Statistical Mechanics: lecture 2

Entropy and relative entropy

Recommended reading: Salinas, chapters 3 and 5

Nielsen and Chuang, 11.1 and 11.2.

Tomé and de Oliveira, chapter 7.

Shannon Entropy

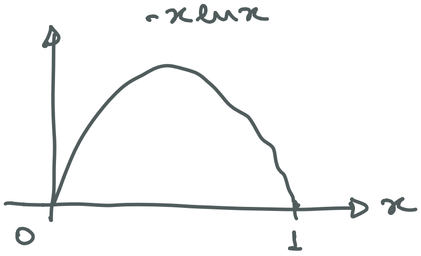
Consider a generic random variable X which can take on a discrete set of values $x = n$, each with probability P_n. The reason why X is random is related to the amount of information we have about it. For instance, let x denote the number of students that attend class today. This number is only random if I (the professor) don't have complete information about the whereabouts of all students. If I knew exactly where they were, where they were going and so on, the number X would not be random.

In classical probability theory, randomness is always informational. The prob. distribution P_n is always conditioned on the information we have about X.

The amount of information we have about X, as captured by the distribution P_n, is quantified by the Shannon entropy:

$$\boxed{S = -\sum_n P_n \ln P_n} \qquad (1)$$

To see why this is the case it is useful to have in mind the following plot



This function is always non-negative and zero if and only if $x=0$ or $x=1$. The entropy is then simply a sum of such functions. Thus, first and foremost, we conclude that the entropy is non-negative:

$$S \geqslant 0 \tag{2}$$

Secondly, if X is a deterministic process then $p_m = 1$ for some m and $p_n = 0$ for all $n \neq m$. Since $x \ln x$ is zero in both cases, we conclude that

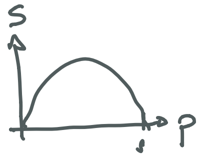
$$S = 0 \quad \text{when the process is deterministic} \tag{3}$$

Since the entropy is a sum of positive quantities, for any other distribution for which $p_m \neq 1$, the entropy will therefore be strictly larger than zero.

This already gives a good justification as to why S quantifies the information we have about X: it is zero when we have complete information and positive when we have some uncertainty about it.

A useful example is the Binary entropy, corresponding to a process where X can take on only two values with probabilities p and $1-p$. The entropy in this case reads

$$S = -p \ln p - (1-p) \ln(1-p) \tag{4}$$



the entropy is zero when $p = 0$ or $p = 1$ and is maximal when $p = 1/2$. the case $p = 1/2$ corresponds to the uniform distribution: in this case we know absolutely nothing about X.

This is a funny fact which most people don't realize. For instance, if the forecast says there is a 50% chance of rain, that is pure crap! It just means they have no idea whatsoever whether it will rain or not!

A more rigorous justification as to why Eq (1) is indeed a good quantifier of information was provided by Shannon in a paper entitled "A mathematical theory of communication".

The idea is as follows. Let us define a surprise function $I(p)$. This quantity is meant to be such that $I(p_m)$ captures the amount of information we gain about X when the event $X=m$ occurs. The idea is that the entropy should then be defined as the average surprise

$$S = \sum_m p_m I(p_m) \tag{5}$$

So, of course, our final result will be $I(p) = -\ln p$. But what Shannon did was to derive this as a consequence of a set of properties any surprise function should satisfy.

In particular, Shannon postulated that $I(p)$ should satisfy:

(1) $I(p) \geqslant 0$

(2) $I(1) = 0$

(3) $I(p)$ should be monotonically decreasing in p

(4) $I(p_1 p_2) = I(p_1) + I(p_2)$

Property (2) means that if an event always occurs, then this gives us no information at all (we are never surprised). Property (3), on the other hand, means that if an event is likely, it conveys little information. If it is rare, then when it happens we learn a lot (surprise!).

Property (4) is a bit less intuitive, but absolutely crucial. when a prob. dist. is a product it means that the two processes are statistically independent. (4) thus means that if two events are statistically independent, the amount of information we learn about them is simply the sum of the individual informations.

which kinds of functions satisfy (4)? we know the answer, of course: the logarithm. Thus

$$I(p) = a \ln p$$

where a is a constant. The base of the log is irrelevant, as it would simply rescale a. To comply with (1) and (3) we must have $a < 0$. The actual value of $|a|$ is unimportant since it will simply rescale S in the end Thus we simply set it to $a = -1$, which finally leads to

$$I(p) = -\ln p. \tag{6}$$

In summary, the entropy quantifies the amount of information encoded in P_n because it quantifies the average surprise $\langle -\ln P_n \rangle$, related to how much information we gain from each realization $X = n$.

