An introduction to Thermodynamics

December 25, 2024

Contents

1   Microstates and macroscopic quantities

Suppose that we want to develop a physical model of a gas. Let us assume that this gas consists of a large number of distinguishable particles (molecules) with an identical behavior, that are confined to some volume V which is large compared to the size of an individual particle, but finite (and might in fact change over time), for instance because the gas is contained in a vessel. Then, one approach to describing the state of this physical system could be to record, at any point in time, the exact position and momentum of each of the particles.

If we know how the particles interact with each other and with the vessel, then we could, at least in theory, calculate the future state of the system at any point in time given its state at an initial time. This state is called the *microstate* of the system.

However, it is obvious that, given the large number of particles in a gas, this approach is in reality useless - we will never be able to actually calculate all the interactions, and even if we had a very powerful computer that could do this, we would not be able to measure the position and momentum of all particles at one specific point in time precisely enough.

Fortunately, it turns out that in practice, there are some quantities like the temperature, the volume of the gas and its energy, that we can measure and that follow some empirically derived laws. Thus we could describe the gas at some point in time by specifying these quantities only and see how they develop over time. Such a state is called a *macrostate*. Formally, we could describe a macrostate as the values of a set of functions defined on the space of microstates, one function for each relevant quantity like the energy or the temperature.

We know from experience that the quantities that make up such a state can fluctuate over time. If, for instance, we allow two vessels containing hot liquids to exchange heat, then the temperatures within these vessels will change over time. At some point, however, the system will reach a state called a *thermal equilibrium state* and these quantities will settle to certain values and remain constant until we make a modification to the system.

Some of these quantities, like the number of particles or the volume, are *extensive quantities*: when we split our system into subsystems (which are still assumed to be large compared to the microscopic scale), these quantities add up. Other macroscopic quantities like the temperature and the pressure do not have this property. However, from what follows, we will see that these are in a certain way derived quantities and that it is useful to start our description of these systems with extensive quantities.

Let us now assume that we have, once and for all, picked a set of extensive quantities like the energy U, the number of particles N and the volume V, that we combine into a vector

$$X = (U, V, N)$$

that we use the fully specify the macroscopic state of our system. Now consider a situation where we have two initially separated systems, say Σ_1 and Σ_2, that are in equilibrium and hence, X is constant over time for each of them if seen in isolation. Let us now bring these two systems together and wait until equilibrium is reached again. Then, we know, as we did assume our macroscopic variables to be extensive, that the energy, number of particles and the volume of the new system in total is simply the sum of the individual quantities. But obviously, the values for the individual systems will have changed. If, for instance, we allow two systems, one with a very high temperature and one with a low temperature, to exchange energy, we expect that in the equilibrium state, the first system will have lost some of its energy which has been absorbed by the second system. So how can we find the new

values of X for each of the systems? The answer is closely related to the concept of thermodynamical entropy.

2   Entropy

One way to introduce the entropy is using a maximum principle - we want to be able to characterize those states that are actually realized after reaching equilibrium as the macrostates that maximize a certain function. Following [4], section 1.10, we phrase the assumption that such a function exists as follows (this is in fact a combination of postulate II and postulate III in [4])

There exists a function - the entropy - of the extensive parameters of any composite system defined for all equilibrium states with the following property: the values assumed by the extensive parameters in the equilibrium state of an isolated system in the absence of an internal constraint are those that maximize the entropy among all states allowed by the external constraints. Moreover, the entropy is additive with respect to subsystems. It is differentiable and a (strictly) monotonically increasing function of the total energy.

In [4], it is not explicitly mentioned that the maximum principle that is expressed by this statement is only valid for isolated systems, though this becomes apparent from the context. However, we will see later that this restriction is in fact necessary (cf. for instance chapter 9 of [6]). A system can, for instance, reduce its entropy by transferring heat to a system with lower temperature, but this process will inevitably increase the entropy of the composite system. In many practical applications, the composite system is made up by the system under interest and some concieved system acting as the environment that essentially represents the total surroundings of the system under investigation, or - if you want - the entire universe outside of our primary system.

To understand how this principle relates to the famous *second law of thermodynamics*, let us consider a hypothetical system consisting of a barrel with a movable wall separating the barrel into two compartments, as in diagram 1, and assume that the two chambers contain gases with different pressures P_1 and P_2, so that the wall experiences a net force of

$$(P_1 - P_2) \cdot A$$

where A is the area of the wall. Initially, the movable wall is somehow fixed, so that the system has an additional constraint fixing the volumes of the two chambers. Some time after this system has been initially established, the gases inside the two chambers and with them the composite system will have reached equilibrium and the entropy has settled to a certain value.
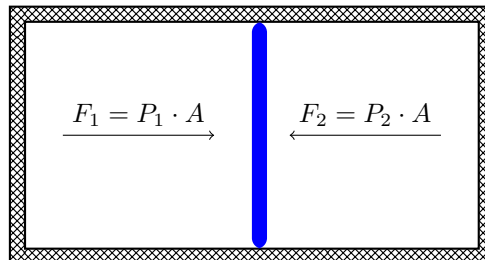

Figure 1: Two chambers in a barrel separated by a movable wall

When we now allow the wall to move freely, we have effectively removed a constraint. The current state of the system is then obviously still allowed by the

remaining constraints, but additional states (i.e. values of the extensive parameters) become accessible. In addition, the system is now no longer in equilibrium. Instead, the wall will start to move until the pressures on both sides are equal and the system settles in a new equilibrium. Our maximum principle assures that the new entropy is maximized in this state. As the initial state was also permissible by the constraints, the new value of the entropy can thus not be smaller than the initial value. In other words, the entropy has increased or remained equal, which is exactly what the famous *second law of thermodynamics* states.

Also note that strictly speaking, the formalism that we are about to develop only assumes the existence of the entropy for equilibrium states. While the wall in the above example is moving, the system is not in equilibrium, so strictly speaking the entropy of the composite system is not even defined during this time. If you think of the states of the system as points in the *configuration space* spanned by the extensive parameters, then intuively speaking, the system will disappear from the configuration space if it moved out of equilibrium, for instance due to some interaction with another system or due to some constraints being altered, and only reappear once it has settled into its new equilibrium. We know that the entropy of this new state is equal to or higher than the entropy of the starting point, but we have lost track of the states of the system while the change took place.

To rectify this, it is common to consider only quasi-static processes, i.e. processes that proceed so slowly that at every point in time, the system is actually in equilibrium. Then the process can be described as a path in configuration space, and we can apply calculus or the idea of infinitesimal changes which we will do heavily in what follows. Note, however, that this is an idealization which however turns out to provide a good approximation to many real processes (see also the discussion in section 4.2 of [4]).

To see the postulates in action and understand entropy further, let us now conduct a thought experiment. Assume that we have two systems that are initially isolated against each other and against the environment and that each of the subsystems is in equilibrium. As the two subsystems cannot interact, the composite system will then also be in equilibrium and will be characterized by extensive quantities like volume, energy and number of particles.

