psi}$

9.1 MaxEP principle

The example presented by Landauer, was however, not, the origin of the so-called maximum entropy production (MaxEP, sometimes called MEP) principle. The conjecture of MaxEP originated from some promising (but controversial) successes in studies of planetary climates [60, 61, 62] and since then was applied in a loose manner to other fields, such as fluid turbulence [63, 64] and crystal growth morphology [65].

The main reasons that drove controversies, e.g. [66], around those successes were the accuracy of the results and the *ad hoc* and unsystematic manner in which MaxEP was applied. For example, the earliest successes of MaxEP applied to Earth's climate were based on a 2-zone model where the energy balance and temperatures were obtained through maximization of entropy production (EP) associated with meridional heat transport in the atmosphere and the oceans, completely ignoring the dominant part of the total EP coming from radiative EP.

There are also different formulations of MaxEP [67, 68, 69, 70], many not really formalised, which were subjects to a pletora of readily available reviews, e.g. [71, 72, 41]. To author's best knowledge, all versions were predicting, a time independent probability distribution function, therefore, a detailed review of individual differences will not be undertaken in this work. Instead, we will comment on the most common formulation from the point of view of dissipation function and then present what seems to be, the most promising idea.

The usual working formulation of MaxEP principle states that *for systems admitting a spectrum of possible steady states, MaxEP says that the system is most likely to be found in steady state with the greatest entropy production.* Directly from this statement the relation between MaxEP and Prigogine's MinEP becomes clear.

9.1.1 Relation to MinEP

For linear, near equilibrium systems that only admit a single steady state, MinEP states that system's steady state produces less entropy than any possible transient state. MinEP compares steady states with transient states.

For some not yet determined class of nonlinear, far from equilibrium systems that admit multiple steady states, MaxEP states that the system settles in the steady state with the greatest entropy production. MaxEP compares steady states to other steady states, staying silent on transient states.

9.1.2 Extrema of the Dissipation Function and MaxEP

The dissipation function is similar to the entropy production, and although it is not directly connected to a state function, the various fluctuation theorems provide exact, non-equilibrium relations. Given this similarity, it is interesting to see how consistent with it might be the MaxEP principle.

If we consider a special case of equation (96) when the system is driven i.e. dissipation is given by equation (105), then we obtain:

$$f(\mathbf{\Gamma};t) = f(\mathbf{\Gamma};0)\exp\left[-\beta\int_0^t ds\ V\boldsymbol{J}(\mathbf{\Gamma}_{-s})\cdot\boldsymbol{F}_e(t-s)\right] \tag{145}$$

from which it is clear that the probability of observing the phase $\mathbf{\Gamma}$ at time t is increased if the integrated dissipation terminating at phase $\mathbf{\Gamma}$ and time t is large and positive. Furthermore, we suspect that the dissipation term will, at long times, dominate over the initial probability distribution, thus justifying MaxEP as a good approximation.

Moreover, considering equation (99) with $B = \Omega$, one can see, that if, the autocorrelation function $\langle\Omega(\mathbf{\Gamma})\Omega(\mathbf{\Gamma}_t)\rangle$ decays monotonically with time, then the value of the ensemble average of the instantaneous dissipation function, $\langle\Omega(t)\rangle$ will be have a higher value when it reaches its steady state than for any other state it passes through. The system should therefore expose itself to a steady-state of maximium dissipation. The assumption of monotonicity of the autocorrelation function of course doesn't hold in all cases. Through numerical simulations done in [73] it was shown that the dissipation average sometimes peaks to a higher number before reaching a stable steady state value. One should note here, that in the light of formulation of MaxEP from section 9.1, this result doesn't prove or disprove the MaxEP principle as MaxEP compares steady states to steady states. It's simply the instantaneous dissipation that is not necessarily maximized at all times.

The (mostly) negative statements with regards to MaxEP, one finds in the works of Evans [2], seem to be referring to a universal MaxEP principle valid at all times, rather than to some approximation. There are of course valid reasons for scepticism about MaxEP as a general law. In an extensive analysis analysis of family of dissipation functionals of the form $f = D(D_v/D_m)^n$ and $f = D_m(D_v/D_m)^n$ for $n \geq 0$, involving the total dissipation D, dissipation in the mean flow D_m, and dissipation in the fluctuating flow D_v, Kerswell [74] was unable to identify a universal dissipation functional that applied to all flow problems. Finally, the study of generic MaxEP hypotheses through numerical simulations at the microscopic level is rather hard, because systems that generate multiple steady states, such as convection and turbulent flow, are computationally very expensive.

