Thermodynamics and Information Theory

Kirone Mallick
Institut de Physique Théorique
Université Paris-Saclay, CEA, CNRS
CEA Saclay
91191 Gif-sur-Yvette, France

Abstract. The Laws of Classical Thermodynamics and the fundamental principles of Statistical Mechanics are reviewed with an emphasis on their logical structure. We discuss the universality of equilibrium fluctuations (Brownian Motion) and describe some contemporary results such as Jarzynski and Crooks Non-Equilibrium Work Identities. A brief review of Shannon Entropy reveals the formal analogy between Thermodynamics and Information Theory: these two sciences must be coupled in order to understand and to exorcise Maxwell's demon.

1   Introduction

In his treatise on Thermodynamics [1], R. Kubo mentions a small book by the Czech chemist F. Wald (1861-1930), entitled *The Mistress of the World and her Shadow,* a metaphor alluding to Energy and Entropy. Quoting Robert Emden, Kubo notes that 'in the huge manufactory of natural processes, the principle of entropy occupies the position of manager, for it dictates the manner and method of the whole business, whilst the principle of energy merely does the bookkeeping, balancing credits and debits [2].' Yet, while Energy seems to be familiar to all of us, Entropy remains a mysterious concept, frequently (mis)used in everyday language as a substitute for chaos, noise, disorder, disorganization or even... business inefficiency [3].

Equilibrium statistical mechanics tells us that entropy relates the microscopic realm to the macroscopic world, by enumerating how many micro-configurations of a system are compatible with our sense-data and the measurements performed at our scale. It allows us to quantify the loss of information by coarse-graining from the microscale to the macroscale. To be fair, Entropy should be considered as a source of surprise rather than confusion.

The aim of this article is to recount the subtle ballet of Entropy between Physics and Information Theory, choreographed by the puckish demon imagined by Maxwell in 1867, as an incarnation 'of that force that always wills the evil and always produces the good'. Learning about entropy will entail us to review the Principles of thermodynamics and their underlying statistical basis, with never ending thermal fluctuations, exemplified by the Brownian motion. By modeling nonequilibrium dynamics, we shall relate entropy to stochastic trajectories. This will lead us to the Fluctuation Theorem and to the Nonequilibrium Work Identities. The stage will be set to face information theory, to confront the demon and the various attempts to exorcise him (or her).

2   Thermodynamics: A Brief Review

Thermodynamics describes macroscopic properties of matter (solid, fluid, radiation...) in terms of a small number of macroscopic observables (such as pressure, volume, mass, temperature) assuming that these properties do not vary with time. The Laws of Thermodynamics allow us to derive some general relations amongst these properties irrespective of the structure of matter at the atomic scale. Indeed, the Two Principles were established during the XIXth century before the dawn of atomic physics [4, 5, 6, 7, 8, 9].

Thermodynamics can be viewed as the science of *energy conversions*. In order to establish a correct balance, two guiding principles must be respected:
(i) all forms of energy involved must be identified correctly and accounted for;
(ii) different forms of energy are not equivalent. Some energy conversions are free of cost while others come with a fee and need *compensation* (according to Clausius).

Thermodynamics is one of the most elegant branches of Physics, but it is also notoriously difficult. This feature has been perfectly emphasized by Onsager (see Fig. 1): 'As in other kinds of bookkeeping, the trickiest questions that arise in the application of thermodynamics deal with the proper identification and classification of the entries; the arithmetics is straightforward' (Onsager, 1967).

Figure 1: Lars Onsager (1903-1976) obtained the Nobel Prize in Chemistry 1968 for 'the discovery of the reciprocal relations bearing his name, which are fundamental for the thermodynamics of irreversible processes'.

2.1   The two Principles of Thermodynamics

We shall start by reviewing some elementary conversion problems. The simplest example is the conversion of Mechanical Energy into different forms (kinetic and potential); a ball that falls from a height h reaches the ground with velocity $v^2 = 2gh$ where $g \sim 9.8m/s^2$ is the gravity acceleration. This is the content of the celebrated experiments that Galileo is said to have performed from Pisa's Leaning Tower (see

Fig. 2). In this elementary calculation, the friction of air has been neglected: this conversion of potential energy into kinetic energy occurs without a fee (conservation of the total mechanical energy). The free fall of a body can also be used to perform a work $\mathcal{W}$ (e.g. by attaching it to a pulley), the value of which is given by

$$\mathcal{W} = \mathcal{E}_{\text{initial}} - \mathcal{E}_{\text{final}} = -\Delta\mathcal{E} = mgh, \tag{1}$$

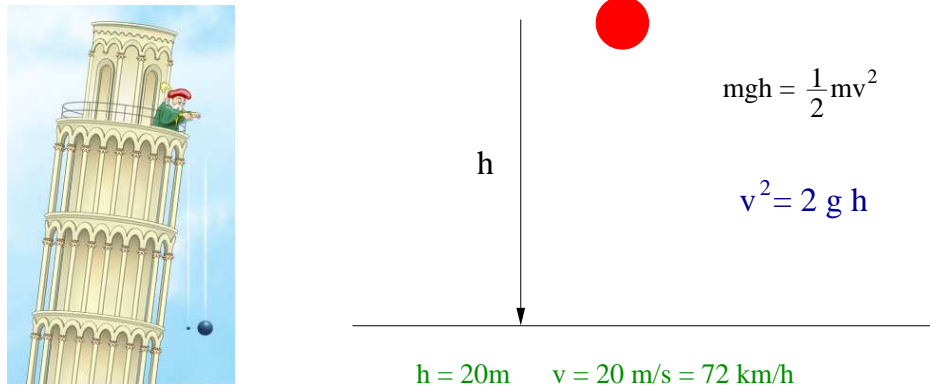where $\mathcal{E}$ represents the potential energy.

Figure 2: The legendary experiment of Galileo in Pisa and its schematic representation.

Using such elementary energy balance arguments one can easily estimate the order of magnitude of the maximal height that a pole-vaulter can jump on the earth (use the fact that that the highest speed a human being can reach is roughly $10m/s$).

The above processes are assumed to be free of dissipation and can be described in purely mechanical terms. In particular, they are perfectly reversible with time: for example, the motion of a pendulum clock allows us to measure time by slicing it into periods of constant duration but it does not tell us what the direction of time is: a movie displaying the oscillations of a dissipation-less pendulum can be run backwards in time without anybody noticing it.

In reality, some dissipation is always present: a ball bouncing on the ground looses energy at each shock and stops after a few bounces. Of course, energy as a whole is conserved because heat is evolved. This is the content of the First Principle of Thermodynamics: Heat must be taken into account when doing energy book-keeping. The work of James Joule established that Work and Heat are two ways of exchanging energy between a system and its environment. This lead to the *First Principle of Thermodynamics:*

$$\boxed{\Delta E = \partial W + \partial Q.} \tag{2}$$

The energy E in this equation is now the total internal energy of the system.

In layman's words, the First Principle states that: **The energy of the universe is constant**.

Exercise (proposed by Jean-Marc Victor): Energy balance problems can sometimes be rather subtle. Consider two perfectly identical spheres at the same temperature and made of the same material (Fig. 3). One sphere lies on the ground
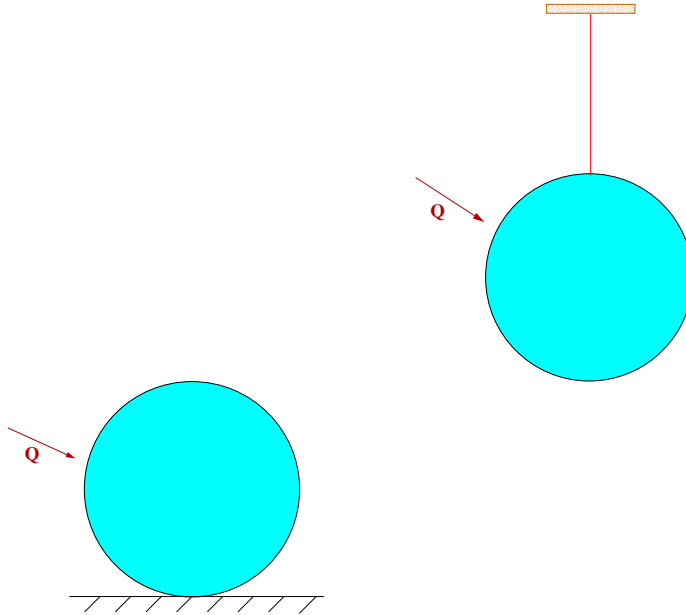
Figure 3: The two spheres puzzle.

whereas the other is hanging, attached by a rigid thread. The same quantity of heat Q is given to each sphere. Which sphere will be hotter? (we suppose that there is no heat transfer from a sphere to its environment, *i.e.*, ground, air, thread...).

In presence of dissipation, time-reversibility at macroscopic scale is lost. Projecting the movie of a ball bouncing on the ground backwards in time would display an impossible process: the ball would appear to bounce spontaneously higher and higher by absorbing heat from the ground. Such a process satisfies the first law of thermodynamics but would clearly never happen in reality.

In short, some processes are possible whereas others are not. This can be a very difficult task to detect the hidden flaw in some highly involved mechanisms. How can one *discriminate* between *possible* and *impossible* processes? The solution to this problem is provided by the Second Law of Thermodynamics, elaborated by Carnot in 1824 (see Fig. 4), Clausius (1850) and Kelvin (1851). Two classical formulations of the second principle are [4]:

- Clausius formulation: *No process is possible whose* **sole** *result is the transfer of heat from a cooler body to a hotter body.*

- Kelvin-Planck formulation: *No process is possible whose* **sole** *result is the absorption of heat from a reservoir and the conversion of that heat into work.*

The Clausius and the Kelvin-Planck formulations present two elementary, universal, archetypical forbidden processes. These two statements can be shown to be equivalent and they cover all possible cases; they provide the answer to the problem of deciding whether a given process is possible or not: by suitable couplings and mappings, an impossible process can always be transformed into a 'machine' that violates Kelvin-Planck's or Clausius' statement. At this stage, thermodynamics acquires a logical beauty akin to that of classical geometry. This elegant structure is
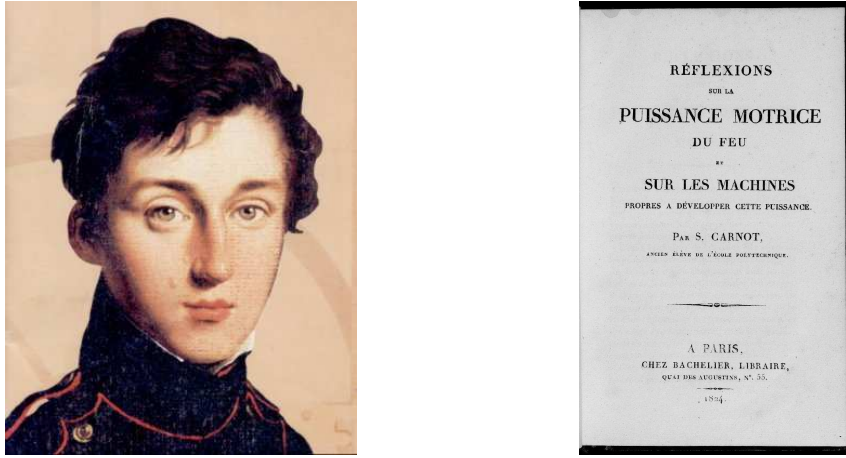
Figure 4: Sadi Carnot (1796-1832).

perfectly demonstrated in the classical books of Fermi or Pippard [5, 6] or in the recent textbook of Don Lemons, *Mere Thermodynamics* [7].

The Second Principle was put on a quantitative basis by Clausius, who introduced, in 1851, the **entropy** state function which measures the degree of irreversibility of a process. This is expressed by Clausius' Inequality, which becomes an equality if and only if the process is reversible

$$S_2 - S_1 \geq \int_{1 \to 2} \frac{\partial Q}{T}. \tag{3}$$

The Clausius or Kelvin-Planck statements of the second law can be reformulated in a more formal way: any process that would result in a decrease of the entropy of an isolated system is impossible.

The inequality (3), when applied to the universe considered as a whole, implies that **the entropy of the universe increases**. This sentence can be considered as a popular (albeit informal) statement of the Second Principle.

Although Energy is a familiar concept that plays a prominent role in many processes [5, 10, 11], one should never forget Entropy that drives secretly many phenomena observed in daily life, for example the melting of ice. Ultimately, thermodynamic effects are due to the interplay of Energy and Entropy: A thermal system seeks to minimize its energy while maximizing its entropy at the same time. The subtle balance between these two state functions is encoded in the **Free Energy** F that plays a fundamental role in statistical physics:

$$F = E - TS. \tag{4}$$

The interpretation of Free Energy as maximum available work is classical. Consider a system that evolves from a state A to a state B, both at the temperature T equal to that of the environment (see Fig. 5). Suppose that the system exchanges heat only with its environment. Then, because of irreversibility, the Work, $\mathcal{W}_{\text{useful}}$, that one can extract from this system is at most equal to to the decrease of free energy:

$$\mathcal{W}_{\text{useful}} \leq F_{\text{initial}} - F_{\text{final}} = -\Delta F. \tag{5}$$

The equality is valid when the process is reversible. Comparing with equation (1) which is purely mechanical (with no heat transfer) we observe that the role of potential energy is now played by the *thermodynamic potential F*, and that the equality is replaced by an inequality because of dissipative effects.

Remark: One often considers the work W *we perform* on the system, which is the *opposite* of the work available *from* the system. The inequality (5) then becomes

$$W \geq F_B - F_A = \Delta F. \tag{6}$$

In other words, in order to increase the free energy of an isothermal system by an amount ΔF one has to perform an amount of work at least equal to ΔF. In general, because of irreversibility, the work performed W must be strictly greater than the free-energy variation.

We briefly recall the derivation of the Maximal Work Inequality, which requires the two Principles of Thermodynamics:

- *The First Principle* states that $\Delta E = W + Q$ (Recall that $\mathcal{W}_{\text{useful}}$ is equal to $-W$)

- We now use the *Second Principle:* $\int_{A \to B} \frac{\partial Q}{T} \leq S_B - S_A = \Delta S$. From which we conclude that $Q \leq T\Delta S$.

We conclude that

$$\Delta F = \Delta E - T\Delta S = W + Q - T\Delta S \leq W.$$

It is useful to define the dissipated work, $W_{\text{diss}} = W - \Delta F$, which from the above equation is given by

$$\frac{W_{\text{diss}}}{T} = \Delta S + \left(\frac{-Q}{T}\right) \equiv \Delta S(\text{universe}) \geq 0.$$

The interpretation of this identity is clear: the dissipated work W_{diss}/T represents the total entropy production, in the universe, by the process. This entropy production must be non-negative.
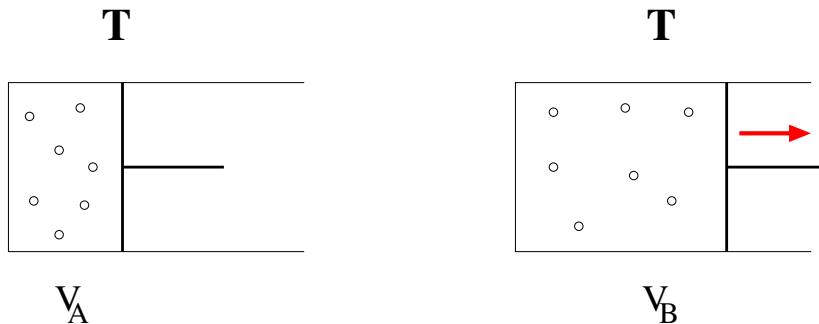

Figure 5: An illustration of the Maximum Work relation on a simple piston-gas system.