We now bring together the two systems such that they are connected by a thin, but rigid wall that allows heat to pass, but cannot be crossed by any particles (see diagram 2). As soon as this happens, the system is no longer in equilibrium. In fact, energy in the form of heat can now be exchanged between the systems. We have effectively removed a constraint, i.e. the constraint that the individual energies are fixed, but we still have the constraint imposed on us by the conservation of energy, i.e. the sum of the energies remains constant. Thus the energies of the subsystems will change until the composite system has again reached an equilibrium state that we want to determine.
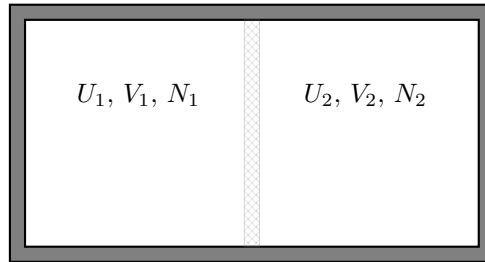

Figure 2: Vessel with a diathermal wall

To apply the principle of maximal entropy, we first have to define a system of coordinates for our configuration space, i.e. a set of extensive parameters of the system such that their combined value characterizes the system completely. The choice of these coordinates system and the associated state space is crucial. In particular, we need to define how we reflect the existing constraints. In our case, a possible choice would be to use the individual energies U_i along with the individual particle numbers N_i and the volumes V_i. However, if we do this, we still have a constraint, namely $U_1 + U_2 = U$, where U is the total energy of the system that we assume to be a constant. This would restrict the possible trajectories of the system in the configuration space to a submanifold. It is easier to adjust coordinates a bit to better reflect this, so we choose U, U_1, N_1, N_2, V_1 and V_2. Then the entropy is a function on the configuration space, i.e.

$$S = S(U, U_1, V_i, N_i)$$

It is very common to include the total energy as a parameter, even though it is usually fixed by the conservation of energy for the overall system. However, we will soon see that the partial derivative with respect to it plays a vital role. More generally, given an arbitrary composite system, one usually choses the total energy U and some extensive parameters of the systems, denoted by X_i, as coordinates for the configuration space. If the energies of all but one subsystem is among the X_i, then we can reconstruct all energies, and the conservation of total energy is expressed as the simple constraint $U = const$.

To extract useful information from the principle of maximum entropy, we now have to relate the total entropy to the entropies of the two subsystems, and this is were the assumption that the entropy is additive comes into play. Specifically, in this example, we have an entropy S^i for each of the subsystems, and additivity implies that

$$S(U, U_1, N_i, V_i) = S^1(U_1, N_1, V_1) + S^2(U - U_1, N_2, V_2)$$

Let us now see what happens if we allow heat to flow between the systems. The principle of maximal entropy tells us that those parameters which are no longer restricted will then settle at a value which maximizes the entropy. In our case, the parameter U_1 is now free to change, and so are the N_i and the V_i. Thus, as function of U_1, the entropy will be maximized. Thus

$$0 = \frac{\partial S}{\partial U_1}$$

at the point of equilibrium. This directly translates into

$$0 = \frac{\partial S^1}{\partial U_1}(U_1, N_1, V_1) - \frac{\partial S^2}{\partial U_2}(U - U_1, N_2, V_2)$$

So the partial derivative of the entropy with respect to the energy seems to control the flow of heat. For a general system with entropy $S(U, V, N, ...)$, we call

$$\beta = \frac{\partial S}{\partial U}$$

the *inverse temperature* and

$$T = \frac{1}{\beta}$$

the *temperature* of the system. Note that this is a partial derivative of the entropy and therefore a function of U, V and N and any other parameters. Thus we find

that stability is reached when the temperatures of both systems reach the same value - a very appealing result.

As part of our postulates, we have assumed that the derivative of S with respect to U is positive, and therefore we can express U as a function of S (and the fixed values for V and N). Then of course

$$\frac{\partial U}{\partial S} = (\frac{\partial S}{\partial U})^{-1} = \frac{1}{\beta} = T$$

When we also take the usual assumption that the entropy is strictly concave, then

$$\frac{\partial \beta}{\partial U} = \frac{\partial^2 S}{\partial U^2} < 0$$

and we can therefore invert β as a function of U, in other words, if, for fixed N and V, we know T, then we know U and vice versa.

Similary, we can find a physical interpretation for the other partial derivatives. For instance, taking the total derivative of

$$S(U(S, V, N), V, N) = S$$

with respect to V - which is zero as the entropy has a minimum at the equilibrium state - gives

$$0 = \frac{d}{dV} S(U(S, V, N), V, N) = \frac{\partial S}{\partial U} \frac{\partial U}{\partial V} + \frac{\partial S}{\partial V}$$

so that we obtain

$$-\frac{\partial U}{\partial V} = T \frac{\partial S}{\partial V}$$

This quantity is called the *pressure* and denoted by P:

$$P = -\frac{\partial U}{\partial V} = T \frac{\partial S}{\partial V}$$

Finally, what is called the *chemical potential* is given by

$$\mu = \frac{\partial U}{\partial N} = -T \frac{\partial S}{\partial N}$$

(see [4], chapter 2). These quantities are called *intensive quantities*. As the entropy is assumed to be homogeneous, they do not change if we pass to a subsystem, i.e. the temperature and pressure of a part of a system are equal to the temperature and pressure of the full system (in equilibrium). As the fundamental relations that we have derived so far will appear over and over again in thermodynamics, let us summarize them once more before moving on.

$$\boxed{\begin{aligned} T &= \frac{1}{\beta} = (\frac{\partial S}{\partial U})^{-1} \\ P &= -\frac{\partial U}{\partial V} = T \frac{\partial S}{\partial V} \\ \mu &= \frac{\partial U}{\partial N} = -T \frac{\partial S}{\partial N} \end{aligned}}$$

So the temperature describes how the energy depends on the entropy, the pressure measures how it depends on the volume and the chemical potential measures the contribution of the particle number to the energy.

This simple example can be generalized as follows. Suppose we have an entropy presented as a function

$$S = S(U, X_1, \ldots, X_n)$$

where the X_i are some arbitrary extensive variables that together with U span the configuration space. We also assume that the total energy is constrained to a specific value U_0 but that the other parameters can vary freely (implicitly, we therefore make the assumption that the configuration is an open subset of $\mathbb{R}^{n+1}$ so that we can apply calculus or at least has a non-empty interior). The maximum entropy principle can then be phrased conveniently in terms of differentials and the Hessian. In fact, the entropy is a smooth function which has a differential dS, and so is the coordinate function U. If we think of a tangent vector as an infinitesimal variation, then the variations which are compatible with the constraint $U = U_0$ are the tangent vectors in the kernel of dU. The maximum entropy principle then states that on the subspace $U = U_0$, the entropy S, as a function of the X_i, has a maximum. If we assume that this is true for every U_0, this implies that, at the equilibrium point, the kernel of dU is contained in the kernel of dS and that the Hessian matrix

$$\mathcal{H}_{X_i}(S) = (\frac{\partial^2 S}{\partial X_i X_j})_{i,j}$$

of partial derivatives with respect to the X_i is negative definite.

It is important to realize that macroscopic variables are only defined for systems that are in equilibrium. If we remove an internal constraint, then, for some time, the system will not be in equilibrium and therefore *not* be represented by a point in the configuration space. Once we reach equilibrium, the system "reappears" in the configuration space and will be located at a (nearby, if the change was small) point. In theoretical considerations, we often work with idealized processed which are supposed to be a sequence of small changes to the system such that points in the configuration space given by the equilibrium states are so close to each other that they are reasonably close to a continuous or even smooth curve γ in the configuration space, so that we can apply the apparatus of calculus to draw conclusions. This is called a *quasi static process*. If we are particularly interested in the state of a specific subsystem, say the first system, we could alternatively work in the configuration space given by X_1 and X, and observe how the equilibrium states of the composite system move in this space to deduce the changes in X_i.

