

## 2

# Statistical Mechanics and Thermodynamics

The fundamental physical laws describe the dynamics of physical systems only at a very microscopic level. In order to make predictions about the phenomena we see around us, we have to make *inferences* about the compounded effect of the physical laws acting on all the microscopic variables in parallel. In general, analytical derivations alone are incapable of deducing the macroscopic phenomena from the underlying laws. Statistical mechanics can be considered the first theory that successfully inferred the stationary macroscopic behavior of such systems as fluids, gases, and solids. It explained thermodynamic concepts such as *heat*, *temperature*, and *entropy* that cannot be simply understood, let alone deduced, from the underlying microscopic laws. It has by now become clear that the essentials of the statistical mechanical approach lie in the method of *statistical inference* that is used to move from the microscopic level of the physical laws to the macroscopic level of the thermodynamic states. This methodology has far more general applicability than equilibrium thermodynamics. Since this methodology will be used in an evolutionary context later on, we will introduce the essentials of this approach in this chapter.

In the course of the last three centuries, the fundamental laws of physics have been formulated at an ever more microscopic level. While Newton's theory of gravitation described such day to day phenomena as the falling of apples and the orbit of the moon, one can hardly imagine concepts further away from the experiences of our daily life than the currency of contemporary physics: electrons, quarks, gauge-fields, and space-time curvatures. Ironically, even those electrons and quarks are, in a certain sense, metaphors for the much more abstract mathematical concepts that feature in the fundamental laws.

What then is their bearing on the experiences of our daily life? On one side, a reductionist may stress that knowledge of the fundamental laws combined with explicit and detailed knowledge of the boundary conditions would *in principle* provide enough information to derive all the macroscopic phenomena from the underlying laws. On the other side, an "emergentist" might argue that it is not the fundamental laws themselves, but "emergent" higher level concepts, which are central to our understanding of the world around us. Both are right, and both are in danger of neglecting the essential facts that might marry their seemingly different viewpoints.

Two obstacles lie between a strong reductionist viewpoint and any useful application. First, the fact that the explicit and detailed knowledge of the boundary conditions, which are necessary to analytically derive macroscopic phenomena from the microscopic laws, are consistently absent in practical situations. Only in very exceptional situations, far

removed from day to day reality, and specifically set up in physics laboratories, are the boundary conditions sufficiently controlled for the fundamental laws to be able to make predictions of behavior that are testable.

Second, even if all boundary conditions are known specifically, it still may be computationally infeasible to derive predictions of macroscopic behavior from the fundamental laws. That is, one might in principle write down the equations whose solutions would give the desired predictions, but solving these equations might be impossible either analytically or numerically. Both these obstacles occur as a rule rather than as an exception. For instance, the common occurrence of deterministic chaos in nonlinear differential equations shows that uncertainties in the boundary conditions, no matter how small, may be amplified so rapidly that detailed predictions from knowledge of the underlying microscopic dynamics are generally impossible.

Considering these facts, the “emergentist” is right that the main problem is to find higher level representations of macroscopic systems that are able of making useful predictions. First, the variables in these higher level representations should be open to direct measurement, unlike, for instance, the joint quantum state of all atoms and electrons in a system. And second, one should be able to make predictions of macroscopic behavior in terms of these variables.

The search for an informative representation is however not independent of the fundamental laws. It would be foolish to ignore that the macroscopic phenomena that we see around us are rooted in the fundamental laws of physics. This is, of course, the sense in which the reductionist is right. Any description on an “emergent” higher level should in the end be understood in terms of the underlying laws. This however does not imply that this higher level *description* is somehow contained in the fundamental laws. The situation is similar to after just being taught the rules of the game of Go, trying to understand the play of two Japanese professionals. From the reductionist point of view, even the most subtle play in the game of Go doesn’t require any other initial information than its rules. On the other hand, the process of taking these rules and *unfolding* them, so to speak, towards an understanding of high level behavior (i.e. professional game play), is so vastly complicated and rich, that it has been studied by Go scholars for centuries, and has filled hundreds and hundreds of books. The contents of these books, for instance, are not contained in the rules of Go themselves, they *emerge* through human investigations of the consequences of the rules in *playing* Go.

The topic of this chapter is then to briefly discuss the methodologies that have been developed to relate underlying microscopic laws to macroscopic behavior.

## 2.1 Macroscopic Laws

Many of the common “laws” of physics taught in high school are very far from the fundamental microscopic laws but are in fact macroscopic “emergent” laws. For the purpose of illustrating the ubiquity of macroscopic laws, I will discuss a couple of such macroscopic laws, how they represent a system’s state, and how they relate to the underlying microscopic laws.

### 2.1.1 Elasticity: Hooke's Law

Imagine a beam of some material that has a length  $L$  when no forces are exerted on it. We then start pulling or pushing on this beam, with a force  $F$ . Hooke's law states that the change in length of this beam  $\Delta L$  is proportional to the force  $F$  that is exerted:

$$F = k \Delta L. \quad (2.1)$$

The constant  $k$  is called the *force constant* and is material dependent. Instead of representing the state in terms of the locations of all molecules in the material, let alone in terms of some joint quantum state, the above formula relates its macroscopic length to a macroscopic "force" that is being exerted. This macroscopic law can be understood in terms of the forces between the molecules in the material that resist being pushed together or pulled apart. This resistance in turn should eventually be explained in terms of the quantum electrodynamics of the electrons in the solid. Deriving Hooke's law, and its limitations, from quantum electrodynamics is hardly possible however. The interactions between the molecules in the material are highly complex in general and Hooke's law is only an approximate and largely phenomenological law. For small forces and deformations it does work to some extent, and it nicely explains the common occurrence of oscillations in mechanical systems.

### 2.1.2 Electric current: Ohm's Law

Ohm's law states that the local current density in a material is proportional to the local electric field. In a more familiar form, it states that the current  $I$  through a wire in a circuit is proportional to the Voltage drop  $V$  over the wire:

$$V = I R, \quad (2.2)$$

where  $R$  is the *resistance* of the wire. Again, Ohm's law doesn't formulate the behavior of the system in terms of its microscopic constituents such as the electrons moving through the wire, or the atoms that make up the wire. Similarly, the resistance  $R$  is a macroscopic quantity that cannot be easily derived from microscopic laws—i.e. quantum electro dynamics in this case. Note also that it isn't obvious at all from the microscopic laws that (for most materials) there is a material dependent *constant* that relates the electron flux through the material to the potential difference over the material. Of course, Ohm's law is only an approximate law. It does not work for many materials, such as semi-conductors, and it is in blatant contradiction with phenomena such as super-conductivity. It is also an oversimplification to call resistance a constant. Resistance is known to be dependent on external conditions such as temperature. But in many situations, Ohm's law works fine, and electrical circuits can be constructed using it. In this sense, Ohm's law is much more useful than the fundamental laws when one is concerned with simple applications.