A last refinement of the Maximal Work Theorem that will become clear in the sequel is the following. Thermodynamics considers macroscopic observables which assume well-defined values. However, Statistical Mechanics predicts that observables

are subject to thermal fluctuations around their average, thermodynamic, values. In particular, the work W should also be considered to be a random variable. Thus, one should be aware that, strictly speaking, thermodynamic identities refer to mean values. In full rigor, the Maximal Work Theorem should be written as

$$\langle \mathcal{W}_{\text{useful}} \rangle \leq F_{\text{initial}} - F_{\text{final}} = -\Delta F \tag{7}$$

and we emphasize again that the equality can only occur for a reversible process. In a later section of this review, we shall explain that there exists an exact identity, known as Jarzynski's relation, which is valid both for reversible and irreversible processes. The classic inequality (7) is a consequence of Jarzynski's relation.

2.2 Molecular theory of heat and the framework of Statistical Mechanics

The works of Maxwell, Boltzmann and Gibbs (fig. 6) have led to a statistical interpretation of entropy and to the foundations of Statistical Mechanics which encompasses classical thermodynamics and provides its microscopic basis.
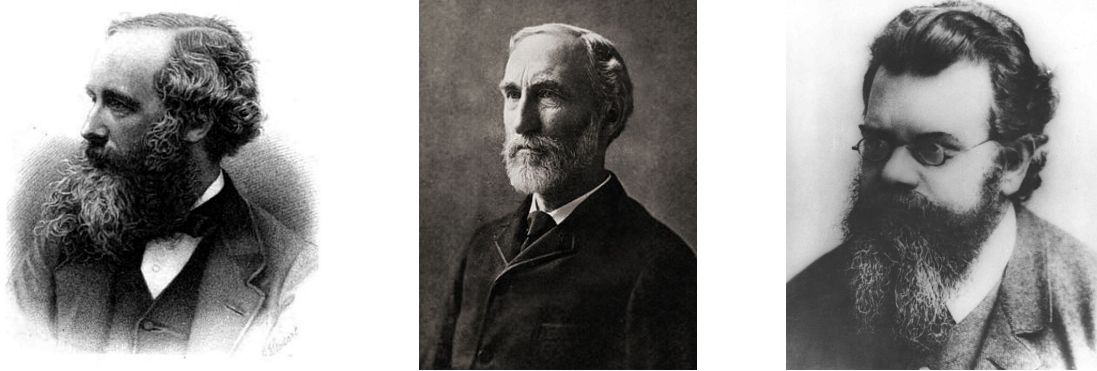

Figure 6: Three Fathers of Statistical Mechanics: James Clerk Maxwell (1831-1879), Josiah Willard Gibbs (1839-1903) and Ludwig Eduard Boltzmann (1844-1906).

The phase space of a macroscopic system that contains typically 10^{23} molecules is huge. Our senses and our measurement devices are sensitive only to a few global and averaged properties. In fact, at the microscopic scale, a system, *even at equilibrium*, evolves continuously from one microstate to another but most of the time *these different microstates are perceived by us as being the same macroscopic state.* We have indeed access only to extremely *coarse-grained* sense-data and a tremendous number of microstates are compatible with this data. Thus, coarse-graining from the microscale to the macroscale implies a huge loss of information: entropy quantifies this fact.

More precisely, consider an isolated system of macroscopic volume V with total energy E. We suppose that all microscopic states of the system having energy E are equiprobable: this assumption, known as the 'microcanonical probability ensemble' is a foundation-stone of statistical mechanics. (This can be proved rigorously for systems that display sufficiently strong ergodicity). The thermodynamic entropy S of a microcanonical system enumerates the total number of such microscopic states,

$\Omega(E, V)$ and the Boltzmann formula (fig. 7) states that

$$\boxed{S = k_B \log \Omega} \qquad \text{with } k_B \simeq 1.3810^{-23}. \qquad (8)$$

The determination of Entropy is fundamentally a counting, a combinatorial, problem.

We shall illustrate this fact by analyzing a simplistic toy model of a perfect gas enclosed in a box of volume V. Consider a gas consisting of N classical distinguishable molecules: a microscopic state is specified by knowing all the positions and the velocities of all molecules. The entropy S of the gas will depend on the volume V and the total energy E of the gas. We simplify the discussion further and analyze only the positional configurations (forgetting about velocities - in fact, impulsions ultimately factor out). Let us discretize the box by supposing that the position of an individual molecule inside the box is known up to a precision $\Delta V \ll V$. The total number of micro-configurations is then given by $\Omega = \left(\frac{V}{\Delta V}\right)^N$. Using Boltzmann's formula, we deduce that, at constant energy, the entropy is given by $S = Nk_B \log V + \text{const}$. The unwanted constant will drop out if we consider entropy variations. Let us suppose that we double the volume of the box by performing an isothermal expansion $V \to 2V$. Because the temperature is constant, the internal energy of the gas is constant and the variation of entropy, $\Delta S = S_2 - S_1$, is due only to the volume change. We have

$$S_2 - S_1 = Nk_B \log 2. \qquad (9)$$

This equation can be understood by saying that each particle contributes to the entropy increase by the amount $k_B \log 2$, which results from the doubling of the available volume. Note that after doubling the volume, a particle can be in two different regions (or 'states'), namely the left part or the right part of the box with equal probabilities $1/2$.

We can emphasize the combinatorial content of the previous model by making it even more abstract and removing all spurious references to gases, particles etc... We consider an assembly consisting of N independent elements, that can occupy m different states that we label $i = 1, 2, \ldots, m$. We assume that state i can be occupied with probability p_i. If $N \gg 1$, we shall have roughly $n_1 = Np_1$ objects in state 1, $n_2 = Np_2$ objects in state 2 etc... The total number or configurations Ω is given by the multinomial coefficient

$$\Omega = \frac{N!}{\prod_{i=1}^{m} n_i!} = \frac{N!}{\prod_{i=1}^{m} (Np_i)!}.$$

The corresponding entropy is evaluated by using Stirling's formula and we obtain

$$S = k_B \log \Omega \simeq -Nk_B \sum_{i=1}^{m} p_i \log p_i. \qquad (10)$$

The contribution of each elementary constituent to the total entropy thus can be evaluated as

$$\boxed{S = -k_B \sum_{i=1}^{m} p_i \log p_i.} \qquad (11)$$

Note that the volume doubling calculation above is retrieved by taking $p_1 = p_2 = 1/2$.

This result has far reaching consequences in fundamental Physics. Suppose that the states $i = 1, 2, \ldots, m$ represent m different energy levels $\epsilon_1, \ldots, \epsilon_m$ that can be occupied by N non-interacting objects. Besides, consider that the average energy $N\epsilon$ of the system is kept constant (for example by connecting it to a heat reservoir). As above, we denote by $n_i = p_i N$ the average occupation of state i. We thus have the following two constraints:

$$\sum_{i=1}^{m} p_i = 1, \tag{12}$$

$$\sum_{i=1}^{m} p_i \epsilon_i = \epsilon. \tag{13}$$

Following Schrödinger [12], one imposes that the total entropy S of the ensemble, given by the formula (10), must be maximal, as required by the Second Principle of Thermodynamics (or by saying that the most probable configurations overwhelmingly dominate over the others when $N \to \infty$). The logic is then reverted as follows: What will be the assignment of the probabilities p_i that maximizes S under the constraints (12) and (13)? This problem is readily solved (for example by using Lagrange multipliers) and one finds

$$p_i = \frac{1}{Z} e^{-\epsilon_i/k_B T}, \tag{14}$$

where the constants Z and T are adjusted to satisfy the constraints. E. Schrödinger then proves explicitly that T is identical to the physical concept of temperature and that the Partition Function Z *(Zustandsumme)* is simply related to the thermodynamic Free Energy (see equation (17) below).

A more standard path to study a system thermalized a given temperature T (i.e. a system in contact with a huge thermal reservoir at temperature T) is to apply Boltzmann's fundamental equation (8) to the total isolated entity consisting of the system + reservoir and to eliminate (trace out) the degrees of freedom of the reservoir [10]. One shows that the probability of observing a microscopic configuration $\mathcal{C}$ of energy $E(\mathcal{C})$ is given by the **Boltzmann-Gibbs canonical law:**

$$P_{\text{eq}}(\mathcal{C}) = \frac{e^{-E(\mathcal{C})/k_B T}}{Z}. \tag{15}$$

The Partition Function Z which insures that all probabilities sum up to 1 (normalization) is given by

$$Z = \sum_{\mathcal{C}} e^{-E(\mathcal{C})/k_B T} = \sum_{E} \Omega(E) e^{-E/k_B T}. \tag{16}$$

The canonical law, which implies a probabilistic description of the microscopic structure of a thermal system is derived from Boltzmann's formula (8) under very general assumptions. The framework of statistical mechanics then is laid out by the following relation, deduced from Eqs. (4, 8, 15) and (16) and which links the Free Energy with the Partition Function:

$$F = -k_B T \log Z. \tag{17}$$

From this relation, the probabilistic expression of the entropy for a system at temperature T is readily obtained as

$$S = -k_B \sum_{\mathcal{C}} P_{\text{eq}}(\mathcal{C}) \log P_{\text{eq}}(\mathcal{C}) \,, \tag{18}$$

which is identical to the combinatorial expression (11).

The precise logical order under which Statistical Mechanics is presented depends on the books and on the lecturer's tastes. The standard point of view [10] or the probabilistic/combinatorial approach of Schrödinger [12, 13, 14] are equally valid but their underlying logic is different.

The important fact is that Statistical Mechanics provides us with a systematic procedure to analyze systems at thermal equilibrium:

- Describe the microstates of the system and find a suitable microscopic Hamiltonian.

- Calculate Z and deduce the Free Energy F.

- Derive from F the thermodynamic properties of the system such as its phase diagram.

Of course, applying this well-defined program to a given problem can be incredibly difficult. Nobody knows how to calculate Z for the three-dimensional Ising model...

Figure 7: Ludwig Boltzmann (1844-1906). The celebrated formula for the entropy is inscribed on Boltzmann's grave in Vienna.

Before concluding this section, we recall the Sakur-Tetrode formula (1912) for the entropy of a mono-atomic classical ideal gas, taking into account quantum indistinguishability and phase-space discreteness [4, 13]

$$S = k_B N \log \left[\frac{V}{N} \left(\frac{m k_B T}{2\pi\hbar^2} \right) \right] + \frac{5}{2} k_B N, \tag{19}$$

where m is the mass of a gas particle and $\hbar$ is Planck's constant. For one mole of Helium at 300K, the total entropy is about 100J/K.

2.3   Brownian Motion: Equilibrium is dynamical

Equilibrium is a dynamical concept: a system in thermal equilibrium keeps on evolving from one microstate to another even if it appears to our imperfect senses to remain in the same macrostate. Thermodynamics deals only with averaged values: it can not account for microscopic fluctuations. Though these fluctuations are usually very minute (of relative order of 10^{-11} for a system containing one mole of matter) they can be detected either by using measuring devices which are becoming finer and finer or by studying very small systems. Statistical Mechanics allows us to calculate the probability distributions of observables (and not only their averages) and perfectly describes the thermal fluctuations.

The paradigm for thermal fluctuations is Brownian Motion discovered by Robert Brown who observed through a microscope, the perpetual, restless, giggling of a pollen grain in water (fig. 8). This phenomenon is the signature, at our scale, of the granular, discontinuous, structure of matter. It is the experimental footprint of the existence of atoms.
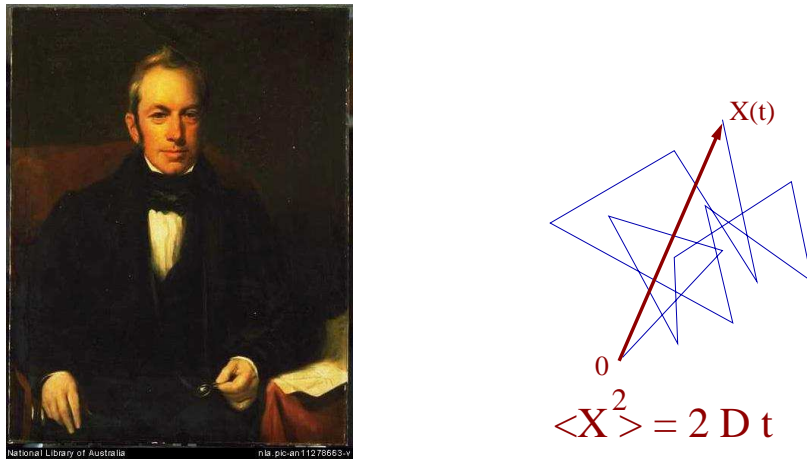

Figure 8: Robert Brown (1773-1858) and a sketch of Brownian motion.

The theory of Brownian Motion was elaborated by Albert Einstein, in 1905. The Brownian particle (for example, a grain of pollen) is restlessly shaken by random shocks with the molecules of water. Because of these shocks, the pollen grain undergoes an erratic motion and diffuses with time around its original position: although the average position of the Brownian particle does not change with time (just because of isotropy of space) the quadratic average (*i.e.*, the variance) of the position grows linearly with time:

$$\langle X^2(t) \rangle = 2Dt \, . \tag{20}$$

For a spherical particle of radius a, immersed in a liquid of viscosity η at (absolute) temperature T, the *diffusion constant* D is, given according to Einstein, by

$$D = \frac{RT}{6\pi \eta a \, \mathcal{N}} \, , \tag{21}$$

where $\mathcal{N} \simeq 6 \, 10^{23}$ is the Avogadro number, and $R \simeq 8.31$ is the perfect gas constant. This formula is extraordinary in the sense that it relates observables D, T, η and

a, which are all *macroscopic* to the number $\mathcal{N}$ of *atoms* in a mole of matter. This relation allowed Jean Perrin to *weigh experimentally an atom of hydrogen* (as he himself stated in his book 'The Atoms'); indeed $1/\mathcal{N}$ is roughly equal to the mass of one atom of hydrogen in grams. In his experiments, Perrin used small latex spheres with $a \sim 0.1\mu m$, immersed in water ($\eta = 10^{-3} kg m^{-1} s^{-1}$) at temperature $T = 300K$. The typical value of D is then $10^{-12} m^2/s$, *i.e.*, the Brownian particles diffuse about one micrometer in one second. All these values, though not strictly macroscopic, were observable with an optical microscope at the beginning of the twentieth century and are much larger than the atomic scale. The theory of Brownian motion and its experimental verification established beyond any doubt the existence of atoms, considered previously to be a mere hypothesis.