Example 2.1 (see 4.1 and 4.4 in [4]). Let us now make our example in diagram 2 more tangible by assuming a specific relation between entropy and the extensive parameters which holds for both subsystems. Specifically, let us assume that our system consists of two chambers, both filled with an ideal gas, with particle numbers N_1 and N_2, so that the energy of each subsystem is given by

$$U_i = cN_iRT_i$$

In a later example, we will derive a formula for the entropy of an ideal gas, namely

$$S = Ns_0 + NR\ln\big[(\frac{U}{U_0})^c\frac{V}{V_0}N^{-c-1}\big]$$

Thus, up to an additive constant that depends on particle number and volume, the entropies of the individual subsystems are given by

$$S_i = cN_iR\ln U_i$$

Now let us again assume that at some point, we allow heat to flow between the two subsystems, but keep their volumes and particle numbers constant. Let U denote the total energy of our system, so that $U = U_1 + U_2$. Using the additivity of the entropy, the total entropy S is then a function of, say, U_1 and U and given by

$$S = cN_2R\ln[U_1^{\frac{N_1}{N_2}}(U - U_1)] = cN_2R\ln[UU_1^{\frac{N_1}{N_2}} - U_1^{\frac{N_1}{N_2}+1}]$$

Let us now fix the value U of the total energy and determine the equilibrium state by maximizing S. The derivative of the entropy with respect to U_1 is given by

$$\frac{dS}{dU_1} = \frac{cN_2R}{U_1^{\frac{N_1}{N_2}}(U-U_1)}[\frac{N_1}{N_2}UU_1^{\frac{N_1}{N_2}-1} - (\frac{N_1}{N_2}+1)U_1^{\frac{N_1}{N_2}}]$$

$$= \frac{cN_2R}{U_1(U-U_1)}[\frac{N_1}{N_2}U - \frac{N_1+N_2}{N_2}U_1]$$

$$= \frac{cR}{U_1(U-U_1)}[N_1U - (N_1+N_2)U_1]$$

which is zero if and only if the final energy of system one is given by

$$U_{1,f} = U\frac{N_1}{N}$$

i.e. the energies are distributed in the same ratio as the particles (which is not quite surprising). Let us now compute the final temperature T_f of subsystem one (and then, by the equilibrium condition, subsystem two as well). We have

$$T_f = \frac{U_{1,f}}{cN_1R} = \frac{U}{(N_1+N_2)cR}$$

However

$$U = U_1 + U_2 = cRN_1T_1 + cRN_2T_2$$

so that we eventually obtain that the final temperature is the weighted arithmetic mean of the initial temperatures:

$$T_f = \frac{N_1T_1 + N_2T_2}{N_1 + N_2}$$

As volume and particle number for each of the subsystems is constant and the energy of each subsystem is proportional to its temperature, the change of the entropy of subsystem i will be given by

$$\Delta S_i = N_iC\ln\frac{T_f}{T_i}$$

Note that this is *not* necessarily positive, i.e the entropy of a subsystem can decrease. If, for instance, $N_1 = N_2$, then T_f will simply be the mean of the T_i, and therefore $T_f > T_i$ for one of the i (unless the initial temperatures are already identical). Then the system i will increase its temperature and its entropy, but the other subsystem will decrease its entropy and its temperature. The important point, however, is that change in the total entropy, given by

$$\Delta S_1 + \Delta S_2 = C[N_1C\ln\frac{T_f}{T_1} + N_2\ln T_fT_2] = C\ln\frac{T_f^{N_1+N_2}}{T_1^{N_1}T_2^{N_2}}$$

is positive. And this is always true, as the numerator is guaranteed to be at least equal to the denominator by the inequality of weighted arithmetic and geometric mean, and therefore the logarithm is never negative.

Also note that in this case, the entropy changes even though no heat is exchanged with the environment. Thus heat flow between a system and its environment is not the only source of a change in entropy, other sources are possible, like - as in this case - an internal flow of heat or, as we will see later, the process of mixing to initially separated sub-systems.

3   Quasistatic and irreversible processes

We have already defined a quasistatic process to be a process during which the system is at every point in time in equilibrium so that macroscopic quantities like temperature and pressure are well defined. As this term is often the source of confusion, let us take some more time to discuss this.

One of the reasons why the concept of quasistatic processes can be challenging is that different authors use that term with different meanings. In [5], a quasi-static process is first defined in section 1.5 for the special case of a slow compression.

> I need to assume that as the gas is compressed it always remains in internal equilibrium, so that its pressure is uniform from place to place (and hence well defined). For this to be the case, the piston's motion must be reasonably slow, so that the gas has time to continously equilibrate to the changing conditions. The technical term for a volume change that is slow in this sense is quasistatic.

So the key point here is that the macroscopic variables are at any point in time well defined, so that it makes sense to speak about temperature and pressure of a system and its entropy (which is usually assumed to be a function of these variables). A bit later, in section 2.6, when reversible and irreversible processes are discussed, Schroeder defines a process to be reversible if its entropy remains unchanged and explicitly points out that while every reversible process is quasi-static (otherwise its entropy would not be well defined) but not every quasi-static process is reversible:

> In Chapter 3 I will prove that every reversible volume change must in fact be quasi-static...a quasi-static process can still be irreversible, however, if there is also heat flowing in or out or if entropy is being created in some other way.

Unfortunately, some authors use the terms quasi-static and reversible interchangeably. An example is [6], where Sweden states in section 11.1 that

> In the limit of infinitesimal steps, the increase of entropy can vanish. Such a series of infinitesimal steps is called a quasi-static process. A quasi-static process is reversible. Since there is no increase in entropy, it could be run backwards and return to the initial state.

We will not adopt this convention but differentiate between quasi-static processes and reversible processes, as it is done by Schroeder. This is also the approach that Callen takes. In [4] section 4.2, he defines a quasistatic process as follows

> ... a process on the hypersurface [the configuration space - added by the author] .... from an initial state to a terminal state. Such a curve is known as a quasi-static locus or a quasi-static process. A quasi-static process is thus defined in terms of a dense succession of equilibrium states.

and proceeds to define a process to be reversible when the entropy does not change along the path in configuration space that the process defines, and to be irreversible if the entropy is increased.

In other words, a quasi-static process is a process which can be sufficiently well approximated by a continous (and in fact smooth curve) in the configuration space spanned by the extensive parameters of the system in question. For every point of this curve, the entropy is then well defined, so that we can apply our formalism to make predictions about the behaviour of the system.

Note however that the term "equlibrium states" is to be taken with a grain of salt. To illustrate this, let us again consider the process of heat flowing between a body and a heat bath. Let us denote the variables referring to the heat bath with an index HB. The heat bath and the body under consideration together then form a composite system. Then we can, as we have done it during our discussion of the identification of the temperature with a partial derivative, use additivity to write the entropy of the overall system as

$$S_{tot}(U_{tot}, U, V_{HB}, V, N_{HB}, N) = S(U, V, N) + S_{HB}(U_{tot} - U, V_{HB}, N_{HB})$$

Let us now assume that the temperature of the heat bath T_{HB} is different from the temperature T of the body. Then we have already seen that the overall system will not be in equilibrium, i.e. $dS_{tot} > 0$. If the heat transfer is sufficiently slow, however, for instance because the body and the heat bath are connected by a wall with very low thermal conductivity, then each system considered in isolation, i.e. the body and the heat bath, will be in internal equilibrium. So each of them has a well defined temperature and entropy, and the above equation makes sense. Thus the process defines a curve in the configuration space and would thus have to be considered as quasi-static, even though the overall system is not in equilibrium.

So we need to interpret the term "equilibrium states" in a pragmatic way (and looking at section 4.4 of Callens book, for instance, where he essentially uses the same approach to discuss heat flow, indicates that he did so as well). Note, however, that the overall system has no physically meaningful temperature. Formally, however, we can of course take the partial derivative of the above expression for the entropy and obtain

$$\frac{\partial S_{tot}}{\partial U_{tot}} = \frac{\partial}{\partial U_{tot}} S_{HB}(U_{tot} - U, V_{HB}, N_{HB}) = \frac{1}{T_{HB}}$$

so that formally, the temperature of the overall system is identical to that of the heat bath. Once the system has settled into an equilibrium, this is the same as the temperature of the body and this expression matches our physical intuition. Until then, however, this does not appear to be meaningful - which should not surprise us, as we already know that the temperature as an intrinsic quantity is only physically meaningful if the system is in equilibrium.