The relative entropy - Kullback-Leibler divergence

Another quantity, intimately related to the entropy, is the relative entropy (also called Kullback-Leibler divergence) defined as

$$\boxed{S(p\|q) = \sum_n p_n \ln p_n/q_n} \qquad (7)$$

where p_n and q_n are two probability distributions. the relative entropy functions as a type of distance between two probability distributions. In fact, it satisfies

$$\boxed{S(p\|q) \geq 0 \quad \text{and} \quad S(p\|q) = 0 \ \text{iff} \ p_n = q_n} \qquad (8)$$

which is known as Gibbs inequality. The proof is based on the inequality

$$\ln x \leq x - 1 \qquad \forall \ x > 0$$

which yields (I write $\ln p_n/q_n = -\ln q_n/p_n$)

$$\begin{aligned} S(p\|q) = -\sum_n p_n \ln q_n/p_n &\geq -\sum_n p_n \left(\frac{q_n}{p_n} - 1\right) \\ &= -\sum_n q_n + \sum_n p_n \\ &= 0 \end{aligned}$$

The entropy and the relative entropy can be related by introducing the uniform distribution

$$\pi_n = \frac{1}{d} \qquad \text{(uniform dist.)} \tag{9}$$

where d is the dimension of the state space in question. This is the most uninformative distribution. It essentially says we know absolutely nothing. In this case

$$S(p\|\pi) = \sum_n p_n \ln p_n/\pi_n = \underbrace{\sum_n p_n \ln p_n}_{-S(p)} + \underbrace{\sum_n p_n}_{1} \ln d$$

$$\therefore \qquad \boxed{S(p) = \ln d - S(p\|\pi)} \tag{10}$$

This result is very nice. It shows, first of all, that due to Eq (8), the largest value of $S(p)$ occurs for the uniform distribution and reads

$$\boxed{\max S(p) = \ln d, \quad \text{occurs for } p = \pi = \text{uniform dist.}} \tag{11}$$

Combining with (2), we now have two bounds for S:

$$\boxed{\begin{array}{ccc} \text{maximally} & & \text{maximally} \\ \text{deterministic.} & 0 \leqslant S \leqslant \ln d & \text{disordered} \\ \text{Know everything} & & \text{know nothing} \\ & & \text{(John Snow)} \end{array}} \tag{12}$$

This analysis holds when the number of states d is finite. When this is not so, it will still be true that $0 \leq S$. But now S will be unbounded from above.

The bound (12) shows that $S=0$ means a lot of information, whereas $S = \ln d$ means no information at all. This motivates us to define the information contained in p_n as

$$\boxed{\mathrm{Info}(p) = \ln d - S(p) = S(p \| \pi)} \tag{13}$$

I love this formula: information is defined as how far p_n is from a state where we don't know anything. Makes sense! We will not use this definition a lot in our course because Info and S only differ by a constant. And also, S is defined when $d=\infty$ whereas d is not. But I nonetheless like to talk about because I think it really clarifies the relation between entropy and information, which is always a bit confusing at first.

Gibbs entropy = Shannon entropy of a thermal state

The Shannon entropy (1) is defined for any probability distribution P_n. Let's check now what happens when we assume that these probabilities are given by the thermal Gibbs formula

$$P_n = \frac{e^{-\beta E_n}}{Z} \tag{14}$$

The trick is to substitute this only in the log part of $P_n \ln P_n$. We then get

$$S = -\sum_n P_n \left[-\beta E_n - \ln Z\right]$$

$$= \beta \underbrace{\sum_n E_n P_n}_{U} + \ln Z \underbrace{\sum_n P_n}_{1}$$

Thus

$$\boxed{S = \beta U + \ln Z} \tag{15}$$

This formula provides a very easy way of computing the entropy of a thermal state. We don't need to compute $-\sum_n P_n \ln P_n$. If we already have Z and U, then S comes for free.