Einstein's formula (21) can be interpreted as the simplest manifestation of the Fluctuation-Dissipation Relation: consider that the pollen grain of size a, immersed in water, is subject to a small *drag force* f_{ext} (suppose for example that it is being pulled by an external operator). Because of this force, the pollen acquires a velocity v, and is subject to a frictional force $-\gamma v$ because of the viscosity η of the surrounding water. The friction coefficient γ was calculated by Stokes at linear order in the velocity and it is given by $\gamma = 6\pi\eta a$ (assuming the pollen to be a perfect sphere). Balancing the drag force with the frictional force, leads to the limiting speed:

$$v_\infty = \sigma f_{\text{ext}} \quad \text{with } \sigma = \frac{1}{6\pi\eta a}. \tag{22}$$

The *susceptibility* σ measures the *linear response* to the external drive f_{ext}. Using this concept of susceptibility, Einstein's relation can be rewritten as:

$$\boxed{D = k_B T \sigma,} \tag{23}$$

$k_B = R/\mathcal{N}$ being Boltzmann's constant. In other words, fluctuations at equilibrium, quantified by D, are proportional to the susceptibility σ which quantifies the Linear Response to a small external perturbation that drives the system out of equilibrium. There are many good books and articles on Brownian motion and linear response. Some useful references are [10, 15, 16, 17, 18, 19].

2.4   Universality of Brownian motion: Feynman's Ratchet and Pawl

One reason why Brownian Motion was so troublesome to 19th century physicists was (apart from the fact that they could not find a suitable explanation for it) that the pollen grain was undergoing a kind of perpetual motion even while remaining in contact with a single heat source (the water bath). Moreover, one could conceive a Gedanken Experiment in which this perpetual motion could be coupled to a mechanical rectifier such as a wheel allowed to rotate only in one direction. Thus, when the Brownian particle would move in one direction, say Eastwards, the wheel would rotate whereas it would stay still if the particle moved Westwards (see Fig. 10). This is in essence the celebrated Ratchet and Pawl model discussed by Feynman in Chap. 46 of his Lectures on Physics, Volume 1 [18]. Feynman rediscovered a model initially proposed by Smoluchowski (see Fig. 9). Thus, the second principle would be in trouble because this rectified motion of the wheel could be used to extract some work from a single heat source.
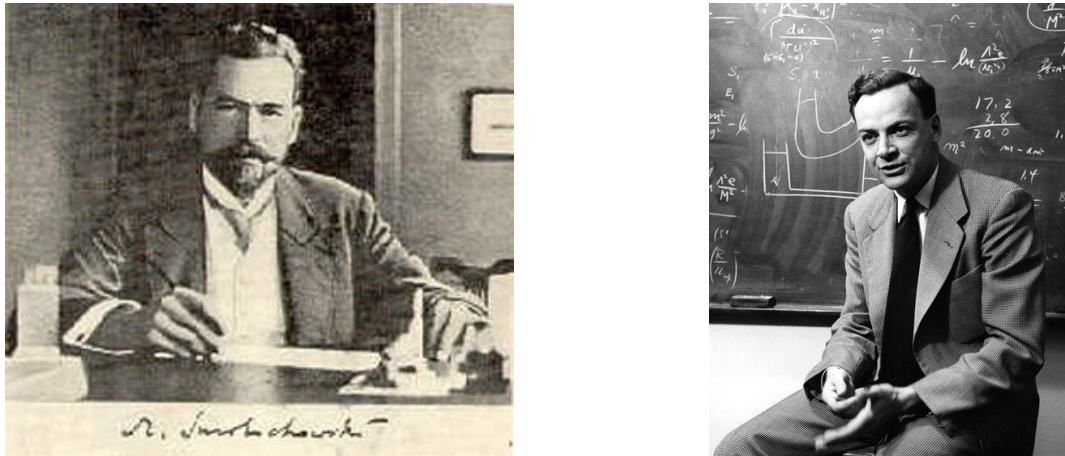
Figure 9: Marian Smoluchowski (1872-1917) was a pioneer of statistical physics. Richard Phillips Feynman (1918-1988) made many fundamental discoveries, including that 'Physics isn't the most important thing. Love is.'

In order for the pollen grain to cause rotation of the wheel in the Gedanken Experiment at a perceptible rate, this wheel must be very small. However, all bodies are subject to thermal fluctuations which typically are inversely proportional to their size. This universal character of thermal fluctuations leads to the resolution of the paradox: the one-way wheel is also subject to intrinsic thermal fluctuations which cause it to move in the forbidden direction. A precise calculation, see, *e.g.*, [22], shows that the two effects (the rotation of the wheel by the Brownian particle versus the spontaneous motion in the forbidden direction) perfectly compensate each other and no net rotation of the wheel occurs: 'the second law is saved'.
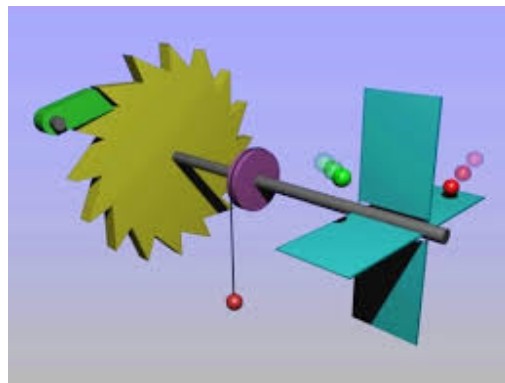

Figure 10: A Smoluchowski-Feynman ratchet.

2.4.1   Application to Molecular Motors

The concept of rectification of thermal fluctuations will be useful in non-equilibrium situations and will provide us a basic model for molecular motors in biological cells.

A significant part of the eucaryotic cellular traffic relies on 'motor' proteins that move in a deterministic way along filaments similar in function to railway tracks or

freeways (kinesins and dyneins move along tubulin filaments; myosins move along actin filaments). The filaments are periodic (of period ~ 10nm) and have a fairly rigid structure; they are also polar: a given motor always moves in the same direction. These motors appear in a variety of biological contexts: muscular contraction, cell division, cellular traffic, material transport along the axons of nerve cells...

Molecular Motors move by using the ratchet effect: they provide an example of *rectification* of Brownian motion (for reviews see, *e.g.*, [21, 23]). This rectification process relies on an external energy source, provided by ATP hydrolysis, that enables the motor to undergo transitions between different states and when these transitions break the detailed balance, a *directed motion* sets in (see Fig. 11). In order to move the motor consumes r ATP fuel molecules per unit time, which are hydrolyzed to ADP + P:

$$ATP \rightleftharpoons ADP + P\,.$$

The relevant chemical potential is thus given by $\Delta\mu = \mu_{\text{ATP}} - \mu_{\text{ADP}} - \mu_{\text{P}}$.
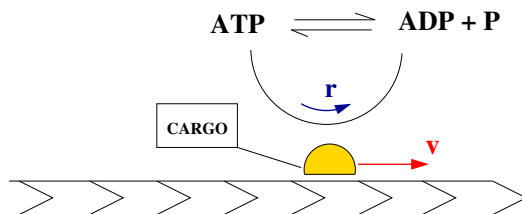

Figure 11: Schematic representation of a molecular motor: by hydrolyzing ATP, the motor proceeds along the polar filament and carries a 'cargo' molecule.

The principle of the motor is shown in figure 12 where the motor is represented by a small particle that can move in a one-dimensional space. At the initial time $t = 0$, the motor is trapped in one of the wells of a periodic asymmetric potential of period a. Between time 0 and t_f, the asymmetric potential is erased and the particle diffuses freely and isotropically at temperature T. At time t_f, the asymmetric potential is re-impressed, the motor slides down in the nearest potential valley and, because of damping, is trapped in one of the wells. The motor has maximal chance to end up in the same well where it was at time $t = 0$. However, it has a small probability to be trapped in the well located to the right and (because of the asymmetry of the potential) an even smaller probability to end up in the left well. In other words, because the potential is asymmetric, the motor has more chances to slide down towards the right: this leads on average to a net total current.

In general, the motor is subject to an external force f_{ext} which tilts the potential. Besides, when ATP is in excess, the chemical potential $\Delta\mu = \mu_{\text{ATP}} - \mu_{\text{ADP}} - \mu_{\text{P}}$ becomes *positive*. A basic problem is then to determine the velocity of the motor $v(f_{\text{ext}}, \Delta\mu)$ (mechanical current) and the ATP consumption rate $r(f_{\text{ext}}, \Delta\mu)$ (chemical current) as functions of the external mechanical and chemical loads [20, 21, 22, 23, 24].
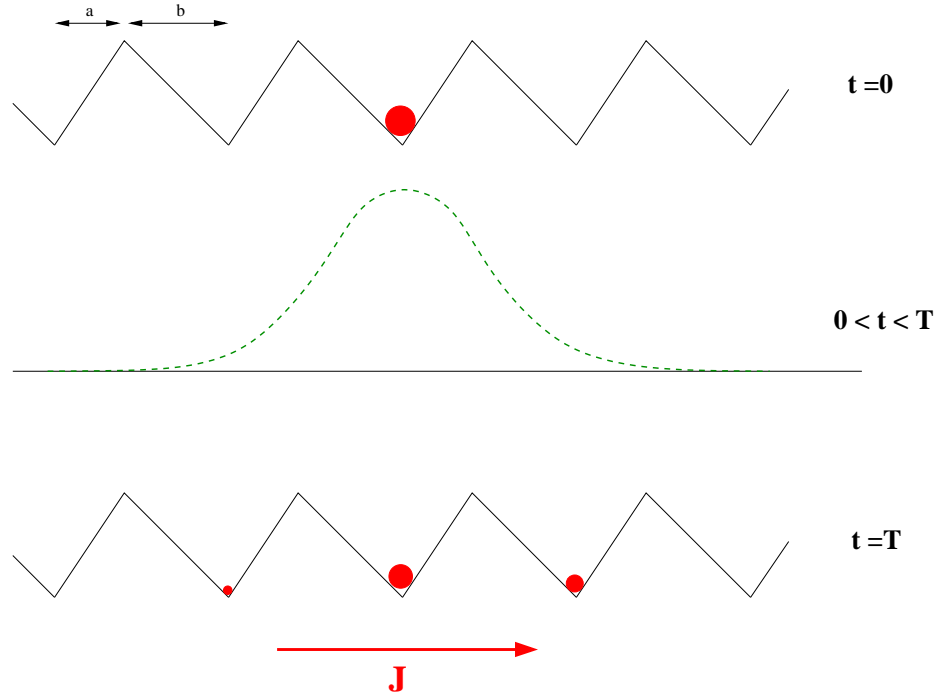
Figure 12: The principle of a Brownian ratchet: by inscribing and erasing periodically an asymmetric potential, a directed motion of the particle is induced. In this example, the potential is a saw-tooth function of period $a = A + B$. Since $B > A$, the slopes are different and the potential is asymmetric. The relative sizes of the probabilities of ending in one of the wells are represented by the sizes of the disks in the lowest picture. The right and left probabilities being different, this leads on average to a net total current J.

3   Equilibrium and Non-Equilibrium Dynamics

3.1   Markovian dynamics

An efficient way to describe systems out of equilibrium is to use a probabilistic approach that goes back to Einstein's 1905 paper and to Smoluchowski work at the same period. The idea is to write an evolution equation for the probability $P_t(\mathcal{C})$ for the system to be in the microstate (or configuration) $\mathcal{C}$ at time t. In order to achieve such a description, one has to:

1. Enumerate the Microstates $\{\mathcal{C}_1, \mathcal{C}_2 \ldots\}$ of the system. These microstates can form a discrete or a continuous set depending on the problem studied.

2. Specify the transition rates between two configurations. An important and common assumption is that these rates do not depend on the previous history of the system but only on the configuration $\mathcal{C}$ at time t and on the target configuration $\mathcal{C}'$ at time $t + dt$: this is the *Markovian hypothesis* which amounts to neglecting short time correlations. Thus, one must specify which transitions $\mathcal{C} \to \mathcal{C}'$ are allowed between times t and $t + dt$ and give their probability $M(\mathcal{C}', \mathcal{C})dt$.

The evolution of the system can thus be viewed as a path on the configurations network and the time evolution of $P_t(\mathcal{C})$ is obtained by writing a balance equation in this *network* between the incoming flux and the outgoing flux at each configuration as shown in Fig. 13.
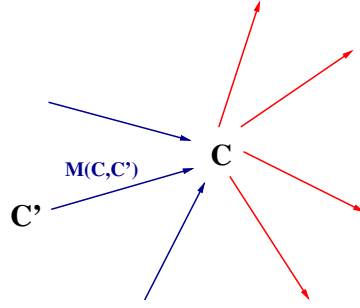
Figure 13: A network representation of the transition rates in a Markovian dynamics.

$$\frac{d}{dt}P_t(\mathcal{C}) = \sum_{\mathcal{C}' \neq \mathcal{C}} M(\mathcal{C}, \mathcal{C}')P_t(\mathcal{C}') - \left\{ \sum_{\mathcal{C}' \neq \mathcal{C}} M(\mathcal{C}', \mathcal{C}) \right\} P_t(\mathcal{C}). \qquad (24)$$

This rather simple looking equation is nothing but the **Master equation** for non-equilibrium statistical mechanics which plays, in this field, a role analogous to Newton's equation in mechanics or Schrödinger's equation in Quantum Mechanics. The Master (or Markov) equation (24) is a linear equation in the probability vector P_t whose components are given by $\{P_t(\mathcal{C})\}$. It is more convenient to write it in a matrix form

$$\frac{dP_t}{dt} = M.P_t \,, \qquad (25)$$

where the **Markov Operator** M has matrix elements given by $M(\mathcal{C}, \mathcal{C}')$ for $\mathcal{C} \neq \mathcal{C}'$ and its diagonal terms are defined as

$$M(\mathcal{C}, \mathcal{C}) = - \sum_{\mathcal{C}' \neq \mathcal{C}} M(\mathcal{C}', \mathcal{C}) \,. \qquad (26)$$

The Markov Operator M fully encodes the microscopic dynamics of the system in this probabilistic approach which already implies a level of coarse-graining. Many physical properties of non-equilibrium systems are obtained by investigating the characteristics of M such as its symmetries or the various types it can assume etc... Markov operators have the following basic properties:

(i) Non-diagonal terms are non-negative.

(ii) Diagonal terms are negative.

(iii) Equation (26) implies that the sum of the elements of M over any column vanishes. This in turn implies the conservation of total probability, *i.e.*, $\sum_{\mathcal{C}} P_t(\mathcal{C})$ is constant.

(iv) A vector P_∞ in kernel of M is a stationary state, *i.e.*,

$$\frac{dP_\infty}{dt} = 0 \,. \qquad (27)$$

Under rather general assumptions, it is possible to prove the Perron-Frobenius Theorem (see, *e.g.*, [16]) which ensures that: (i) the kernel of M is non-degenerate and therefore the stationary state is unique; (ii) all other eigenvalues of M have strictly negative real parts; the inverse of these real parts correspond to the intrinsic

relaxation times of the system towards its stationary state (the imaginary parts characterize the oscillations during relaxation).

Remark: We shall limit our discussion to the Markovian case. The analysis of non-Markovian systems is an active field which requires specific techniques. It often happens, however, that a non-Markovian problem can be embedded into a larger Markovian system by taking into account supplementary degrees of freedom.