Still, we can work with the formula for the entropy and use it to derive predictions. We could, for instance, relate the temperature in equilibrium to the heat capacities of the heat bath and the body, as Callen does it in section 4.4 of [4]. It is also instructive to look at the change in total entropy. We have, using the fact that by definition, $T_{tot} = T_{HB}$,

$$T_{tot}dS_{tot} = T_{HB}(dS_{HB} + dS)$$
$$= T_{HB}dS_{HB} + (T_{HB} - T)dS + TdS$$

The first term represents the heat flowing into the heat bath. The third term represents the heat flowing into the hot body. The second term, however, is an additional contribution to the entropy. If conservation of energy holds for the composite system and no mechanical work is done, then

$$0 = dU_{tot} = dU_{HB} + dU = T_{HB}dS_{HB} + TdS$$

and combining these two equations shows that

$$T_{tot}dS_{tot} = (T_{HB} - T)dS$$

So this term represents the excess entropy that is generated due to the heat transfer between two system with different temperatures. We have seen that, unless $T_{HB} = T$, $dS_{tot} > 0$, so this process is irreversible and entropy is created out of nothing.

4 The energy representation

So far we have worked with the entropy, considered its derivatives and used them to draw conclusions on equilibrium states. Essentially, we have seen that the entire behavior of the macroscopic variables of a thermodynamical system is governed by the entropy. This is called the *entropy representation*. In this representation, the energy U is a variable and the entropy a function on the configuration space.

However, we have also seen that the relation between entropy and energy for given V and N can be inverted, so that we can express the energy as

$$U = U(S, V, N)$$

and we have used the chain rule to relate the partial derivatives of U to those of S. Clearly, U as a function of S contains the same information, as we can always revert to S as a function of U. In many cases, however, it is more convenient to work with the energy as a primary quantity. This is called the *energy representation*, and in this representation, the entropy is a coordinate and the energy a function on the configuration space.

Let us now write down the total differential of the energy U, as a function of S, V and N. Using what we have learned about the partial derivatives of U in the previous section, we find that

$$\boxed{dU = TdS - PdV + \mu dN}$$

This equation is so important that it is often referred to as the *fundamental thermodynamic equation*. It corresponds to three different ways to add energy to a system. First, we can exercise mechanical force by decreasing the volume. This will require an amount of work given by

$$dW = Fdx = PAdx = -PdV$$

We therefore call the quantity (mathematically this is a one-form on the state space)

$$dW = -PdV$$

the *work* done on the gas. Note that this is in general not a closed one-form and not the total differential of any function W, so the notation has to be taken with care. Similarly, we call the one-form

$$dQ = TdS$$

the *heat*. When we consider a thermodynamical process, i.e. a path in the state space, which leaves the number of particles constant, we then find that

$$dU = dQ + dW$$

In other words, the change of energy, assuming a constant number of particles, is the sum of the heat added to the system and the work done on the system. This is the famous *first law of thermodynamics*. We note that - as discussed before - our formalism applies only to quasi-static processes, which is why some sources call TdS the quasi-static heat and $-PdV$ the quasi-static work.

The identification of the transferred heat dQ with TdS can be inverted to arrive at the relation

$$dS = \frac{dQ}{T}$$

To illustrate the power of this seemingly simple relation, let us consider a *heat engine* which is a device that can exchange heat with a reservoir at temperate T_h and a

second reservoir at a lower temperature T_l in order to produce some mechanical work W. This engine is supposed to operate in cycles, so that after one cycle, the state of the engine is restored to its original state.

Let us assume that during one cycle, the engine consumes a certain amount Q_h out of the reservoir with the higher temperature and transfers some heat Q_l into the reservoir with the lower temperature. As the internal energy of the engine does not change during a cycle, we can conclude that

$$W = Q_h - Q_l$$

or

$$\frac{W}{Q_h} = 1 - \frac{Q_l}{Q_h}$$

This quantity is often called the *efficiency* of the engine and denoted by e. On the other hand, using the form of the first law stated above, we can conclude that the entropy of the first reservoir is decreased by $\frac{Q_h}{T_h}$ while that of the reservoir with the lower temperature is increased by $\frac{Q_l}{T_l}$. Thus the total change of entropy of the entire system (the engine plus the reservoirs) is

$$\Delta S = \frac{Q_l}{T_l} - \frac{Q_h}{T_h}$$

As this is non-negative, we can conclude that

$$\frac{Q_l}{T_l} \geq \frac{Q_h}{T_h}$$

or

$$\frac{Q_l}{Q_h} \geq \frac{T_l}{T_h}$$

Combining this with our expression for the efficiency, we find that

$$e = 1 - \frac{Q_l}{Q_h} \leq 1 - \frac{T_l}{T_h}$$

Thus we have derived a universal upper boundary for the efficiency of a heat engine, without even making any assumptions about the exact nature of the engine or having a closed expression for the entropy. In many cases, however, the entropy can be expressed as an expression of the extensive parameters, either derived from statistical considerations, which we will do in a later chapter, or based on empirical considerations. To illustrate this, let us consider an example.

5   The model of an ideal gas

An *ideal gas* is a thermodynamical system which is given by the following fundamental relations:

$$PV = NRT$$
$$U = cNRT$$

Here R is the *gas constant* which is related to Avogadros number N_A and the Boltzmann constant k_B by $R = N_A k_B$ and c is typically equal to 3/2. As the *heat capacity at constant volume* is given by

$$c_V = \frac{\partial U}{\partial T}\big|_V = cNR$$

we see that up to the constant R, the number c is the heat capacity of the gas per mole of substance. Also note that the *equipartition theorem* relates c to the number f of degrees of freedom by

$$f = 2c$$

Let us now see whether we can derive the entropy of the ideal gas from these two fundamental relations. We know that the entropy is homogeneous. Thus we can define an entropy density s by

$$s(U, V) = S(U, V, N = 1)$$

and have that

$$S(U, V, N) = NS(\frac{U}{N}, \frac{V}{N}, 1) = Ns(\frac{U}{N}, \frac{V}{N})$$

Of course s has the same partial derivatives with respect to U and V as S, and therefore we obtain the following expression for its total differential:

$$ds = \frac{\partial S}{\partial U} dU + \frac{\partial S}{\partial V} dV$$

We can express this by the intensive quantities that we have defined and obtain

$$ds = \frac{1}{T} dU + \frac{P}{T} dV$$

Using the fundamental relations above with $N = 1$, we have

$$\frac{1}{T} = cR\frac{1}{U}$$
$$\frac{P}{T} = \frac{R}{V}$$

so that

$$ds = cR\frac{1}{U} dU + \frac{R}{V} dV$$

and therefore we obtain in a bit of a shorthand notation

$$ds = cRd\ln U - Rd\ln V$$

This allows us to find an expression for s, namely

$$s = s_0 + cR\ln\frac{U}{U_0} - R\ln\frac{V}{V_0} = s_0 + R\ln\left[\left(\frac{U}{U_0}\right)^c \frac{V}{V_0}\right]$$

with an unknown constant s_0, a reference volume V_0 and a reference energy U_0. By multiplying by N again, we find that (note that we need to use $\frac{U}{N}$ and $\frac{V}{N}$ according to the defining relation for s)

$$S = Ns_0 + NR\ln\left[\left(\frac{U}{NU_0}\right)^c \frac{V}{NV_0}\right] = Ns_0 + NR\ln\left[\left(\frac{U}{U_0}\right)^c \frac{V}{V_0} N^{-c-1}\right]$$

If we want, we can now express the unknown constant in terms of the entropy of a reference state - in fact, we find that this is the entropy of a single particle at volume V_0 and energy U_0. Conversely, if we start with an assumed entropy of this type, we can compute

$$\frac{1}{T} = \frac{\partial S}{\partial U} = \frac{cRN}{U}$$

which gives us

$$U = cRNT$$

and similarly

$$PV = -TV\frac{\partial S}{\partial V} = TV\frac{RN}{V} = NRT$$

Thus we can recover the laws of an ideal gas from the entropy.

