Non-equilibrium systems and growth of complexity

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

July 21, 2017

Contents

1 Introduction

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 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.

After that, around 1850 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.

More important developments came after the recognition by Rudolf Clausius and James Clerk Maxwell in 1850s (first noticed by Daniel Bernoulli in 1738) 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, Rudolf Clausius introduced the notion of entropy, defined as the ratio of heat and temperature and redefined the Second Law stating that for isolated systems this quantity can only increase in time.

In 1872 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 objection was poised known as the Loschmidt paradox which stated basically, 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 by Gibbs as well[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 to the later

part of this paragraph 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 around 1876 that in a gas 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 1900, 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. It was 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. 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, the development of thermodynamics was largely overshadowed by quantum theory and little fundamental work was done on it. Nevertheless, the Second Law had become to be regarded as a fundamental principle, whose foundations should be questioned only as a curiosity[3].

2.2 The information era and Schrödinger's influence on physics

In the 1940s Claude Shannon introduced the notion of information quantity [**?**] 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 coincided with the discovery of the structure of DNA by James Watson and Francis Crick and together with written by Erwin Schrödinger, influential book titled "What is life?" sparked enthusiasm and inspired generations of physicists to answer the alluring (though not easy) question of the role of physics in biological processes.

In any event it probably wouldn't be an exaggeration to say that Schrödinger himself (as he admits), was inspired by the work of German-American physicists Max Delbrück; who helped launch the molecular biology research program in the late 1930s and explained (in main part) the mechanism of heredity and mutation. Regardless, Schrödinger makes some very essential observations on the nature of living organisms.

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 stay unaffected

(in special environments) by random molecular motion known as heat 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 it's 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. Perhaps, worth mentioning is his hypothesis of "life intensity" the term which ought to parallel with 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 it's 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 noticed that the way in which he derived the H-Theorem from his equation, implicitly assumed that the particles are uncorrelated before the collisions (through Stosszahlansatz), but become correlated after the collision [4] [5], thus causing the time-reversal asymmetry[2] .

[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 or the chaos hypothesis of Onsager

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 state macrostate. The final state is obtained by taking the *sum* of the probabilities over the indistinguishable microstates, 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 are able to study all degrees of freedom at the final state, the arrow of time would indeed disappear as is the case with microscopic or 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 [6] [7] [3] [8] [9] this explanation of irreversibility noticed long time ago by Clausius, Boltzmann[3], Kelvin, Maxwell[10] 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 retarded potentials of electromagnetism, randomness of the radiative process, quantum mechanics, CP violation, fluctuations, cellular automata and even gravity. We now will shortly discuss 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: "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?" [11].

His personal dissatisfaction with lack of irreversibility and unitary evolution at the microscopic level led him to postulate the existence of microscopic representation of entropy in the equations of motion. In essence, the idea was to change the fundamental laws of physics as to make them irreversible at the microscopic level. The details of this approach, known in literature as Misra-Prigogine-Courbage theory[9], will not be presented here, as in so far no evidence for its validity has been found[12].

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 disagrees. 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 explanation[13].

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

[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

manifestations of the irreversibility - such as friction - are clearly macroscopically large.

A different school of thought [**?**][14], 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 to 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.

The practice of MaxEnt approach makes frequent use of the formula for maximization of relative entropy (negative Kullback-Leibler divergence), which uncovers it's essence,

$$H(p \vee q) = -\sum_i p_i \ln \frac{p_i}{q_i} \tag{1}$$

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.

2.6 Physics of computation

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 1970s, Charles Bennett pointed out that it is not the case[15], laying some early groundwork for relating computational and thermodynamic ideas.

3 Treatments of entropy in the standard contexts

3.1 Gibbs entropy

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

$$S_G = k_B \int \rho \log \rho, \tag{2}$$

and generalizes both results of Boltzmann: Boltzmann H, which is useful only for description of systems of non-interacting molecules[16] 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 volume of microstates or measure of our degree of ignorance as to the true unknown microstate.

Moreover, it can also be demonstrated[16] that the change of Gibbs entropy over a reversible path is equal to Clausius entropy:

$$\Delta S_G = \int_1^2 \frac{d\langle K + B\rangle + \langle P\rangle d\Omega}{T} \quad = \frac{dQ}{T}. \tag{3}$$

Here 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]. This alleged arbitrariness of this procedure was subject to critique [2].

One can propose a thought experiment, leading to paradox and immediately resolving it to see the case more clearly. Consider a case of an simple gas 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 localize in the volume $L/2$, we would probably not 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. In this scenerio, the final phase space is of course larger than the initial one and Gibbs entropy increases, as expected.

The means of practical use of still non-coarse-grained entropy in closed systems subject to adiabatic change was explained by MaxEnt approach of Jaynes[16]. 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 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 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.

Moreover, even the exact parameters of entropy depend on the situation, so does their number. We can increase their number as far as we wish and in doing so, we move toward classical deterministic description and the notion of entropy collapses (this is not the case for von Neumann entropy).

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

$$S_{vN} = -\operatorname{Tr}(\hat{\rho} \log \hat{\rho}). \tag{4}$$

For which the general form of the density matrix operator is

$$\hat{\rho} = \sum_k p_k \left|\psi_k\right\rangle \left\langle\psi_k\right| \tag{5}$$

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

$$\hat{\rho} = \left|\psi\right\rangle \left\langle\psi\right| \tag{6}$$

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 therefore do not change with time.

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 generalization 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

$$|2,0,0,...\rangle \tag{7}$$

which entropy is of course zero. In 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

$$\hat{\Psi}(x)\,|2,0,0,...\rangle = \phi_1(x)\,|1,0,0,...\rangle\,, \tag{8}$$

which entropy is zero. It is important to notice though that our system lost a particle and therefore the systems before and after measurement are not equivalent! 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 state using a linear combination of states with *any* value of momentum:

$$c_1\ \ |2,,0,...\rangle + c_2\ \ |1,1,0,...\rangle + c_3\ \ |1,0,1,...\rangle + ... \tag{9}$$

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 entropy of course 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. $ket1,1,...$ 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. However 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

$$\hat{\rho}^A = \mathrm{Tr}_B\,\hat{\rho} \tag{10}$$

then in general the following identities are satisfied

$$\begin{aligned} S(\rho) &\leq S_A + S_B \\ S(\rho) &\geq |S_A - S_B| \end{aligned} \tag{11}$$

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 [in progress]

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 there are obeyed. The resulting dynamics are quite generally termed hydrodynamics in condensed-matter physics, in analogy to the dynamic equations which are valid in this limit for the flow of gases and liquids.

Its usage can be usually justified by assuming analyticity of thermodynamic state functions arbitrarily close to equilibrium - then, local equilibrium is obtained from first order expansion of thermodynamic properties in the irreversible fluxes $\{X_i\}$ [17].

An approached pioneered by Onsager, for the entropy for open systems, is an extension of Clausius entropy for isolated systems:

$$dS = d_i S + d_e S \tag{12}$$

Where $d_i S$ is connected with entropy produced within the system and $d_e S$ is the entropy transferred across the boundaries of the system.

One then develops an explicit expression for entropy production, assuming that even outside equilibrium (but near) entropy depends only on the same variables as at equilibrium ("local" equilibrium)

$$P = \frac{d_i S}{dt} = \sum_\rho J_\rho X_\rho \geq 0 \tag{13}$$

where J_ρ are the rates of the various irreversible processes involved (chemical reactions, heat flow, diffusion...) and X_ρ are the corresponding, generalized forces (affinities, gradients of temperature, of chemical potentials...).

In near equilibrium regime, where local thermodynamic equilibrium is expected to be valid, the theory predicts that there will be a 'spontaneous production of entropy' in non-equilibrium systems. This spontaneous production of entropy is given by the entropy production per unit volume σ by the following expression [18]

$$\int d\boldsymbol{r}\sigma(\boldsymbol{r}, t) = \int d\boldsymbol{r}(\sum_i J_i(\boldsymbol{r}, t) X_i(\boldsymbol{r}, t)) > 0, \tag{14}$$

where $J_i(\boldsymbol{r}, t)$ are the Navier-Stokes hydrodynamic fluxes (e.g. the stress tensor, heat flux vector,...) at position $\boldsymbol{r}$ and time t and X_i is the thermodynamic force which is

conjugate to $J_i(\boldsymbol{r}, t)$ (e.g. strain rate tensor divided by temperature or the gradient of the reciprocal of temperature,... respectively).

Problems: 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 example highlights a serious problem for linear, irreversible thermodynamics based on the concept of entropy production. As we will later see, this is not the case for *dissipation*.

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

$$\delta X(\lambda f(t), t) = \lambda \delta X(f(t), t) \tag{15}$$

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 $X = \delta X$):

$$C(t) = \langle X(t)X(0) \rangle = \frac{\mathrm{Tr}\{X(t)X(0)\rho_{eq}\}}{\mathrm{Tr}\{\rho_{eq}\}}. \tag{16}$$

where ρ_{eq} is the equilibrium density function.