The archetypical model of Markovian dynamics is the the random walk on a discrete lattice, which is used to explain diffusive behaviour (a classic reference is [15]). We consider a random walker on a one-dimensional discrete lattice with spacing a where the walker can jump from a site to its two neighbours with probability dt between time t and $t + dt$. The Master equation for the random walker is given by

$$\frac{dP_t(na)}{dt} = P_t((n+1)a) + P_t((n-1)a) - 2P_t(na)\,, \tag{28}$$

and the corresponding Markov Operator is nothing but the discrete Laplacian:

$$M = \begin{pmatrix} \ddots & \ddots & \ddots & \ddots & & & \\ 0 & 1 & -2 & 1 & 0 & & \\ 0 & 0 & 1 & -2 & 1 & 0 & \\ & 0 & 0 & 1 & -2 & 1 & \ddots \\ & & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

3.2   Connection to Thermodynamics

In the probabilistic approach of non-equilibrium dynamics that we have described above, we made no mention of equilibrium statistical mechanics. Up till now, the discussion has been purely mathematical. However, this description has to be related to thermal physics. The *connection* with thermodynamics is obtained by imposing the condition that the stationary state of the Markovian dynamics P_∞ (Eq. (27)) be given by the Boltzmann-Gibbs canonical formula (15) which describes a system in thermal equilibrium:

$$P_\infty(\mathcal{C}) = P_{\text{eq}}(\mathcal{C}) = \frac{e^{-E(\mathcal{C})/k_B T}}{Z}\,. \tag{29}$$

This *matching condition* with thermodynamics imposes a *global constraint* on the rates in the Markov operator:

$$\sum_{\mathcal{C}' \neq \mathcal{C}} M(\mathcal{C}, \mathcal{C}') e^{-E(\mathcal{C}')/k_B T} = e^{-E(\mathcal{C})/k_B T} \left\{ \sum_{\mathcal{C}' \neq \mathcal{C}} M(\mathcal{C}', \mathcal{C}) \right\}. \tag{30}$$

This condition implies that the system described by the stochastic dynamics reaches ultimately a state of thermodynamic equilibrium. For systems far from equilibrium with a non-thermodynamic stationary state, this relation does not hold.

3.3   Time-Reversal Invariance and Detailed Balance

The master equation (24) can be rewritten as follows:

$$\frac{d}{dt}P_t(\mathcal{C}) = \sum_{\mathcal{C}'} \{M(\mathcal{C},\mathcal{C}')\,P_t(\mathcal{C}') - M(\mathcal{C}',\mathcal{C})\,P_t(\mathcal{C})\} = \sum_{\mathcal{C}'} J_t(\mathcal{C},\mathcal{C}')\,, \qquad (31)$$

where $J_t(\mathcal{C},\mathcal{C}')$ represents the net local current flowing between the configurations $\mathcal{C}'$ and $\mathcal{C}$ at time t;

$$J_t(\mathcal{C},\mathcal{C}') = M(\mathcal{C},\mathcal{C}')\,P_t(\mathcal{C}') - M(\mathcal{C}',\mathcal{C})\,P_t(\mathcal{C})\,. \qquad (32)$$

The local current encodes a balance of in going and out going currents between $\mathcal{C}'$ and $\mathcal{C}$ that is represented graphically in figure 14.

The local current is anti-symmetric with respect to the configurations:

$$J_t(\mathcal{C},\mathcal{C}') = -J_t(\mathcal{C}',\mathcal{C})\,. \qquad (33)$$

Besides, total probability conservation leads to

$$\sum_{\mathcal{C},\mathcal{C}'} J_t(\mathcal{C},\mathcal{C}') = 0\,. \qquad (34)$$
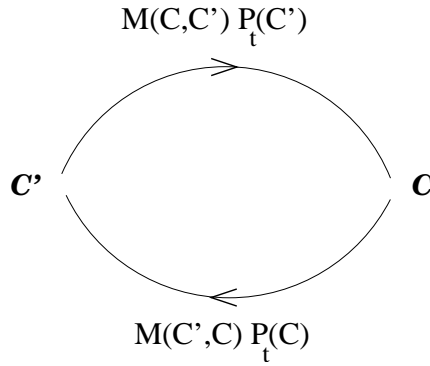
This relation is valid for any time t.

Figure 14: The local current between two configurations in a Markovian system.

When $t \to \infty$, the system reaches a steady-state with a stationary probability density. This implies that the total current flowing *to* $\mathcal{C}$ compensates exactly the total current flowing *from* $\mathcal{C}$:

$$\sum_{\mathcal{C}'} J_\infty(\mathcal{C},\mathcal{C}') = 0\,. \qquad (35)$$

This global balance condition is nothing but a reformulation of the stationarity condition (27).

Taking into account the fact that the microscopic dynamics of the system which is represented by an effective Markovian model is Hamiltonian and that Hamiltonian dynamics is in general time-reversible, Onsager derived the following, much stronger, constraint:

$$M(\mathcal{C},\mathcal{C}')P_{\text{eq}}(\mathcal{C}') = M(\mathcal{C}',\mathcal{C})P_{\text{eq}}(\mathcal{C})\,. \qquad (36)$$

This remarkable relation, known as **Detailed Balance**, is a consequence of the time-reversal symmetry of the microscopic dynamics of the system. Detailed balance implies that the net local current $J_t(\mathcal{C}, \mathcal{C}')$ between two configurations $\mathcal{C}$ and $\mathcal{C}'$ *vanishes at thermodynamic equilibrium.* This very strong condition implies of course the vanishing of the global currents (Eq. (35)). The detailed balance condition is a fundamental dynamic property of equilibrium systems that goes beyond the laws of classical thermodynamics. (Magnetic fields and overall rotation effects have been excluded from our discussion for sake of simplicity. A modified version of detailed balance allows to take into account these effects [16].)

There is an elegant way to express detailed balance as a Hermiticity property of the Markov operator M. Indeed, detailed balance (36) can be rewritten as

$$P_{\text{eq}}(\mathcal{C})^{-1/2} M(\mathcal{C}, \mathcal{C}') P_{\text{eq}}(\mathcal{C}')^{1/2} = P_{\text{eq}}(\mathcal{C}')^{-1/2} M(\mathcal{C}', \mathcal{C}) P_{\text{eq}}(\mathcal{C})^{1/2}. \tag{37}$$

Introducing now the diagonal matrix

$$Q = \begin{pmatrix} \ddots & & \\ & P_{\text{eq}}(\mathcal{C})^{1/2} & \\ & & \ddots \end{pmatrix}, \tag{38}$$

we observe that detailed balance implies that the operator $\mathcal{W} = Q^{-1}MQ$ is Hermitian:

$$\mathcal{W} = \mathcal{W}^\dagger. \tag{39}$$

A system in thermodynamic equilibrium satisfies detailed balance. Conversely, if the steady state of a system breaks the detailed balance condition then this steady state is necessarily a Non-Equilibrium Stationary State (NESS). Hence, from a technical point of view, violating detailed balance ensures that the system is out of equilibrium. Moreover, the mathematical description of systems far from equilibrium typically involves Non-Hermitian evolution operators.

3.4 Physical Interpretation of Detailed Balance

We have said above that detailed balance was derived by Onsager as a consequence of the time-reversibility of the underlying Hamiltonian dynamics. Here, we shall prove the converse of this assertion: the equilibrium state of a system which satisfies the detailed balance relation (36) is necessarily invariant by time-reversal.

We first recall two properties of any Markovian dynamics which will be useful in the following:

1. The probability of remaining in $\mathcal{C}$ during a time interval τ is given by:

$$\lim_{dt \to 0} (1 + M(\mathcal{C}, \mathcal{C}) dt)^{\frac{\tau}{dt}} = \mathrm{e}^{M(\mathcal{C}, \mathcal{C})\tau}. \tag{40}$$

2. The probability of going from $\mathcal{C}$ to $\mathcal{C}'$ during dt is: $M(\mathcal{C}', \mathcal{C})dt$.

Let us study a system in its equilibrium state. As we have already emphasized, macroscopic equilibrium is in fact a dynamical state in which the system continuously evolves between microscopic configurations. The frequency with which a given configuration is visited is proportional to its stationary probability. We consider a trajectory $\mathcal{C}(t)$ of the system between time $t = 0$ and a final observation time T (see figure 15). We suppose that at $t = 0$ the system is in configuration $\mathcal{C}_0$, that

it jumps from $\mathcal{C}_0$ to $\mathcal{C}_1$ at time $t_1 > 0$ and remains in $\mathcal{C}_1$ till time t_2; at $t_2 > t_1$, it jumps from $\mathcal{C}_1$ to $\mathcal{C}_2$ and remains in $\mathcal{C}_2$ till t_3 etc... More generally, the system jumps from $\mathcal{C}_k$ to $\mathcal{C}_{k+1}$ at time t_{k+1}, for $k = 1, \ldots, n$. The final jump from configuration $\mathcal{C}_{n-1}$ to $\mathcal{C}_n$ occurs at t_n and the system remains in $\mathcal{C}_n$ till the final time T. What is the probability $\Pr\{\mathcal{C}(t)\}$ of observing the trajectory $\mathcal{C}(t)$? Using recursively the two properties recalled above we obtain:

$$\Pr\{\mathcal{C}(t)\} = e^{M(\mathcal{C}_n,\mathcal{C}_n)(T-t_n)} \mathbf{M(\mathcal{C}_n,\mathcal{C}_{n-1})dt_n} \, e^{\mathbf{M(\mathcal{C}_{n-1},\mathcal{C}_{n-1})(t_n-t_{n-1})}} \ldots$$

$$e^{M(\mathcal{C}_2,\mathcal{C}_2)(t_3-t_2)} \mathbf{M(\mathcal{C}_2,\mathcal{C}_1)dt_2} \, e^{\mathbf{M(\mathcal{C}_1,\mathcal{C}_1)(t_2-t_1)}} \mathbf{M(\mathcal{C}_1,\mathcal{C}_0)dt_1} \, e^{\mathbf{M(\mathcal{C}_0,\mathcal{C}_0)t_1}} \mathbf{P_{eq}(\mathcal{C}_0)} \quad (41)$$
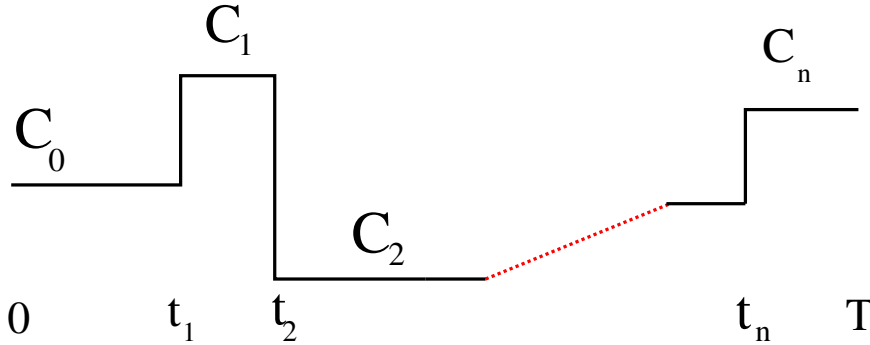
TRAJECTORY C(t)

Figure 15: A typical trajectory with discrete jumps in a Markovian dynamics.

We now calculate the probability of observing the time-reversed trajectory $\hat{\mathcal{C}}(t) = \mathcal{C}(T - t)$ (see figure 16). The system starts at $t = 0$ in configuration $\mathcal{C}_n$ and remains in that configuration till the time $T - t_n$ at which it jumps into $\mathcal{C}_{n-1}$. The next jump from $\mathcal{C}_{n-1}$ to $\mathcal{C}_{n-2}$ occurs at date $T - t_{n-1}$. More generally, the system jumps from $\mathcal{C}_k$ to $\mathcal{C}_{k-1}$ at time $T - t_k$, for $k = n, n - 1, \ldots, 1$. At date $T - t_1$, the system reaches the configuration $\mathcal{C}_0$ and remains in it till the final time T. The probability of this trajectory is given by:

$$\Pr\{\hat{\mathcal{C}}(t)\} = e^{M(\mathcal{C}_0,\mathcal{C}_0)t_1} \mathbf{M(\mathcal{C}_0,\mathcal{C}_1)dt_1} \, e^{\mathbf{M(\mathcal{C}_1,\mathcal{C}_1)(t_2-t_1)}}$$

$$\ldots e^{M(\mathcal{C}_{n-1},\mathcal{C}_{n-1})(t_n-t_{n-1})} \mathbf{M(\mathcal{C}_{n-1},\mathcal{C}_n)dt_n} \, e^{\mathbf{M(\mathcal{C}_n,\mathcal{C}_n)(T-t_n)}} \mathbf{P_{eq}(\mathcal{C}_n)}. \quad (42)$$

The ratio of the probability of observing a given trajectory (41) to the probability of the time-reversed trajectory (42) is thus given by

$$\frac{\Pr\{\mathcal{C}(t)\}}{\Pr\{\hat{\mathcal{C}}(t)\}} = \frac{M(\mathcal{C}_n,\mathcal{C}_{n-1})M(\mathcal{C}_{n-1},\mathcal{C}_{n-2})\ldots M(\mathcal{C}_1,\mathcal{C}_0) \, P_{eq}(\mathcal{C}_0)}{M(\mathcal{C}_0,\mathcal{C}_1) \, M(\mathcal{C}_1,\mathcal{C}_2) \, \ldots \quad M(\mathcal{C}_{n-1},\mathcal{C}_n) \, P_{eq}(\mathcal{C}_n)}. \quad (43)$$

If, in the numerator of this expression, we use recursively the detailed balance condition (36)

$$M(\mathcal{C}_{k+1},\mathcal{C}_k)P_{eq}(\mathcal{C}_k) = P_{eq}(\mathcal{C}_{k+1})M(\mathcal{C}_k,\mathcal{C}_{k+1}) \quad \text{for} \quad k = 0, 1 \ldots n-1 \,,$$

we find that

$$\frac{\Pr\{\mathcal{C}(t)\}}{\Pr\{\hat{\mathcal{C}}(t)\}} = 1 \,. \quad (44)$$
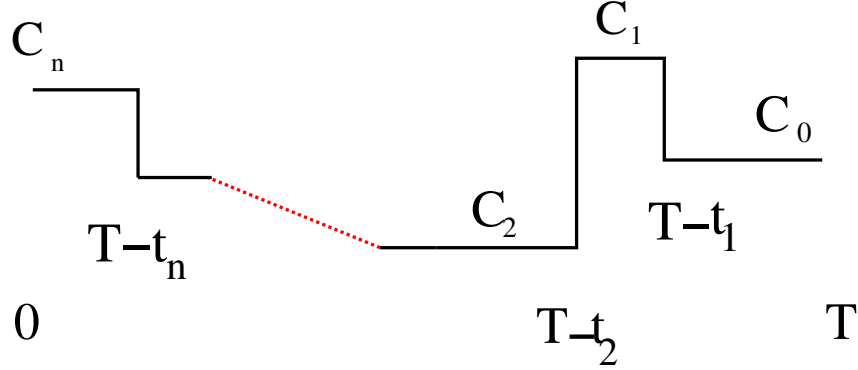
TIME–REVERSED TRAJECTORY C(T–t)

Figure 16: The time-reversed trajectory of the trajectory drawn in Figure 15.

We have thus shown that detailed balance implies that the dynamics in the stationary state is time reversible.