9.1.3 MaxEnt based formulations of MaxEP

Perhaps the most promising interpretation of MaxEP principle, put forward by Dewar[75, 76, 77], assumes that MaxEP is not a physical principle at all. Instead, by analogy to Jaynes MaxEnt it is an inference method i.e. a method for deducing the most unbiased predictions from an incomplete set of statistical data. In the light of this interpretation, problems with finding maximum entropy production in a given model, may be related to *the way* the principle was applied, i.e. finding the right constraints and not to the principle itself. Instances of apparent failure were noted in phenomenological models of heat flow in plasma/fluid system [69], where maximum as well as minimum was observed

depending on how the system was driven, and also in the climate general circulation models (GCM), where no extremum of EP was not found [41]. Whether those failures can indeed be justified by unlucky selection of constraints remains to be shown. Nevertheless, a precocious procedure was proposed by Dewar which involves three steps.

The first step is done by narrowing the scope of validity. We notice that systems that are weakly driven have (neglecting the set of measure zero) only one steady state available.[18] Therefore there is no room for MaxEP to operate and we instead focus on systems driven strongly in far for equilibrium regime. Two classic examples of far from equilibrium systems involve *shear turbulence* with Reynolds numbers greater than the critical value necessary for the onset of turbulence and *Rayleigh-Bénard* cell with Rayleigh numbers greater than the critical value necessary for the onset of convection. In those scenarios both examples exhibit many flow solutions allowed when we apply only a restricted set of stationary conditions rather than the full dynamics.

Secondly, Dewar introduces information theoretical measure of the distance from equilibrium, or *irreversibility* I (similar to defined in section 149), defined in terms of the relative probabilities of forward and reverse fluxes. The first demand (dynamical instability) is then reformulated as a strong inequality constraint $I > I_{min}$. Then using procedures known from MaxEnt it is shown that I adopts it's maximum possible value under the stationarity constraints.

The final step consists of reinterpretation of I as thermodynamic entropy production. In this derivation of MaxEP, entropy production depends on the applied constraints.

The just outlined procedure (which is Dewar's third iteration of the principle) is presented in the proceeding section in detail.

9.2 MaxEP principle as an inference principle

We will consider a general open system (volume V, bondary Ω) which exchanges both matter and energy with it's surroundings. The system may consist of several components. The presence of fluxes, both within the system, and between the system and its environment is the primary characteristics of the non-equilibrium stationary states. We'll denote the instantaneous value of those fluxes, by the vector $\boldsymbol{f}$. The flux vector $\boldsymbol{f}$ may be related to some local density ρ with the use of the continuity equation $\partial\rho/\partial t = -\nabla\cdot\boldsymbol{f}+h$, where h denotes a local source; alternatively (for example in case of Navier-Stokes equations), the components of $\boldsymbol{f}$ might themselves be identified with local densities. Macroscopic state of the system is then described by $\boldsymbol{f}$ (or ρ) with the stationarity condition given by equation

[18] Of course, in principle, there is always only one stationary state when our knowledge about the system is full, however, by assumption, in case of strongly driven systems our ignorance is much greater and there is more room for an inference principle.

(42). Similarly to the MaxEnt we maximize the relative entropy

$$H = -\int p(\boldsymbol{f}) \ln \frac{p(\boldsymbol{f})}{q(\boldsymbol{f})} d\boldsymbol{f}, \tag{146}$$

with respect to probability distribution function $p(\boldsymbol{f})$, subject to given dynamical constraints C, where $q(\boldsymbol{f})$ is a prior probability distribution function, with the symmetry $q(\boldsymbol{f}) = q(-\boldsymbol{f})$ which corresponds to zero flux state $\boldsymbol{F} = \int q(\boldsymbol{f})\boldsymbol{f} d\boldsymbol{f} = 0$. By Gibbs' inequality, $H \leq 0$ with equality if and only if $p(\boldsymbol{f}) = q(\boldsymbol{f})$. The constraints represented by C are written in the generic form of functionals of fluxes $\phi_m(\boldsymbol{f})$ and labeled by m:

$$\int p(\boldsymbol{f})\phi_m(\boldsymbol{f}) d\boldsymbol{f} = 0 \tag{147}$$

which we can demand without loss of generality to be zero. We also have the normalization constraint

$$\int p(\boldsymbol{f}) d\boldsymbol{f} = 1. \tag{148}$$

Now, in order to enforce the multiplicity of stationary states we introduce the *irreversibility* defined by the Kullback-Leibler (KL) divergence of $p(\boldsymbol{f})$ and $p(-\boldsymbol{f})$:

$$I = \int p(\boldsymbol{f}) \ln \frac{p(\boldsymbol{f})}{p(-\boldsymbol{f})} d\boldsymbol{f}. \tag{149}$$