With 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:

$$|C(t)| \leq C(0) \tag{17}$$

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

$$\lim_{t \to \infty} C(t) = 0. \tag{18}$$

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[19], thus leading to [4]:

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

$$\langle X_i(t+\tau)X_j(t)\rangle = \langle X_i(t-\tau)X_j(t)\rangle = \langle X_i(t)X_j(t+\tau)\rangle \tag{19}$$

Then dividing τ and going with it to the limit $\tau \to 0$ we obtain

$$\langle \dot{X}_i(t)X_j(t)\rangle = \langle X_i(t)\dot{X}_j(t)\rangle \tag{20}$$

Now one might perform an analysis from 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 i.e. $t > 0$.

From experience one then postulates a set of phenomenological coupled equations bringing the system back to equilibrium state:

$$\dot{\bar{X}}_i = -\sum_j \lambda_{ij}\bar{X}_j \tag{21}$$

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 (see 3.1):

$$P \propto \exp\left(\frac{S(\bar{X}_1, ..., \bar{X}_N) - S_0}{k_B}\right) \tag{22}$$

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

$$S - S_0 = -\sum_{ij} S_{ij}\bar{X}_i\bar{X}_j \tag{23}$$

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

$$F_i = -\frac{\partial S}{\partial \bar{X}_i} = \sum_j S_{ij}\bar{X}_j \tag{24}$$

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

$$\bar{X}_j = \sum_i (S^{-1})_{ji}F_i \tag{25}$$

Inserting those back to equation (21) one gets

$$\dot{\bar{X}}_i = -\sum_j \lambda_{ij}\sum_k (S^{-1})_{jk}F_k = \sum_k \gamma_{ik}F_k \tag{26}$$

4.3 Onsager relations and hypothesis

In 1931, Onsager[20] shown that γ_{ik} from the previous paragraph is in fact symmetric.

We can now do just that by combining equation (20) with equation (26), thus obtaining Onsager relations

$$\gamma_{ij} = \gamma_{ji}. \tag{27}$$

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[5] term $H' = -fX$. Let's now consider the expression for $\bar{X}$ for time $t < 0$:

$$\bar{X}(0) = \frac{\mathrm{Tr}\{X(0)e^{-\beta(H-fX)}\}}{\mathrm{Tr}\{e^{-\beta(H-fX)}\}} \approx \beta f \langle X(0)X(0)\rangle = \beta f C(0) \tag{28}$$

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

$$\bar{X}(t) = \frac{\langle X(t)e^{-\beta(H-fX)}\rangle}{\langle e^{-\beta(H-fX)}\rangle} \approx \beta f \langle X(t)X(0)\rangle = \beta f C(t) \tag{29}$$

Onsager hypothesis can now be seen as simply

$$\frac{\bar{X}(t)}{\bar{X}(0)} = \frac{C(t)}{C(0)} \tag{30}$$

As a practical note on application of Onsager relations, we quote Charles Kittel [21]:

"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 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

$$\bar{Y} = \frac{\mathrm{Tr}\{Ye^{-\beta(H-fX)}\}}{\mathrm{Tr}\{e^{-\beta(H-fX)}\}} = \mathrm{Tr}\{Ye^{-\beta(H-F-fX)}\} \tag{31}$$

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

where F denotes the free energy coming from the partition function. This linear response defines the static isothermal susceptibility χ^T_{BA} by[6]:

$$\bar{Y} = \chi^T_{YX} f \tag{32}$$

One then uses the following identity[22]:

$$e^{\beta(a+b)} = e^{\beta a}(1 + \int_0^\beta d\lambda \; e^{-\lambda a} b e^{\lambda(a+b)}), \tag{33}$$

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

$$\begin{aligned} \bar{Y} &= \int_0^\beta d\lambda \; \mathrm{Tr}\{Y e^{-\lambda(H-F)} X e^{\lambda(H-F-fX)}\} f \approx \int_0^\beta d\lambda \; \mathrm{Tr}\{Y e^{-\lambda(H-F)} X e^{\lambda(H-F)}\} f \\ &= \langle YX \rangle f \end{aligned} \tag{34}$$

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 case:

$$\chi^T_{YX} = \langle YX \rangle. \tag{35}$$

4.5 Steady states [in progress]

The most basic of non-equilibrium conditions, the steady state is already difficult to describe as many of the basic state functions (including temperature and entropy), are undefined for far from equilibrium states and the distribution function of a steady state is fractal and non-analytic. One however can impose a formal 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 $p(\boldsymbol{f})$:

$$\begin{aligned} \langle \frac{\partial \rho}{\partial t} \rangle_{p(\boldsymbol{\rho})} &= 0 \\ \langle \boldsymbol{f} \rangle_{p(\boldsymbol{f})} &\neq 0 \end{aligned} \tag{36}$$

4.6 Definition of temperature [in progress]

When the temperature differences are "smooth" enough, i.e., locally there is a reasonable definition of temperature (local equilibrium), then the temperature gradient determines the heat flux. In the opposite case, it is molecular kinetics who determines the energy transfer. The latter happens much faster and local equilibrium gets established quickly.

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

On the other hand, far from equilibrium there might be a problem defining temperature and also Clausius entropy which depends on it. One of the solutions provided by Evans at al in is to define temperature of non equilibrium state by the temperature of the underlying equilibrium state to which the system would otherwise relax.

4.7 MinEP

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

$$S = S_0 + \delta S + \frac{1}{2}\delta^2 S \tag{37}$$

This quantity is than used as a Lyapunov function and has 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 steady states near-equilibrium. It is only near-equilibrium that the quantity $(\delta^2 S)_{ss}$ generates probability of fluctuations, as Prigogine insisted in response to criticism [23].

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 structures, called poetically "dissipative structures".

An often given example of instabilities leading to formation of structures are the Rayleigh-Bénard convection cells, which simplified non-equilibrium (not 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 consider a simple model consisting of three elements: the cooler C, the heater H and the 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 picture 1.

Treating the heater and the cooler as the environment, we can think of our system S as an open. Further on we'll analyze the system S from the perspective of internal (i)

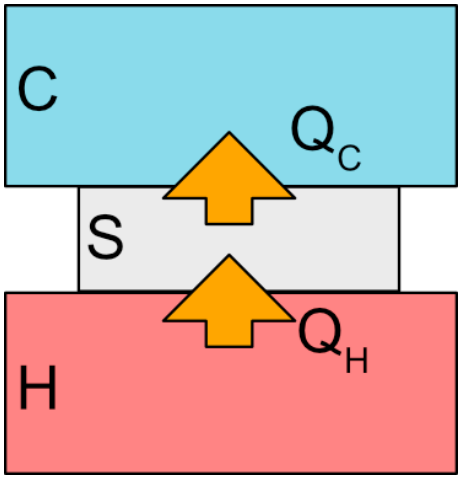

Figure 1: System (S) model

entropy production and external (e) entropy flux, flowing *to* the system S. Of course the change in entropy will be the sum of those two contributions:

$$dS_S = dS_i + dS_e. \tag{38}$$

In the current analysis let's consider a 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:

$$dS_e = \frac{dQ_H}{T_H} + \frac{dQ_C}{T_C} = dQ_H \left(\frac{1}{T_H} - \frac{1}{T_C} \right) = dQ_H \left(\frac{T_C - T_H}{T_H T_C} \right) < 0. \tag{39}$$

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:

$$\sigma_e \equiv \frac{dS_e}{dt}. \tag{40}$$

In the considered scenerio, the σ_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:

$$\sigma_i \equiv \frac{dS_i}{dt}. \tag{41}$$

When $T_H = T_C$, i.e. the system is in equilibrium with constant entropy S_{EQ} then it follows that $\sigma_i = 0$. Therefore the rate of internal entropy production σ_i should be a function of system's entropy S_S, i.e. $\sigma_i = \sigma_i(S_S)$ with the boundary condition $\sigma_i(S_S = S_{EQ}) = 0$.

Near the equilibrium state $S_S = S_{EQ}$, we can Taylor expand the function $\sigma_i(S_S)$ to it's linear term

$$\sigma_i(S_S) = \sigma_i\left(S_{EQ}\right) + \left(S_S - S_{EQ}\right) C_1 + \mathcal{O}\left(S_S^2\right), \tag{42}$$

fulfilling $\sigma_i\left(S_{EQ}\right) = 0$.

The dimensional and stability analysis tells us that C_1 has the dimension of inverse time and in the case of $\sigma_e = 0$ 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 (38) we get

$$\frac{dS_S}{dt} = \sigma_i\left(S_S\right) = \left(S_S - S_{EQ}\right) C_1, \tag{43}$$

The solution of the equation (43) is then

$$S_S(t) = S_{EQ} + (S_0 - S_{EQ}) e^{-t/\tau}, \tag{44}$$

where the initial condition was set $S_S(0) = S_0$.