For later reference, let us also express the entropy in terms of temperature, volume and particle number. We can use the ideal gas relations to write

$$\left(\frac{U}{U_0}\right)^c\frac{V}{V_0}N^{-c-1} = N_0^{-c}\left(\frac{T}{T_0}\right)^c\frac{V}{V_0}$$

so that, absorbing $-cR\ln N_0$ into the constant s_0 and splitting the logarithm, we find that

$$S = Ns_0 + cNR\ln\frac{T}{T_0} + NR\ln\frac{V}{V_0}$$

Let us now use the formula that we have found for the entropy to derive a similar expression for the energy U. First, let us calculate the chemical potential μ. The partial derivative of S with respect to N is given by

$$\frac{\partial S}{\partial N} = \frac{\partial}{\partial N}\left(Ns_0 + NR\ln\left[\left(\frac{U}{U_0}\right)^c\frac{V}{V_0}N^{-c-1}\right]\right)$$

$$= s_0 + R\ln\left[\left(\frac{U}{U_0}\right)^c\frac{V}{V_0}N^{-c-1}\right] + NR\frac{\partial}{\partial N}\ln\left[\left(\frac{U}{U_0}\right)^c\frac{V}{V_0}N^{-c-1}\right]$$

The first two terms are recognized as $\frac{S}{N}$. The last term can be calculated as follows

$$\frac{\partial}{\partial N}\ln\left[\left(\frac{U}{U_0}\right)^c\frac{V}{V_0}N^{-c-1}\right] = \frac{\partial}{\partial N}\ln\left[N^{-c-1}\right]$$

$$= -(c+1)\frac{\partial}{\partial N}\ln N = -(c+1)\frac{1}{N}$$

Putting all this together yields

$$\frac{\partial S}{\partial N} = \frac{S}{N} - (c+1)R$$

When we combine this expression with the definition of μ as $-T$ times the partial derivative, we obtain

$$\mu = -Ts_0 - TR\ln\left[\left(\frac{U}{U_0}\right)^c\frac{V}{V_0}N^{-c-1}\right] + (c+1)RT = -T\frac{S}{N} + (c+1)RT$$

which we can also write as

$$\mu N = -TS + cNRT + PV$$

Taking a look at the equations that we started with, the second term is simply the energy and the third term is PV. We therefore find that

$$\mu N = -TS + U + NRT$$

or

$$U = TS - PV + \mu N$$

as an expression for the internal energy in terms of entropy, volume and particle number plus the intensive variables temperature, pressure and particle number - which, of course, is exactly the expression predicted by our discussion of homogeneous functions in the appendix.

The ideal gas that we have discussed so far was assumed to only contain one type of substance, so you would call this a *monatomic ideal gas*. Let us now see what happens if we consider a mixture of several ideal gases. As the gases do not interact, they will all occupy the same volume V, and we also assume equilibrium, so that the gases all have the same temperature T. However, each of them will exert a certain pressure P_i on the walls of the vessel in which the gas is contained, called the *partial pressure*, and the total pressure is given by

$$P = \sum_i P_i$$

a fact which is sometimes called Dalton's law. As we assume that the gases do not interact, the ideal gas laws hold for each of them separately, so that

$$P_i V = N_i R T$$
$$U_i = c_i N_i R T$$

Note that the first relation implies that

$$P_i = N_i \frac{RT}{V}$$

and

$$P = N \frac{RT}{V}$$

so that

$$P_i = \frac{N_i}{N} P$$

Let us now again derive an equation for the entropy of the mixture. To do this, we start with the relation

$$TS = U + PV - \sum_i \mu_i N_i$$

We know that the total energy is the sum of the U_i and the total pressure is the sum of the P_i. Thus

$$TS = \sum_i U_i + \sum_i P_i V - \sum_i \mu_i N_i$$

Let us now assume that the chemical potentials of the components are the same as it would be in isolation (which seems reasonable as we think of the components as non-interacting, we will later justify this assumption with a bit more formal argument). Thus we assume that

$$\mu_i = -T s_{0,i} - TR \ln\left[\left(\frac{U_i}{U_0}\right)^c \frac{V}{V_0} N_i^{-c-1}\right] + (c_i + 1)RT$$

Now the ideal gas laws in the form written above imply that

$$\frac{U}{U_0} = \frac{N_i}{N_0} \frac{T}{T_0}$$

so that

$$\left(\frac{U_i}{U_0}\right)^c_i N_i^{-c_i-1} = N_0^{-c_i}\left(\frac{T}{T_0}\right)^{c_i} \frac{1}{N_i}$$

By redefining the constant $s_{0,i}$ we can therefore write the chemical potential as

$$\mu_i = -T s_{0,i} - TR \ln\left[\left(\frac{T}{T_0}\right)^c \frac{V}{N_i V_0}\right] + (c_i + 1)RT$$

Therefore

$$N_i \mu_i = -N_i T s_{0,i} - N_i T R \ln \left[\left(\frac{T}{T_0}\right)^c \frac{V}{N_i V_0}\right] + N_i(c_i + 1)RT$$

Now the last term is

$$N_i(c_i + 1)RT = N_i c_i RT + N_i RT = U_i + P_i V$$

so that

$$N_i \mu_i = -N_i T s_{0,i} - N_i T R \ln \left[\left(\frac{T}{T_0}\right)^c \frac{V}{N_i V_0}\right] + U_i + P_i V$$

Plugging this into our equation for TS above, the last two terms cancel with $\sum_i U_i$ and $\sum_i P_i V$, and we obtain

$$TS = \sum_i N_i T s_{0,i} + \sum_i N_i T R \ln \left[\left(\frac{T}{T_0}\right)^c \frac{V}{N_i V_0}\right]$$

so that the entropy is

$$S = \sum_i N_i s_{0,i} + \sum_i c_i N_i R \ln \frac{T}{T_0} + \sum_i N_i R \ln \frac{V}{N_i V_0}$$

in accordance with the expression in [4], section 3.4. Note that a gas consisting of N_i particles (or moles) of constituent i at temperature T and pressure P_i would occupy the volume

$$\frac{N_i RT}{P_i} = \frac{N_i RT}{P \frac{N_i}{N}} = \frac{RT}{P} = V$$

and its entropy would therefore be

$$N_i s_{0,i} + c_i N_i R \ln \frac{T}{T_0} + N_i R \ln \frac{V}{N_i V_0}$$

Therefore we find that *the entropy of a mixture of ideal gases* is the sum of the entropies that each constituent would have at particle number N_i, temperature T and pressure P_i (or equivalently volume V). Put differently, if we think of our system as a composite system where each subsystem has volume V, temperature T and consists of all atoms of one type, we could appeal to the additivity of entropy to calculate the entropy of the overall system and would arrive at exactly the same expression. This alternative approach to deriving the entropy also shows that our earlier assumption that the chemical potentials are independent of each other is not only justified but in fact a direct consequence of the additivity of the entropy. Note, however, that for this to work, we need to make sure that the two gases are distinguishable, i.e. that for each molecule we can tell to which gas it belongs, as otherwise our subsystems would not be well defined (see also section 2.6 of [5] for a discussion of the subtleties related to this).