By Gibbs' inequality, $I \geq 0$ with equality if and only if $p(\boldsymbol{f}) = p(-\boldsymbol{f})$, so that $I = 0$ corresponds to the equilibrium state $\boldsymbol{F} = \boldsymbol{0}$. Making the first step of the procedure, we demand that there is a state characterized by minimal irreversibility, $I > I_{min}(C) > 0$, the value of which depends on the stationarity conditions C of equation (147). Those conditions should also determine the upper bound $I \leq I_{max}(C)$ by assumption.

Additionally we introduce a trial mean flux $\boldsymbol{F}$ (which is subsequently relaxed) via the auxiliary constraint

$$\int p(\boldsymbol{f})\boldsymbol{f} d\boldsymbol{f} = \boldsymbol{F}. \tag{150}$$

The motivation for introducing $\boldsymbol{F}$ is to allow one to establish an extremal principle whereby $\boldsymbol{F}(C)$ is determined by varying the trial solution $\boldsymbol{F}$[19]. Under given constraints the MaxEnt solution for $p(\boldsymbol{f})$ is given by

$$p(\boldsymbol{f})^* = q(\boldsymbol{f}) Z^{-1} \exp\left[\boldsymbol{\lambda} \cdot \boldsymbol{f} + \boldsymbol{\alpha} \cdot \boldsymbol{\phi}(\boldsymbol{f}) - \mu(d(\boldsymbol{f}) - e^{-d(\boldsymbol{f}))}\right], \tag{151}$$

where $d(\boldsymbol{f}) = \ln p(\boldsymbol{f})/p(-\boldsymbol{f})$, $\phi(\boldsymbol{f})$ denotes the vector with components $\phi_m(\boldsymbol{f})$, $Z = Z(\lambda, \boldsymbol{\alpha}, \mu)$ is a normalisation factor (partition function) and $\boldsymbol{\lambda}$, $\boldsymbol{\alpha}$ and μ are Lagrange

[19]This approach is analogous to the way in which equilibrium variational principles (e.g. minimum free energy) can be derived from MaxEnt by enlarging the set of fixed macroscopic variables X to include one or more free unconstrained variables Y, then maximizing $S = H_{max}(X, Y)$ with respect to Y with X held fixed.

multipliers for (150), (147) and the upper-bound inequality $I < I_{max}$ respectively. The maximized relative entropy is given by:

$$S(\boldsymbol{F}, I_0, C) \equiv H_{max} = \ln Z(\lambda, \boldsymbol{\alpha}, \mu) - \boldsymbol{\lambda} \cdot \boldsymbol{F} + \mu(I_0 - 1). \tag{152}$$

The next step involves maximizing $S(\boldsymbol{F}, I_0, C)$ with respect to the trial flux solution $\mathbf{F}$ with I_0 held fixed. In the absence of the dynamic instability condition $I > I_{min}(C)$, MaxEnt predicts the basal state $I = I_{min}(C)$, i.e. minimal irreversibility, but when the basal state is excluded by the dynamical instability condition MaxEnt predicts a probability distribution function for which $I = I_{max}(C)$ i.e. maximal irreversibility which is characterized by $\mu = 0$ and equation (151) becomes:

$$p(\boldsymbol{f})^* = q(\boldsymbol{f}) Z^{-1} \exp[\boldsymbol{\lambda} \cdot \boldsymbol{f} + \boldsymbol{\alpha} \cdot \boldsymbol{\phi}(\boldsymbol{f})]. \tag{153}$$