### 2.1.3 Diffusion: Fick's Law

Fick's law relates the time derivative of the concentration of particles at a certain location to the Laplacian of the concentration at that location:

$$\frac{\partial C}{\partial t} = D \nabla^2 C, \quad (2.3)$$

where  $D$  is the *diffusion coefficient*. The above macroscopic law describes how concentrations of particles smear out over a volume over time. Relating this macroscopic law to the underlying microscopic motions of the particles is somewhat less straightforward than in the case of Ohm's law. There it was still possible *in principle* to derive Ohm's law from quantum electro dynamics. In contrast, the original derivation of the above law, given by Einstein, might be considered one of the first examples of *statistical mechanics*. It involves assuming that collisions between a reference particle and background particles randomize the movements of this reference particle. One then calculates the probability distributions of where a reference particle might end up after a short period of time. Notice that this derivation involves the introduction of *random* influences on the reference particle. Although the laws of Newtonian mechanics in principle give a deterministic description of the movement and collisions of all particles, the macroscopic law can only be derived by introducing some averaging procedure that effectively treats the movements of a reference particle as random. This is an essential ingredient of the statistical mechanical approach: details of the microscopic state are modeled as *random* in order to obtain the macroscopic behavior. This is something different than what could be called a derivation of the macroscopic behavior from the underlying microscopic laws. It additionally involves somehow taking into account the locations and velocities of all particles without any specific knowledge of them. A general procedure for doing this will be described below.

### 2.1.4 Ideal Gas: Equation of State

Lastly, the ideal gas law

$$PV = nkT, \quad (2.4)$$

relates the macroscopic variables pressure  $P$ , volume  $V$ , temperature  $T$ , and particle number  $n$  to each other. The constant  $k$  is Boltzmann's constant. In this equation, all variables that appear are clearly macroscopic variables. The particle number  $n$  and the pressure  $P$  can both be understood in a straightforward way from the microscopic variables, pressure just being the sum of the forces applied to the walls by all the microscopic particles per unit of area. The temperature  $T$ , however, has a much more complicated interpretation—for instance, it is often interpreted as the average energy per degree of freedom in the system. Again, deriving this law from the microscopic intuition of a large number of particles bouncing around in a box involves assuming some *randomness*. Doubling the size of the box makes the particles hit the walls half as often, making the pressure half as large. This for instance assumes that particles don't cluster together in some complicated distribution but spread randomly through the box. For particles of an ideal gas, the only degrees of freedom are their velocities. Since the particle's energy is proportional to the *square* of its speed, doubling the temperature roughly amounts to

multiplying the particle speeds by  $\sqrt{2}$ <sup>1</sup>. Thus, particles hit the walls  $\sqrt{2}$  times as often, but also  $\sqrt{2}$  times as hard, leading to a doubling of the pressure.

## 2.2 Equilibrium Statistical Mechanics

We have seen above that many of the (older) laws of physics are in fact macroscopic laws that relate macroscopic quantities to each other. For most practical cases, these macroscopic (approximate) laws are of much more use than stating the detailed microscopic laws. If one builds a circuit one is not generally interested in the quantum wave functions of all the electrons that provide the current. Moreover, it is experimentally entirely infeasible to measure the state of the system in such detail. Similar considerations of course apply to gas particles in a box. The essential problem then becomes how to deal with a lack of *information*. That is, how is one to infer predictions of macroscopic quantities given that one lacks knowledge of the detailed microscopic state of a system but only has knowledge of the values of some macroscopic quantities? Stated in this way, it becomes clear that the problem is not that of constructing some new *physical theory* but rather simply a problem of *inference* from incomplete information. What is needed is a combination of the information available from the microscopic laws and the information about the state of its system in terms of some macroscopic quantities to predict other macroscopic quantities of interest. It turns out that there is a definite and very straightforward procedure that describes how this statistical inference is to be performed. This procedure goes under the name of the “maximum entropy method”. Below we will introduce this general method and show how equilibrium statistical mechanics provides an example of this method.

### 2.2.1 The Maximum Entropy Method

Consider a system with  $N$  possible microscopic states  $i \in \{1, 2, \dots, N\}$ . We want to make predictions about measurable macroscopic quantities of this system, given that we only have some very limited knowledge of the system’s state. For instance, we only know that some quantity  $f$ —such as the total current through a wire, or the pressure of a gas—is fixed at a certain value. Alternatively, we even might not know that  $f$  is fixed, but only know its average value  $\langle f \rangle$ . We further know that for each possible microstate  $i$  the quantity  $f$  takes on some value  $f_i$ . For instance, we know the pressure of the gas given the exact locations and velocities of all particles contained in it. Other than that, we have no information whatsoever about which of the microscopic states  $i$  the system is more or less likely to be in. We now want to predict the value  $\langle g \rangle$  of some other observable quantity  $g$ . To this end we will have to construct a distribution  $\{p_i\}$  of probabilities to find the system in a microscopic state  $i$ . Since  $N$  tends to be very large for physical systems, we have to determine a huge number of probabilities  $\{p_i\}$  and this may seem an impossible task given that we only have *one* constraint (the value  $\langle f \rangle$ ) to go by. This would indeed be a problem if we were to analytically *deduce* the probabilities  $\{p_i\}$ , but what we want to do amounts to statistically inferring the most

---

<sup>1</sup>Note that this reasoning implicitly assumes some features of the distribution of particle speeds.

reasonable assignment of probabilities  $\{p_i\}$  given our knowledge  $\langle f \rangle$ . The maximum entropy method solves this problem by prescribing that one **choose that assignment of probabilities  $\{p_i\}$ , which is consistent with our knowledge about  $f$ , and has the highest “uncertainty” associated with it.**

The uncertainty of a distribution  $\{p_i\}$  is quantified by its *entropy*  $H$  which is defined as

$$H = -k \sum_{i=1}^N p_i \log(p_i), \quad (2.5)$$

where  $k$  is a proportionality constant. The entropy  $H$  is the central quantity in information theory, as originally developed by Shannon [127, 128]. It quantifies our degree of ignorance, or lack of knowledge or uncertainty, about the specific microscopic state  $i$  of the system given the probability distribution  $\{p_i\}$ . Viewed slightly differently, it quantifies the amount of information one would obtain if one were to be told what microscopic state  $i$  the system is actually in. One can think of it as proportional to the average number of independent yes/no questions that need to be answered before one knows the exact microscopic state of the system. If one, for instance, knows the exact microscopic state  $i$  of the system, the entropy is zero. If one knows nothing whatsoever about the system’s state, the entropy equals  $k \log(N)$ . In appendix A entropy and some related information theoretic concepts are introduced. There it will also be shown that the functional form (2.5) is unique given a small set of axioms.

If one accepts this interpretation of entropy, the prescription to maximize it becomes nothing other than common sense. If one were to choose a distribution  $\{p_i\}$  that does *not* maximize the entropy, one would choose a distribution that has less uncertainty associated with it than the maximum entropy distribution. Such a choice amounts to making additional implicit assumptions about the state of the system which are not warranted based on the information that we have. The maximum entropy prescription is nothing else than demanding that we honestly confess our ignorance to the fullest of its extent: Know thy ignorance!