3.5 Entropy Production in Markovian systems

By analogy with equation (18) which gives an expression of the Entropy in the canonical ensemble of Statistical Mechanics, it is possible to define formally a time-dependent 'entropy' function for any Markovian system [16]:

$$S(t) = -\sum_{\mathcal{C}} P_{\mathrm{t}}(\mathcal{C}) \log P_{\mathrm{t}}(\mathcal{C}) \,. \tag{45}$$

Using the Markov equation (31) in terms of the local currents, the time derivative of this function is given by

$$\frac{dS(t)}{dt} = -\sum_{\mathcal{C}} \frac{dP_t(\mathcal{C})}{dt} \left(\log P_{\mathrm{t}}(\mathcal{C}) + 1\right) = -\sum_{\mathcal{C},\mathcal{C}'} J_t(\mathcal{C},\mathcal{C}') \log P_{\mathrm{t}}(\mathcal{C})$$

$$= \sum_{\mathcal{C},\mathcal{C}'} J_t(\mathcal{C},\mathcal{C}') \log P_{\mathrm{t}}(\mathcal{C}') \,, \tag{46}$$

where we have used the global conservation (34). The last equality is obtained by using the antisymmetry of the local currents (33) and by exchanging the role of the dummy variables $\mathcal{C}$ and $\mathcal{C}'$. The expression for the time derivative of $S(t)$ can be written in a more elegant manner by taking the half-sum of the last two equalities:

$$\frac{dS(t)}{dt} = \frac{1}{2} \sum_{\mathcal{C},\mathcal{C}'} J_t(\mathcal{C},\mathcal{C}') \log \frac{P_{\mathrm{t}}(\mathcal{C}')}{P_{\mathrm{t}}(\mathcal{C})} \,. \tag{47}$$

Transforming this expression, we obtain

$$
\begin{aligned}
\frac{dS(t)}{dt} &= \frac{1}{2} \sum_{\mathcal{C},\mathcal{C}'} J_t(\mathcal{C},\mathcal{C}') \left(\log \frac{P_t(\mathcal{C}')}{P_t(\mathcal{C})} + \log \frac{M(\mathcal{C},\mathcal{C}')}{M(\mathcal{C}',\mathcal{C})} - \log \frac{M(\mathcal{C},\mathcal{C}')}{M(\mathcal{C}',\mathcal{C})} \right) \\
&= \frac{1}{2} \sum_{\mathcal{C},\mathcal{C}'} J_t(\mathcal{C},\mathcal{C}') \log \frac{M(\mathcal{C},\mathcal{C}')P_t(\mathcal{C}')}{M(\mathcal{C}',\mathcal{C})P_t(\mathcal{C})} - \frac{1}{2} \sum_{\mathcal{C},\mathcal{C}'} J_t(\mathcal{C},\mathcal{C}') \log \frac{M(\mathcal{C},\mathcal{C}')}{M(\mathcal{C}',\mathcal{C})} \\
&\equiv \frac{d_i S}{dt} + \frac{d_e S}{dt}\, ,
\end{aligned}
\tag{48}
$$

where the first expression in the last equality is called the *entropy production term*, also denoted by σ_i, and the second term the *entropy flux*. The entropy production can be proved, using convexity relations, to be always positive, $\sigma_i \geq 0$, whereas the entropy flux can be positive or negative. At equilibrium, σ_i vanishes identically *because of detailed balance* (Eq. (36)). In the vicinity of equilibrium the linear response theory can be used to show that σ_i decreases towards 0 when the system relaxes to equilibrium. For a system in a non-equilibrium stationary state that does not satisfy detailed balance, σ_i does not vanish but entropy production and entropy flux compensate each other exactly.

4   The Gallavotti-Cohen Fluctuation Theorem for Markovian Thermodynamics

Systems in a non-equilibrium stationary state typically exhibit a non-vanishing macroscopic current J (*e.g.*, a current of particles, or a heat flux...). Therefore, time-reversal and detailed balance are broken in the mathematical description of the system. The stationary state, which is in the kernel of the Markov operator M, is in general not given by a Boltzmann-Gibbs law (which satisfies detailed balance). In fact, there is no general rule at present which would allow us to calculate the stationary state knowing the external constraints applied to the system. As opposed to the case of thermal equilibrium, there is no general theory of non-equilibrium statistical mechanics. Some general results on systems far from equilibrium have however been found that now we review.

4.1   Generalised Detailed Balance

Violation of detailed balance is the source of macroscopic currents which maintain the system far from equilibrium. This violation can be due to different factors: (i) the existence of an external driving force that pushes the particles in a given direction; (ii) the presence of reservoirs of unequal chemical potential (or temperature) that generates a current. The second case is particularly important to model the interaction of a system with its environment and the fluxes that are induced by this interaction. For a system connected to reservoirs there often exists a relation which plays a role similar to that of detailed balance and implies some fundamental properties of the stationary state. This relation is called generalised detailed balance. We shall discuss it in the case of a discrete Markovian system [28] which can undergo an elementary transition between two configurations during the interval $(t, t + dt)$. We shall suppose that we are studying an observable Y_t which varies by y at each elementary transition. For each elementary transition, we can specify how

Y_t changes:

$$\mathcal{C} \to \mathcal{C}' \text{ and } Y_t \to Y_t + y \quad \text{with probability } M_y(\mathcal{C}', \mathcal{C}) dt. \tag{49}$$

By *time reversal*, the transition occurs from $\mathcal{C}' \to \mathcal{C}$. Assuming that y is odd (*i.e.*, it changes its sign), we have $Y_t \to Y_t - y$. Finally, we suppose that there exists a constant γ_0 such that transition rates satisfy **the generalised detailed balance condition**:

$$\boxed{M_{+\mathbf{y}}(\mathcal{C}', \mathcal{C}) P_{\text{stat}}(\mathcal{C}) = M_{-\mathbf{y}}(\mathcal{C}, \mathcal{C}') \, e^{\gamma_0 \mathbf{y}} \, P_{\text{stat}}(\mathcal{C}').} \tag{50}$$

For $\gamma_0 = 0$, usual detailed balance is recovered. This relation holds, under general assumptions, for a system in contact with reservoirs that drive it out of equilibrium (in fact, it can be shown to be a consequence of usual detailed balance for the global model obtained by taking into account the system plus the reservoirs).

4.2   Time Reversal and the Gallavotti-Cohen Symmetry

We now investigate the relation between generalised detailed balance and time reversal [28]. We have to modify the calculations done in equations (41, 42, 43 and 44) by taking into account all the factors of the type $e^{\gamma_0 y}$ that appear at each jump in the ratio between the probabilities of forward and time-reversed trajectories.
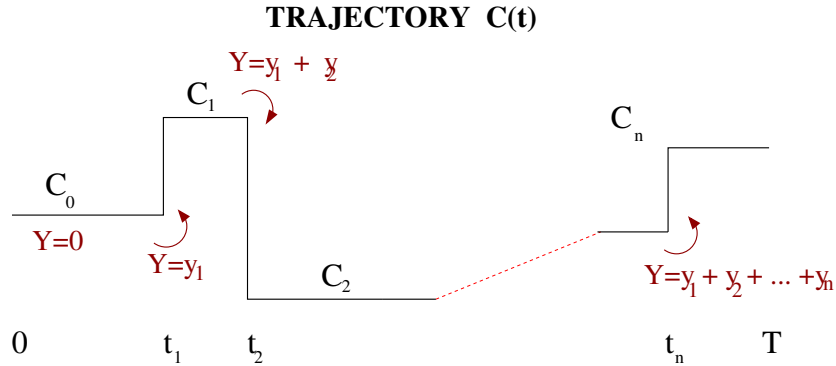

TRAJECTORY C(t)

Figure 17: A trajectory in a Markovian system taking into account the variation of the observable Y_t at each jump.

Following the same steps as in section 3.4, we finally obtain

$$\frac{\Pr\{\mathcal{C}(t)\}}{\Pr\{\hat{\mathcal{C}}(t)\}} = e^{\gamma_0 Y\{\mathcal{C}(t)\}}, \tag{51}$$

where $Y\{\mathcal{C}(t)\} = y_1 + y_2 + \ldots y_n$ is the cumulated value when the system follows the trajectory $\mathcal{C}(t)$ between 0 and t (see fig. 17).

We now recall that Y is odd under time-reversal and therefore we have $Y\{\hat{\mathcal{C}}(t)\} = -Y\{\mathcal{C}(t)\}$. Summing equation (51) over all possible histories between times 0 and t, and taking γ to be an arbitrary real number, we obtain

$$\sum_{\mathcal{C}(t)} e^{(\gamma - \gamma_0)Y\{\mathcal{C}(t)\}} \Pr\{\mathcal{C}(t)\} = \sum_{\hat{\mathcal{C}}(t)} e^{-\gamma Y\{\hat{\mathcal{C}}(t)\}} \Pr\{\hat{\mathcal{C}}(t)\}. \tag{52}$$

Because the relation between $\mathcal{C}(t)$ and $\hat{\mathcal{C}}(t)$ is one-to-one, we deduce that

$$\left\langle \mathrm{e}^{(\gamma-\gamma_0)Y_t} \right\rangle = \left\langle \mathrm{e}^{-\gamma Y_t} \right\rangle . \tag{53}$$

Using the fact that $\left\langle \mathrm{e}^{\gamma Y_t} \right\rangle \simeq \mathrm{e}^{\mathcal{E}(\gamma)t}$, we obtain

$$\mathcal{E}(\gamma - \gamma_0) = \mathcal{E}(-\gamma) . \tag{54}$$

By Legendre transform of this equation, the **Gallavotti-Cohen Fluctuation Theorem** for the Large Deviation Function $\Phi(j)$ (see Appendix A for a brief introduction to large deviations) is derived [25, 26, 27, 28, 29]:

$$\boxed{\Phi(j) = \Phi(-j) - \gamma_0 j .} \tag{55}$$

From the definition of the Large Deviations Function, the Fluctuation Theorem implies that in the long-time limit

$$\frac{\Pr\left(\frac{Y_t}{t} = j\right)}{\Pr\left(\frac{Y_t}{t} = -j\right)} \simeq \mathrm{e}^{\gamma_0 j} . \tag{56}$$

This symmetry property of the Large Deviations Function is valid *far from equilibrium*. This fact has been proved rigorously by various authors in many different contexts (chaotic systems, Markovian dynamics, Langevin dynamics...).

Remark: In the original works (see *e.g.*, [28] and references therein), the authors studied the Large Deviations Function for the entropy production σ. Recalling that the entropy flow is given by (48)

$$\frac{d_e S}{dt} = \frac{1}{2} \sum_{\mathcal{C},\mathcal{C}'} J_t(\mathcal{C},\mathcal{C}') \log \frac{M(\mathcal{C}',\mathcal{C})}{M(\mathcal{C},\mathcal{C}')} ,$$

one defines an entropy transfer for each jump as

$$y = \log \frac{M(\mathcal{C}',\mathcal{C})}{M(\mathcal{C},\mathcal{C}')} .$$

The increment of the entropy flow at each jump is thus given by y. A property similar to generalised detailed balance is *tautologically true* for y:

$$M_y(\mathcal{C}',\mathcal{C}) = M_{-y}(\mathcal{C},\mathcal{C}')\mathrm{e}^{\gamma_0 y} \quad \text{with} \quad \gamma_0 = 1 .$$

This relation implies a Fluctuation Theorem, given by

$$\Phi(\sigma) - \Phi(-\sigma) = -\sigma ,$$

where $\Phi(\sigma)$ is the large deviations function associated with entropy flow.

5   Jarzynski and Crooks Non-Equilibrium Identities

In this section, we describe a remarkable recent result in non-equilibrium statistical physics, which came as a surprise when it was first published by C. Jarzynski (Fig. 18).

In the section about classical thermodynamics, we have recalled, (see Eq. (6)), that the work performed on a system in contact with a heat reservoir at temperature T satisfies the relation

$$\langle W \rangle \geq F_B - F_A = \Delta F \,, \tag{57}$$

where F_A is the free energy of the initial state and F_B that of the final state. We point out that in classical thermodynamics the work performed really means an average over many experiments (*e.g.*, an operator pulling a piston enclosing a perfect gas from volume V_A to V_B, see figure 5). To emphasize this fact we have rewritten here Equation (6) with the notation $\langle W \rangle$ instead of simply W.

Figure 18: Christopher Jarzynski. His first paper on the celebrated Work Identity appeared in Physical Review Letters in 1997 [30].

Two decades ago, Christopher Jarzynski found that this classical inequality, well-known since the 19th century, can be deduced from an underlying *remarkable identity* valid for non-equilibrium systems. In the beginning, this identity was proved only for Hamiltonian systems, but Jarzynski (and others) have been extending its validity to more and more cases (such as Markovian dynamics or Langevin systems) [31, 32] and have verified it on exactly solvable models (see for example [33, 34, 35] and references therein). Experimental results [36, 37, 38] have also confirmed the Jarzynski relation which is now firmly established and is considered to be one of the few exact results in non-equilibrium statistical mechanics.

Jarzynski's Identity states that

$$\left\langle e^{-\frac{W}{k_B T}} \right\rangle = e^{-\frac{\Delta F}{k_B T}}. \tag{58}$$

The average in this equation is taken over a non-equilibrium ensemble of individual trajectories of finite duration t_f.

The precise set up of Jarzynski's identity is as follows. The system has been prepared in a canonical equilibrium state A at temperature T and is in this state for $-\infty \leq t \leq 0$ At time $t = 0$, the state is modified by an external operator according to a well-defined protocol $\lambda(t)$ that lasts for a finite period of time $0 \leq t \leq t_f$ (see Fig. 19).

- At $t = 0$, For $t \leq 0$, $\lambda(0) = \lambda_A$ and the system is *in equilibrium* in the state A.

- Between 0 and t_f, the operator acts on the system by changing a control parameter $\lambda(t)$ according to a fixed well-defined protocol which does not have to be quasi-static and which drives the system far from equilibrium.

- At t_f, the operator stops to act and the control parameter is fixed to a value $\lambda(t) = \lambda(t_f) = \lambda_B$ for $t \geq t_f$. We emphasize that the system is *not at equilibrium* at time t_f.

During the whole process, the system remains in contact with a heat-bath at temperature T. After an infinite time, it will reach the thermal equilibrium state B at temperature T. We emphasize that the protocol $\lambda(t)$ is not assumed to be 'slow'.

Jarzynski's Identity connects data related to a non-equilibrium process (the exponential work average on the left hand side of the identity) with thermodynamics (the free energy on the right hand side).
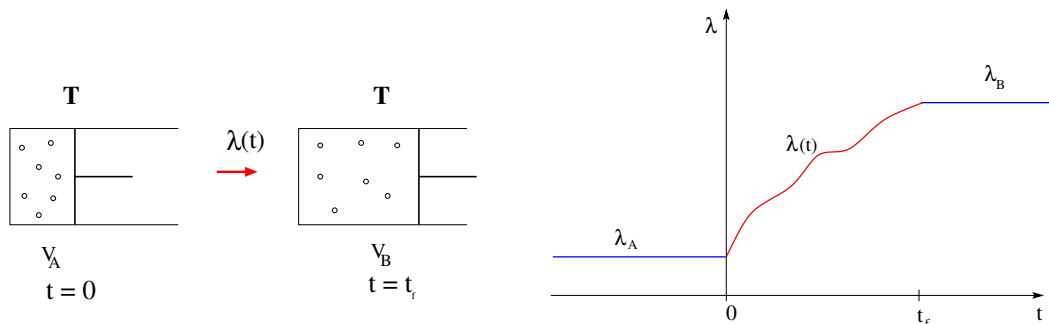

Figure 19: Set-up of Jarzynski's formula for the case of a gas in a cylinder. Here $\lambda(t)$ corresponds to the volume: the operator moves the piston according to a well-defined protocol and stops at time t_f when the volume has reached V_B.