The entropy also turns out to be directly related to the specific heat. Write (15) as $U = T(S - \ln Z)$. Then

$$C = \frac{\partial U}{\partial T} = S - \ln Z + T\left(\frac{\partial S}{\partial T} - \frac{\partial}{\partial T}\ln Z\right)$$

$$= \underbrace{S - \ln Z}_{\beta U} + T\frac{\partial S}{\partial T} + \underbrace{T\frac{1}{T^2}\frac{\partial}{\partial \beta}\ln Z}_{-\beta U}$$

Thus we see that

$$\boxed{C = \frac{\partial U}{\partial T} = T\frac{\partial S}{\partial T}} \tag{16}$$

This shows a special property of equilibrium: when the system is in equilibrium, changes in energy also imply changes in entropy. This emphasizes the fact that thermal states $\rho_n \propto e^{-\beta E_n}$ form a very special set of states. If you are dealing with generic non-equilibrium states, you can change the state's energy without changing the entropy and vice-versa. But if you are restricted to the manifold of thermal states, then changing U will automatically also change S.

Examples

Let's look at some examples using the table in page 31 of lecture 1. For the qubit we get

(Qubit) $$\boxed{S = \frac{\beta\epsilon}{e^{\beta\epsilon}+1} + \ln(1+e^{-\beta\epsilon})} \qquad (17)$$

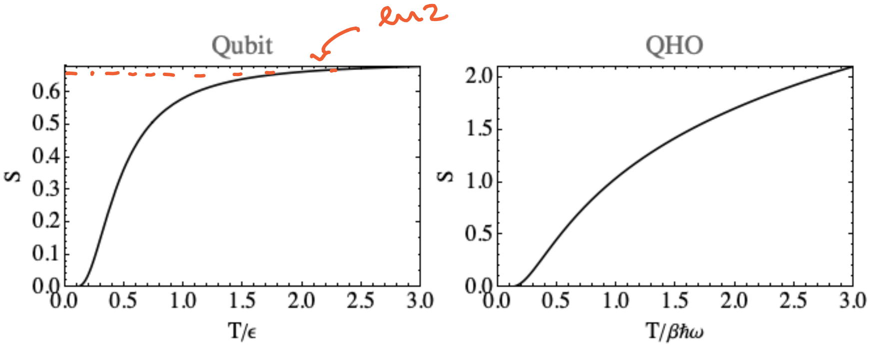
whereas for the QHO we get

$$S = \beta\hbar\omega\left(\langle n\rangle + 1/2\right) + \ln\left(\frac{e^{-\beta\hbar\omega}}{1-e^{-\beta\hbar\omega}}\right)$$

Note how the zero point fluctuations cancel out, leaving us with

(QHO) $$\boxed{S = \frac{\beta\hbar\omega}{e^{\beta\hbar\omega}-1} - \ln(1-e^{-\beta\hbar\omega})} \qquad (18)$$

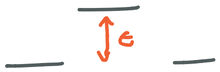
The similarity between (17) and (18) is noteworthy. Here is how they look like:



For $T = 0$, both tend to zero because the probabilities become deterministic. For $T \to \infty$ the qubit case tends to $\ln 2$ (because $d = 2$). The QHO is unbounded and grows like $S \sim -\ln \beta\hbar\omega$ for large T (small $\beta\hbar\omega$).

Let us also look at the Λ-atom studied in lecture 1. The partition function was

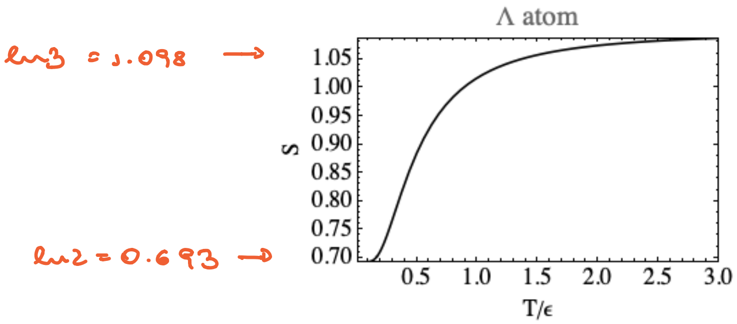
$$Z = 2 + e^{-\beta\epsilon}$$



where the factor of 2 comes from the fact that the ground state is two-fold degenerate. The entropy in this case is