We can now apply this method to the problem of predicting the expected value  $\langle g \rangle$  of a quantity  $g$ . If we know that the system has some fixed value  $f_c$  for the variable  $f$ , we will find a maximum entropy distribution  $\{p_i\}$  which is zero for all  $i$  that have  $f_i \neq f_c$ , and *uniform* over all  $i$  that have  $f_i = f_c$ . As was of course to be expected, all microstates  $i$  that have a value  $f_i$  which does not equal  $f_c$  are excluded. That one should take a uniform distribution over the other states is of course intuitive but not entirely obvious. Equation (2.5) has the feature that it is maximized when all (nonzero)  $p_i$  are taken equal. Our prediction  $\langle g \rangle$  for the quantity  $g$  becomes in this case:

$$\langle g \rangle = \sum_{\{i|f_i=f_c\}} \frac{g_i}{N(f_c)}, \quad (2.6)$$

where  $N(f_c)$  is the number of microstates  $i$  that have  $f_i = f_c$ , and  $g_i$  is the value of the quantity  $g$  for microstate  $i$ .

In the case where we only know the average value  $\langle f \rangle$  of the quantity  $f$ , the maximum entropy distribution  $\{p_i\}$  can be found using the method of Lagrange multipliers

## 2.2 Equilibrium Statistical Mechanics

---

which is explained in appendix B. The answer is

$$p_i = \frac{1}{Z(\lambda)} e^{-\lambda f_i}, \quad (2.7)$$

where the normalization constant  $Z(\lambda)$  is given by

$$Z(\lambda) = \sum_i e^{-\lambda f_i}. \quad (2.8)$$

The parameter  $\lambda$  is determined by demanding that

$$\sum_i f_i p_i = \langle f \rangle. \quad (2.9)$$

Our prediction for  $\langle g \rangle$  now becomes

$$\langle g \rangle = \frac{1}{Z(\lambda)} \sum_i g_i e^{-\lambda f_i}. \quad (2.10)$$

In a certain sense, this is all there is to statistical mechanics. For example, an isolated physical system cannot exchange matter or energy with its surroundings. Its energy  $E$  is therefore constant. In this case, given that the energy of the system is  $E$ , the maximum entropy method prescribes that *all* microstates  $i$  with this energy  $E$  are equally likely to occur. The maximum entropy distribution is a uniform distribution over all states  $i$  that have energy  $E$ . In statistical mechanics, this is known as the *micro-canonical* distribution.

In the case where a system is closed (no exchange of matter but possible exchange of energy) one generally only knows the average energy  $\langle E \rangle$  of the system. In this case, the maximum entropy distribution is given by equation (2.7), with  $f_i$  in the role of the energy of state  $i$ . The parameter  $\lambda$  is inversely proportional to *temperature* in this case. This distribution is commonly referred to as the Boltzmann distribution. In cases where the system can exchange matter as well as energy, one would generally condition the distribution on the average energy  $\langle E \rangle$  and the average number of particles  $\langle n \rangle$ . In this case, the distribution looks like

$$p_i = \frac{1}{Z(\beta, \mu)} e^{-\beta(E_i + \mu n)}, \quad (2.11)$$

where the standard notation  $\beta = 1/(kT)$  for inverse temperature and  $\mu$  for *chemical potential* is being used. Essentially all distributions over microstates that are commonly used in statistical mechanics can be derived in this almost trivial way. One specifies the macroscopic variables that are used to describe a system and derives the maximum entropy distribution given their average values. Other quantities (such as specific heats etc.) can then be derived by using the maximum entropy distribution.

Just to mention some more examples of maximum entropy distributions: If a quantity  $x$  can take on values in  $\mathbb{R}$ , and we only know its mean  $\mu$  and variance  $\sigma^2$ , its maximum entropy distribution is a *Gaussian* distribution:  $p(x) \propto \exp(-(x - \mu)^2/\sigma^2)$ . More

generally, if we only know the average values of functions  $f_i(x)$  of the variable  $x$ , the maximum entropy distribution is given by  $p(x) \propto \exp(\sum_i c_i f_i(x))$ , where the  $c_i$  are chosen such as to obey the constraints  $\langle f_i(x) \rangle = f_i$ . For example, if we only know the average *scale*  $\langle \log(x) \rangle = s$  of a positive variable  $x$ , the maximum entropy distribution is a power-law distribution:  $p(x) \propto x^\alpha$ , where the exponent  $\alpha$  is chosen such that  $\langle \log(x) \rangle = s$ . Appendix B describes a general method for deriving such maximum entropy distributions.

Why would predictions based on maximum entropy distributions be successful? Note that the maximum entropy methodology does not guarantee success. It might fail for several reasons. First of all, we predict the average  $\langle g \rangle$  of a quantity  $g$  from the maximum entropy distribution, but if the variance  $\langle g^2 \rangle - \langle g \rangle^2$  (calculated from the maximum entropy distribution as well) is large compared to the mean, we cannot expect our prediction for  $\langle g \rangle$  to be accurate for any particular case. Thus, accurate predictions for  $\langle g \rangle$  can only be expected when the variance is much smaller than the mean. For the systems that are typically studied in statistical mechanics the size of the variance relative to the mean scales (away from phase transition points) with the inverse of the square root of the number of particles. Since the number of particles in these systems is very large (order  $10^{23}$ ), the predictions are typically very accurate.

However, even if the relative variance is small, it may still turn out that the predictions give wrong answers. In this case, we are forced to conclude that our counting of the microstates is wrong. For instance, if an experimenter derives maximum entropy distributions for a box of gas based on volume and particle number alone, his predictions are not likely to agree with experiments. Moreover, the outcomes on hot summer days will turn out to be consistently different from the outcomes on cold winter days. His predictions failed because he has taken into account microstates with very different mean energies (temperatures) while the mean energy of his system is a well defined macroscopic quantity that varies with temperature. If he had taken the mean energy or temperature into account, he would have obtained correct predictions.

One might think of this as a weakness of the theory but in fact it is one of its strengths. The failure of the predictions, even when the relative variance was small, points us to additional macroscopic variables that have to be taken into account. In fact, new laws of physics have been discovered in the past as the result of such discrepancies. It was found, for instance, that the number of different microstates with  $N$  identical particles each having the same energy is not  $N!$  but 1 for certain types of particles. This counterintuitive counting of states was eventually explained by the theory of quantum mechanics. Even the birth of quantum mechanics can be traced back to the realization that in order to obtain correct predictions, one has to assume that certain variables cannot take on continuous values but are “quantized”.

In summary, we use our microscopic knowledge of the system under study to construct the set of all possible microstates and to express macroscopic variables as functions of the microstates. That is, we assume that we know all microstates that are possible, and that given a specific microstate we can calculate any macroscopic variable associated with that microstate. We then choose a set of macroscopic variables and derive the maximum entropy distribution conditioned on the values of these macroscopic variables. After that, we can make predictions of other variables by averaging over the maximum entropy distribution.

### 2.2.2 Is that all?

At this point, it is helpful to note that this view of statistical mechanics is by no means standard in the scientific community in general and the physics community in particular. In the view presented here, the maximum entropy distributions are obtained as a result of *statistical inference* based on partial information (for instance, the mean energy). This informational view of statistical mechanics contends that, given only knowledge of the mean energy of the system, we cannot do anything other than choose the maximum entropy distribution for making predictions about other macroscopic quantities. The maximum entropy distribution corresponds to the “most ignorant” distribution given the partial information. Any other distribution over the microstates would introduce unwarranted biases.

Most physicists would however probably subscribe to the view that the success of statistical mechanics merely shows that *apparently* the maximum entropy distribution gives an accurate description of the actual distribution over microstates that occurs in physical systems. From this point of view, one obviously has to justify the use of the maximum entropy distribution by independent arguments. From the statistical inference point of view, no such justification is necessary; the maximum entropy distribution is simply the least biased way to deal with a lack of information. But from the more commonly held point of view, one would have to show that the maximum entropy distribution is the actual distribution with which systems visit their microstates.