Remarks:

1. From convexity (Jensen's inequality), we have

$$\left\langle e^{-\frac{W}{k_B T}} \right\rangle \geq e^{-\frac{\langle W \rangle}{k_B T}} \ .$$

Hence, Jarzynski's Work Theorem yields the classical inequality for the Maximum Available Work.

2. However, in order to have an equality, *there must be* individual trajectories that *do not satisfy* the classical inequality (57) i.e., there must be some realizations for which

$$W < \Delta F \quad i.e. \quad \mathcal{W}_{\text{useful}} > -\Delta F.$$

Such special occurrences are called 'Transient Violations of the Second Principle'. It must be emphasized that the Second Principle is **not** violated because the Second Principle concerns averages and states that the *average* of the performed work is greater than the free energy difference, which remains true. The Second Principle does not say anything about individual behaviour. However, in thermodynamics, we are so used to the fact that individual measurements usually reflect the typical average behaviour that we forget the fact that these two quantities can be different.

The 'transient violations' of the Second Principle can be quantified thanks to an identity due to G. Crooks [39, 40, 41] which is even more precise than Jarzynski's relation.
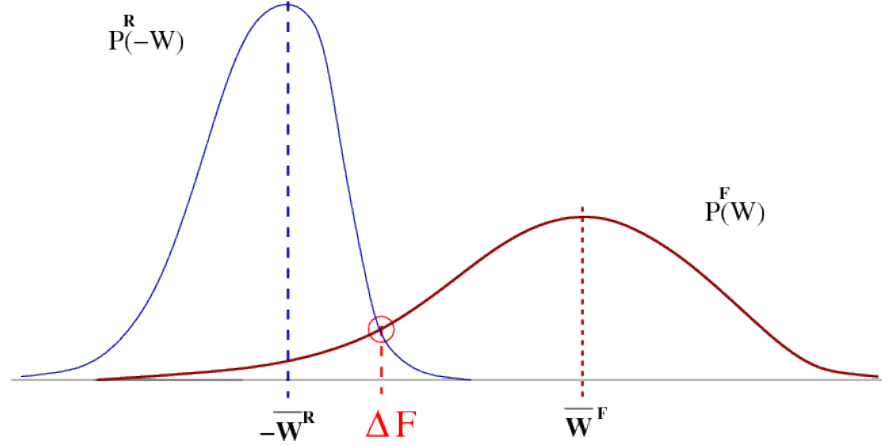

Figure 20: Graphical representation of Crooks' relation.

Let $\lambda^{\mathcal{F}}(t)$ be a protocol of duration t_f that drives the system from V_A to V_B and let $\lambda^{\mathcal{R}}(t) = \lambda^{\mathcal{F}}(t_f - t)$ be the *time reversed* protocol. It is then possible to measure the work done during the Forward process and the work done during the Reversed process. These quantities are both random variables with probability distributions $\mathrm{P}^{\mathcal{F}}$ and $\mathrm{P}^{\mathcal{R}}$, respectively. The following identity is satisfied by these probability distributions (Crooks, 1999):

$$\frac{\mathrm{P}^{\mathcal{F}}(W)}{\mathrm{P}^{\mathcal{R}}(-W)} = \mathrm{e}^{\frac{W - \Delta F}{k_B T}} . \tag{59}$$

Note the similitude between this identity and the Fluctuation Theorem in the form given in equation (56). In fact, the proof of Eq. (59) that we now sketch follows similar lines [34]. The main difference with the analysis of Section 4.2 is that the Markov matrix depends on time through the protocol $\lambda(t)$. In order to calculate the ratio between forward and backward trajectories, we need a local balance condition between the rates. We shall assume that each transition at a given time (and therefore for a given value of λ) satisfies a detailed balance condition:

$$\frac{M_\lambda(\mathcal{C}, \mathcal{C}')}{M_\lambda(\mathcal{C}', \mathcal{C})} = \mathrm{e}^{-\frac{E_\lambda(\mathcal{C}) - E_\lambda(\mathcal{C}')}{k_B T}},$$

where $E_\lambda(\mathcal{C})$ is the energy of configuration $\mathcal{C}$ for a fixed value of the parameter λ. Calculating, as in Eq. (43), the ratio of the probability of a trajectory to that of the time-reversed trajectory, we obtain, using the local detailed balance condition,

$$\frac{\mathrm{Pr}\{\mathcal{C}(t)\}}{\mathrm{Pr}\{\hat{\mathcal{C}}(t)\}} = \frac{P_{\mathrm{eq}, \lambda_0}(\mathcal{C}_0)}{P_{\mathrm{eq}, \lambda_n}(\mathcal{C}_n)} \mathrm{e}^{-\frac{1}{k_B T} \sum_{i=1}^{n} (E_{\lambda_i}(\mathcal{C}_i) - E_{\lambda_i}(\mathcal{C}_{i-1}))}, \tag{60}$$

where, following Crooks [39], the total heat transferred is defined as

$$Q = \sum_{i=1}^{n} [E_{\lambda_i}(\mathcal{C}_i) - E_{\lambda_i}(\mathcal{C}_{i-1})].$$

Besides, we have

$$\frac{P_{\mathrm{eq},\lambda_0}(\mathcal{C}_0)}{P_{\mathrm{eq},\lambda_n}(\mathcal{C}_n)} = \mathrm{e}^{-\frac{\Delta F}{k_B T} + \frac{\Delta E}{k_B T}},$$

where the variation of energy between the final and the initial configurations is $\Delta E = E_{\lambda_n}(\mathcal{C}_n) - E_{\lambda_0}(\mathcal{C}_0)$. From the First Principle $\Delta E = Q + W$, we conclude that Eq. (60) reduces to

$$\frac{\Pr\{\mathcal{C}(t)\}}{\Pr\{\hat{\mathcal{C}}(t)\}} = \mathrm{e}^{\frac{W - \Delta F}{k_B T}}, \tag{61}$$

where W represents the work done along the forward trajectory $\mathcal{C}(t)$. Summing as in over all trajectories that correspond to the same amount of work and using the fact that W is odd under time reversal, Crook's relation (59) is derived.

We now discuss some consequences of equation (59). Note that Crooks' equation implies Jarzynski's relation is a direct consequence:

$$\left\langle \mathrm{e}^{-\frac{W}{k_B T}} \right\rangle = \int \mathrm{e}^{-\frac{W}{k_B T}} \mathrm{P}^{\mathcal{F}}(W)\,\mathrm{d}W = \mathrm{e}^{\frac{-\Delta F}{k_B T}} \int \mathrm{P}^{\mathcal{R}}(-W)\,\mathrm{d}W = \mathrm{e}^{\frac{-\Delta F}{k_B T}},$$

where, to derive the last equality we have used the fact that $\mathrm{P}^{\mathcal{R}}$ is a normalized probability distribution.

We can also calculate the order of magnitude of the probability of a transient violation of amplitude $\zeta > 0$:

$$\mathrm{Prob}^{\mathcal{F}}(W < \Delta F - \zeta) = \int_{-\infty}^{\Delta F - \zeta} \mathrm{P}^{\mathcal{F}}(W)\,\mathrm{d}W = \int_{-\infty}^{\Delta F - \zeta} \mathrm{P}^{\mathcal{R}}(-W)\,\mathrm{e}^{\frac{W - \Delta F}{k_B T}}\,\mathrm{d}W$$

$$= \int_{-\infty}^{0} \mathrm{P}^{\mathcal{R}}(\zeta - \Delta F - v)\,\mathrm{e}^{\frac{v}{k_B T}}\mathrm{e}^{-\frac{\zeta}{k_B T}}\,\mathrm{d}v \leq \mathrm{e}^{-\frac{\zeta}{k_B T}}. \tag{62}$$

To derive the last inequality, we have used the fact that $\mathrm{e}^{\frac{v}{k_B T}} \leq 1$ for $v \leq 0$ and also that $\int_{-\infty}^{0} \mathrm{P}^{\mathcal{R}}\mathrm{d}v \leq 1$. The probability of a violation of amplitude $\zeta > 0$ is exponentially small with ζ, but again such violations are necessary to ensure the validity of Crooks' and Jarzynski's relations. We observe that for a transient violation to have non-vanishing probability, ζ must be of the order of $k_B T$. On the other hand, $\Delta F \sim N k_B T$, where N is the number of degrees of freedom in the system, which usually are of the order of the Avogadro number. Therefore, transient violations are one part in 10^{23} for macroscopic systems: this is totally unobservable... One has to work with very small systems, such as biophysical objects, to have the chance to observe anything.

Jarzynski and Crooks' Identities allow us to determine Equilibrium Free Energy differences by doing *Non-Equilibrium experiments*. Many experimental results have been obtained using single-molecule manipulations [36, 37, 38]. The idea behind these experiments is shown in figure 20, which represents graphically Crooks' relation (59). The forward work distribution and the distribution of *minus* the reversed work cross each other at ΔF. These distributions can be measured and plotted using non-equilibrium experimental conditions. This provides a method for determining Free Energy variations, which characterize equilibrium states, through non-equilibrium data (see Fig. 20).

6   Information Theory

Since the pioneering work of C. Shannon, Information Theory has become a major field of the Modern Age [42]. There are many books to guide the reader in the fundamentals of this subject, from the classic monograph [43] of Shannon and Weaver (Fig 21) to very recent lectures notes of E. Witten [44], focused on quantum-information. A selection of well-known references is [45, 46, 47, 48, 49, 50, 51, 52, 53]. A book that I found particularly enthusiasming is *An introduction to Information Theory. Symbols, Signals and Noise* by J.R. Pierce, a colleague and friend of Shannon, who was himself a polymath.
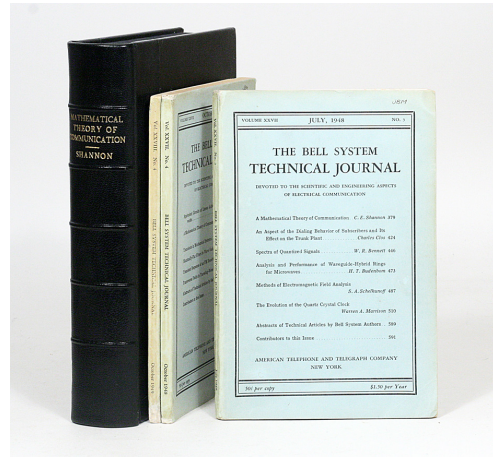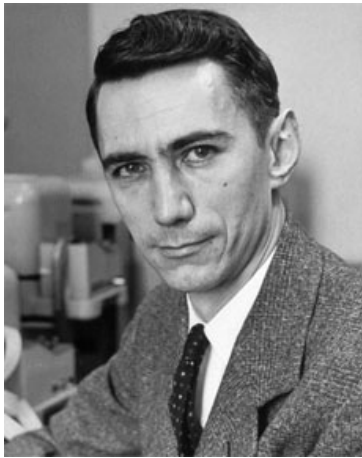

Figure 21: Claude Shannon (1916-2001), Father of the Information Age.

Because the goal of the present chapter is to focus on the interplay between Thermodynamics and Information Theory, we shall just briefly mention some basic facts useful to our aim. A detailed introduction to Information Theory will be found in the contribution of Olivier Rioul in this volume.

- Let us play a little game. Suppose I have 4 objects A, B, C and D. I choose of them (with the same probability 1/4) and I hide it. You have to determine which one I have selected by asking me *binary questions* (to which I can only answer by yes/no). What is the minimal number Q_{min} of questions you need to ask?

  Answer: $Q_{min} = 2$. Devise an optimal strategy and convince yourself that had I started with 2^N objects, you would need to ask $Q_{min} = N$ questions.

- Let's make the problem harder: I still have 4 objects A, B, C and D, but I select them with unequal probabilities: A with probability 1/2, B with 1/4, C or D with 1/8. What is the minimal number of binary questions that you need to ask me *on average*? Can you devise a strategy? *Answer:* $\langle Q_{min} \rangle = 7/4$.

- More generally, I have m objects A_i that I select with probability p_i. Is there a lower bound to the average number of binary questions $\langle Q_{min} \rangle$? Can one devise an optimal a strategy?

- I have two coins: a fair well-balanced coin with a head and a tail, and a fake coin with two heads. I take randomly one of these coins and throw it twice. I record the number of heads that occurred: what information can I deduce about the chosen coin?

- Suppose we want to transmit a message using an alphabet made of m letters, $a_1, \ldots, a_m$, where the letter a_i appears with frequency p_i (think about a natural language, English or French). Each letter is coded by a binary string and we assume that the transmission channel is perfect (noiseless). Can we devise a method to make the code words as short as possible?

  Hint: Recall the Morse code, where the most frequent letter E, was represented by a dot •

- Is it possible to transmit information through a noisy channel at a finite rate (we loosely define the rate by the ratio of the length of the message transmitted by the operator with the length of the original message) with an arbitrary small probability of error?

  Answer: YES: this is Shannon's Fundamental Theorem of Information Theory (1948). The lower bound of the transmission rate is given by a characteristic property of the channel, called its *capacity*.

The solutions to these problems and to many other questions – that would superficially appear to be totally unrelated to one another – involve a common key-concept that Shannon discovered in the late 40's. Shannon was able to quantify the concept of uncertainty, indeterminacy, (lack of) information, or surprise carried by a probability distribution $p_1, \ldots, p_m$. He proved that uncertainty can be measured in a quantitative manner. Moreover, the measure of information could be expressed by a function $H(\{p_1, \ldots, p_m\})$ that, under very reasonable assumptions, is *unique* (up to an overall normalization constant).

Shannon's H function is given by

$$H(\{p_1, \ldots, p_m\}) = - \sum_{i=1}^{m} p_i \log_2 p_i \,, \tag{63}$$

where $\log_2$ is the base 2 logarithm. H is expressed in bits.

As reported in a famous anecdote, the name of H was suggested to Shannon by von Neumann, in Shannon's words: *My greatest concern was what to call it. I thought of calling it 'information,' but the word was overly used, so I decided to call it 'uncertainty.' When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage.'*

<div align="center">**The Pandora box was open.**</div>

7   Thermodynamics and Information: The Maxwell Demon

The mathematical identicalness between Shannon's H function (63) and the formula for the Entropy in Statistical Physics (11) craves for a deep explanation. Jaynes [14]

proposed an reformulation of Statistical Mechanics in which the Shannon Entropy (63), re-expressed in the correct units, as taken as the definition of the Thermodynamic Entropy. From this point of view, the formula (11) is not the result of a combinatorial calculation à la Schrödinger but a starting point. The maximization of the Shannon Entropy under various constraints allows one to retrieve the various Gibbs ensembles and potentials of Statistical Mechanics. Furthermore, Jaynes endeavoured to extend this approach by proposing a minimal entropy production principle to analyze systems far from equilibrium [54]; however, this is only an approximate variational theory that can be considered as a linearization of more accurate dynamical fluctuation principles [55, 56, 57]. Despite its elegance, Jaynes work remained an academic curiosity. Statistical physicists and information theorists continued working on their specific problems without caring much about the other side's entropy.