$$S = \frac{\beta\epsilon}{2e^{\beta\epsilon}+1} + \ln(2+e^{-\beta\epsilon})$$

Here is how this looks like

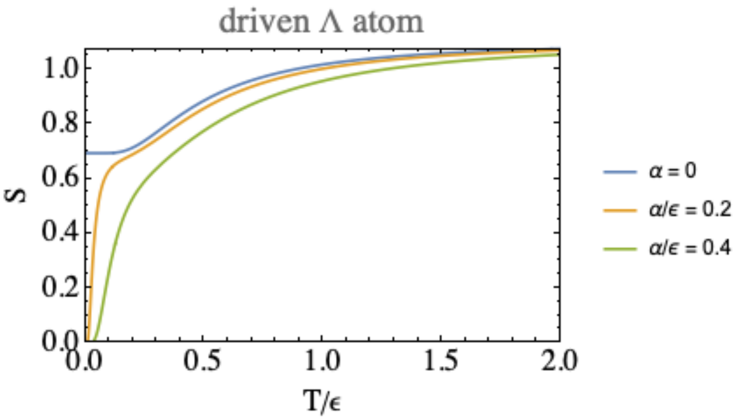


When $T \to \infty$, $S \to \ln 3$. This makes sense since the dimension of the Hilbert space is 3. But when $T \to 0$ the entropy now tends to $\ln 2$ and not zero! That can actually be seen directly from (18) by setting $\beta\epsilon \to \infty$.

Now here is something even more intriguing. Consider the driven Λ atom studied in page 33 of lecture 1:

$$Z = 1 + 2e^{-\beta\epsilon/2}\cosh\left(\frac{\beta\sqrt{\epsilon^2+8\alpha^2}}{2}\right) \tag{19}$$

Here is how the entropy looks like for different α:



Focus on $T \to 0$.

when $\alpha = 0$, we get $S = \ln 2$

But when $\alpha \neq 0$, we always get $S = 0$. It only goes actually to 0 when T is very small. But it always eventually gets there, for any $\alpha \neq 0$.

The reason behind this strange behaviour is the degeneracy of the ground state. when $\alpha = 0$ the ground state is doubly degenerate. But any $\alpha \neq 0$ breaks this degeneracy. Let's now understand how the entropy is sensitive to this feature.

The 3rd law of thermodynamics

Consider the ratio of two Gibbs probabilities

$$\frac{p_m}{p_n} = e^{-\beta(E_m - E_n)}$$

Assume $E_m > E_n$. Then when $T \to 0$ $(\beta \to \infty)$ the quantity $\beta(E_m - E_n)$ becomes very large and thus $e^{-\beta(E_m - E_n)}$ becomes very small. Hence

$$\text{If } E_m > E_n \text{ then } p_m \ll p_n \text{ when } T \to 0$$

In the limit $T \to 0$ then the only state which remains populated is the ground state.

Now suppose the ground-state is g-fold degenerate. Equal energies have equal probabilities, so all ground states must have the same populations. Because of normalization, these probabilities will have to be $1/g$. Thus

$$\lim_{T \to 0} p_m = \begin{cases} 1/g & \text{if } E_m = E_{gs} \\ 0 & \text{otherwise} \end{cases}$$

Now recall that $p \ln p$ is zero when $p = 0$. Thus, the only terms which will contribute to S are the ground states:

$$S = -\sum_n \mathscr{P}_n \ln \mathscr{P}_n = -\sum_{\substack{\text{ground}\\\text{states}}} \frac{1}{g} \ln 1/g$$

$$= -\frac{g}{g} \ln 1/g$$

since there are g ground states. Simplifying:

$$\boxed{\lim_{T\to 0} S = \ln g = \ln(\text{degeneracy of the ground state})} \qquad (20)$$

This is called the 3rd law of thermodynamics. It is a quantum feature related to the discreteness of the energy levels. If you construct stat. mech. classically you find an entropy which diverges when $T \to 0$.

So far we are talking only about systems composed of a single particle. Soon we will generalize all this to systems composed of N particles, where N is very large. In this case it is customary to talk about the entropy per particle $s = S/N$. Eq (20) then becomes

$$\lim_{T\to 0} s = \frac{1}{N} \ln g(N) \qquad (21)$$

where I wrote $g(N)$ to emphasize that now the degeneracy may itself depend on N.