The intermediate solutions not predicted by this procedure would then correspond to transient states, between equilibrium state and the non-equilibrium stationary state. The predicted irreversibility measure (equation 149) is then given by a functional of $\mathbf{F}$:

$$I(\boldsymbol{F}) = 2\ \lambda(\boldsymbol{F})\boldsymbol{F} + 2\ \boldsymbol{\alpha}(\boldsymbol{F})\Phi^A(\boldsymbol{F}), \tag{154}$$

where

$$\Phi^A(\boldsymbol{F}) \equiv \frac{1}{2} \int p(\boldsymbol{f})[\boldsymbol{\phi}(\boldsymbol{f}) - \boldsymbol{\phi}(-\boldsymbol{f})] d\boldsymbol{f}, \tag{155}$$

is the expectation value of the anti-symmetric part of $\boldsymbol{\phi}(\boldsymbol{f})$. One may now notice that if the vector $\boldsymbol{\phi}$ is anti-symmetric or symmetric then (after adding the condition given by equation (147)) the $\boldsymbol{\Phi}^A = 0$, and the irreversibility measure reduces to:

$$I(\boldsymbol{F}) = 2\boldsymbol{\lambda}(\boldsymbol{F})\boldsymbol{F}. \tag{156}$$

A result consistent with the conclusion that there are no general expressions for thermodynamic entropy independent of the constraints.

9.2.1 Relation to fluctuation theorem

In his first approaches, Dewar claimed to derive [75, 77] the fluctuation theorem from MaxEP. The claims in the more recent works are more modest [41]. MaxEP there (and here) described, is in fact an approximation to the fluctuation theorem. Conclusion immediately obvious, regarding the fact that the fluctuation theorem also describes the non-steady states. The irreversibility measure I defined for MaxEP is related to the dissipation function Ω_t of equation (83) through $I = \langle \Omega_t \rangle$, if we interpret the flux vector $\boldsymbol{f}$ as a time-average over time t.

9.2.2 Example application for planetary atmospheres

In this section we will describe an example application of MaxEP to planetary climates, therefore justifying ad-hoc hypotheses made in [61][60]. The constraints in this model 1D box model are given by a simple radiative balance between total incoming short-wave (SW) irradiance and the total outgoing long-wave (LW) irradiance:

$$\sum_i F_{SW,i} = \sum_i F_{LW,i}, \tag{157}$$

where the sum goes over all latitudinal zones. The meridional heat fluxes are identified as $\boldsymbol{f} = \{f_i\}$ where f_i is the flux from zone $i-1$ to zone i. The local (for each zone) equation has then the form

$$F_{SW,i} - F_{LW,i} + \Delta F_i = 0. \tag{158}$$

where $\Delta F_i = F_i - F_{i+1}$ and $\int p(\boldsymbol{f})\boldsymbol{f}d\boldsymbol{f} = \boldsymbol{F} = \{F_i\}$. The basal equilibrium state corresponds then to radiative equilibrium $F_{SW,i} = F_{LW,i}$ with $\boldsymbol{F} = \boldsymbol{0}$ and $I = I_{min}$. Having understood the constraints we can now rewrite the equation (157) as a condition

$$\int p(\boldsymbol{f})\phi_1(\boldsymbol{f})d\boldsymbol{f} = 0, \tag{159}$$

with antisymmetric functional $\phi_1(\boldsymbol{f}) = \sum_i \Delta f_i$. One can show that the Lagrange multiplier $\lambda = -\Delta(\frac{1}{T_i}) = \frac{1}{T_{i+1}} - \frac{1}{T_i}$ by considering a slight modification (non-steady state) of equation (153) in which in zone i energy $u(\tau)$ is stored for some finite time τ [41]. Using the equation (156) we get our final result:

$$I(\boldsymbol{F}) = 2\boldsymbol{\lambda}(\boldsymbol{F})\boldsymbol{F} \propto -\sum_i F_i \Delta\left(\frac{1}{T_i}\right) = \sum_i \left(\frac{\Delta F_i}{T_i}\right), \tag{160}$$

which has the form of entropy production function used by [61, 60]. Further examples can be found in [41].

10 Summary

After over 100 years, the arguments of Boltzmann and others on the Loschmidt paradox have become more tangible and very refined through the sub-macroscopic description provided by the fluctuation theorem and it's 'big-brother' The research on fluctuation theorems has been immensely productive, providing us with better understanding of thermostated systems, steady states, dissipation and relaxation processes, while keeping the correspondence with linear response theory. Nevertheless, the limits of its generality are still being researched, while the applications of fluctuation theorem are already giving us new methods of calculating free energy of large molecules and some early understanding of self replicators and adaptation. The full understanding will almost surely, demand

the inclusion of information theory into the picture and perhaps a shift to the quantum description [78].