Now we include the term σ_e into our considerations. In this case the equation (38) results in the following

$$\frac{dS_S}{dt} = \sigma_e + \sigma_i\left(S_S\right) = \sigma_e + \frac{S_{EQ} - S_S}{\tau}. \tag{45}$$

Given a boundary condition $S_S(0) = S_{EQ}$ it has a solution

$$S_S(t) = S_{EQ} + \sigma_e \tau \left(1 - e^{-t/\tau}\right), \tag{46}$$

where σ_e is a negative constant (graph of this function is presented on 3). In the limit $t \to \infty$ the entropy of the system falls to the minimal value

$$S_{min} = S(t \to \infty) = S_{EQ} + \sigma_e \tau < S_{EQ}. \tag{47}$$

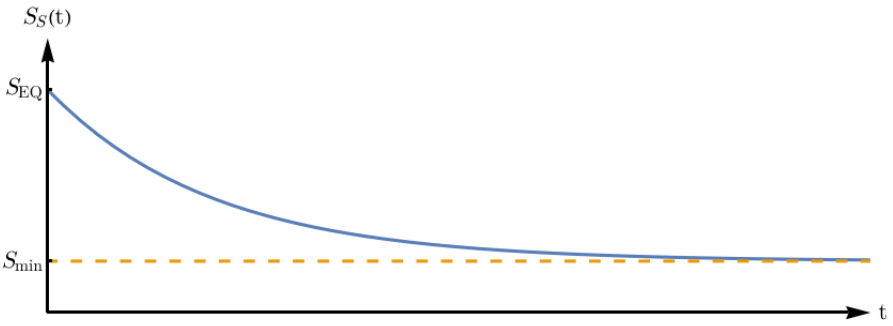

Figure 2: Inducing lower entropy with heat flow.

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 σ_e without the non-equilibrium condition would not form.

6 Measure of irreversibility and the Second Law

In the following section we present a general protoplast of the Fluctuation Theorem. It is derived using nothing else than simple probability calculus and the hypothesis of equal apriori probabilities. In fact, one can obtain from it the exact form of the entropy produced in terms of microscopic transition probabilities for macroscopic objects.

Let's consider a generic system statistical mechanical system a two times, an initial time t_0 and a 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 macrostate consists of some number of corresponding microstates denoted by M_i for the initial macrostates and N_j for the final macrostates. 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.

The probability of the forward transition $P(B_j|A_i)$ is then equal to

$$P(B_j|A_i) = \frac{K_{ij}}{M_i} \tag{48}$$

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

$$P(A_i|B_j) = \frac{P(B_j|A_i)P(A_i)}{P(B_j)}, \tag{49}$$

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 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

$$\sum_i P(A_i|B_j) = 1 \tag{50}$$

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

$$\begin{aligned} P(A_i|B_j) &= \frac{K_{ij}}{M_i} P(A_i) \left(\sum_k \frac{K_{kj}}{M_k} P(A_k) \right)^{-1} \\ &= \frac{K_{ij}}{M_i} \left(\sum_{k,m} \frac{K_{kj}}{K_{km}} \right)^{-1} \end{aligned} \tag{51}$$

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 stochasticly-statistical 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:

$$\frac{P(B_j|A_i)}{P(A_i|B_j)} = e^{\ln \sum_{k,m} \frac{K_{kj}}{K_{km}}}. \tag{52}$$

The interpretation of the term in the exponent, can be done by noticing that there's only one known macroscopic quantity[7] that is strictly positive and can change signs after reversing the left side, namely the standard measure of irreversibility - entropy[8].

We have thus, described 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. Also this is perhaps the most straight-forward argument against Loschmidt.

With the use real, positive random matrices, satisfying the conditions $\sum_j K_{kj} = 1$ for any k, to obtain the distributions for the possible values of entropy produced in the transition $S = \ln \sum_{k,m} \frac{K_{kj}}{K_{km}}$. The results, for $N = 10000$ random matrices are presented on figure 3.

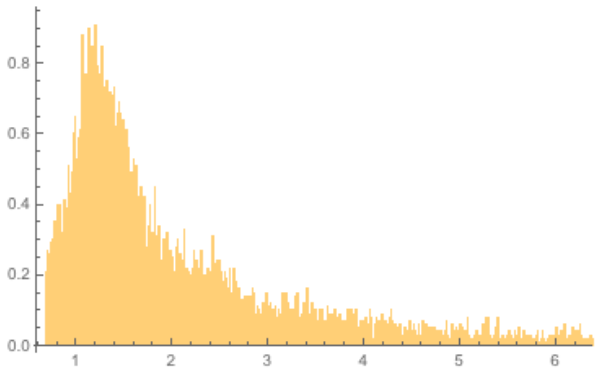

Figure 3: Distribution of entropy produced for $N_A = 2$ and $N_B = 2$

A remarkable feature of this results is that the obtained distributions are not gaussian and that most likely value is proportional to the number of states.

[7]Up to a numerical factor - k_B^{-1}

[8]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.

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 on final macrostate the analysis would collapsed to a time-reversable case. 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 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 line of attack 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 [24], optically trapped colloids [25], shearing systems [26], pendulums [27], molecular motors [28] and various quantum mechanical systems [29].

There exist many approaches for deriving fluctuation theorems, which can be roughly described as deterministic[17][2] or stochastic [30]

In the deterministic framework of Evans, irreversibility finds its origins in non-linear terms which provide a contraction of phase space, in contrast to the more direct irreversibility of the equations of motion found in stochastic descriptions which also has the merit of fewer technical difficulties [31].

Those technical difficulties have their source in ergodicity. 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. 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. The Fluctuation Relation in the deterministic case holds only if the system has certain 'ergodic' properties[32].

Despite this difficulties the deterministic approach of Searles and Evans will be described for reasons of generality, for aesthetic reasons 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 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 first time-reversible, deterministic thermostats and ergostats (homogeneous thermostats) were invented in the early 1980s by Hoover, Ladd and Moran. Prior to this development there was no satisfactory mathematical way of modelling thermostatted non-equilibrium steady states[33].

The construction of thermostated time reversible systems is usually done inserting some time-reversible, but non-Hamiltonian terms into part of 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 in the remote boundaries.

The construction of deterministic thermostats is usually accomplished by including a term $-S_i \alpha \boldsymbol{p}_i$ to the EOM. This 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) through introduction of an extra degree of freedom described by α.

7.1.2 Time reversibility

Let's consider an isolated Hamiltonian system of interacting particles. In the microscopic picture the systems phase space $\{\boldsymbol{q_1}, ..., \boldsymbol{q_N}, \boldsymbol{p_1}, ..., \boldsymbol{p_N}\} \equiv (\boldsymbol{q}, \boldsymbol{p}) \equiv \boldsymbol{\Gamma}$, $\boldsymbol{q_i}, \boldsymbol{p_i}$ denoting position and conjugate momenta of the particle i evolve according to Hamiltonian equations of motion

$$\begin{aligned} \dot{q}_i &= \frac{\partial H(\boldsymbol{q}, \boldsymbol{p})}{\partial \boldsymbol{p_i}} \\ \dot{p}_i &= -\frac{\partial H(\boldsymbol{q}, \boldsymbol{p})}{\partial \boldsymbol{p_i}.} \end{aligned} \tag{53}$$

We now define a time reversal mapping M^T to be an operator acting on phase space (the brackets indicate that the operator acts here on the phase exclusively)

$$M^T[\Gamma] \equiv (\boldsymbol{q}, -\boldsymbol{p}) \tag{54}$$

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

$$\dot{\boldsymbol{\Gamma}} \equiv iL(\boldsymbol{\Gamma})\boldsymbol{\Gamma} \tag{55}$$

which is given by

$$S^t\mathbf{\Gamma} \equiv exp[iL(\mathbf{\Gamma})t]\mathbf{\Gamma}. \tag{56}$$

we will further refer to $exp[iL(\mathbf{\Gamma})t]$ as the p-propagator or the phase space propagator. An easy to check property useful in later part of this work is

$$\frac{d}{dt}(S^t\mathbf{\Gamma}) = iL(\mathbf{\Gamma})\exp[iL(\mathbf{\Gamma})t]\mathbf{\Gamma} = S^t\dot{\mathbf{\Gamma}} \tag{57}$$

The time reversal dynamics satisfies an easy to check equation (with a simple physical intuition behind it)

$$M^T S^t M^T S^t \ \mathbf{\Gamma} = \mathbf{\Gamma} \tag{58}$$

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

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 $p(\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}(t),t))$, that a phase $\mathbf{\Gamma}$, will be observed within an infinitesimal phase space volume of size $\delta V_{\mathbf{\Gamma}}$ about $\mathbf{\Gamma}(t)$ at time t, is given by,

$$p(\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}(t),t)) = f(\mathbf{\Gamma}(t),t)\delta V_{\mathbf{\Gamma}}. \tag{59}$$

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

$$\langle A(t)\rangle = \lim_{N_{\mathbf{\Gamma}}\to\infty}\sum_{j=1}^{N_{\mathbf{\Gamma}}} A(S^t\mathbf{\Gamma}_j)/N_{\mathbf{\Gamma}} \tag{60}$$

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

$$\langle A(t)\rangle = \int d\mathbf{\Gamma}A(\mathbf{\Gamma})f(\mathbf{\Gamma};t) = \int d\mathbf{\Gamma}A(S^t\mathbf{\Gamma})f(\mathbf{\Gamma};0) \tag{61}$$

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

$$\langle A(t)\rangle = \langle A(t+\Delta)\rangle \tag{62}$$

for any $\Delta > 0$.