Many systems in Nature are such that g is either independent of N or grow polynomially with N. In this case, in the limit $N \to \infty$ $\frac{1}{N} \ln g = 0$. For this reason, in the past it was common to formulate the 3rd law as saying that $s \to 0$ when $T \to 0$. However, we nowadays know that there are systems for which g actually grows exponentially with N, so that $\frac{1}{N} \ln g$ remains finite when $N \to \infty$. A famous example are the so-called spin glasses. Eq (20) is therefore the correct way of stating the problem.

Let's now see what the 3rd law says about the heat capacity. Recall Eq (16): $C = T\, \partial S / \partial T$. This means that

$$S(T) - S(T_0) = \int_{T_0}^{T} dT\, \frac{C(T)}{T} \tag{22}$$

According to the 3rd law, S should tend to a finite constant, so $S(T) - S(T_0)$ should tend to zero when T and T_0 are small. If $C(T) \sim e^{-\Delta/T}$, as happens for gapped systems, then this will be definitely true since $e^{-\Delta/T}$ goes to zero super super fast.

Let's then assume something a bit softer. For instance, suppose $C(T) \sim T^{x}$, where x is some exponent. Then (22) reduces to

$$S(T) - S(T_0) \sim \frac{T^{x} - T_0^{x}}{x}$$

This has to vanish when T, T_0 get small. Thus, we must have $x > 0$. This therefore implies that $C(T) \to 0$ as $T \to 0$:

$$\boxed{S \to \text{const} \quad \Longrightarrow \quad C(T) \to 0}$$

The heat capacity should thus vanish as $T \to 0$.

Helmholtz free energy

Looking back at (15), let us define a new quantity

$$\boxed{F = -T \ln Z} \tag{21}$$

called the Helmholtz free energy. Then we can write (15) as

$$\boxed{S = \beta U + \ln Z = \beta(U - F)} \tag{22}$$

The free energy turns out to be a very important quantity in thermodynamics, since it is pretty much the partition function. For instance $U = -\frac{\partial}{\partial \beta} \ln Z$ can be written as

$$\begin{aligned} U = -\frac{\partial}{\partial \beta} \ln Z &= -T^2 \frac{\partial}{\partial T}\left(\frac{F}{T}\right) \\ &= F - T\frac{\partial F}{\partial T} \end{aligned} \tag{23}$$

Plugging this in (22) then yields

$$\boxed{S = -\frac{\partial F}{\partial T}} \tag{24}$$

Combining this with (16) then yields

$$\boxed{C = T\frac{\partial S}{\partial T} = -T\frac{\partial^2 F}{\partial T^2}} \tag{25}$$

Since $C > 0$, we see that F must be concave in T.

Free energy and relative entropy

The free energy actually turns out to be connected to the relative entropy (7). Consider a system prepared in some generic non-equilibrium state with probabilities p_m (not thermal!). Based on Eq.(22), let us now define the non-equilibrium free energy $F(p)$ associated with the distribution p, as

$$F(p) = U(p) - TS(p) \tag{26}$$

where $U(p) = \sum_m E_m p_m$ is the average energy in the distribution p_m and $S(p) = -\sum_m p_m \ln p_m$ is the corresponding Shannon entropy. We will now compare $F(p)$ with another quantity, namely $S(p\|p_{th})$, the relative entropy between p_m and the thermal state $p_m^{th} = e^{-\beta E_m}/Z$. We have

$$\begin{aligned} S(p\|p_{th}) &= \sum_m p_m \ln p_m / p_m^{th} \\ &= \sum_m p_m \ln p_m - \sum_m p_m \ln p_m^{th} \\ &= -S(p) - \sum_m p_m \left[-\beta E_m - \ln Z\right] \\ &= -S(p) + \beta \underbrace{\sum_m E_m p_m}_{U(p)} + \ln Z \end{aligned}$$

Thus

$$S(p\|p_{th}) = \beta U(p) - S(p) + \ln Z \tag{27}$$

But $\ln Z$ is related to the equilibrium free energy (21):

$$F_{th} = -T \ln Z.$$

If we multiply (27) by T on both sides, we also see the appearance of the non-eq. free energy (26):

$$T\, S(\rho \| \rho_{th}) = U(\rho) - T S(\rho) + T \ln Z$$
$$= F(\rho) - F_{th}.$$

Thus, we reach the really really important conclusion that

$$\boxed{F(\rho) = F_{th} + T\, S(\rho \| \rho_{th})} \qquad (28)$$

this relates eq. and non-eq. free energies with the distance between the non-equilibrium state and the thermal state.