In the present work we conclude, that the presented principle of MaxEP, pioneered by Dewar, stays in reasonable agreement with, the more general, fluctuation theorem. The proof of its universal validity, however, remains a challenging task, given its non-physical nature and difficulties in finding the appropriate constraints.

References

[1] J. W. Gibbs, "The collected works of J. Willard Gibbs," 1928.

[2] D. J. Evans, D. J. Searles, and S. R. Williams, *Fundamentals of classical statistical thermodynamics: dissipation, relaxation, and fluctuation theorems*. Berlin: John Wiley & Sons, 2016.

[3] A. Einstein, "Investigations on the Theory of Brownian Motion.," 1956.

[4] L. Onsager, "Reciprocal relations in irreversible processes. I.," *Physical review*, vol. 37, no. 4, pp. 405–426, 1931.

[5] L. Onsager, "Reciprocal Relations in Irreversible Processes. II.," *Phys. Rev.*, vol. 38, pp. 2265–2279, Dec. 1931.

[6] R. Kubo, "Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems," *Journal of the Physical Society of Japan*, vol. 12, pp. 570–586, June 1957.

[7] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 623–656, 1948.

[8] Schrodinger, Erwin, "What is Life? The Physical Aspect of the Living Cell.," *The American Naturalist*, vol. 79, no. 785, pp. 554–555, 1945.

[9] F. Schwabl, *Statistical mechanics*. Advanced Texts in Physics, Berlin: Springer, 2002.

[10] J. R. Dorfman, "An Introduction to Chaos in Nonequilibrium Statistical Mechanics, 1999."

[11] R. O. Doyle, "The Origin of Irreversibility," *Information Philosopher*.

[12] D. Layzer, "Cosmic evolution and thermodynamic irreversibility," *Pure and Applied Chemistry*, vol. 22, no. 3-4, 1970.

[13] S. Wolfram, *A new kind of science*. Urbana-Champaign, IL: Wolfram Media, 2002.

[14] C. Rovelli, "Is Time's Arrow Perspectival?," *arXiv.org*, May 2015.

[15] M. Courbage and I. Prigogine, "Intrinsic randomness and intrinsic irreversibility in classical dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 80, pp. 2412–2416, Apr. 1983.

[16] J. C. Maxwell, "On the Dynamical Theory of Gases.," in *Proceedings of the Royal Society of . . .*, 1866.

[17] I. Prigogine, "Time, structure and fluctuations," *Nobel Lectures in Chemistry 1971-1980*, 1993.

[18] K. Gustafson, "Microscopic irreversibility," *Discrete Dynamics in Nature and Society*, vol. 2004, no. 1, pp. 155–168, 2004.

[19] J. Bricmont, "Science of chaos of chaos in science?," *Annals of the New York Academy of Sciences*, vol. 775, pp. 131–175, June 1995.

[20] J. G. Fox, "Evidence Against Emmision Theories," *American Journal of Physics*, vol. 33, pp. 1–17, Jan. 1965.

[21] C. Kiefer, "Quantum cosmology and the arrow of time," *Brazilian Journal of Physics*, vol. 35, pp. 296–299, June 2005.

[22] J. Barbour, T. Koslowski, and F. Mercati, "Identification of a Gravitational Arrow of Time," *Physical Review Letters*, vol. 113, pp. 181101–5, Oct. 2014.

[23] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM journal of research and development*, vol. 5, no. 3, pp. 183–191, 1961.

[24] C. H. Bennett, "Logical reversibility of computation," *IBM journal of research and development*, vol. 17, no. 6, pp. 525–532, 1973.

[25] W. G. Hoover, A. J. C. Ladd, and B. Moran, "High-Strain-Rate Plastic Flow Studied via Nonequilibrium Molecular Dynamics," *Physical Review Letters*, vol. 48, pp. 1818–1820, June 1982.

[26] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, "Probability of second law violations in shearing steady states," *Physical Review Letters*, vol. 71, pp. 2401–2404, Oct. 1993.

[27] G. Gallavotti and E. G. D. Cohen, "Dynamical Ensembles in Nonequilibrium Statistical Mechanics," *Physical Review Letters*, vol. 74, pp. 2694–2697, Apr. 1995.

[28] C. Jarzynski, "Nonequilibrium Equality for Free Energy Differences," *Physical Review Letters*, vol. 78, pp. 2690–2693, Apr. 1997.