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.

$$\lim_{t\to\infty}\langle A(t)\rangle = \lim_{t\to\infty}\frac{1}{t}\int_0^t ds\ A(S^s\mathbf{\Gamma}) \tag{63}$$

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 (in later introduced) dissipation function.

7.1.3 Liouville equation

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

$$\frac{df(\mathbf{\Gamma};t)}{dt} = -f(\mathbf{\Gamma};t)\frac{\partial}{\partial\mathbf{\Gamma}}\cdot\dot{\mathbf{\Gamma}}(\mathbf{\Gamma}) = -f(\mathbf{\Gamma};t)\Lambda(\mathbf{\Gamma}) \tag{64}$$

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

$$\frac{\partial f(\mathbf{\Gamma};t)}{\partial t} = -\frac{\partial}{\partial\mathbf{\Gamma}}\cdot[\dot{\mathbf{\Gamma}}(\mathbf{\Gamma})f(\mathbf{\Gamma};t)] = -(\frac{\partial}{\partial\mathbf{\Gamma}}\cdot\dot{\mathbf{\Gamma}} + \dot{\mathbf{\Gamma}}\cdot\frac{\partial}{\partial\mathbf{\Gamma}})f(\mathbf{\Gamma};t) \tag{65}$$

where by moving the last term to the left side we get back equation (64).

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

$$\Lambda \equiv (\partial/\partial\mathbf{\Gamma})\cdot\dot{\mathbf{\Gamma}} = 0 \tag{66}$$

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 $\Lambda = -(3N_{th}-4)\alpha$.

If we make the following substitution $\mathbf{\Gamma}\to S^t\mathbf{\Gamma}$ in equation (64) this first-order ordinary differential equation is solved by

$$f(S^t\mathbf{\Gamma};t) = \exp\left[-\int_0^t ds\ \Lambda(S^s\mathbf{\Gamma})\right]f(\mathbf{\Gamma};0) \tag{67}$$

The measure of an infinitesimal phase space volume $dV_{\mathbf{\Gamma}}(S^s\mathbf{\Gamma})$ 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):

$$dV_{\mathbf{\Gamma}}(S^t\mathbf{\Gamma}) = \exp\left[\int_0^t ds\ \Lambda(S^s\mathbf{\Gamma})\right] dV_{\mathbf{\Gamma}}(\mathbf{\Gamma}) \tag{68}$$

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

7.1.4 Dissipation

The dissipation function serves as mathematical replacement for the 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 can a new quantity is that in some cases dissipation can be negative, but strictly speaking entropy interpreted in the light of information theory, should be positive. This should not come up as a surprise since we're getting closed to scales at which the notion of thermodynamic entropy emerge.

Dissipation function was first properly (not implicitly) defined in 2000 by Searles and Evans[34]. The dissipation function 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 $\delta V_{\mathbf{\Gamma}}$ and $\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}^*) \equiv \delta V_{\mathbf{\Gamma}}(M^T S^t \mathbf{\Gamma})$:

$$\frac{p(\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}; 0))}{p(\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}^*; 0))} = \frac{f(\mathbf{\Gamma}; 0)\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma})}{f(\mathbf{\Gamma}^*; 0)\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}^*)} \tag{69}$$

now by noting that the Jacobian for the time reversal map is unity,
$\delta V_{\mathbf{\Gamma}}(M^T\mathbf{\Gamma}^*)/\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}^*) = 1$ together with equation (68) we get

$$\frac{p(\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}; 0))}{p(\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}^*; 0))} = \frac{f(\mathbf{\Gamma}; 0)}{f(\mathbf{\Gamma}^*; 0)} \exp\left[-\int_0^t ds\ \Lambda(S^s\mathbf{\Gamma})\right] \tag{70}$$

where the logarithm of the right side we'd like to use as the definition of the time integral of dissipation Ω:

$$\int_0^t ds\ \Omega(S^s\mathbf{\Gamma}) \equiv \ln\left(\frac{f(\mathbf{\Gamma}; 0)}{f(\mathbf{\Gamma}^*; 0)}\right) - \int_0^t ds\ \Lambda(S^s\mathbf{\Gamma}) \equiv \Omega_t(\mathbf{\Gamma}) \tag{71}$$

One should perhaps underline that this is the place in which we postulate that some time-asymmetry takes place. Usually one defines an auxiliary quantity $\bar{\Omega}_t$ called *time-averaged dissipation* defined through relation $\Omega_t(\mathbf{\Gamma}) \equiv \bar{\Omega}_t(\mathbf{\Gamma})t$.

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.

7.1.5 Evans-Searles Fluctuation Theorem

If we now choose our volume elements 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 we get the Evans-Searles Fluctuation Theorem (ESFT):

$$\frac{p(\bar{\Omega}_t = A)}{p(\bar{\Omega}_t = -A)} = \exp[A\ t]. \tag{72}$$

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 are required. It was also verified experimentally, see Wang(2002)[35], Carberry(2007)[25]

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

Second Law Inequality 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.

$$\langle \Omega_t \rangle \geq 0,\ \forall t > 0 \tag{73}$$

The proof follows from simple integration of equation (72):

$$\begin{aligned}
\langle \Omega_t \rangle &= \int_{-\infty}^{\infty} dB\ p(\Omega_t = B)B \\
&= \int_{0}^{\infty} dB\ p(\Omega_t = B)B + \int_{\infty}^{0} dB\ p(\Omega_t = B)B \\
&= \int_{0}^{\infty} dB\ p(\Omega_t = B)B - \int_{0}^{\infty} dB\ p(\Omega_t = -B)B \\
&= \int_{0}^{\infty} dB\ p(\Omega_t = B)B(1 - \exp[-B]) \geq 0
\end{aligned} \tag{74}$$

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:

$$\bar{\Omega}_{eq,t}(\mathbf{\Gamma}) = 0, \forall \mathbf{\Gamma} \in D, \forall t > 0 \Rightarrow \langle \Omega_t \rangle = 0, \forall t > 0 \tag{75}$$

Kawasaki identity also known as Non-equilibrium Partition Identity (NPI) was first implied for Hamiltonian systems by Yamada and Kawasaki (1967)[36] and is stated as:

$$\langle \exp\left[-\bar{\Omega}_t t\right] \rangle = 1 \tag{76}$$

This result can also be derived from ESFT given by equation (72):

$$\begin{aligned} \langle \exp\left[-\bar{\Omega}_t t\right] \rangle &= \int_{-\infty}^{\infty} dA \; p(\bar{\Omega}_t = A) \exp[-At] \\ &= \int_{-\infty}^{\infty} dA \; p(\bar{\Omega}_t = -A) \\ &= \int_{-\infty}^{\infty} dA' \; p(\bar{\Omega}_t = A') = 1 \end{aligned} \tag{77}$$

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 $f(\mathbf{\Gamma}; 0)$ and their initial time has to be the same.

One could get an equation independent of this initial time by differentiation of equation (71):

$$\begin{aligned} \frac{\partial}{\partial t} \int_0^t ds \; \Omega(S^s\mathbf{\Gamma}) &= \Omega(S^t\mathbf{\Gamma}) \\ &= \frac{\partial}{\partial t}[\ln f(\mathbf{\Gamma}; 0) - \ln f(S^t\mathbf{\Gamma}; 0) - \int_0^t ds \; \Lambda(S^s\mathbf{\Gamma})] \\ &= -\frac{1}{f(S^t\mathbf{\Gamma}; 0)} \frac{\partial f(S^t\mathbf{\Gamma}; 0)}{\partial t} - \Lambda(S^t\mathbf{\Gamma}) \\ &= -\frac{1}{f(S^t\mathbf{\Gamma}; 0)} \frac{\partial (S^t\mathbf{\Gamma})}{\partial t} \frac{\partial f(S^t\mathbf{\Gamma}; 0)}{\partial (S^t\mathbf{\Gamma})} - \Lambda(S^t\mathbf{\Gamma}) \\ &= -\frac{1}{f(S^t\mathbf{\Gamma}; 0)} S^t\dot{\mathbf{\Gamma}} \frac{\partial f(S^t\mathbf{\Gamma}; 0)}{\partial (S^t\mathbf{\Gamma})} - \Lambda(S^t\mathbf{\Gamma}) \end{aligned} \tag{78}$$

where the last one was obtained using equation (57). If we now set $t = 0$ we obtain the expression for the *instantaneous dissipation function*:

$$\Omega(\mathbf{\Gamma}) = -\frac{1}{f(\mathbf{\Gamma}; 0)} \dot{\mathbf{\Gamma}}(\mathbf{\Gamma}) \frac{\partial f(\mathbf{\Gamma}; 0)}{\partial \mathbf{\Gamma}} - \Lambda(\mathbf{\Gamma}) \tag{79}$$

7.1.7 Dissipation Theorem

Starting from the solution of the Lagrangian form of the Liouville equation (67) we can use dissipation function equation (71) to derive

$$
\begin{aligned}
f(S^t\mathbf{\Gamma};t) &= \exp\left[-\int_0^t ds\ \Lambda(S^s\mathbf{\Gamma})\right] f(\mathbf{\Gamma};0) \\
&= \exp\left[-\int_0^t ds\ \Lambda(S^s\mathbf{\Gamma})\right] f(S^t\mathbf{\Gamma};0)\exp\left[\int_0^t ds\ \Omega(S^s\mathbf{\Gamma})+\int_0^t ds\ \Lambda(S^s\mathbf{\Gamma})\right] \\
&= f(S^t\mathbf{\Gamma};0)\exp\left[\int_0^t ds\ \mathbf{\Omega}(S^s\mathbf{\Gamma})\right],
\end{aligned}
\tag{80}
$$

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

$$
f(\mathbf{\Gamma};t) = f(\mathbf{\Gamma};0)\exp\left[\int_0^t ds\ \mathbf{\Omega}(S^{-s}\mathbf{\Gamma})\right], \tag{81}
$$

which states that the forward in time propagator for the N-particle distribution function is given by the exponential time integral of the dissipative function. Important thing to note that this equation applies to systems with no field or a constant field, the case for time dependent field was presented in [37].