From the Gibbs inequality (8), $S(\rho \| \rho_{th}) \geq 0$, from which we conclude that

$$\boxed{F(\rho) \geq F_{th}, \qquad \forall \text{ dist } \rho.} \qquad (29)$$

This gives a really interesting way of defining what equilibrium is: equilibrium is the state which minimizes the free energy.

When there is no temperature involved, equilibrium is the state which minimizes the energy U; i.e., the ground state. But when $T \neq 0$, this is no longer true. Instead, now, equilibrium will be the state minimizing the free energy $F = U - TS$. Energy U competes with thermal disorder $-TS$ to give the equilibrium state.

We can also look at (28) in a different way by writing

$$U(p) - TS(p) = U_{th} - TS_{th} + TS(p \| p_{th})$$

or, rearranging,

$$\boxed{S(p) = S_{th} + \beta\left[U(p) - U_{th}\right] - S(p \| p_{th})} \qquad (30)$$

Instead of saying that equilibrium is the state which minimizes the free energy, this expression allows us to view equilibrium as the state which maximizes the entropy. But it can't just be some naive maximization, as we already know that the state with the largest entropy is the uniform distribution (compare with Eqs (10) and (11)).

But looking at (30) we see that if we restrict the distributions p_m to those for which $U(p) = U_{th}$, then the 2^{nd} term vanishes and Gibbs' inequality immediately gives p_m^{th} as the largest possible entropy.

To summarize:

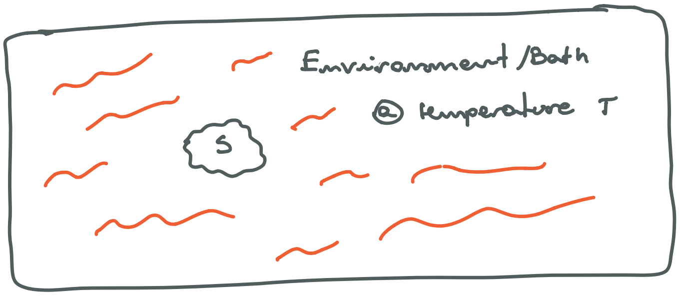
- Equilibrium is the state which minimizes the free energy.
- Equilibrium is the state, within the set of states having $U(\rho) = U_{th}$, which maximize the entropy.

These two statements are, of course, entirely equivalent.

Relaxation to equilibrium

When we talk about thermal equilibrium, it does not matter how the system equilibrated. All that matters are the populations $p_n^{th} = e^{-\beta E_n}/Z$ characterizing the equilibrium state. Relaxation towards equilibrium and equilibrium are two different problems. And this is good because the former is much more difficult to describe.

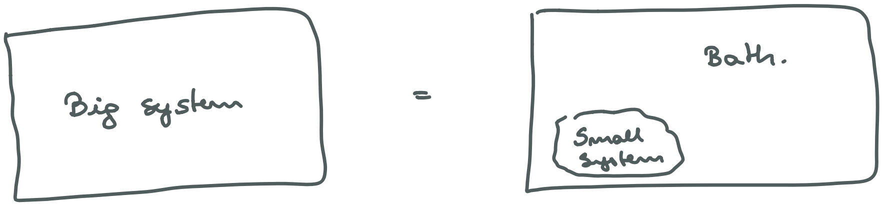
There are two main ways through which a system may reach thermal equilibrium. One way is when it is in contact with a heat bath:



A bath is any very large, very complex, very chaotic system. A bucket of water is a good example. The idea is that the bath is so complex that its fine details don't really matter. All that matters are some of its more coarse-grained properties like temperature, volume and so on.

When a system is put in contact with a bath it will usually thermalize at the same temperature as the bath. This is the reason why we can talk about temperature when we deal with small systems like a single atom, for example. We don't try to define the temperature of a single atom. Instead, we can simply imagine that the β in $e^{-\beta E}$ is actually a property of the bath with which the system is coupled.

On the other hand, when our system is itself very large, it doesn't really need a bath. Instead, the system will equilibrate even if left by itself. The idea is that in this case one part of the system will act as a bath to the other parts.



In this argument, chaos and complexity become very important. If the "small system" shakes a little bit, this will cause ripples in the "bath". Chaos ensures that these ripples never ever come back to the system. They are just lost in all the chaotic mess of the bath.