It is interesting to analyze that expression a bit further. For that purpose, let us imagine that we have a collection of separated gases, each with a particle number N_i and each at the same pressure P and temperature T. Let N again denote the total particle number. The volume of each gas is then given by

$$V_i = \frac{N_i RT}{P}$$

and the total volume is

$$V = \sum_i \frac{N_i RT}{P} = \frac{NRT}{P}$$

The total entropy of the system given by this initially separated collection of gases is then

$$S_{initial} = \sum_i N_i s_{0,i} + \sum_i c_i N_i R \ln \frac{T}{T_0} + \sum_i N_i R \ln \frac{V_i}{N_i V_0}$$

As

$$\frac{V_i}{N_i} = \frac{V}{N}$$

this is the same as

$$S_{initial} = \sum_i N_i s_{0,i} + \sum_i c_i N_i R \ln \frac{T}{T_0} + \sum_i N_i R \ln \frac{V}{N V_0}$$

When we now allow the gases to mix, no mechanical work is done as the systems already have the same pressure, and no heat is exchanged, as they already have the same temperature. Once equilibrium is reached, we can apply the formula that we have found above to conclude that the final entropy of the system is

$$S_{final} = \sum_i N_i s_{0,i} + \sum_i c_i N_i R \ln \frac{T}{T_0} + \sum_i N_i R \ln \frac{V}{N_i V_0}$$

The difference between these two terms is given by

$$\Delta S = S_{final} - S_{initial} = \sum_i N_i R \ln \frac{V}{N_i V_0} - \sum_i N_i R \ln \frac{V}{N V_0}$$

which is the same as

$$\Delta S = \sum_i N_i R (\ln \frac{N}{N_i})$$

This is inherently positive, as $N > N_i$, and represents a gain in entropy that is called the *entropy of mixing*. We will meet this term again when we calculate the chemical potential of a solute in a later section. This is another example for a process which is not quasi-static and for which our identification of the heat transfer (which is zero in this case) with $T\Delta S$ (which is not zero as we have just seen) breaks down.

6   The principal of minimal internal energy

We have seen that in the entropy representation, the principle of maximum entropy allows us to derive statements about the equilibrium states of a thermodynamical system. It expresses the equilibrium state as the maximum of some function on the configuration space.

   We also know that the energy representation is fully equivalent to the entropy representation. Thus it appears natural that there is an extremum principle in the energy representation which is equivalent to the principle of maximum entropy. It turns out that this is in fact the case, and that the equilibrium states are those states that minimize the total energy subject to the existing constraints.

   To derive this principle, let us again assume that we are given a system by a configuration space with coordinates $U, X_1, \ldots, X_n$ in the entropy representation. We do not make any specific assumptions on the X_i, thus they could be any parameters of a composite system, including the energies of subsystems, but we assume that they are unconstrained. The entropy is then a function

$$S = S(U, \{X_i\})$$

defined on the configuration space. We also know that the temperature is positive and therefore that everywhere on the configuration space

$$\frac{\partial S}{\partial U} = \frac{1}{T} > 0$$

Thus we can apply the implicit function theorem and find - at least locally - a function

$$U = U(S, \{X_i\})$$

such that

$$S = S(U(S, \{X_i\}), \{X_i\})$$

Note that the S on the left hand side of this equation is a coordinate of the configuration space in the energy representation, while the S on the right hand side is a function on the configuration space of the entropy representation. In a more neutral notation, the map

$$U \times \mathrm{id} \colon (s, \{x_i\}) \mapsto (U(s, \{x_i\}), x_i)$$

is a diffeomorphism which maps the configuration space in the energy representation to the configuration space in the entropy representation.

Now let γ be an arbitrary smooth curve in the energy representation, given - in coordinates - by

$$\gamma(t) = (s(t), \{x_i(t)\})$$

and let

$$\bar{\gamma}(t) = (u(t), \{x_i(t)\})$$

be the corresponding curve in the entropy representation. We write

$$\gamma(0) = (s_0, \{x_i(0)\})$$

and

$$\bar{\gamma}(0) = (u_0, \{x_i(0)\})$$

and assume that this is an equilibrium point. Moreoever, we assume that

$$s(t) = s_0$$

for all t. We want to show that $U \circ \bar{\gamma}$ has a minimum at $t = 0$. Differentiating the relation

$$s(t) = S(\bar{\gamma}(t))$$

gives

$$0 = \dot{s}(t) = \langle dS, (\dot{u}(t), \{\dot{x}_i(t)\}) \rangle \tag{1}$$

for all t. Applied at $t = 0$, this yields in turn

$$0 = \frac{\partial S}{\partial U}\dot{u}(0) + \sum_i \frac{\partial S}{\partial X_i}\dot{x}_i(0)$$

But by assumption, $\bar{\gamma}(0)$ is an equilibrium point and the X_i are unconstrained. Therefore, by the principle of maximum entropy, the partial derivatives

$$\frac{\partial S}{\partial X_i}$$

are all zero at the equilibrium point. Thus we find that

$$0 = \frac{\partial S}{\partial U}\dot{u}(0) = \frac{1}{T}\dot{u}(0)$$

from which we can conclude that

$$\dot{u}(0) = 0$$

Thus we have found an extremum. We still have to show that this is a minimum. For that purpose, we differentiate equation (1) at $t = 0$. Observing that $\dot{u}(0) = 0$ and the partial derivatives of S with respect to the X_i are zero at the equilibrium point, only a few terms survive and give

$$0 = \frac{\partial S}{\partial U}\ddot{u}(0) + \sum_i \dot{x}_i(0)\frac{d}{dt}|_{t=0}\frac{\partial S}{\partial X_i}(\bar{\gamma}(t))$$

Applying the chain rule once more, we have that

$$\frac{d}{dt}|_{t=0}\frac{\partial S}{\partial X_i} = \sum_j \dot{x}_j(0)\frac{\partial^2 S}{\partial X_i \partial X_j} + \dot{u}(0)\frac{\partial^2 S}{\partial U \partial X_j}$$

Now the second term is again zero, as the derivative of $u(t)$ is zero at $t = 0$. Putting all the remaining terms together, we therefore obtain

$$0 = \frac{1}{T}\ddot{u}(0) + \sum_i\sum_j \frac{\partial^2 S}{\partial X_i \partial X_j}\dot{x}_j(0)\dot{x}_i(0)$$

But the sum over i and j is nothing but

$$\langle \mathcal{H}_{X_i}\dot{x}, \dot{x}\rangle$$

i.e. the Hessian with respect to the X_i, applied to the derivatives $\dot{x}_i$. By the maximum entropy principle, this Hessian is, at the equilibrium, negative definite, so the sum is negative (unless $\dot{x} = 0$ which can only happen for a constant curve). We therefore immediately obtain

$$\ddot{u}(0) > 0$$

which qualifies the extremum as a minimum, and the proof is complete.

To summarize, we have demonstrated that in the energy representation, the principle of maximal entropy turns into the principle of minimal energy, which is phrased as follows in [4]: *the equilibrium of any unconstrained internal parameter is such as to minimize the energy for the given entropy.*

Example 6.1. Let us consider the system that we have studied in example 2.1 once more and try to derive a condition for the equilibrium using the minimum energy principle. The first step is to obtain the energy representation that expresses the total energy as a function of the total entropy and the remaining parameters. Thus we start with our relation between the total entropy and U_1 and solve this for U. Our formula for the entropy that we have derived in example 2.1 yields

$$\exp(\frac{S}{cN_2 R}) = U_1^{\frac{N_1}{N_2}}(U - U_1)$$

so that we obtain an expression for U in terms of entropy and our parameter U_1.

$$U(S, U_1) = U_1 + \exp(\frac{S}{cN_2 R})U_1^{-\frac{N_1}{N_2}}$$

The principle of minimum energy now tells us to minimize this for fixed S. So let us take the derivative with respect to U_1.

$$\frac{\partial U}{\partial U_1} = 1 - \frac{N_1}{N_2}\exp(\frac{S}{cN_2R})U_1^{-\frac{N_1}{N_2}-1}$$

This looks a bit intimidating, but simplifies greatly if we use our expression for the exponential that we have already used once.

$$\frac{\partial U}{\partial U_1} = 1 - \frac{N_1}{N_2}\exp(\frac{S}{cN_2R})U_1^{-\frac{N_1}{N_2}-1}$$

$$= 1 - \frac{N_1}{N_2}(U - U_1)U_1^{-\frac{N_1}{N_2}-1}$$

$$= 1 - \frac{N_1}{N_2}(U - U_1)U_1^{-1}$$

$$= 1 - \frac{N_1}{N_2}(\frac{U}{U_1} - 1)$$

This is zero if and only if

$$1 = \frac{N_1}{N_2}(\frac{U}{U_1} - 1)$$

i.e.

$$N_2 U_1 = N_1(U - U_1)$$

or

$$NU_1 = N_1 U$$

yielding the exact same condition

$$U_1 = U\frac{N_1}{N}$$

for the equilibrium as before. Thus we arrive at the same condition, regardless of whether we vary the entropy for fixed total energy or whether we vary the energy for fixed entropy.