This is true to no field or a constant field - Beyond the second law page 37

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: $f_{ne}(\mathbf{\Gamma};t)$ and cannot be written in a closed, time-stationary form - contrary to statements found in Jaynes (1980) As with ESFT, this result can be applied to any initial ensemble and time-reversible dynamics satisfying $AI\mathbf{\Gamma}$.

From equation (81) we can calculate non-equilibrium ensemble averages of any physical phase function $B(t)$:

$$
\begin{aligned}
\langle B(t)\rangle &= \int_D d\mathbf{\Gamma}\ B(\mathbf{\Gamma})\exp\left[\int_0^t ds\ \Omega(S^{-s}\mathbf{\Gamma})\right] f(\mathbf{\Gamma};0) \\
&= \langle B(0)\exp\left[\int_0^t ds\ \Omega(S^{-s}\mathbf{\Gamma})\right]\rangle_{f(\mathbf{\Gamma};0)}
\end{aligned}
\tag{82}
$$

Differentiating the last equation with respect to time we get

$$
\begin{aligned}
\frac{d\langle B(t)\rangle}{dt} &= \int_D d\mathbf{\Gamma}\ B(\mathbf{\Gamma})\Omega(S^{-t}\mathbf{\Gamma})f(\mathbf{\Gamma};t) \\
&= \int_D d\mathbf{\Gamma}\ B(S^t\mathbf{\Gamma})\Omega(\mathbf{\Gamma})f(\mathbf{\Gamma};t) \\
&= \langle B(t)\Omega(0)\rangle_{f(\mathbf{\Gamma};0)}
\end{aligned}
\tag{83}
$$

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

$$
\langle B(t)\rangle_{f(\mathbf{\Gamma};0)} = \langle B(0)\rangle_{f(\mathbf{\Gamma};0)} + \int_0^t ds\langle B(s)\Omega(0)\rangle_{f(\mathbf{\Gamma};0)}, \tag{84}
$$

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.

A simple consequence of this theorem can be read of immediately that is, for equilibrium (lack of dissipation) the ensemble averages stay constant.

In systems in which the external field drives the system out of equilibrium in a linear manner (weak field), eq. (84) reduces to 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:

$$\lim_{t\to\infty}\langle A(0)B(t)\rangle_\mu = \langle A\rangle_\mu\langle B\rangle_\mu \tag{85}$$

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

$$\lim_{t\to\infty}\langle A(0)B(t)\rangle = 0 \tag{86}$$

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 it's TTCFs must converge to finite values:

$$\left|\int_0^t ds\langle A(0)B(s)\rangle\right| = const < \infty \tag{87}$$

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

$$\left|\int_0^t ds\langle B(s)\Omega(0)\rangle\right| = const < \infty \tag{88}$$

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 such that the non-equilibrium distribution is of the form

$$f(\mathbf{\Gamma};t) = \exp[-\beta H(\mathbf{\Gamma}) + \lambda(t)g(\mathbf{\Gamma})]Z^{-1} \tag{89}$$

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 says 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.

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

This result is exact arbitrarily far from equilibrium and independent of system size.

7.1.11 Driven systems

Driven systems are a subcategory of non-equilibrium systems which are subject to an external dissipative field F_e. For driven systems we define so-called *dissipative factor*, $[\beta J](\mathbf{\Gamma})]$, using the equation,

$$\Omega(\mathbf{\Gamma}) \equiv -[\beta J](\mathbf{\Gamma})VF_e \tag{90}$$

where V is the volume of the system. Even though in this definition dissipation is a linear functional of the field, we can always hide higher order dependence under F_e.

If the system that is driven was initially at equilibrium, then equation (84) can be rewritten as:

$$\langle B(t)\rangle_{f(\mathbf{\Gamma};0)} = \langle B(0)\rangle_{f(\mathbf{\Gamma};0)} - V\int_0^t ds\ \langle[\beta J](0)B(s)\rangle_{f(\mathbf{\Gamma};0)}F_e \tag{91}$$

and at zero field reduces to Green-Kubo expression for the linear response.

If we consider a simple, nonequilibrium, thermostatted system of volume, V consisting of charged particles driven by an external field F_e and consider a time-average of the current density along a trajectory as $J_{c,t} = \frac{1}{t}\int_0^t J_c(s)\ ds$, the fluctuation relation of equation (72) can then be stated:

$$\frac{p(J_{c,t} = A \pm dA)}{p(J_{c,t} = -A \pm dA)} = \exp[A\beta F_e V t] \tag{92}$$

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 definit sign and the second law of thermodynamics is retrieved. In obtaining there results, nothing is assumed about the form of the distribution of current density (it deos not have to be Gaussian). Moreover, in the weak field limit, the rate of entropy production, $\dot{S}$, is given by linear irreversible thermodynamics as $\dot{S} \equiv \sum \langle J_i \rangle V X_i / T$ where the sum is over the product of all conjugate thermodynamic fluxes, J_i and thermodynamics forces, X_i divided by the temperature of the system, T. The relation stands out simply as:

$$\lim_{F_e \to 0} \dot{S}(t) = k_B \langle \Omega(t) \rangle. \tag{93}$$

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. The change in entropy for a process will be similarly related to the time integral of the dissipation

$$lim_{F_e \to 0} \Delta S = k_B \langle \Omega_t \rangle. \tag{94}$$

7.1.12 Non-equilibrium Steady States

From equation (62) we see that stationarity of a system implies that it's 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 (In fact we saw that in a non-equilibrium stationary state the Gibbs entropy diverges at a constant rate toward negative infinity). The time independent values of physical properties, can however, be dependent on the initial phase $\mathbf{\Gamma}$, if it is not we call it peNESS (physically ergodic non-equilibrium steady state):

$$\lim_{t \to \infty} \langle A(t) \rangle_0 = \lim_{t \to \infty} \frac{1}{t} \int_0^t ds \; A(S^s \mathbf{\Gamma}) \tag{95}$$

where $\langle ... \rangle_0$ denotes an ensemble over the initial time $t = 0$ and ensemble $f(\mathbf{\Gamma}; 0)$. Contrary to intuition, not all NESSs are physically ergodic. An example is the Rayleigh-Benard 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.

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 connect two equilibrium states.

In order to get the connection with Crooks fluctuation theorem, we'll 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

$$\begin{aligned}\exp[\Delta X_\tau(\mathbf{\Gamma})] &\equiv \lim_{\delta V_{\mathbf{\Gamma}}\to 0} \frac{p_{eq,1}(\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma};0))Z(\lambda_1)}{p_{eq,2}(\delta V_{\mathbf{\Gamma}}(S^\tau\mathbf{\Gamma};0))Z(\lambda_2)} \\ &= \frac{f_{eq,1}(\mathbf{\Gamma})d\mathbf{\Gamma}Z(\lambda_1)}{f_{eq,2}(S^\tau\mathbf{\Gamma})d(S^\tau\mathbf{\Gamma})Z(\lambda_2)}\end{aligned} \tag{96}$$

where $Z(\lambda_i)$ is the partition function for the system and 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 it's 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 byt starting from the equilibrium distribution given by $f_{eq,2}(\mathbf{\Gamma})$ of system 2.