Master equation

The biggest difficulty in describing the relaxation towards equilibrium is that it is model specific: it depends on the details of the system and how it is coupled to the bath. this is to be contrasted with equilibrium, which is universal. Notwithstanding, I would like to describe to you one simple, but widely used model of relaxation known as a master equation (or MEq).

The idea is to model the dynamics of the populations $P_n(t)$ by a simple first order equation of the form

$$\boxed{\frac{dP_n}{dt} = \sum_m W_{n,m} P_m(t)} \qquad (31)$$

where $W_{n,m}$ represent transition rates from m to n. These elements cannot be arbitrary because the variables in question are probabilities and thus should satisfy $P_n(t) \in [0,1]$ and $\sum_n P_n(t) = 1$ for all times.

Summing (31) over n yields

$$\sum_n \frac{dP_n}{dt} = \underbrace{\frac{d}{dt} \underbrace{\sum_n P_n}_{1}}_{0} = \sum_{n,m} W_{n,m} P_m(t)$$

Thus, the W_{nm} must be such that

$$\sum_{n,m} W_{nm} P_m = 0$$

But this must be true for any p_m. Thus, we conclude that

$$\boxed{\sum_n W_{nm} = 0} \tag{32}$$

In words, the columns of W add up to 1.

If we discretize time in (31) we may write

$$p_n(t+\Delta t) \simeq p_n(t) + \Delta t \sum_m W_{n,m}\, p_m(t) \tag{33}$$

The left-hand side has to be non-negative, since it is a probability. And this must be true for all p_m. Thus, we conclude that $W_{nm} \geqslant 0$. The only exception is W_{nn}, which can be negative because of the $p_n(t)$ in Eq (33). In fact, because of (32) we may write

$$W_{nn} = -\sum_{k\neq n} W_{kn} \tag{34}$$

which clearly shows that $W_{nn} < 0$ and, in fact, that W_{nn} is fully determined by the off-diagonal entries.

Using this we can also write (31) as

$$\begin{aligned}\frac{dp_n}{dt} &= \sum_{m\neq n} W_{nm}\, p_m + W_{nn}\, p_n \\ &= \sum_{m\neq n} W_{nm}\, p_m - \Big(\sum_{m\neq n} W_{mn}\Big) p_n\end{aligned}$$

which can be written more nicely as

$$\boxed{\frac{dp_n}{dt} = \sum_{m\neq n} \left\{ W_{nm}\, p_m - W_{mn}\, p_n \right\}} \qquad (35)$$

This clearly shows how the MEq is nothing but a balance equation. Probability enters p_n from p_m at a rate W_{nm}. And it leaves p_n towards p_m at a rate W_{mn}. The rate at which p_n changes (dp_n/dt) is thus determine by the net flow of probability comming from all other states minus the net flow of probability leaving p_n. Quite nice eh?

The values of the rates $W_{n,m}$, however, are highly model specific, so to push further we have to start making some assumptions. First and foremost, we are interested in using (35) to model the relaxation towards equilibrium. So we will assume that the W_{nm} are such that the Gibbs state

$$p_n^{th} = e^{-\beta E_n}/Z \qquad (36)$$

is a fixed point of (35). That is

$$\sum_{m\neq n} \left\{ W_{nm}\, p_m^{th} - W_{mn}\, p_n^{th} \right\} = 0 \qquad (37)$$

We will also assume that this fixed point is unique.

You can view (37) as a kind of global balance: it essentially says that in equilibrium the net flow of probability in and out of each state n balances out.

It turns out, however, that the vast majority of systems in nature satisfy an even stronger property known as detailed balance. Namely, each term in (37) actually vanishes in equilibrium

$$\boxed{W_{nm} P_m^{th} = W_{mn} P_n^{th}} \tag{38}$$

Detailed balance means that it is not only the net flows which vanish, but the individual flows between each pair of states (n, m). Detailed balance is a very physical property. It depends on the details of how the system interacts with the bath. Notwithstanding, it turns out to actually be quite frequent in Nature. And what is more important, it dramatically simplifies the analysis.

Another way of writing (38) is as

$$\boxed{\frac{W_{nm}}{W_{mn}} = e^{-\beta(E_n - E_m)}} \tag{39}$$

This emphasizes an important fact about the W_{nm}: in general the details of W_{nm} depend on the specifics of the problem. But the ratios between $m \to n$ and $n \to m$ only depends on the energy difference between these two states.