Therefore, the proponents of this “objective” view are engaged in a program which tries to show that the sort of physical systems that are studied in statistical mechanics generally evolve towards the maximum entropy distribution. One for instance tries to show that particles bouncing around in a box form an *ergodic* system. That is, for long times, the time average of the microstates that such systems visits is equal to the maximum entropy distribution independent of the initial distribution over microstates. This *ergodic program* has not been entirely successful due to some fundamental difficulties. For instance, since the laws of physics are reversible, one can uniquely determine the microstate of the system arbitrary amounts of time in the past or future given the microstate of the system at any particular time. This implies that the entropy of a distribution over microstates can only be *constant* over time. One of the ways of dealing with this problem is to calculate the entropy with respect to some *coarse graining* of the space of microstates. Although one assumes that the momenta and positions of particles can still take on a continuum of values, one only calculates the entropy with respect to the frequencies of particles in discrete *bins* of momentum and position. However, even under these ad hoc simplifications, only a small class of simplified systems can be shown to evolve towards the maximum entropy distribution with respect to these bin-frequencies.

In contrast to the objective view, the informational view provides a conceptually simple justification for use of maximum entropy distributions. It however does not specify how one should choose the set of macroscopic variables with respect to which the maximum entropy distribution is calculated; it can be applied to any set of macroscopic variables. The question: “Which macroscopic variables should one choose?” does not have a simple general answer. There are, however, some constraints and qualitative considerations to go by. First, one generally chooses macroscopic variables that can be easily measured or controlled. Additionally, the macroscopic variables have to be suf-

ficient in the sense that definite predictions of other variables can be obtained from the maximum entropy distributions. For instance, if we don't condition the maximum entropy distribution on macroscopic variables at all, we obtain a uniform distribution over all microstates. In that case, our predictions might have a variance associated with them that is not small at all relative to the mean and we cannot expect our predictions to be accurate.

In equilibrium statistical mechanics one is interested in the properties of the system that are constant over time. A macroscopic variable that fluctuates wildly over time is therefore not very suitable; measuring it at one time doesn't provide a good prediction for its value at some other time. The microscopic laws tell us that energy is a conserved quantity. This implies that it cannot change over time when the system is isolated and therefore provides a useful macroscopic quantity. Momentum and angular momentum are other well known conserved quantities. Why are these not included in the common description of systems that are typically studied in statistical mechanics? The reason is simply that they are zero for systems that are not moving or rotating macroscopically, and this is typically true for the systems studied. However, some systems studied in astrophysics do in fact move and rotate macroscopically. For these systems the momenta and angular momentum do appear as macroscopic variables in the statistical mechanical description.

Choosing an appropriate set of macroscopic variables is further complicated by the fact that a physical system does not generally have *one* set of macroscopic variables that describes its behavior adequately for all parameter settings. That is, under a change of external parameters, the same physical system may occur in different "phases" that require different sets of macroscopic variables to describe it. This phenomenon will be discussed in the next section.

## 2.3 Phase Transitions

It is of course well known that systems in nature can exhibit different "phases" and that by changing some macroscopic variables, a system can be made to undergo phase transitions between these different phases. The boiling and freezing of water are probably the most commonly observed examples of such phase transitions. There are many fascinating phenomena involved with phase transitions but in this section we will only focus on those features of phase transitions that bear on the evolutionary phenomena studied later on.

In the previous sections it was pointed out that under certain conditions the fluctuations in a macroscopic quantity as calculated from the maximum entropy distribution might become large relative to its mean, calculated from the same maximum entropy distribution. In general, a phase transition can be associated with such large fluctuations in one or more macroscopic quantities. Under a change in some macroscopic control variable (such as temperature) the fluctuations in some other macroscopic variables suddenly become very large. The control parameters for which the fluctuations "blow up" define the points where the phase transition occurs. As was also pointed out in the previous sections, the fact that fluctuations in some macroscopic quantity are large indicates that we need additional macroscopic variables to describe the state of the system. In

## 2.3 Phase Transitions

---

statistical mechanics, these new macroscopic quantities that emerge through a phase transition are generally referred to as *order parameters*. As will be shown below, they are often associated with the spontaneous *breaking of a symmetry* that was present in the system before the phase transition.

The simplest and most illustrious example of the breaking of symmetry and the appearance of a new macroscopic variable through a phase transition is probably the transition from a paramagnetic to a ferromagnetic phase in ferromagnetic materials. A well studied prototype model for such ferromagnetic systems is the Ising ferromagnet model. It consists of a one, two, or three dimensional square lattice with a microscopic magnetic spin located at each site. Each magnetic spin can occur in two states only: up or down. The microscopic state of the system is given by specifying which spins are up and which spins are down. Spins only interact with their neighbors on the lattice. Each pair of neighboring spins has a contribution to the energy of the system that depends on the relative orientation of the spins. A pair of aligned spins contributes less than a pair of anti-aligned spins. The energy  $E_s$  of a microscopic state  $s$  is thus only dependent on the number of aligned and anti-aligned pairs of neighboring spins. The equilibrium (maximum entropy) distribution at a temperature  $T = 1/(k\beta)$  is given by the Boltzmann distribution  $p_s = \exp(-\beta E_s)/Z$ . This equation in principle contains all necessary information regarding the equilibrium properties of the Ising ferromagnet.

With respect to the symmetries of this system, one obviously notes that the system possesses a spin flip symmetry. That is, the energy  $E_s$  of a state  $s$  stays the same when all spins are reversed  $s \rightarrow \bar{s}$ . Since the probability  $p_s$  of a microstate  $s$  depends only on its energy  $E_s$ , it follows that a state  $s$  has the same probability of occurrence as its spin reverse  $\bar{s}$ , i.e.  $p_s = p_{\bar{s}}$ . This symmetry guarantees that the expectation value of the macroscopic magnetization  $M$ , which can be roughly defined as the difference between the number  $N_+$  of spin “up” pairs and the number of spin “down” pairs  $N_-$  is zero:

$$\langle M \rangle \equiv \left\langle \frac{N_+ - N_-}{N_+ + N_-} \right\rangle = 0. \quad (2.12)$$

For every state  $s$  with a magnetization  $M_s$ , there is an equally probable opposite state  $\bar{s}$  with the opposite magnetization  $M_{\bar{s}} = -M_s$ .

One might expect that equation (2.12) implies that the Ising ferromagnet will never exhibit any macroscopic magnetization. For high temperatures (small  $\beta$ ), this is indeed the case. However, lowering the temperature to some critical temperature  $T_c$ , the fluctuations  $\langle M^2 \rangle$  can suddenly become very large<sup>2</sup>. When this happens, we can no longer expect our theoretical predictions ( $\langle M \rangle = 0$ ) to be accurate in the sense that any particular example of an Ising ferromagnet should have no macroscopic magnetization. What happens is that for high temperatures, the Boltzmann distribution is concentrated around microstates  $s$  that have no macroscopic magnetization. There will be regions where most spin pairs are up, but there will be roughly equally many regions where spin pairs are mostly down making the total magnetization roughly zero. For low temperatures however, the Boltzmann distribution has two peaks; one at a magnetization  $M = +1$  and one at a magnetization  $M = -1$ . The mean magnetization  $\langle M \rangle$  is still zero, but its variance is large, i.e.  $\langle M^2 \rangle = 1$ .