The probability that the phase variable ΔX_τ takes the value B for a forward evolved trajectories is given by

$$p_{eq,1}(\Delta X_{\tau,f} = B \pm dB) = \int_{\Delta X_{\tau,f}=B\pm dB} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma}) \tag{97}$$

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

$$p_{eq,2}(\Delta X_{\tau,r} = -B \mp dB) = \int_{\Delta X_{\tau,r}=-B\mp dB} d\mathbf{\Gamma} f_{eq,2}(\mathbf{\Gamma}) \tag{98}$$

Now looking at the ration of those probabilities we get (to simplify the notion we'll suppress $\pm B$ and instead use $+/-$ in the superscript of ΔX)

$$\frac{p_{eq,1}(\Delta X^+_{\tau,f})}{p_{eq,2}(\Delta X^-_{\tau,r})} = \frac{\int_{\Delta X^+_{\tau,f}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma})}{\int_{\Delta X^-_{\tau,r}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,2}(\mathbf{\Gamma})} \tag{99}$$

Now using the definition of generalized work, equation (96), two times first get $f_{eq,2}(\mathbf{\Gamma})d\mathbf{\Gamma} = \exp[\Delta X_{r,\tau}(\mathbf{\Gamma})]f_{eq,1}(S^T\mathbf{\Gamma})d(S^T\mathbf{\Gamma})Z(\lambda_1)/Z(\lambda_2)$ and also (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:

$$
\begin{aligned}
\frac{p_{eq,1}(\Delta X^{+}_{\tau,f})}{p_{eq,2}(\Delta X^{-}_{\tau,r})} &= \frac{\int_{\Delta X^{+}_{\tau,f}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma})}{\int_{\Delta X^{-}_{\tau,r}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,2}(\mathbf{\Gamma})} \\
&= \frac{\int_{\Delta X^{+}_{\tau,f}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma}) Z(\lambda_2)/Z(\lambda_1)}{\int_{\Delta X^{+}_{\tau,f}(M^T S^\tau \mathbf{\Gamma})} \exp\left[-\Delta X^{+}_{\tau,f}(M^T S^\tau \mathbf{\Gamma})\right] d(M^T S^\tau \mathbf{\Gamma}) f_{eq,1}(M^T S^\tau \mathbf{\Gamma})} \\
&= \frac{\int_{\Delta X^{+}_{\tau,f}(\mathbf{\Gamma})} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma}) Z(\lambda_2)/Z(\lambda_1)}{\int_{\Delta X^{+}_{\tau,f}(\mathbf{\Gamma}')} d\mathbf{\Gamma}' \exp\left[-\Delta X^{+}_{\tau,f}(\mathbf{\Gamma}')\right] f_{eq,1}(\mathbf{\Gamma}')} \\
&= \exp[B] \frac{Z(\lambda_2)}{Z(\lambda_1)}
\end{aligned}
\tag{100}
$$

Rewriting this again in full notion

Not sure if that would hold for $\tau \neq 0$

$$
\frac{p_{eq,1}(\Delta X_{\tau,f} = B \pm dB)}{p_{eq,2}(\Delta X_{\tau,r} = -B \mp dB)} = \exp[B] \frac{Z(\lambda_2)}{Z(\lambda_1)} \tag{101}
$$

we obtain the generalized Crooks fluctuation relation (GCFR).

In order to use GCFT we specialize it to an actual statistical mechanical ensemble and system of dynamics, obtaining the canonical forms of CFT between initial and final equilibrium states with the same values of temperature, volume and number of particles (T,V,N). We will consider T-mixing, Nose-Hoover isothermal systems in which the Hamiltonian is subject to a parametric transformation. Equilibrium distribution function $f(\mathbf{\Gamma}; 0)$ and the related free energy $F(\lambda)$ and partition function Z_c are given by

$$
\begin{aligned}
f(\mathbf{\Gamma}; 0) &= Z_c^{-} 1 \exp[-\beta H_0(\mathbf{\Gamma})] \\
F(\lambda) \equiv -k_B T \ln Z_c(\lambda) &= -k_B T \ln[\int d\mathbf{\Gamma} \exp[-\beta H_0(\mathbf{\Gamma}, \lambda)]]
\end{aligned}
\tag{102}
$$

The Hamiltonian given by

$$
H_0(\mathbf{\Gamma}, \lambda(t)) = \sum_{i=1}^{N} \frac{p_i^2}{2m} + \Phi(\boldsymbol{q}, \lambda t) \tag{103}
$$

is varied parametrically from $\lambda_1 = \lambda(0)$ to the final, unique equilibrium state $\lambda_2 = \lambda(\tau)$ (to which it will relax thanks to the property of T-mixing). In systems coupled to additional degrees of freedom, the phase space volume changes and so the work with the Hamiltonian H_E for the extended system becomes

$$
\begin{aligned}
\Delta X_\tau &= \beta(H_E(S^\tau \mathbf{\Gamma}, \lambda(\tau)) - H_E(\mathbf{\Gamma}, \lambda(0))) + \ln\left[\frac{d\mathbf{\Gamma}}{d(S^\tau \mathbf{\Gamma})}\right] \\
&= \beta(H_E(S^\tau \mathbf{\Gamma}, \lambda(\tau)) - H_E(\mathbf{\Gamma}, \lambda(0))) + \int_0^\tau ds\ \Lambda(S^s \mathbf{\Gamma}) \\
&= \beta(H_E(S^\tau \mathbf{\Gamma}, \lambda(\tau)) - H_E(\mathbf{\Gamma}, \lambda(0)) + \Delta Q_\tau) \\
&= \beta \Delta W_\tau
\end{aligned}
\tag{104}
$$

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

Discussion of the extended Hamiltonian + the isothermal result

$$\frac{p_1(\Delta W_\tau = W)}{p_2(\Delta W_\tau = -W)} = exp[\beta(W - \Delta F)] \tag{105}$$

7.3 Jarzynski Equality

In ordinary statistical physics when transitions between two equilibrium states are performed infinitely slowly along some path between the initial point A and the final point B, 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.

In 1996, Christopher Jarzynski[38] derived a very useful equality for the equilibrium free energy differences between two configuration of a system in terms of an ensemble finite-time measurements of the work performed during parametric switching in between those two configurations.

$$\langle \exp(-\beta W) \rangle = \exp(-\beta \Delta F) \tag{106}$$

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. In fact a generalized Jarzynski equality (GJE) follows from:

$$\begin{aligned} \langle \exp[-\Delta X_\tau(\mathbf{\Gamma})] \rangle_{eq,1} &= \int_{-\infty}^{\infty} dB \; p_f(\Delta X_\tau = B) \exp[-B] \\ &= \int_{-\infty}^{\infty} dB \; p_r(\Delta X_\tau = -B) \frac{Z(\lambda_2)}{Z(\lambda_1)} \\ &= \frac{Z(\lambda_2)}{Z(\lambda_1)} \end{aligned} \tag{107}$$

where the brackets $\langle ... \rangle_{eq,1}$ denote an equilibrium ensemble average over the initial equilibrium distribution.

The usual practice is is use the inequality $e^x \geq 1 + x$ to rewrite it in a form of an inequality:

$$\begin{aligned}\frac{Z(\lambda_2)}{Z(\lambda_1)} &= \langle \exp[-\Delta X_\tau] \rangle_1 \\ &= exp[-\langle \Delta X_\tau \rangle_1] \langle \exp[-\Delta X_\tau + \langle \Delta X_\tau \rangle_1] \rangle \\ &\geq exp[-\langle \Delta X_\tau \rangle_1] \langle 1 - \Delta X_\tau + \langle \Delta X_\tau \rangle_1 \rangle \\ &= exp[-\langle \Delta X_\tau \rangle_1]\end{aligned} \tag{108}$$

or taking the logarithm

$$\langle \Delta X_\tau \rangle \geq \ln \left[\frac{Z(\lambda_1)}{Z(\lambda_2)} \right] = \beta \Delta F_{21} \tag{109}$$

here the right side has is the free energy difference. Now it's time to specialize our generalized work with:

$$\Delta X = \beta \int_0^\tau ds \; W(s) \tag{110}$$

where W denotes the work. This inequality implies $\Delta W_{21} \geq \Delta F_{21}$, so the minimum work is expended if the path is reversible or quasi-static, in which case the work is, in fact, the difference in the free energies divided by $k_B T$.

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 implies

$$\oint ds \langle X(s) \rangle = \oint \langle dS \rangle \geq 0 \tag{111}$$

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 we have to complete many cycles until the system settles into a periodic response to the cyclic protocol before we can apply the cyclic integral of the heat. Not all systems do settle into a cyclic response. The reason for this disrepancy is that after the cycle is complete no more work is being done, but the long relaxation process of course includes heat exchange.

8 Application to self-replication and adaptation

Now we wish to zoom out a bit from our theoretical considerations about the foundations of fluctuation theorems and just take them as a fact in some recent and interesting applications[39][40]. The main new truth obtained from them 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.

We proceed by switching to a stochastic description in which the equation (105) takes the form

$$\frac{\pi(j \to i; \tau)}{\pi(i \to j; \tau)} = \langle \exp\left[-\beta \Delta Q^\tau_{i \to j} \right] \rangle_{i \to j}, \tag{112}$$

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.

Now 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[9]:

$$\pi(I \to II) = \int_{II} dj \int_{I} di \; \pi(i \to j) p(i|I) \tag{113}$$

and

$$\pi(II \to I) = \int_{I} di \int_{II} dj \; \pi(j \to i) p(j|II) \tag{114}$$

Now let's investigate the ratio

$$\begin{aligned} \frac{\pi(II \to I)}{\pi(I \to II)} &= \frac{\int_I di \int_{II} dj \; \pi(j \to i) \frac{p(j|II)}{p(i|I)} p(i|I)}{\int_{II} dj \int_I di \; \pi(i \to j) p(i|I)} \\ &= \frac{\int_I di \int_{II} dj \; \pi(i \to j) \langle \exp\left[-\beta \Delta Q^{\tau}_{i \to j}\right] \rangle_{i \to j} \frac{p(j|II)}{p(i|I)} p(i|I)}{\int_{II} dj \int_I di \; \pi(i \to j) p(i|I)} \\ &= \langle \langle e^{-\beta \Delta Q^{\tau}_{i \to j}} \rangle_{i \to j} e^{\ln[\frac{p(j|II)}{p(i|I)}]} \rangle_{I \to II} \end{aligned} \tag{115}$$

where in the last step we made use of equation (112) 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.