[29] G. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," pp. 1–7, Feb. 2008.

[30] J. England, "Dissipative adaptation in driven self-assembly," vol. 10, pp. 919–923, Nov. 2015.

[31] J. England and G. Haran, "To fold or expand–a charged question.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 14519–14520, Aug. 2010.

[32] E. T. Jaynes, "Gibbs vs Boltzmann Entropies," *American Journal of Physics*, vol. 33, no. 5, pp. 391–398, 1965.

[33] E. T. Jaynes, "Information Theory and Statistical Mechanics. II," *Phys. Rev.*, vol. 108, no. 2, pp. 171–190, 1957.

[34] P. Glansdorff, I. Prigogine, and R. N. Hill, *Thermodynamic theory of structure, stability and fluctuations*, vol. 41. American Journal of Physics, 1973.

[35] S. R. De Groot and P. Mazur, *Non-equilibrium thermodynamics*. 2013.

[36] T. Gilbert and J. R. Dorfman, "Entropy Production: From Open Volume-Preserving to Dissipative Systems," *Journal of Statistical Physics*, vol. 96, no. 1-2, pp. 225–269, 1999.

[37] S. Goldstein, J. L. Lebowitz, and Y. Sinai, "Remark on the (Non)convergence of Ensemble Densities in Dynamical Systems," *arXiv.org*, pp. 393–395, Apr. 1998.

[38] R. Landauer, "Stability and entropy production in electrical circuits," *Journal of Statistical Physics*, vol. 13, no. 1, pp. 1–16, 1975.

[39] H. B. G. Casimir, "On Onsager's Principle of Microscopic Reversibility," *Reviews of Modern Physics*, vol. 17, pp. 343–350, Apr. 1945.

[40] C. Kittel, *Introduction to Solid State Physics; 8th ed.* Hoboken, NJ: Wiley, 2005.

[41] R. C. Dewar, C. H. Lineweaver, R. K. Niven, and K. Regenauer-Lieb, *Beyond the Second Law: An Overview*. Understanding Complex Systems, Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

[42] G. Nicolis and I. Prigogine, "Irreversible processes at nonequilibrium steady states and Lyapounov functions," *Proceedings of the National Academy of Sciences*, vol. 76, no. 12, pp. 6060–6061, 1979.

[43] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, "Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies," *arXiv.org*, pp. 231–234, Dec. 2005.

[44] D. M. Carberry, M. Baker, and G. M. Wang, "An optical trap experiment to demonstrate fluctuation theorems in viscoelastic media," *Journal of Optics A: ...*, vol. 9, no. 8, pp. S204–S214, 2007.

[45] S. Joubaud, N. B. Garnier, and S. Ciliberto, "Fluctuation theorems for harmonic oscillators," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, pp. P09018–P09018, Sept. 2007.

[46] U. Seifert, "Fluctuation theorem for a single enzym or molecular motor," *EPL (Europhysics Letters)*, vol. 70, no. 1, pp. 36–41, 2005.

[47] T. Monnai, "Unified treatment of the quantum fluctuation theorem and the Jarzynski equality in terms of microscopic reversibility," *Physical Review E*, vol. 72, pp. 19–4, Aug. 2005.

[48] D. J. Evans and L. Rondoni, "Comments on the Entropy of Nonequilibrium Steady States," *Journal of Statistical Physics*, vol. 109, no. 3/4, pp. 895–920, 2002.

[49] J. Kurchan, "Fluctuation theorem for stochastic dynamics," *Journal of Physics A: Mathematical and General*, vol. 31, pp. 3719–3729, Apr. 1998.

[50] D. J. Searles and D. J. Evans, "Fluctuation theorem for stochastic systems," *Physical Review E*, vol. 60, pp. 159–164, July 1999.

[51] J. Kurchan, "Six out of equilibrium lectures," *arXiv.org*, Jan. 2009.

[52] D. J. Searles and D. J. Evans, "The fluctuation theorem and Green–Kubo relations," *The Journal of Chemical Physics*, vol. 112, pp. 9727–9735, June 2000.

[53] 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," *Physical Review Letters*, vol. 89, pp. 128–4, July 2002.

[54] T. Yamada and K. Kawasaki, "Nonlinear effects in the shear viscosity of critical mixtures," *Progress of Theoretical Physics*, 1967.