---

<sup>2</sup>For a 1-d Ising ferromagnet this strictly only occurs at  $T_c = 0$ , which makes the 1-d case somewhat degenerate.

For real ferromagnets one observes that as the temperature falls below  $T_c$ , the system spontaneously acquires either a positive or a negative magnetization. What happens is that the regions in which most spin are aligned become as large as the magnet itself. That is, either the spin up regions take over the entire magnet, or the spin down regions will take over the entire magnet. The system thus undergoes a phase transition from a state with magnetization zero, to a state with either positive or negative magnetization. At any point in time, and for any particular system, one will find the system *either* in the positive magnetization state or the negative magnetization state. The macroscopic prediction  $M = 0$  no longer describes the physical situation accurately as was to be expected from the large variance ( $M^2$ ). However, if we introduce one additional macroscopic parameter we can again accurately describe the system's macroscopic behavior. We simply need to add to the description a *directional* variable that tells whether the system has positive or negative magnetization. For systems with for instance positive magnetization, all states with negative magnetization are excluded and the Boltzmann distribution will thus again be concentrated in a single peak.

Obviously, this phase transition has a spontaneous symmetry breaking associated with it. The symmetry between positive and negative magnetization states gets broken. Below the critical temperature, tiny fluctuations in magnetization away from zero will be amplified into a large macroscopic magnetization that gets fixed in the ferromagnet. Once the system has developed a large magnetization, it is very unlikely to spontaneously revert to the opposite magnetization. At that point, we need a new macroscopic variable (magnetization direction) to accurately describe the macroscopic behavior of the system. This is typical of phase transitions (especially "continuous" phase transitions): Fluctuations in a macroscopic quantity become macroscopic themselves, a symmetry of the system is broken, and a new macroscopic variable, often called an order parameter, appears. Of course, the situation is reversed when the phase transition is approached from the opposite direction, i.e. increasing the temperature in the case of the ferromagnet. In more complicated cases, more than one new macroscopic parameter may appear at the same time. That is, more complicated symmetries might be broken. There are even cases in which infinitely many new macroscopic variables may appear at once.

Finally, it should be noted that the symmetry breaking and phase transition are induced by a change in a control parameter of the system, such as temperature. In the evolutionary setting to be described later on, the symmetry breaking occurs *dynamically* due to the internal dynamics of the system. In that context, the occurrence of a new macroscopic variable corresponds to the appearance of a new (functional) phenotypic trait in the evolving population.

## 2.4 Non-equilibrium Statistical Mechanics

Until now, the discussion of statistical mechanics has focused on the behavior of systems in *thermodynamic equilibrium*. Thermodynamic equilibrium is characterized by a stationary state of the system in which no macroscopic flows between the system and its environment are present. The methodology of statistical mechanics, as explained in the previous sections, has proven to be able to accurately describe the macroscopic behavior

of systems in equilibrium. However, as soon as one moves away from the equilibrium situation, things start to become rather complicated if one takes the common view that the maximum entropy distribution somehow corresponds to the actual “objective” distribution over microstates.

One will be faced with the problem of having to somehow *derive* new distributions over microstates for systems out of equilibrium. For equilibrium, it is known that the maximum entropy distributions provide accurate predictions, but for non-equilibrium systems it is not known which distributions to use. Apart from first-order perturbations around an equilibrium state [30], almost no general methods have been developed to derive non-equilibrium microstate distributions. Further, the ergodic hypothesis is entirely unsuited for non-equilibrium situations, since one is generally interested in *dynamics*. For equilibrium situations, the ergodic hypothesis justifies using the maximum entropy distribution by assuming that averages calculated with the maximum entropy distribution give the same results as *time averages* over the stationary macrostate. However, this justification precludes calculation of the *dynamics* of macroscopic variables. In short, the “objective” view that the maximum entropy distribution corresponds to the actual physical distribution over microstates, renders non-equilibrium problems almost inapproachable.

In contrast, if one takes the informational view advocated in section 2.2.1, non-equilibrium situations cause considerably fewer problems. For example, there are no conceptual difficulties. The method remains essentially unchanged: One assumes complete knowledge of the microscopic state space and microscopic dynamics of the system under study, and only partial information on the specific state that the system is in. In the dynamic, non-equilibrium situation, this information may be in the form of the averages of some macroscopic quantities at different times, or knowledge of certain macroscopic flows or gradients in the system. One then again constructs the maximum entropy distribution  $p_s(t)$ , which now depends on time as well. Other macroscopic quantities of interest can then again be predicted using this maximum entropy distribution. This general approach to non-equilibrium statistical mechanics was introduced in the 1970s by Jaynes [84], similar treatments can be found in [5, 55, 131, 132].

To make the ideas more precise, consider the case in which the average values  $f(t_i)$  of some macroscopic variable  $f$  are given for a set of times  $\{t_i\}$  with  $i = \{1, 2, \dots, n\}$ . We set the origin at the first time  $t_1 = 0$ . For each of these times, the average  $\langle f \rangle$  is thus given by some number  $f(t_i)$ . Additionally, for any microstate  $s$ , we can uniquely predict its microscopic time behavior

$$s(t) = G[s(0), t], \quad (2.13)$$

where the operator  $G$  represents the microscopic equations of motion of the system. From this information we derive a maximum entropy distribution  $p_s(0)$  over microstates at time  $t = t_1 = 0$ . By using equation (2.13), we can then obtain the distribution  $p_s(t)$  for arbitrary times. Note that this distribution  $p_s(t)$  is generally *not* a maximum entropy distribution with respect to the macroscopic variable  $f(t)$  at  $t$ . That is, if we had constructed the maximum entropy distribution based on the information that  $\langle f \rangle$  equals  $f(t)$  at time  $t$ , we would have generally obtained a different maximum entropy distribution  $p'_s(t)$ .

Constructing the maximum entropy distribution  $p_s(0)$  is done in complete analogy to the methods described in section 2.2.1. Using the equations of motion (2.13), the constraints take on the form

$$\sum_s p_s(0) f_{G[s,t_i]} = f(t_i), \quad (2.14)$$

where  $f_s$  denotes the value of the quantity  $f$  for microstate  $s$ , such that  $f_{G[s,t_i]}$  denotes the value of  $f$  at time  $t_i$  given that the system was in state  $s$  at time  $t = 0$ . Thus, given the initial distribution  $p_s(0)$  at time  $t = 0$ , we demand that the average values  $\langle f \rangle$  at the different times  $t_i$  are given by  $f(t_i)$ . But in this form, the problem is exactly the same as the ones described in the appendix B. From a mathematical point of view, averages of the same quantity at different times are equivalent to a set of constraints given by averages of different quantities at one particular time. The averages  $\langle f \rangle$  at different times can be inferred from  $p_s(0)$  using the equations of motion (2.13). Therefore, the maximum entropy distribution is simply given by

$$p_s(0) = \frac{1}{Z} \exp \left( - \sum_i \lambda_i f_{G[s,t_i]} \right), \quad (2.15)$$

where  $Z$  is again a normalization constant and the Lagrangian multipliers  $\lambda_i$  are chosen such that the constraints (2.14) are all satisfied.