One can rewrite the equation (115) as

$$\frac{\pi(II \to I)}{\pi(I \to II)} = \langle e^{-\beta \Delta Q^{\tau}_{i \to j} + \ln[\frac{p(j|II)}{p(i|I)}]} \rangle_{I \to II} \tag{116}$$

remembering that $\Delta Q_{i \to j}$ contains a path ensemble average, and compare it with equation (52) to see the essential difference between those two equations, namely the partial knowledge about microscopic trajectories and their dissipated heat.

Now moving the left side of (116) to the right side and under the ensemble one gets

$$\langle e^{-\beta \Delta Q^{\tau}_{i \to j}} e^{\ln[\frac{p(j|II)}{p(i|I)}]} e^{-\ln[\frac{\pi(II \to I)}{\pi(I \to II)}]} \rangle_{I \to II} \tag{117}$$

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

$$\beta \langle \Delta Q^{\tau}_{i \to j} \rangle_{I \to II} + \langle \ln[\frac{p(i|I)}{p(j|II)}] \rangle_{I \to II} + \ln \frac{\pi(II \to I)}{\pi(I \to II)} \geq 0. \tag{118}$$

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

$$\beta \langle \Delta Q^{\tau}_{i \to j} \rangle_{I \to II} + \ln \frac{\pi(II \to I)}{\pi(I \to II)} + \Delta S_{int} \geq 0. \tag{119}$$

This is a very 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 [40].

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

8.1 Self-replication

The obtained result can be applied to a simple model of self-replicators. Let's suppose we have a master equation for $n \gg 1$ governing the population

$$\dot{p}_n(t) = g\ n\,(p_{n-1}(t) - p_n(t)) - \delta\ n\,(p_n(t) - p_{n+1}(t)) \tag{120}$$

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 dead state with macrostate I, then naturally we assign $\pi(I \to II) = g\ \Delta t$ and $\pi(II \to I) = \delta\ \Delta t$ for some time Δt.

The equation then (119) dictates

$$\Delta S_{int} + \beta\langle \Delta Q^{\tau}_{i\to j}\rangle_{I\to II} \geq \ln \frac{g}{\delta}. \tag{121}$$

which can interpreted as a general bound on self replication. An important thing to notice here is that the ΔS_{int} is expected to be negative, because the self-replicator exists in non-equilibrium, living state.

If one fixes all the terms other than the growth rate, than one can get a bound on the growth rate, by simple algebraic manipulation

$$g \leq g_{max} = \delta \exp\big[\Delta S_{int} + \beta\langle \Delta Q^{\tau}_{i\to j}\rangle_{I\to II}\big]. \tag{122}$$

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 from work done on the system by some external driving field, such as through the absorption of light.

Another comment can be made 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 complexity inner complexity (ΔS_{int}) low.

8.2 Traversal of energy landscape

Let's now consider a case of driven thermostatted system with two possible target macrostates II, III and we are now interested in the propability ratio between those two. From equation (116) we get

$$\ln\frac{\pi(I\to II)}{\pi(I\to III)} = \ln\frac{\pi(II\to I)}{\pi(III\to I)} - ln\frac{\langle e^{-\beta\Delta Q_{i\to j}+\ln\frac{p_f^{II}}{p_i^I}}\rangle_{I\to II}}{\langle e^{-\beta\Delta Q_{i\to k}+\ln\frac{p_f^{III}}{p_i^I}}\rangle_{I\to III}} \tag{123}$$

where the initial and final microstates where noted 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:

$$W_{i\to j} = \Delta Q_{i\to j} + H_* - H_I \tag{124}$$

where the $*$ is the final state indicator, 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:

$$\ln\frac{\pi(I\to II)}{\pi(I\to III)} = \ln\frac{\pi(II\to I)}{\pi(III\to I)} - ln\frac{\langle e^{\beta(H_{II}-H_I)+\ln\frac{p_f^{II}}{p_i^I}}\rangle_{I\to II}}{\langle e^{\beta(H_{III}-H_I)+\ln\frac{p_f^{III}}{p_i^I}}\rangle_{I\to III}} - ln\frac{\langle e^{-\beta W}\rangle_{I\to II}}{\langle e^{-\beta W}\rangle_{I\to III}} \tag{125}$$

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

$$H_* - H_I = -\beta^{-1}(\ln p^*_{eq}Z^*_{eq} - \ln p^I_{eq}Z^I_{eq}) = \beta^{-1}(\ln\frac{p^I_{eq}}{p^*_{eq}} + \ln\frac{Z^I_{eq}}{Z^*_{eq}}) \tag{126}$$

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

$$\ln\frac{\pi(I\to II)}{\pi(I\to III)} = \ln\frac{\pi(II\to I)}{\pi(III\to I)} - ln\frac{\langle\frac{p_f^{II}}{p_{eq}^{II}}\rangle_{II}}{\langle\frac{p_f^{III}}{p_{eq}^{III}}\rangle_{III}} - ln\frac{\langle e^{-\beta W}\rangle_{I\to II}}{\langle e^{-\beta W}\rangle_{I\to III}} + \ln\frac{Z^{II}_{eq}}{Z^{III}_{eq}} \tag{127}$$

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

$$W_d = W - Z^*_{eq} + Z^I_{eq} \tag{128}$$

thus obtaining

$$\ln\frac{\pi(I\to II)}{\pi(I\to III)} = \ln\frac{\pi(II\to I)}{\pi(III\to I)} - ln\frac{\langle\frac{p_f^{II}}{p_{eq}^{II}}\rangle_{II}}{\langle\frac{p_f^{III}}{p_{eq}^{III}}\rangle_{III}} - ln\frac{\langle e^{-\beta W_d}\rangle_{I\to II}}{\langle e^{-\beta W_d}\rangle_{I\to III}} \tag{129}$$

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

$$-\ln\langle\exp(-\beta W_d)\rangle = \beta\langle W_d\rangle - \Phi \tag{130}$$

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 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.[41]

One might now argue that $\langle W_d \rangle$ can be connected with drift through oscillatory energy barriers depending on the model, this in fact recently demonstrated by Nikolay Perunov et al[39].

We have now seen how recent progress in microscale non-equilibrium statistical physics has allowed us to make plausible predictions about inanimate matter. Indeed the fruitfulness of this approach were noticed by other groups who seek to further generalize it - this 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. [42]

This approach has also gained a lot of criticism. According to Kondepudi [43], and to Grandy[44], there is no general rule that provides an extremum principle that governs the evolution of a far-from-equilibrium system to a steady state.

There seems to be a theoretical relationship between maximum irreversibility and dynamic stability, a link suggested here by the fact that MaxEnt/MaxEP predicts the same dissipation functional as Malkus's instability criterion in shear turbulence[42].

9.1 Rayleigh's insight

Studying jets of water from a nozzle, Rayleigh [45] noted that when a jet is in a state of conditionally stable dynamical structure, the mode of fluctuation most likely to grow to its full extent and lead to another state of conditionally stable dynamical structure is the one with the fastest growth rate. In other words, a jet can settle into a conditionally stable state, but it is likely to suffer fluctuation so as to pass to another, less unstable, conditionally stable state. He used like reasoning in a study of Benard convection[46]. These physically lucid considerations of Rayleigh seem to contain the heart of the distinction between the principles of minimum and maximum rates of dissipation of energy and

entropy production, which have been developed in the course of physical investigations on so-called MEP principle by later authors.

9.2 MEP principle

Recently, a common working formulation of the maximum entropy production (MEP) principle surfaced stating roughly that *for systems admitting a spectrum of possible steady states, MEP says that the system is most likely to be found in steady state with the greatest entropy production*

The conjecture of MEP has shown some promising (but controversial) success in studies of planetary climates [47], fluid turbulence [48] [49], crystal growth morphology, biological adaptation as well as earthquake dynamics. The reason main reason that drove controversies around those successes has been an *ad hoc* and unsystematic manner in which MEP was applied.

For example, the earliest successes of MEP 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. At the same time general circulation models (GCM), found no extremum in EP. MEP was also not found in phenomenological models of heat flow in plasma/fluid system [50], where maximum as well as minimum was observed depending on how the system is driven. Nevertheless, possible adjustments to MEP principle are still being researched.

9.2.1 Relation to MinEP

For linear, near-equilibrium systems that only admit a single steady state, MinEP says that all of the system's transient states have a higher entropy production than the steady state. A transient state is a temporary state that is not a steady state. MinEP compares steady states with non-steady states.

For some yet-to-be-determined class of non-linear, far-from-equilibrium systems that admit a continuum of possible steady states, MEP says that the system is most likely to be found in the steady state with the greatest entropy production. MEP compares steady states to other steady states, but says nothing about transient states.