It is instructive to study a graphical plot of a typical entropy and visualize how and why both principles - maximal entropy and minimal energy - work and yield the same result[1]. In figure 3, we have displayed the entropy of the composite system considered in example 2.1. The entropy S is plotted along the z-axis, and the grey surface displays the value of the entropy depending on the total energy U and the energy U_1 of the first subsystem. The blue solid lines represent values of the entropy for four different, but fixed values of the energy. The black dotted line shows those points in the state space that are equilibrium points. Thus we see that given a fixed value of U, the point of the blue curve that corresponds to the equilibrium is actually the point with the largest value of S along the curve, which corresponds to the principle of maximum entropy.

Similary, the dotted blue line represents a curve with a fixed value of S, i.e. the intersection of the surface given by the entropy with a plane perpendicular to the S-axis. We see that, among the points on this curve, the equilibrium point is the one with the lowest value of the total energy U, illustrating the principle of minimum entropy. Thus it is the special shape of the entropy surface that forces these two extremum principles to determine the same point in the configuration space.

Finally, we note that of course, as the energy representation and the entropy representation are completely equivalent, both principles hold in both representations, the only difference being that the principle is more easily exploited if the constraint fixes the value of a coordinate and not the value of some function.

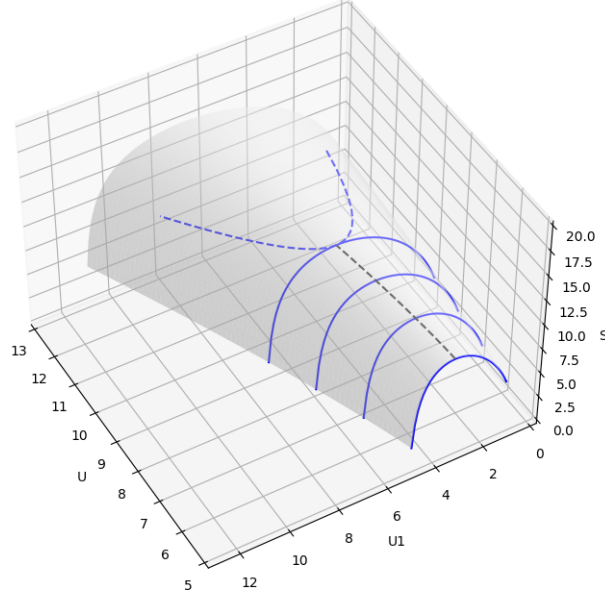[1]This is in fact a Python generated version of an illustration in [4]

Figure 3: Plot of an entropy function

7   The maximal work theorem

To further illustrate the formalism developed so far and to introduce some terminology that we will need in the sequel let us now consider special types of thermodynamical subsystems.

A *reversible work source* is a subsystem which can change its energy only via changes in its geometry. In other words, it is a system surrounded by impermeable (so that the particle number is constant) and adiabatic (so that no heat transfer is possible) movable walls. Thus the change of energy for this system is

$$dU = -PdV = dW$$

Strictly speaking we would have to phrase this as follows: for every path $\gamma(s)$ in the configuration space that is compatible with the constraints

$$dU(\gamma(s)) = -P(\gamma(s))dV(\dot{\gamma}(s))$$

for all s. We will, however, continue to use the simpler notation above if the meaning is obvious. Note that this implies that for any process, the entropy is of this subsystem is actually constant:

$$dS = \frac{P}{T}dV + \frac{1}{T}dU = \frac{1}{T}(PdV + dU) = 0$$

Similarly, we can define a *reversible heat source* to be a subsystem surrounded by rigid walls, which do however allow heat to pass. Thus, the only energy change that can occur in this subsystem is due to heat transfer, i.e.

$$dU = TdS = dQ$$

Thus the change of entropy of such a system is given by

$$dS = \frac{1}{T}dU = \frac{1}{T}dQ$$

It is important to note that T in this expression is itself a function of the extensive variables and not necessarily a constant. The *heat capacity* of the heat source is defined to be

$$C(T) = \frac{\partial U}{\partial T}$$

so that

$$dQ = C(T)dT$$

A *thermal reservoir* or *heat bath* is a reversible heat source with an infinite heat capacity, so that, no matter how large the transfer of heat actually is, the temperature of the reservoir remains constant. Similary a *pressure reservoir* is a reversible work force with constant pressure.

Let us now consider a composite system which is made up of three subsystems, namely a reversible work source, a reversible heat source and a third subsystem, which we call the *primary subsystem*, and which is the one which is actually of interest. Let us denote its state by (U, V, N). We assume that the primary subsystem is connected to the reversible work source and the reversible heat source in such a way that it can exchange heat with the heat source and work with the work source.

This is a simple model of an engine that extracts heat from the heat source (which we do not assume to have infinite heat capacity) and transfers a part of the energy received in this way to the work source, for instance by driving some cyclinder. We want to know under which conditions the work delivered to the work source (which we should actually call a work sink in this case) is maximal.

To do this, let us consider the change of energy and entropy that occur during a process. We have seen that the source of work does not undergo any change in entropy, so the total change of entropy is

$$dS_{tot} = dS + dS_{RHS} = dS + \frac{1}{T_{RHS}}dQ_{RHS} \geq 0$$

where S is the entropy of the primary system and S_{RHS} is the entropy of the heat source. In particular, we obtain that

$$dQ_{RHS} \geq -T_{RHS}dS$$

Now we can use the conservation of energy and the fact that the change in energy for the heat source and work source is entirely due to heat and work by definition to obtain

$$0 = dU_{tot} = dU + dQ_{RHS} + dW_{RWS}$$

Thus we find that

$$dW_{RWS} = -dU - dQ_{RHS} \leq -dU + T_{RHS}dS$$

with equality if and only if the total entropy does not change at all. A process for which the total entropy is constant is called a *reversible process*. Thus we find that the work transferred to the reversible work sink is maximized if the process is reversible. This fact is known as the *maximal work theorem*.

Now let us specialize a bit further and assume that the heat source is actually a thermal reservoir, i.e. its temperature is constant. As we assume equilibrium, and as any difference in the temperatures would cause heat to flow further, this implies that in a quasistationary process, the temperature T of the primary subsystem is

constant as well and equal to T_{RHS}. Then our equations simplify further. In fact, we have

$$-dQ_{RHS} \leq TdS$$

and therefore

$$dW_{RWS} \leq TdS - dU$$

Now, as the temperature does not change, this is - for this particular type of process - the same as

$$dW_{RWS} \leq d(TS - U)$$

Integrating this inequality along a path in the configuration that describes a process subject to the defined constraints, we can conclude that the work that can be delivered by the primary subsystem by drawing heat from a heat reservoir with infinite capacity is bounded by the difference of the function

$$F = U - TS$$

between initial and final state. This quantity - which can be defined for any system - is called the *Helmholtz free energy* or *Helmholtz potential*. We will now see how this potential can be obtained using a Legendre transform.