Intuitively, the maximum entropy method distributes probability as uniformly as possible over all states that are consistent with the macroscopic information that is provided. In this time-dependent case, this means that the distribution  $p_s(0)$  gives high (uniform) probability to all states  $s$  that exhibit values of the macroscopic quantity  $f$  which are consistent with the values  $f(t_i)$  at each of the times  $t_i$  for which the macroscopic quantity  $f$  was specified.

When the maximum entropy distribution  $p_s(0)$  has been determined, it can be used to predict the dynamics of other macroscopic quantities. For instance, for any time  $t$ , the predicted value  $\langle g(t) \rangle$  for the average of some macroscopic variable  $g$  is given by

$$\langle g(t) \rangle = \sum_s p_s(0) g_{G[s,t]}. \quad (2.16)$$

Using such equations, we may derive macroscopic equations of motion for macroscopic variables such as  $g(t)$  from the maximum entropy distribution.

Notice that in this formalism, predictions for future time are not necessarily only dependent on the values of the macroscopic quantities at the current time, but may depend on the values of the macroscopic variables in the past as well. This contrasts with the microscopic dynamics as generated by the operator  $G$ . There, the future microscopic state of a system is always only dependent on the current microscopic state<sup>3</sup>. It may seem counterintuitive that the dynamics on a macroscopic level exhibits such *memory effects* while the underlying microscopic laws do not. The maximum entropy method is however not a law of physics, it is a law of statistical inference. Significantly, this may easily lead to memory effects. In particular, it will typically be impossible to derive

<sup>3</sup>This holds in particular for all the fundamental laws of physics.

differential equations for the macroscopic quantities such as  $g(t)$ . If one wants to obtain memoryless equations of motion for the macroscopic variables, one has to make further approximations. This will be done in the next section.

### 2.4.1 Memoryless Approximation

In many cases, the formal microscopic equations of motion (2.13) can be rather complicated, since  $G$  is often nonlinear. Moreover, deriving macroscopic equations of motion using equations such as (2.16) can be rather tedious. In those cases, we might want to make some additional assumptions to simplify the analysis.

One possible approach is to assume that the dynamics for the macroscopic variables of interest is slow compared to the dynamics on the microscopic level. Under that assumption, one can construct the dynamics on the level of the macroscopic variables by assuming that the microscopic maximum entropy distribution remains valid at all times. That is, given that we know the value  $f(t)$  of the macroscopic variable  $f$  at time  $t$ , we have a maximum entropy distribution  $p_s \propto \exp(-\lambda' f_s)$  at time  $t$ , with mean  $\langle f \rangle = f(t)$ . From this, we derive the value  $f$  at the next instant  $t + dt$ , by integrating the equations of motion (2.13) for a very small time interval  $dt$ . From this we find a small change  $df$  in the macroscopic variable  $f$ , i.e.  $f(t) \rightarrow f(t) + df$  in the time  $dt$ . This is essentially the same procedure as outlined before. However, at this point we reassume a new maximum entropy distribution  $p_s \propto \exp(-\lambda' f_s)$  with mean  $\langle f \rangle = f(t) + df$ . We then repeat the procedure and, in this way, iteratively derive the macroscopic dynamics of the variable  $f$ . One can think of this approximation as neglecting any memory of the state at time  $t$  in the distribution at time  $t + dt$  apart from the value  $f(t) + df$  of the macroscopic variable  $f$ . This absence of memory makes this approximation appealing. In particular, we are able to derive phenomenological equations of motion for the macroscopic variable  $f(t)$ . In this thesis, we will take this memoryless approximation in predicting macroscopic variables for an evolving population.

It is important to note once more that it is not necessary that the maximum entropy distributions somehow accurately describe the “real” distributions of the non-equilibrium dynamical systems. In general, we are not specifically interested in the actual microscopic distributions but are only interested in predicting some macroscopic observables. The maximum entropy distribution might be inaccurate in the sense of not predicting the actual microscopic distribution but might still accurately predict the macroscopic observables.

The same argument holds for assuming no memory and reassuming maximum entropy after each time increment. This assumption is typically provably wrong from a microscopic point of view. But this does not necessarily mean that it produces inaccurate predictions for macroscopic observables. Roughly speaking, the maximum entropy distribution at time  $t + dt$  might differ from the distribution obtained by direct integration of the microscopic equations of motion from the maximum entropy distribution at time  $t$  by assigning too much probability to certain states and too little to others. If, with respect to predicting a macroscopic quantity  $g$ , the states with “high  $g$ ” and the states with “low  $g$ ” are spread roughly equally among the states that obtain too much and too little probability, the predictions for the quantity  $g$  will not be affected. In order for the predictions to fail, the states with “high  $g$ ” should consistently fall in the set of states that

obtains too much probability or fall consistently in the set of states that obtains too little probability. The fact that this does not often occur in practice accounts for the robustness of the maximum entropy approach. Notably: in cases where the approach does fail, it often points the way to interesting relations between certain macroscopic variables.

## 2.5 The Second Law

One of the most frequently discussed laws of physics is the second law of thermodynamics. It comes in many different varieties and is often presented as an extremely fundamental law that describes a natural tendency common to all processes in nature. It is not uncommon to find a description of the second law along the lines of: Whatever happens in the universe, its entropy can only go up. Such a colloquial description can even be found in the famous *Lectures on Physics* by Richard Feynman [43]. These generalized interpretations of the second law have led to conceptual misunderstandings, especially regarding its relevance for biology and evolution. One can think of another authority, Erwin Schrödinger, and the concept of negative entropy, in this regard [125]. In combination with the conceptualization of entropy as representing disorder, these ideas have even led some to claim that evolution as a whole is at odds with the second law of thermodynamics, since order seems to spontaneously arise through evolution—something which in the popular view is prohibited by the second law.

Since evolution is the topic of this thesis, and since some methodology borrowed from statistical mechanics is used in this thesis as well, it seems appropriate to briefly discuss these misconceptions. The second law is, in fact, far more restricted and of limited applicability than is often suggested.

In the original version of the second law, which will be discussed below, it is stated that under a certain class of operations on a system, the thermodynamic entropy of the system at the end of the operations cannot be smaller than its initial entropy. The first thing to note here is that thermodynamic entropy as such is only defined for systems in *thermodynamic equilibrium*. This immediately implies that the second law applies only to processes for which the system is in equilibrium both at the beginning and end. For systems that are out of equilibrium, and continue to remain out of equilibrium, the second law simply does not apply. Living systems in particular are very far from equilibrium—since they, for instance, are sustained by energy fluxes—and the second law therefore does not speak to them. The same holds for the biosphere as a whole. The periodic heating due to sunlight and cooling at night guarantees that the biosphere can never reach a state of equilibrium as long as the sun is burning.

We could stop our discussion of the second law here, but it is instructive to briefly review how it arises and what its general implications for non-equilibrium dynamics and evolution might be. In the following, we will again present the informational view on these matters in which the entropy of a system measures the amount of uncertainty about the precise microstate of a system, given knowledge of some of its macroscopic variables. This contrasts with the “objective” view that the entropy of a system corresponds to a *physical property* of that system.