However, there was one venerable problem where entropy and information had to come face to face, a puzzle that dated from the very beginning of Statistical Mechanics, Maxwell's demon (see Fig. 22). In his book *Theory of Heat* (1872), Maxwell imagined a thought experiment in which an intelligent being (that Lord Kelvin later christened as a 'demon') is able to violate the Second Law of thermodynamics by sorting out fast and slow molecules of gas in a box initially at uniform temperature:

If we conceive of a being whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are as essentially finite as our own, would be able to do what is impossible to us. For we have seen that molecules in a vessel full of air at uniform temperature are moving with velocities by no means uniform, though the mean velocity of any great number of them, arbitrarily selected, is almost exactly uniform. Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower molecules to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics.
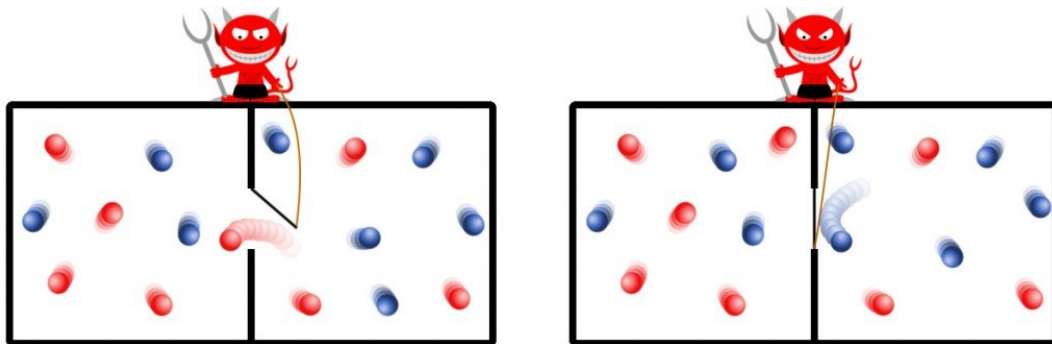

Figure 22: Maxwell's demon analyzes the speed of every particle inside the container. By allowing hot particles to enter the left-half of the box and keeping the cold particles to the right side, the demon creates raises the temperature of one chamber in comparison of the other, without expending any work, in contradiction with Clausius' statement of the Second Principle.

For the last 150 years, scientists have struggled to exorcise Maxwell's demon.

The various attempts have shed light on the relations between Thermodynamics and Information [58]. Although the debate is not fully settled yet, significant progress has been achieved and, recently, Maxwell's Demons have appeared in the laboratories, to paraphrase the words of C. Bennett [59], one of the major players in this field. Experimental aspects of the Maxwell's Demon are reviewed in [60]. Here, we shall outline some major theoretical steps in the understanding the demon puzzle.

If the demon is an autonomous and inert device, located within the gas, then it is subject to thermal fluctuations and to Brownian motion. This demon could be a trap that opens only in one direction (if it is hit by energetic molecules coming from the right side of the box, for example). Such a mechanical or electrical rectifier is nothing but a ratchet. As analyzed by Smoluchowski and Feynman (see section 2.4 and references therein) the universality of Brownian motion would prevent this demon to operate at thermal equilibrium; silly demons are easy to exorcise.

An issue much more subtle and interesting arises if one supposes that the demon is capable, somehow, to extract information from the gas from some physical measurement, to record this information and act accordingly, and, finally, to reset itself anew by discarding the acquired information once it has been used. In some sense, the demon has to act in an 'intelligent' way, while remaining analyzable by the Laws of Physics. In 1929, Leo Szilard wrote a classic paper, *On the decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings* [61], in which he proposes a simplified model for a intelligent Maxwell demon, that has become the cornerstone of most of the subsequent studies (see Fig. 23).

A Szilard engine consists in a single gas particle in a box. The demon determines which half of the box the particle is in (Step A) an inserts a partition (a movable wall) inside the box (Step B). He also attaches a pulley and a mass to the move-able wall. The particle bounces on the wall, pushes it away and performs work by raising the mass (Step C). Because the box is connected to a heat reservoir, the expansion of the gas is isothermal. When the wall reaches the left side of the box, it is removed and the cycle is completed (Step D). All told, it seems that the single-particle gas has extracted heat from the its environment to perform a total work of $k_B T \log 2$, in apparent contradiction with the Second Principle.
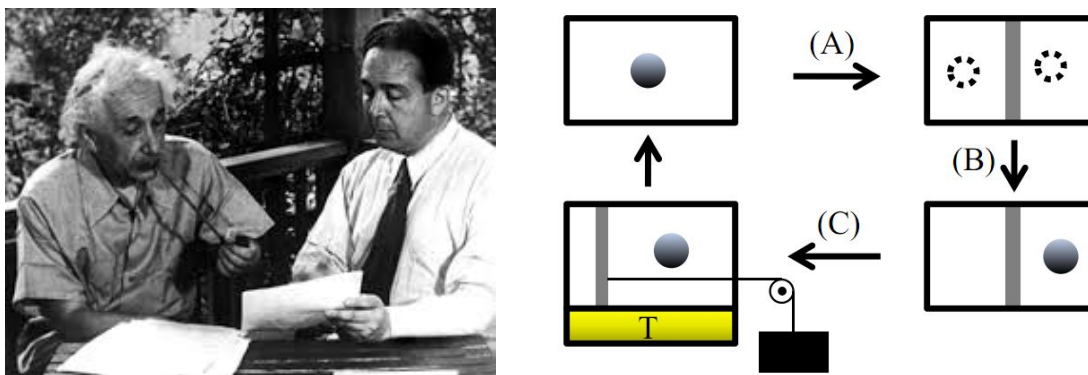

Figure 23: Leo Szilard (1898-1964) discussing with a co-worker of his. On the right: a sketch of Szilard's engine.

Szilard's engine works by acquiring a binary information (a *bit*) on the position of the particle: left or right. Szilard understood that, to save Second Principle, the

acquisition of this information must cost an entropy production, at least equal to the 'fundamental amount' $k_B \log 2$. However, it was not clear in Szilard's analysis where this thermodynamic expense was located: in the measurement process, in the recording of the information or in the erasing procedure that closes the cycle?
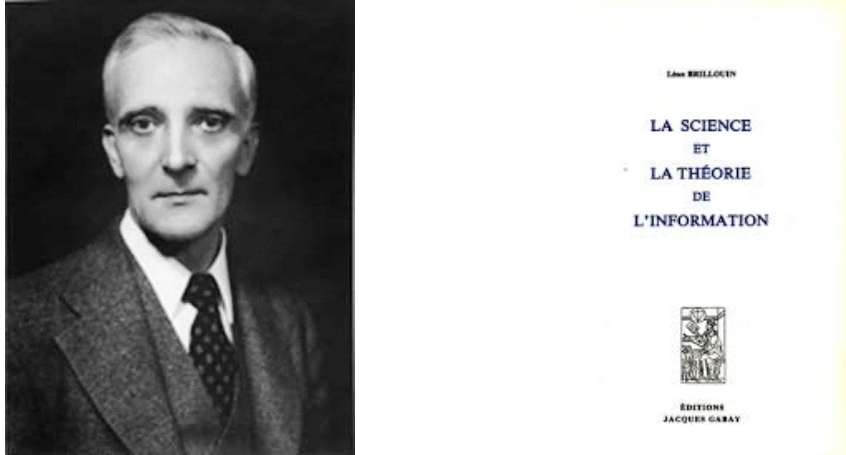

Figure 24: Léon Brillouin (1889-1969) associated Thermodynamics with Information Theory in his attempt to exorcise Maxwell's demon.

In 1951, Léon Brillouin (see fig. 24), and also, independently, Dennis Gabor, analyzed thoroughly the measurement process by postulating that the demon uses photons to acquire information, that are distributed according to Planck's blackbody law. Brillouin made a bold step: he added, in his entropy balance equations, a contribution coming from the newly discovered Shannon Information entropy to the usual Thermodynamic entropy. In other words, he argued that Information entropy and Thermodynamic entropy were directly connected and should be treated on equal footing. The conversion coefficient between information and physical entropy is given by 1 bit $= k_B \log 2$, as found by Szilard. In Brillouin's interpretation, the Second Law has to be generalized in order to exorcise Maxwell's demon. Thus, after one cycle of the Szilard engine, the enclosed gas returns to its initial state while the environment has lost a total heat $\Delta Q = -k_b T \log 2$ (corresponding to the work done by the engine) implying a decrease of the total entropy of the universe of $\Delta S = \Delta Q/T = -k_b \log 2$ in contradiction to the Second Principle. However, in step B (see Fig. 23) an information ΔI has been gained on the position of the molecule. The Shannon entropy corresponding to this information is given by $\Delta I = H = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$bit. If all sources of entropy are duly recorded in the balance sheet (with the correct conversion factor), we realize that

$$\Delta S + \Delta I \geq 0.$$

The entropy loss of the universe is accompanied by an information gain that compensates it [45, 62, 63].

One important point that was not in Brillouin's analysis was the *physical origin* of the increase of entropy due to acquisition of information. Brillouin and Gabor attributed it to the measurement process which used photons distributed according to the blackbody radiation. It appeared however that reversible measurement schemes

could be devised [67, 68] and that Brillouin's exorcism was not sufficient to get rid of the demon.

In 1961, Rolf Laudauer (see fig. 25) put forward a new Principle, relying on the fact that *'Information is Physical'*. Laudauer introduced the concept of logical irreversibility [64, 65] and analyzed *the process of erasure of information* (a similar idea was explored by O. Penrose in [66]). After completing one cycle, the demon has to set back its memory to its original state before starting afresh. According to Landauer, memory erasure is a source of heat and entropy that can never be avoided. **Landauer Principle** states that the minimum possible amount of energy required to erase one bit of information, known as the Landauer limit is given by $k_B \log 2$. Landauer's surmise can now be tested in the laboratory thanks to modern techniques. Some seminal experiments have been carried out by S. Ciliberto and coworkers at ENS Lyon [60]; their results and the contributions of other groups are described in this volume.

Figure 25: Rolf Landauer (1927-1999) discovered the Principle that bears his name in 1961: 'Information is Physical'. Charles Bennett (born in 1943) is one of the founding fathers of quantum information theory.

The story of the demon is not complete yet. Stimulated by the discoveries of Jarzynski and Crooks, the concepts of energy and entropy were applied to stochastic dynamical systems, along the lines sketched in Section 5. This new Stochastic Thermodynamics [71, 72, 73, 74] provides us with an efficient framework to formulate classical measurement and feed-back processes and to generalize the Second Principle in various settings [75, 76, 77, 78, 79, 80]. Besides, Landauer Principle can derived by using calculations of path probabilities akin to the one presented above to derive the Gallavotti-Cohen Theorem and the Crooks work relation [69, 70]. We recommend the book of T. Sagawa, based on his PhD work, for a review of the most recent discoveries [81].

8   Conclusion

The aim of this presentation is to arouse the curiosity of a non-specialist reader to the various faces of entropy and to provide him or her with fairly up to date entries to the literature. Only classical systems have been discussed here. The field of Quantum Information is huge and, of course, Maxwell's demon has a quantum

twin [82] even more impish and 'subtle', but (hopefully) 'not malicious'. Quantum aspects are reviewed in details in other contributions to this volume: one may also refer to [44] for a 'Mini-Introduction' and [83] for a classic reference. Last but not least, let us mention an ultimate haunting spirit, straddling over a box that could swallow our whole universe and our understanding: the solution of S. Hawking's Black hole Information Paradox would require a combination of quantum mechanics and general relativity, that the reader is left to conceive [84].

8.1   APPENDIX A: Large deviations and cumulant generating functions

The concept of a large-deviation function is a useful and well-known tool in probability. It will be illustrated by the following example.

Let $\epsilon_1, \ldots, \epsilon_N$ be N binary variables where $\epsilon_k = \pm 1$ with probability $1/2$ for $k = 1, \ldots N$. Suppose that the ϵ_k's are independent and identically distributed. Their sum is denoted by $S_N = \sum_1^N \epsilon_k$. We recall:

1. The *Law of Large Numbers* implies that $S_N/N \to 0$ (almost surely).

2. The *Central Limit Theorem* implies that $S_N/\sqrt{N}$ becomes a Gaussian variable of unit variance.

We now quantify the probability that S_N/N takes a non-typical value r, with $-1 < r < 1$. One can show (using the Stirling formula) that in the large N limit

$$\Pr\left(\frac{S_N}{N} = r\right) \sim \mathrm{e}^{-N\,\Phi(r)}, \tag{64}$$

where $\Phi(r)$ is given by

$$\Phi(r) = \frac{1+r}{2} \log\left(\frac{1+r}{2}\right) + \frac{1-r}{2} \log\left(\frac{1-r}{2}\right) + \log 2 \,. \tag{65}$$

The function $\Phi(r)$ is called a **Large Deviations Function**. Because of the Law of Large Numbers, we know that $\Pr(\frac{S_N}{N} = 0)$ tends to 1 when $N \to \infty$. The large deviations function must therefore vanish at $r = 0$, which is indeed the case.

More generally, let Y_t be a random variable (for example the total charge transported through a system) that depends on the time t. We assume that when $t \to \infty$, we have $\frac{Y_t}{t} \to J$, i.e. $\frac{Y_t}{t}$ converges towards its mean-value. The random variable Y_t satisfies a Large Deviations Principle if the following identity holds in the large time limit:

$$P\left(\frac{Y_t}{t} = j\right) \sim e^{-t\Phi(j)}.$$

The function $\Phi(j)$ is a *large deviations function* of the rate of production of Y_t. Note that $\Phi(j)$ is positive and vanishes at $j = J$.

Another useful quantity to consider is the moment-generating function of Y_t defined as the average value $\left\langle \mathrm{e}^{\mu Y_t} \right\rangle$. Expanding with respect of the parameter μ, we get

$$\log \left\langle \mathrm{e}^{\mu Y_t} \right\rangle = \sum_k \frac{\mu^k}{k!} \langle\langle Y^k \rangle\rangle_c \,,$$

where $\langle\langle Y^k \rangle\rangle_c$ is the k-th cumulant of Y_t. In many cases, one can show that in the long time limit, we have

$$\left\langle \mathrm{e}^{\mu Y_t} \right\rangle \simeq \mathrm{e}^{E(\mu)t} \qquad \text{when} \quad t \to \infty.$$

The function $E(\mu)$ is the cumulant generating function. The previous identity shows that all cumulants of Y_t grow linearly with time and their values are given by the successive derivatives of $E(\mu)$

It is readily shown that the large deviations function $\Phi(j)$ and the cumulant generating function $E(\mu)$ are related by Legendre transform:

$$\boxed{E(\mu) = \max_j \left(\mu j - \Phi(j) \right).}$$

Indeed, using saddle-point, we obtain

$$\left\langle \mathrm{e}^{\gamma Y_t} \right\rangle = \int \mathrm{Pr}(Y_t) \mathrm{e}^{\gamma Y_t} dY_t = t \int \mathrm{Pr}\left(\frac{Y_t}{t} = j \right) \mathrm{e}^{\gamma t j} dj \sim \int \mathrm{e}^{\gamma t j - t \Phi(j)}.$$

9   APPENDIX B: Proof of the Jarzynski Formula for Hamiltonian dynamics

We present here the original proof given by C. Jarzynski (PRL, 1997) [30]. Suppose that, for $t \leq 0$, the system is in a state A at thermal equilibrium with its environment at temperature T. Between $0 \leq t \leq t_f$, the coupling to the thermal bath is plugged out. The system is isolated and evolves deterministically according to a Hamiltonian dynamics, with Hamiltonian $H_{\lambda(t)}(p, q)$, which depends on time through the protocol $\lambda(t)$. For $t \geq t_f$, the Hamiltonian remains fixed at $H_{\lambda_B}(p, q)$. The system is reconnected to the thermal bath at T and evolves towards the equilibrium thermodynamic state B. We shall denote by $z = (p, q)$ the phase-space coordinate.