The 2^{nd} law of thermodynamics

The 2^{nd} law is the most important law in thermodynamics since it establishes what you can and cannot do. It can be formulated in many ways. Now we will see the first one of them.

Consider the quantity

$$\Pi = -\frac{d}{dt} S(\rho(t) \| \rho_{th}) \qquad (40)$$

where $S(\rho(t)\|\rho_{th})$ is the relative entropy between the state of the system at time t and the equilibrium state to which the system will eventually relax to.

Using (28) we can also write

$$\begin{aligned} S(\rho\|\rho_{th}) &= \beta\left(F(\rho) - F_{th}\right) \\ &= \beta U(\rho) - S(\rho) - \beta F_{th} \end{aligned}$$

Thus (40) can be written as

$$\boxed{\Pi = -\frac{d}{dt} S(\rho(t) \| \rho_{th}) = -\beta \frac{dF}{dt} = \frac{dS}{dt} - \beta\frac{dU}{dt}} \qquad (41)$$

We will come back to the interpretation of this formula in a second.

But before doing so, let's open it up and write

$$\Pi = -\frac{d}{dt} \sum_n \left\{ p_n(t) \ln p_n(t) - p_n(t) \ln p_n^{th} \right\}$$

$$= - \sum_n \left\{ \frac{dp_n}{dt} \ln p_n + p_n(t) \frac{1}{p_n(t)} \frac{dp_n}{dt} - \frac{dp_n}{dt} \ln p_n^{th} \right\}$$

The term in the middle vanishes:

$$\sum_n \frac{dp_n}{dt} = \frac{d}{dt} \sum_n p_n = 0$$

thus

$$\Pi = - \sum_n \frac{dp_n}{dt} \ln p_n / p_n^{th}$$

Using (35) we now get

$$\Pi = - \sum_{n, m \neq n} \left\{ W_{nm} p_m - W_{mn} p_n \right\} \ln p_n / p_n^{th}$$

Now comes a naughty trick. Since we are summing over both n and m, let us exchange n ↔ m:

$$\Pi = - \sum_{n, m \neq n} \left\{ W_{mn} p_n - W_{nm} p_m \right\} \ln p_m / p_m^{th}$$

Instead of using one form or the other, we use an average of the two:

$$\Pi = -\frac{1}{2} \sum_{n, m \neq n} \left\{ W_{nm} p_m - W_{mn} p_n \right\} \ln \frac{p_n}{p_m} \frac{p_m^{th}}{p_n^{th}}$$

Finally, we use the detailed balance property (38) to write

$$\frac{p_m^{th}}{p_n^{th}} = \frac{W_{mn}}{W_{nm}}$$

We then get

$$\Pi = -\frac{1}{2} \sum_{n, m \neq n} \left\{ W_{nm} p_m - W_{mn} p_n \right\} \ln \frac{W_{mn} p_n}{W_{nm} p_m}$$

or, getting rid of the minus sign,

$$\boxed{\Pi = \frac{1}{2} \sum_{n, m \neq n} \left\{ W_{mn} p_n - W_{nm} p_m \right\} \ln \frac{W_{mn} p_n}{W_{nm} p_m}} \qquad (42)$$

The relevance of this expression lies in the fact that it is a sum of functions of the form

$$(x - y) \ln x/y \geqslant 0 \qquad (43)$$

which is always non-negative for any $x, y > 0$. Thus we conclude that

$$\boxed{\Pi \geqslant 0} \qquad 2^{nd} \text{ law!} \qquad (44)$$

This simple expression is the 2^{nd} law. Let me tell you why.

We now go back to (41) and write it as

$$\boxed{\frac{dS}{dt} = \Pi + \beta \frac{dU}{dt}} \qquad (45)$$

The fact that $\Pi \geqslant 0$ then implies that

$$\boxed{\frac{dS}{dt} \geqslant \beta \frac{dU}{dt}} \qquad (46)$$

which is known as Clausius' inequality. It relates the change in entropy of a system with the corresponding heat exchanged with the bath. According to Clausius, when a process is done reversibly

$$\Delta S = \beta \Delta Q$$

and when it is irreversible

$$\Delta S \geqslant \beta \Delta Q.$$

Eq (44) now explains why: the process is irreversible when $\Pi \geqslant 0$.

The quantity Π [Eq. (40)] is called the entropy production rate.