### 2.5.1 Definitions

Let's first make a precise statement of the second law. We will use the traditional and commonly taught statement due to Clausius. Before the law can be stated, we will need a definition of entropy. Historically, the discovery of entropy as a state variable can be traced back directly to Carnot's principle that the most efficient heat engine must be a reversible engine, as lucidly explained in [85]. If a reversible engine wasn't the most efficient engine, one could use a combination of a more efficient engine and a reversible engine which is running backwards to create a perpetuum mobile. Kelvin then showed how the efficiency of such reversible engines is a function only of the temperatures of the hot and cold reservoirs between which it operates. One can generalize this, and Kelvin did so, to the case in which the system exchanges amounts  $dQ_i$  of heat with different heat baths  $i$  at different temperatures  $T_i$ . The fact that the efficiency of an arbitrarily complicated heat engine, which exchanges heat with a variety of heat baths is still bounded by the efficiency of a reversible one, then yields the inequality

$$\sum_i \frac{dQ_i}{T_i} \leq 0, \quad (2.17)$$

where the ratio  $dQ_i/T_i$  can be thought to measure the efficiency of the energy exchange with reservoir  $i$ . The above form invites one to take the limit to an infinite number of heat baths with which the engine, or system in general, exchanges heat. In this way, the notion of a thermodynamic *path* is introduced. As the system exchanges heat with the different reservoirs, its macroscopic state variables, such as volume and temperature, trace out a path in the space of macroscopic variables. One then has from equation (2.17) that the heat exchange  $dQ$  divided by temperature  $T$ , integrated over a closed path, obeys the inequality

$$\oint \frac{dQ}{T} \leq 0, \quad (2.18)$$

where the equality holds only for a reversible path. Reversible refers to reversibility of the macroscopic state, i.e. slowly moving a piston induces a reversible change in the volume  $V$ . Reversibility is thus fundamentally defined in terms of the possibility to *control* certain macroscopic variables in the laboratory. Notice also that the temperature in equation (2.18) refers to the temperatures of the heat baths with which the system interacts. One can view this collection of heat baths as an *environment* with which the system interacts.

Once one realizes that the integral along a closed reversible path is zero, one realizes immediately that the integral from some point  $A$  to some point  $B$  is independent of whatever reversible path is taken:

$$\int_A^B \frac{dQ}{T} = H_B - H_A. \quad (2.19)$$

That is, one can uniquely associate a number  $H_P$  with each point  $P$  in the space of macroscopic variables. This number, is of course, the thermodynamic *entropy* of the system. Say that we move from  $A$  to  $B$  along a reversible path, and then back from  $B$  to  $A$  along an *arbitrary* path, either reversible or not. According to equation (2.18), the

sum of the integrals along both paths must be smaller than or equal to zero, and since the reversible path from  $A$  to  $B$  integrates to  $H_B - H_A$ , we have

$$\int_B^A \frac{dQ}{T} \leq H_A - H_B, \quad (2.20)$$

for the arbitrary path leading back from  $B$  to  $A$ . By the definition of the entropy,  $H_A - H_B$  is the change  $dH_{\text{sys}}$  in the system's entropy due to its (possibly irreversible) movement from  $B$  to  $A$ . Note further that the left hand side of equation (2.20) is precisely the negative change  $dH_{\text{env}}$  in the entropy of the environment. Moving the left hand side to the right we have:

$$dH_{\text{sys}} + dH_{\text{env}} \geq 0. \quad (2.21)$$

This is Clausius' form of the second law. It states that when a system and its environment both start in equilibrium, undergo some changes, and eventually both end up in some new equilibrium state, that the total entropy of these equilibria is greater than or equal to zero. It is only zero when the changes in system and environment can be reversed.

In this form, the second law hardly has any implications for systems that occur consistently out of equilibrium. It demands that we compare the entropies of the initial equilibrium states and final equilibrium states. Furthermore, the interactions that are considered between the system and its environment are fairly limited. Only heat is exchanged. Note also that these entropies are defined thermodynamically, namely as an integral over a function of temperature and heat exchange. However, through the work of Gibbs, Shannon, and Jaynes we now know that this entropy is equal (up to a constant) to a measure of our ignorance regarding a system's microstate, given knowledge of the macroscopic variables that define its (equilibrium) state. In the next section, we will discuss the second law from the information theoretic viewpoint, see how it arises, and how it bears on more general situations than the ones originally considered by Clausius.

## 2.5.2 Information Theoretic Viewpoint

As mentioned before, the thermodynamic definition of entropy in the previous section appears very different from the information theoretic definition given in equation (2.5). In principle they *are* different, simply because they have different definitions. It was Gibbs' contribution to postulate that the form (2.5) and the definition (2.19) are in fact the same for systems in thermodynamic equilibrium. Gibbs postulated that the equilibrium distributions and their entropies can be obtained by maximizing the form (2.5) within the constraints set by the macroscopic variables that define the macroscopic equilibrium state. Although Gibbs' methods work and have been (almost exclusively) used in statistical mechanics, it has been unclear until quite recently *why* they should work. However, after the work of Shannon [127, 128] had shown that the form (2.5) is a measure of uncertainty, we have now come to understand through the work of Jaynes [84, 83] that the method introduced by Gibbs is not a law of physics, it is a law of statistical inference. So how would the second law arise from this point of view?

As might perhaps have been expected, the second law turns into almost a triviality if one adopts the informational view of entropy as uncertainty. We consider the same

## 2.5 The Second Law

---

setup as in the previous section: A system and an environment are both found in some thermodynamic equilibrium states. We then let the system and environment interact, typically exchanging some heat, after which both are found in some other equilibrium states. Let's assume that initially, the system is in a macroscopic state which has an entropy  $H_{\text{sys}}^i$ . That is, using Gibbs' method of choosing the maximal entropy distribution over microstates consistent with the system's macrostate gives an initial entropy  $H_{\text{sys}}^i$ . Likewise, the initial entropy of the environment, maximized conditioned on its macrostate is given by  $H_{\text{env}}^i$ . Before interaction, the joint entropy  $H$  of the system and its environment is simply given by the sum of the two entropies

$$H = H_{\text{sys}}^i + H_{\text{env}}^i. \quad (2.22)$$

The essential ingredient at this point is the realization that this *joint entropy*  $H$  cannot change due to the interaction between system and environment. The reason for this is that the laws of physics are *reversible*. That is, on the microscopic level, the laws are invariant under time reversal; any microscopic dynamics that is possible forward in time is equally possible backward in time. A consequence of this reversibility is known as *Liouville's theorem*, which states that the *phase space volume* cannot change under reversible dynamics. If we start with a volume in phase space and evolve it forward in time, the volume of this volume cannot change. It can bend and twist in all kinds of ways, but its total volume will remain unaltered. If our maximum entropy distribution is a uniform distribution over some volume in phase space, then the entropy  $H$  is proportional to the logarithm of this volume. We immediately see that Liouville's theorem implies that  $H$  is constant. For non uniform distributions in phase space it can also be shown that  $H$  is still a constant under the microscopic equations of motion. Summarizing, we initially had an uncertainty  $H_{\text{sys}}^i$  as to the precise microstate of the system and an uncertainty  $H_{\text{env}}^i$  as to the microstate of the environment. We thus had an uncertainty  $H = H_{\text{sys}}^i + H_{\text{env}}^i$  as to the joint state of system and environment. From Liouville's theorem, this uncertainty  $H$  cannot change through time.