[55] S. R. Williams and D. J. Evans, "Time-dependent response theory and nonequilibrium free-energy relations," *Physical Review E*, vol. 78, pp. 021119–7, Aug. 2008.

[56] N. Perunov, R. A. Marsland, and J. England, "Statistical Physics of Adaptation," *Physical Review X*, vol. 6, no. 2, 2016.

[57] J. England, "Statistical physics of self-replication," vol. 139, no. 12, pp. 121923–9, 2013.

[58] C. Jarzynski, "Rare events and the convergence of exponentially averaged work values," *Physical Review E*, vol. 73, pp. P09005–10, Apr. 2006.

[59] S. Machlup and L. Onsager, "Fluctuations and Irreversible Process. II. Systems with Kinetic Energy," *Phys. Rev.*, vol. 91, pp. 1512–1515, Sept. 1953.

[60] R. D. Lorenz, "Titan, Mars and Earth: entropy production by latitudinal heat transport.," *onlinelibrary.wiley.com.*

[61] G. W. Paltridge, G. D. Farquhar, and M. Cuntz, "Maximum entropy production, cloud feedback, and climate change," *Geophysical Research ...*, vol. 34, p. 3445, July 2007.

[62] A. Kleidon and R. D. Lorenz, *Non-equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond*. Understanding Complex Systems, Berlin: Springer, 2005.

[63] H. Ozawa, "The second law of thermodynamics and the global climate system: A review of the maximum entropy production principle," *Reviews of Geophysics*, vol. 41, no. 4, pp. 1018–24, 2003.

[64] W. V. R. MALKUS, "Borders of disorder: in turbulent channel flow," *Journal of Fluid Mechanics*, vol. 489, pp. 185–198, July 2003.

[65] L. M. Martyushev and S. V. Serebrennikov, "Morphological stability of a crystal with respect to arbitrary boundary perturbations," *Technical Physics Letters*, vol. 32, pp. 614–617, July 2006.

[66] R. Goody, "Maximum Entropy Production in Climate Theory," *Journal of the Atmospheric Sciences*, vol. 64, pp. 2735–2739, July 2007.

[67] N. Virgo, "From Maximum Entropy to Maximum Entropy Production: A New Approach," *Entropy*, vol. 12, pp. 107–126, Jan. 2010.

[68] P. Županović, D. Kuić, Ž. B. Lošić, D. Petrov, D. Juretić, and M. Brumen, "The Maximum Entropy Production Principle and Linear Irreversible Processes," *Entropy*, vol. 12, pp. 996–1005, May 2010.

[69] Y. Kawazura and Z. Yoshida, "Entropy production rate in a flux-driven self-organizing system," *Physical Review E*, vol. 82, no. 6, 2010.

[70] R. C. Dewar, "4 Maximum Entropy Production and Non-equilibrium Statistical Mechanics," in *Non-equilibrium Thermodynamics and the Production of Entropy*, pp. 41–55, Berlin/Heidelberg: Springer, Berlin, Heidelberg, 2005.

[71] S. Bruers, "Classification and discussion of macroscopic entropy production principles," *arXiv.org*, Apr. 2006.

[72] P. Županović, D. Kuić, D. Juretić, and A. Dobovišek, "On the Problem of Formulating Principles in Nonequilibrium Thermodynamics," *Entropy*, vol. 12, pp. 926–931, Apr. 2010.

[73] S. J. Brookes, J. C. Reid, and D. J. Evans, "The fluctuation theorem and dissipation theorem for Poiseuille flow," *Journal of Physics: ...*, vol. 297, p. 012017, 2011.

[74] R. Kerswell, "Upper bounds on general dissipation functionals in turbulent shear flows: revisiting the 'efficiency' functional," *Journal of Fluid Mechanics*, vol. 461, pp. 1–37, July 2002.

[75] R. Dewar, "Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states," *J. Phys. A*, vol. 36, no. 3, pp. 631–641, 2003.

[76] R. C. Dewar, "Maximum Entropy Production as an Inference Algorithm that Translates Physical Assumptions into Macroscopic Predictions: Don't Shoot the Messenger," *Entropy*, vol. 11, pp. 931–944, Dec. 2009.

[77] R. C. Dewar, "Maximum entropy production and the fluctuation theorem," *Journal of Physics A: Mathematical and General*, vol. 38, pp. L371–L381, May 2005.

[78] J. Kurchan, "A Quantum Fluctuation Theorem," *arXiv.org*, July 2000.