The initial distribution on phase space is the canonical distribution with Hamiltonian H_{λ_A} and $\beta = 1/kT$:

$$\mathrm{P_A}\left(z_0, t = 0 \right) = \frac{\mathrm{e}^{-\beta H_{\lambda_A}(z_0)}}{Z_A}.$$

In the time interval $[0, t_f]$, the system evolves as

$$\dot{p} = -\frac{\partial H_{\lambda(t)}}{\partial q}, \qquad \dot{q} = \frac{\partial H_{\lambda(t)}}{\partial p}$$

the initial condition $(p, q) = (p_0, q_0)$ being sampled according to P_A.

During this evolution, the work received by the system is given by

$$W = \int_0^{t_f} \mathrm{d}t\, \frac{\partial H_{\lambda(t)}(z(t))}{\partial t} = \int_0^{t_f} \mathrm{d}t\, \dot{\lambda}(t) \frac{\partial H_{\lambda(t)}(z(t))}{\partial \lambda}.$$

Note that the evolution is deterministic: the only randomness comes from the initial condition.

For a Hamiltonian dynamics, we have

$$\frac{\partial H_{\lambda(t)}}{\partial t} = \frac{d H_{\lambda(t)}}{dt}$$

and therefore

$$W = \int_0^{t_f} \mathrm{d}t\, \frac{\partial H_{\lambda(t)}(z(t))}{\partial t} = H_{\lambda_B}(z(t_f)) - H_{\lambda_A}(z(0)).$$

For a given value z_0 of $z(0)$, $z_f = z(t_f)$ is uniquely determined. Let us now calculate the exponential average of the work:

$$
\begin{aligned}
\langle \mathrm{e}^{-\beta W} \rangle &= \int \mathrm{d}z_0\, \mathrm{P}_{\mathrm{A}}(z_0) \mathrm{e}^{-\beta(H_{\lambda_B}(z_f)) - H_{\lambda_A}(z_0))} \\
&= \int \mathrm{d}z_0\, \frac{\mathrm{e}^{-\beta H_{\lambda_A}(z_0)}}{Z_A} \mathrm{e}^{-\beta(H_{\lambda_B}(z_f)) - H_{\lambda_A}(z_0))} \\
&= \int \mathrm{d}z_0\, \frac{\mathrm{e}^{-\beta H_{\lambda_B}(z_f)}}{Z_A}.
\end{aligned}
\tag{66}
$$

To conclude, we must make the change of variables $z_0 \to z_f$. This is a one-to-one mapping by the Hamiltonian flow. The key remark is that the Jacobian of this transformation is equal to 1, because of the *Liouville Theorem*, we have $\mathrm{d}z_0 = \mathrm{d}z_f$.

This concludes the original proof given by Jarzynski:

$$
\langle \mathrm{e}^{-\beta W} \rangle = \frac{1}{Z_A} \int \mathrm{d}z_f \left\| \frac{dz_0}{dz_f} \right\| \mathrm{e}^{-\beta H_{\lambda_B}(z_f)} = \frac{Z_B}{Z_A} = \mathrm{e}^{-\beta \Delta F}.
$$

This Hamiltonian proof is mathematically rigorous but from the physics point of view, plugging/unplugging the thermal environment seems artificial. In order to overcome this difficulty, one has to give a mechanical representation of heat exchanges. Here are some ways for achieving this task:

- Take the ensemble 'System + Thermostat' as a big isolated system, governed by a full Hamiltonian.

- Represent the heat-exchanges between the System and Thermostat by a specific (non-Hamiltonian) dynamics: Nosé-Hoover...

- Add to the System's Hamilton equations, a thermal Langevin noise and a dissipative friction term, related by the FDT that represent the effect of the heat bath.

- Model thermal interaction by a Markov Dynamics.

One should be aware that Work Theorems are 'metatheorems': you have to formulate Jarzynski's identity precisely in different contexts and then to give proofs for various settings (more than 15 different proofs are available).

Acknowledgments

References

[1] R. Kubo, *Thermodynamics*. (North Holland, Amsterdam, 1968).

[2] R. Emden, *Why do we Have Winter Heating?* Nature **141**, 908 (1938).

[3] A. Fuchs, *La thermodynamique et l'effet Montesquieu : les usages de l'entropie hors du champ de la thermodynamique.* Publication du Service Enseignements Supérieurs - Didactique de la Chimie (1988).

[4] M. W. Zemansky and R. H. Dittman, *Heat and Thermodynamics.* (McGraw Hill Book Company, 1981).

[5] E. Fermi, *Thermodynamics.* (Dover, New-York, 1956).

[6] A. B. Pippard, *The Elements of Classical Thermodynamics.* (Cambridge University Press, Cambridge, 1964).

[7] D. Lemons, *Mere Thermodynamics.* (The John Hopkins University Press, Baltimore 2009).

[8] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics.* (John Wiley and Sons, New Jersey 1985).

[9] A. M. Steane, *Thermodynamics.* (Oxford University Press, Oxford 2017).

[10] F. Reif, *Statistical Physics: Berkeley Physics Course Vol. 5.* (McGraw-Hill Book Company, 1967).

[11] A. Greven, G. Keller and G. Warnecke Eds., *Entropy.* (Princeton University Press, Princeton, 2003).

[12] E. Schrödinger, *Statistical Thermodynamics.* (Dover, New-York, 1989).

[13] J. Machta, Entropy, information, and computation. Am. J. Phys. **67**, 1074 (1999).

[14] E. T. Jaynes, Information Theory and Statistical Mechanics. Phys. Rev. **106**, 620 (1957); Phys. Rev. **108**, 171 (1957).

[15] S. Chandrasekhar, Stochastic Problems in Physics and Astronomy. Rev. Mod. Phys. **15**, 1 (1943).

[16] N. Van Kampen, *Stochastic Processes in Physics and Chemistry.* (North Holland, Amsterdam, 1992).

[17] B. Duplantier, *Brownian Motion, 'Diverse and Undulating'.* In: Damour T., Darrigol O., Duplantier B., Rivasseau V. (eds) Einstein, 1905–2005. Progress in Mathematical Physics, vol 47. Birkhuser Basel, arXiv:0705.1951.

[18] R. P. Feynman, R. B. Leighton and M. Sands, *Feynman's Lectures in Physics.* Vol.1, Chapter 46. (Basic Books; New Millennium ed., 2011).

[19] H. Spohn, *Large Scale Dynamics of Interacting Particles.* (Springer, Berlin, 1991).

[20] M. O. Magnasco, Szilard's heat engine. Europhys. Lett. **33**, 583 (1996).

[21] F. Jüllicher, A. Ajdari, J. Prost, Modeling Molecular Motors. Rev. Mod. Phys. **69**, 1269 (1997).

[22] C. Jarzynski and O. Mazonka, Feynman's Ratchet and Pawl, an exactly solvable model. Phys. Rev. E **59**, 6448 (1999).

[23] R. Dean Astumian and P. Hänggi, *Brownian Motors.* Phys. Today p.33 (Nov. 2002).

[24] A. Lau, D. Lacoste, K. Mallick, Nonequilibrium Fluctuations and Mechanochemical Couplings of a Molecular Motor. Phys. Rev. Lett. **99**, 158102 (2007); Fluctuation theorem and large deviation function for a solvable model of a molecular motor. Phys. Rev. E **78**, 011915 (2008).

[25] G. Gallavotti and E. D. G. Cohen, Dynamical ensembles in stationary states. J. Stat. Phys. **80**, 931 (1995).

[26] D. J. Evans and D. J. Searles, The fluctuation theorem, Adv. Phys. **51**, 1529 (2002).

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

[28] J. L. Lebowitz and H. Spohn, A Gallavotti-Cohen Type Symmetry in the Large Deviation Functional for Stochastic Dynamics. J. Stat. Phys. **95**, 333 (1999).

[29] G. Gallavotti, Entropy production in nonequilibrium thermodynamics: a point of view. Chaos **14**, 680 (2004).

[30] C. Jarzynski, Nonequilibrium equality for free energy differences. Phys. Rev. Lett. **78**, 2690 (1997).

[31] C. Jarzynski, Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. Phys. Rev. E **56**, 5018 (1997).

[32] C. Jarzynski, Rare Events and the convergence of exponentially averaged work values. Phys. Rev. E **73**, 046105 (2006).

[33] C. Jarzynski, Nonequilibrium work relations: foundations and applications. Eur. Phys. J. B **64**, 331 (2008).

[34] E. Boksenbojm, B. Wynants and C. Jarzynski, *Nonequilibrium thermodynamics at the microscale: work relations and the second law.* Physica A **389**, 4406 (2010).

[35] C. Jarzynski, Equalities and Inequalities: Irreversibility and the Second Law of thermodynamics at the Nanoscale. Annu. Rev. Cond. Mat. Phys. **2**, 329 (2011).

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

[37] C. Bustamante, J. Liphardt, F. Ritort, *The Non-Equilibrium Thermodynamics of Small Systems.* Phys. Today p.43 (July 2005).

[38] F. Ritort, Sem. Poincaré **2**, 193 (2003); cond-mat/0401311.

[39] G. E. Crooks, Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. J. Stat. Phys. **90**, 1481 (1998).

[40] G. E. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys. Rev. E **60**, 2721 (1999).

[41] G. E. Crooks, Path-ensemble averages in systems driven far from equilibrium. Phys. Rev. E **61**, 2361 (2000).

[42] J. Gleick, *Information: A History, a Theory, a Flood.* (Fourth Estate Ltd, 2012).

[43] C. E. Shannon and W. Weaver, *Mathematical Theory of Communication.* (University of Illinois Press, 1963).

[44] E. Witten, A Mini-Introduction To Information Theory. arXiv:1805.11965

[45] L. Brillouin, *La Science et la Théorie de l'information.* (Masson, Paris, 1959).

[46] J. R. Pierce, *An introduction to Information Theory. Symbols, Signals and Noise.* (Dover, New-York, 1980).

[47] S. Kullback, *Information Theory and Statistics.* (Dover, New-York, 1997).

[48] R. B. Ash, *Information Theory.* (Dover, New-York, 1990).

[49] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* (John Wiley and Sons, New Jersey, 2006).

[50] O. Rioul, *Théorie de l'information et du codage.* (Lavoisier, Paris, 2007).

[51] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms.* (Cambridge University Press, Cambridge, 2004).

[52] M. Zinsmeister, *Thermodynamic Formalism and Holomorphic Systems.* SMF/AMS Texts and Monographs, Vol. 2 (1996).

[53] M. Mézard and A. Montanari, *Information, Physics and Computation.* (Oxford University Press, Oxford, 2009).

[54] E. T. Jaynes, The minimum entropy production principle. Ann. Rev. Phys. Chem. **31**, 579 (1980).

[55] C. Maes and K. Netocny, *Minimum entropy production principle.* Scholarpedia.

[56] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio and C. Landim, Macroscopic fluctuation theory. Rev. Mod. Phys. **87**, 593 (2015).

[57] B. Derrida, Non-equilibrium steady states: fluctuations and large deviations of the density and of the currents. J. Stat. Mech. P07023 (2007).

[58] H. S. Leff and A. F. Rex, *Maxwell's Demon 2, Entropy, Classical and Quantum Information, Computing.* (IOP Publishing, Bristol, 2003).

[59] C.H. Bennett and B. Schumacher, *Maxwell's Demons Appear in the Lab..* Nikkei Science (Aug 2011).

[60] E. Lutz and S. Ciliberto, Information: From Maxwell's demon to Landauer's eraser. Physics Today **68**, 30, (Sept. 2015).

[61] L. Szilard, On the decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings. Z. Phys. **53**, 840 (1929), [Reprinted in [58]].

[62] P. Rodd, Some Comments on Entropy and Information. Am. J. Phys. **32**, 333 (1964).

[63] K. Maruyama, F. Nori and V. Vedral, Colloquium: The physics of Maxwell's demon and information. Rev. Mod. Phys. **81**, 1 (2009).

[64] R. Landauer, Information is Physical. Physics Today **44**, 23, (Mai 1991).

[65] R. Landauer, Irreversibility and Heat Generation in the Computing Process. IBM J. Res. Dev. **5**, 183 (1961).

[66] O. Penrose, *Foundations of Statistical Mechanics.* (Pergamon Press, Oxford, 1970).

[67] C. H. Bennett, Demons, Engines and the Second Law. Sci. Am. **257**, 108 (Nov. 1987).

[68] C.H. Bennett, The thermodynamics of computation, a review. Int. J. Theor. Phys. **21**, 905 (1982).

[69] B. Piechocinska, Information erasure. Phys. Rev. A **61**, 062314 (2000).

[70] D. Mandal and C. Jarzynski, Work and information processing in a solvable model of Maxwell's demon. PNAS **109**, 11641 (2012).

[71] K. Sekimoto, Langevin Equation and Thermodynamics. Prog. Theor. Phys. Supp. **130**, 17 (1998).

[72] U. Seifert, Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem. Phys. Rev. Lett. **95**, 040602 (2005).

[73] U. Seifert, Stochastic thermodynamics, fluctuation theorems, and molecular machines. Rep. Prog. **75**, 126001 (2012).

[74] C. Van Den Broeck, *Stochastic thermodynamics: A brief Introduction.* Proc. Int. School of Physics Enrico Fermi, Volume 184: Physics of Complex Colloids, C. Bechinger, F. Sciortino and P. Ziherl Eds, (IOS, Amsterdam; SIF, Bologna, 2014).

[75] J. M. R. Parrondo, J. M. Horowitz and T. Sagawa, Thermodynamics of Information. Nature Phys. **11**, 131 (2015).

[76] T. Sagawa and M. Ueda, Second Law of Thermodynamics with Discrete Quantum Feedback Control. Phys. Rev. Lett. **100**, 080403 (2008).

[77] T. Sagawa and M. Ueda, Minimal Energy Cost for Thermodynamic Information Processing: Measurement and Information Erasure. Phys. Rev. Lett. **102**, 250602 (2009); Phys. Rev. Lett. **106**, 189901(E) (2011).

[78] T. Sagawa and M. Ueda, Generalized Jarzynski Equality under Nonequilibrium Feedback Control. Phys. Rev. Lett. **104**, 090602 (2010).

[79] S. Toyabe, T. Sagawa, M. Ueda, E Muneyuki and M. Sano, Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. Nature Phys. **6**, 988 (2010).

[80] T. Sagawa and M. Ueda, Fluctuation Theorem with Information Exchange: Role of correlations in Stochastic Thermodynamics. Phys. Rev. Lett. **109**, 180602 (2012).

[81] T. Sagawa, *Thermodynamics of Information Processing in Small Systems.* (Springer Verlag, Japan, 2013).

[82] R. J. Scully, *The Demon and the Quantum.* (Wiley-VCH, Singapore 2010).

[83] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information.* (Cambridge University Press, Cambridge, 2000).

[84] J. Preskill, Do Black Holes Destroy Information? arXiv:hep-th/9209058; L. Susskind, Three Lectures on Complexity and Black Holes, arXiv:1810.11563.