At this point it seems that we have just disproved what we wanted to establish in the first place, since we wanted to show that entropy increases but we have just shown it to be constant. However, the entropy which is constant is the *joint entropy*  $H$  of both system and environment while the second law makes a statement about the sum of the final entropies  $H_{\text{sys}}^f$  and  $H_{\text{env}}^f$  of system and environment respectively. Since the system and environment interacted, they became entangled in such a way that their respective microstates became correlated. That is, there is now some information about the environment's microstate in the system and vice versa. In information theory, this phenomenon is expressed most transparently by the following general equality

$$H = H_{\text{sys}} + H_{\text{env}} - I_{\text{sys,env}}, \quad (2.23)$$

where  $I_{\text{sys,env}}$  is the *mutual information* between system and environment, see appendix A.

The mutual information measures how much the uncertainty about the environment's microstate is reduced by knowing the microstate of the system. Mutual information is, of course, always a positive quantity. That is, if it is zero initially, it can only go up. This is precisely what happens in our case. Before the system and environment interacted their

mutual information was zero; knowing the microstate of the environment didn't reduce our uncertainty about the system's microstate. But as soon as system and environment start interacting, their mutual information can only go up. Let's assume that after the environment and system have come to equilibrium again, they will have evolved an amount  $I$  of mutual information. We then have for the final entropies:

$$H = H_{\text{sys}}^f + H_{\text{env}}^f - I. \quad (2.24)$$

Combining this with equation (2.22), we obtain a version of the second law:

$$dH_{\text{sys}} + dH_{\text{env}} = I \geq 0. \quad (2.25)$$

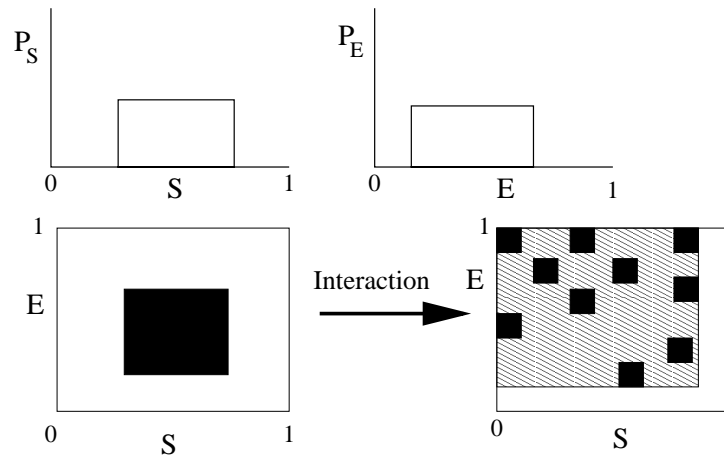


Figure 2.1: Intuitive illustration of the second law. A system and environment both start out in some macroscopic state that restrict their possible microstates—denoted by  $S$  and  $E$  respectively—to some regions of their phase spaces. The maximum entropy distributions  $P_S$  and  $P_E$  over these regions are uniform and shown in the upper two plots. The joint distribution forms a rectangle in the joint phase space shown as the black rectangle in the lower left plot. After interaction between system and environment, the *same volume* of microstates has spread through phase space as shown by the small black squares. The new entropies of system and environment, i.e. the maximal entropy consistent with the new macrostates, are effectively given by the shaded envelope. The sum of entropies of system and environment increases because they develop positive mutual information.

This result is illustrated in figure 2.1. For simplicity, we have assumed that the maximum entropy distributions are all uniform distributions. The possible microstates of the system are denoted by  $S$ , and the possible microstates of the environment by  $E$ . The initial macroscopic states of the system and environment both define some regions in their microscopic state spaces that are consistent with the macroscopic state. The maximum entropy distributions  $P_S$  and  $P_E$  are then uniform over those regions, as shown in the upper plots of figure 2.1. The initial joint state of system and environment

covers a rectangular region in their joint state space shown as the black rectangle in the lower left plot.

After system and environment have interacted, we obtain the situation in the lower right plot. The original rectangular region has broken up and spread over a large part of the joint phase space. Notice that the total volume of the black boxes in the lower right plot is the *same* as the volume of the rectangle in the lower left plot. The joint entropy has not changed. However, if we project the joint distribution on either the horizontal or vertical axis, we see that the separate entropies  $H_{\text{sys}}^f$  and  $H_{\text{env}}^f$  of system and environment have gone up. If we determine their macroscopic states, there is a much larger volume of phase space that is consistent with these new macroscopic states. This larger volume is indicated by the shaded region in the lower right plot. Although we *know* from the reversibility of the laws that the combination of system and environment can only occur in a subset of the shaded region, our inability to assess or control the precise correlations between system and environment effectively forces us to consider the whole shaded region as potential microstates for system and environment. In this way, the sum of their entropies has increased.

### 2.5.3 Implications

Finally, we draw out several consequences of the mutual information version of the second law just derived. First of all, this version is *not* restricted to equilibrium states only. For any combination of a system and environment that are initially without mutual information, their interaction can only lead to an increase in their mutual information. When this happens, the sum of their separate entropies must go up. This holds in general for reversible dynamics. Thus if we let two systems interact, and we notice from the evolution of their macrostates that the sum of their separate entropies (as determined from maximizing entropy with respect to their macroscopic states) has gone up, we know that they must have evolved some mutual information. That is, there are interdependencies to be discovered between them. Or, in other words, there are variables that are correlated between them which, as yet, we cannot measure or control. From this point of view, irreversible phenomena provide a potential for discovering additional “variables”.

In connection with this, it is interesting to point out that considerations of engine efficiencies, which lie at the historical roots of the second law, can be simply described from an informational point of view as well. If one wants to use heat baths to drive an engine, one is essentially obtaining energy from a system whose microscopic state has a great deal of uncertainty. That is, we only know the temperature of the heat bath, which can be considered proportional to the average energy per degree of freedom of the heat bath system. With such enormous uncertainties, it is not surprising that we cannot control exactly *where* the energy is going to be delivered to our engine. That is, some of the energy will end up in degrees of freedom of our engine that do not play a role in its functionality—such as, the vibrations of its walls.

If, on the other hand, we use as an energy source some system other than a heat bath, whose state we know much better, we might be able to set up the interaction between the engine and this system in a much more controlled way, such that more of the energy will be delivered to those degrees of freedom that carry its functionality—such as, the

movement of a piston. In a certain sense, this is why microwave ovens manage to heat up your drink much more quickly than the fire on the stove does (since the microwaves make “use” of knowledge such as the resonance frequency of water). Another example is the surprisingly high efficiency of muscle fiber, which was beautifully analyzed along information theoretic lines in [86].

With respect to biological evolution, it is interesting to speculate that through evolution organisms acquire more and more “control” over the energy sources around them. That is, evolutionary innovations might be viewed in terms of organisms discovering new ways of channeling the energy stored in certain degrees of freedom of the environment. These channels may then deliver the energy to places where the organism “wants it”, whereas the energy may previously have only excited degrees of freedom which were of no “use” to the organism. Maybe somewhat in the way that sunlight shining on a roof of solar cells provides more useful energy to a person than when it shines on their head and makes them sweat.