9.2.2 Extrema of the Dissipation Function and MEP

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 consider whether there exists a principle akin to MaxEnt (maximum entropy) for equilibrium systems, and MEP

(maximum entropy production rate) for non-equilibrium systems. There were several papers on the topic, including Williams and Evans [37], concluding that MEP cannot be applied rigorously to non-equilibrium systems in general, as the distribution function at any time (including steady state) is not just a function of the dissipation at that time. However it might provide a good approximation in some cases.

Considering a system that is driven from an initially equilibrium state to a steady-state one finds that the ensemble average of the instantaneous dissipation function stays positive, and that the average of the total dissipation function will approach infinity at long times. Considering equation (84) with $B = \Omega$, one can see *then if* the autocorrelation function $\langle \Omega(\mathbf{\Gamma}(0))\Omega(\mathbf{\Gamma}(t))\rangle$ decays monotonically with time, then the value of the ensemble average of the instantaneous dissipation funciton, $\langle \Omega(t)\rangle$ will be higher when it reaches its steady state than for any other state it passes through. Therefore the system would find itself in a steady-state that maximises the dissipation function (rate of entropy production).

However this is a special case, under the assumption of a monotonous autocorrelation function of the dissipation. Some numerical studies has been performed to study the behaviour of the instantaneous dissipation function more generally; examining whether it is a maximum in the steady state or if a transient state has a higher average instantaneous dissipation value.

In fact it was shown in [51] through numerical simulations, that the dissipation average peaks before dropping to a steady state for most field values. Therefore it is clear that the instantaneous dissipation is not a maximum in the steady state - the system evolves through unstable (transient) maximum. For stronger fields, the steady state reaches a higher instantaneous dissipation which is also maximum.

One should note here that in the light of formulation of MEP from the previous paragraph this result doesn't prove or disprove the MEP principle as it compares steady states to transient states, instead of steady states to steady states. One might conclude however that the supposed theorem of maximum instantaneous dissipation for steady states does not generally apply.

In order to provide further information on the behaviour of the dissipation, one should consider a system where multiple steady state solutions are known to exist. One such model was investigated by Zhang[52] who considered heat flow in a one-dimensional lattice. The two possible steady states exist - a soliton or a diffusive heat flow - and depend on the initial conditions and the field strength. For a given field strength, there is a certain set of trajectories from which a soliton emerge spontaneously. There's also a critical value of the field strength after which the probability of forming a soliton is equal to unity. This chaos-soliton transition becomes sharper as the size of the system (number of particles) is increased. One might think that it may very well be that after increased simulation time the transition becomes sharper as well, but it turns out for zero-field there exists a set of initial conditions (of zero measure) that also forms a soliton. This peculiar points of the phase space form a basin of attraction which grows after the external field is increased.

This non-uniqness of the steady state stands in contrast to other 1D and 2D numerical simulations like [53].

This might suggest that the strong conjecture of MEP might not hold for all steady state or that the field strength should also be considered a constrain of the system. Perhaps, MEP requires the external forcing to be sufficiently large that low entropy states are unstable. It is clear that the results will depend on the constraints imposed, and therefore the problem could be reformulated as a problem in identification of the appropriate constraints.

It is difficult to study MEP numerically at the microscopic level, the main reason for that is because systems that generate multiple steady states, such as convection and turbulent flow, are computationally very expensive. One would then obtain MEP as an objective way of finding the appropriate constrains.

One should perhaps mention, before committing to such enterprise, that there exists a discouraging example seen in the derivation of the dissipation theorem for driven system from an extremum principle [54]. In that case the number of constraints (chaining constrains of the dissipative flux for each time step) required in order to obtain the exact answer turned out to be infinite.

9.2.3 MaxEnt based formulations of MEP

Perhaps the most promising interpretation of MEP principle, put forward by Dewar , assumes that MEP is not a physical principle at all. Instead, by analogy to Jaynes MaxEnt it is to be interpreted as an inference method i.e. a method for deducing the most unbiased predictions from an incomplete set of statistical data.

First step is done by narrowing the scope of validity. We notice that in systems that are weakly driven have (neglecting the set of measure zero) only one steady state available. Therefore there's no room for MEP to operate and we instead focus on systems driven strongly in far for equilibrium regime. 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.

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, we introduce an information theoretical measure of the distance from equilibrium, or *irreversibility* I, defined in terms of the relative probabilities of forward and reverse fluxes. Our first demand (dynamical instability) is then reformulated as a string

inequality constraint $I > I_{min}$. Then using procedures known from MaxEnt we show 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 MEP, entropy production depends of applied constraints.

In the proceeding section we demonstrate this procedure in detail.

9.3 MEP principle as an inference principle

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 $\boldsymbol{f}$, which may in principle be infinite. 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 the described by $\boldsymbol{f}$ (or ρ) with the stationarity condition given by equation (36)

Similarly to the MaxEnt we maximize the relative entropy

$$H = -\int p(\boldsymbol{f}) \ln \frac{p(\boldsymbol{f})}{q(\boldsymbol{f})} d\boldsymbol{f} \tag{131}$$

with respect to $p(\boldsymbol{f})$, subject to given dynamical constraints (assumed relevant dynamics, typically involving a restricted number of stationary conditions) C, where $q(\boldsymbol{f})$ is a prior p.d.f, with the symmetry $p(\boldsymbol{f}) = p(-\boldsymbol{f})$ 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:

$$\int p(\boldsymbol{f})\phi_m(\boldsymbol{f}) d\boldsymbol{f} = 0 \tag{132}$$

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

$$\int p(\boldsymbol{f}) d\boldsymbol{f} = 1. \tag{133}$$

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

$$I = \int p(\boldsymbol{f}) \ln \frac{p(\boldsymbol{f})}{p(-\boldsymbol{f})} d\boldsymbol{f}. \tag{134}$$

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 $\boldsymbol{F} = \boldsymbol{0}$. The irreversibility I is thus a natural

information-theoretic measure of the distance from equilibrium, or time-reversal symmetry breaking, since it measure the extent to which $p(\boldsymbol{f})$ differs from $p(-\boldsymbol{f})$.

Now we make the first step from the previous paragraph, by demanding 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 (**??**);.

We assume that those conditions also determine the upper bound $I \leq I_{max}(C)$.

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

$$\int p(\boldsymbol{f})\boldsymbol{f}d\boldsymbol{f} = \boldsymbol{F}. \tag{135}$$

The motivation for introducing $\boldsymbol{F}$ this way is that $\boldsymbol{F}$ represents a trial estimate of the actual fluxes, $\boldsymbol{F}(C)$ selected under C. Introducing $\boldsymbol{F}$ in this wa allows one to establish an extremal principle whereby $\boldsymbol{F}(C)$ is determined by varying the trial solution $\boldsymbol{F}$. 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.

The MaxEnt solution for $p(\boldsymbol{f})$ is given by

$$p(\boldsymbol{f})^* = q(\boldsymbol{f})Z^{-1}\exp\left[\boldsymbol{\lambda}\cdot\boldsymbol{f} + \boldsymbol{\alpha}\cdot\boldsymbol{\phi}(\boldsymbol{f}) - \mu(d(\boldsymbol{f}) - e^{-d(\boldsymbol{f}))}\right] \tag{136}$$

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 multipliers for (135), (132) and the upper-bound inequality $I < I_{max}$ respectively. The maximized relative entropy is

$$S(\boldsymbol{F}, I_0, C) \equiv H_{max} = \ln Z(\lambda, \boldsymbol{\alpha}, \mu) - \boldsymbol{\lambda}\cdot\boldsymbol{F} + \mu(I_0 - 1) \tag{137}$$

10 Summary

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

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

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

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

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

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

[9] 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.

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

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

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

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

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

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

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

[17] 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.

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

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

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

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[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] 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.

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

[29] 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.

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

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

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

[33] 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.

[34] 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.

[35] 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.

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

[37] 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.

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

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

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

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

[42] 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.

[43] D. Kondepudi, *Introduction to modern thermodynamics*. Chichester: Wiley, 2008.

[44] W. T. Grandy, *Entropy and the Time Evolution of Macroscopic Systems*. International series of monographs on physics, Oxford: Oxford Univ. Press, 2008.

[45] L. Rayleigh, "On The Instability Of Jets," *Proceedings of the London Mathematical Society*, vol. s1-10, pp. 4–13, Nov. 1878.

[46] L. Rayleigh, "LIX.On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side," *Philosophical Magazine Series 6*, vol. 32, no. 192, pp. 529–546, 1916.

[47] 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.

[48] 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.

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

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

[51] 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.

[52] F. Zhang, D. J. Isbister, and D. J. Evans, "Multiple nonequilibrium steady states for one-dimensional heat flow," *Physical Review E*, vol. 64, pp. 1645–5, July 2001.

[53] A. Maeda and T. Munakata, "Lattice thermal conductivity via homogeneous nonequilibrium molecular dynamics," *Physical Review E*, vol. 52, pp. 234–239, July 1995.

[54] D. J. Evans, "Response theory as a free-energy extremum," *Physical Review A*, vol. 32, pp. 2923–2925, Nov. 1985.