8   Legendre transform and free energy

Let us quickly recall some basic facts about the Legendre transform of a convex function. Assume that we are given a (strictly) convex and (for the sake of simplicity) smooth function

$$f : I \to \mathbb{R}$$

defined on some interval $I \subset \mathbb{R}$. The Legendre transform of f is the function f^* given by

$$f^*(p) = \sup_x \{px - f(x)\}$$

which is defined for all p for which the supremum exists. As f is supposed to be differentiable, we can maximize this expression by setting the derivative to zero and find that we can alternatively write

$$f^*(p) = px - f(x)$$

where x is chosen such that $f'(x) = p$. As we assume that the function is convex and therefore that the second derivative is not zero, the equation

$$p = f'(x) \tag{2}$$

can be solved for x if p is in the range of f', i.e. we can write $x = x(p)$ and express the Legendre transform a bit more precisely as

$$f^*(p) = px(p) - f(x(p))$$

as a function on the range of f'. Note that the function $x(p)$ does actually depend on the choice of f, so care must be taken when using this notation.

Let us now calculate the derivative of the Legendre transform. Applying the product rule and the chain rule, we find that

$$\frac{df^*}{dp} = \frac{d}{dp}(px(p)) - \frac{d}{dp}f(x(p))$$

$$= p\frac{dx}{dp} + x(p) - f'(x(p))\frac{dx}{dp} = x$$

where we have used equation 2 in the last line to see that the first and third term cancel. Thus we obtain the fundamental relations

$$p = \frac{df}{dx}$$

$$x = \frac{df^*}{dp}$$

Example 8.1. Suppose that we are given functions f and g and know that g is the inverse of f. Let us try to express the Legendre transform of g in terms of f^*. We now have two functions $x_1(p)$ and $x_2(p)$ given by the relations

$$f'(x_1(p)) = p$$
$$g'(x_2(p)) = p$$

Now, as f is the inverse of g, we have

$$g'(f(x_1)) = \frac{1}{f'(x_1)} = \frac{1}{p} = g'(x_2(\frac{1}{p}))$$

As g' is invertible, we obtain

$$x_2(\frac{1}{p}) = f(x_1(p))$$

We therefore find that

$$
\begin{aligned}
g^*(\frac{1}{p}) &= \frac{1}{p}x_2(\frac{1}{p}) - g(x_2(\frac{1}{p})) \\
&= \frac{1}{p}f(x_1(p)) - g(x_2(\frac{1}{p})) \\
&= \frac{1}{p}[px_1(p) - f^*(p)] - g(f(x_1(p))) \\
&= \frac{1}{p}[px_1(p) - f^*(p)] - x_1(p) = -\frac{1}{p}f^*(p)
\end{aligned}
$$

or

$$f^*(p) = -pg^*(\frac{1}{p})$$

Now assume that we are given a function $f = f(x, y)$ of two variables. We can apply a Legendre transform to one of the two variables, say x, while keeping y fixed. Let us call this Legendre transform f^*. We then have that

$$\frac{\partial f^*}{\partial p} = x$$

and

$$\frac{\partial f^*}{\partial y} = -\frac{\partial f}{\partial y}$$

Thus, if we write the total differential of f as

$$df = pdx + vdy$$

then

$$df^* = xdp - vdy$$

Finally, we remark that the Legendre transform can also be done for a concave function, with the only difference that then,

$$f^*(p) = \inf_x\{px - f(x)\}$$

The notation f^* that we have used so far is simple, but not quite adequate for functions of several variables as it does not capture the information to which of the variables we have applied the transform. To rectify this, we follow [4], and, given a function

$$f = f(x_1, \ldots, x_n)$$

denote the Legende transform with respect to the coordinate x_i as

$$f[x_i](x_1, \ldots, x_{i-1}, p, x_{i+1})$$

We will now see how this sort of transformation is applied in thermodynamics. Note that in many books, for instance in [4], the term Legendre transform is used for minus one times the Legendre transform as we have introduced it, a convention which we will not adopt.

We have seen that a thermodynamical system is, as long as only macroscopic states in equilibrium are concerned, fully described by the energy as a function of entropy, volume and particle number, i.e. by a relation of the type

$$U = U(S, V, N)$$

In some applications, we do have some information on some derived quantities like the temperature. It would then be helpful to be able to express the fundamental relation above in terms of the intensive parameters, i.e. the derivatives of the energy. This is exactly what the Legendre transform is doing.

Consider, for instance, a system for which we have some information on the temperature that we want to utilize. We can then apply the Legendre transform with respect to the entropy to obtain a function of temperature, volume and particle number that - as the Legendre transform can be reversed - contains the same amount of information as the original relation. Note, however, that, by convention, a minus sign is typically inserted at this point. We obtain the quantity

$$F(T, V, N) = -U[S] = U - TS \tag{3}$$

which is exactly the *Helmholtz free energy* considered before, but now interpreted as a function of T, V and N. Thus this is not a pure reorganisation of known quantities, but comes with a change of parameters. If we are given a tuple (T, V, N), we first need to find the entropy S such that

$$\frac{\partial U}{\partial S}|(S, V, N) = T$$

and then use this value to calculate U and F. By the general theory developed above, we can also immediately write down an expression for the total differential of F:

$$dF = -SdT - PdV + \mu dN$$

We have already seen above that the Helmholtz energy has an interesting physical interpretation: the maximal work that can be delivered by a system connected to a thermal reservoir is the decrease of the Helmholtz energy. In that sense, the Helmholtz energy is the potential of a system to deliver work at fixed temperature, which explains why this quantity is often called the *Helmholtz potential*.

The Helmholtz potential has another important property, which we will now derive. Suppose again that we are considering a system which is in contact with a thermal reservoir of temperature T, and that, as usual, the heat bath can only adjust its energy by exchanging heat with the system. Let us also assume that the overall system, i.e. the combination of the system under consideration and the reservoir, is sufficiently isolated so that its energy is constant. Let us now try

to relate the fact that the total energy U_{tot} is constant to the thermodynamical variables of our system. We have

$$
\begin{aligned}
0 = dU_{tot} &= dU + dU_R \\
&= dU + T_R dS_R \\
&= dU + T dS_R \\
&= dU - T dS + T dS_{tot} \\
&= d(U - TS) + T dS_{tot} = dF + dS_{tot}
\end{aligned}
$$

where we have used the fact that the reservoir can only exchange heat with our system in the second line, the fact that the temperatures of the system and the reservoir are the same in the third line, and the fact that the temperature is constant in the last line. Thus we find that

$$
dF = -T dS_{tot}
$$

As the composite system is isolated, we know that its entropy settles at a maximum when we remove internal constraints. Thus we obtain a minimum principle for the Helmholtz energy - *the equilibrium state of a system which is in diathermal contact with a thermal reservoir minimizes the Helmholtz potential if the overall system is isolated.*
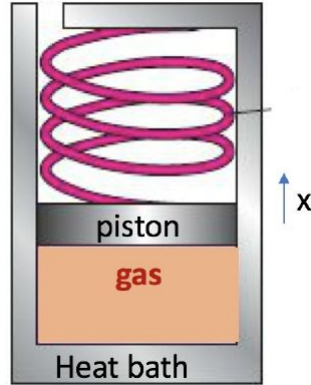

Figure 4: A piston pressing against a spring in a heat bath

To illustrate this, let us look at an example taken from [10]. Imagine a cylinder containing an ideal gas that is, via a piston, in contact with a spring and fully immersed in a heat bath, as shown in diagram 4.

If x denotes the displacement of the spring from its equilibrium position, its potential energy is given by

$$
E_s = \frac{1}{2} c x^2
$$

with a constant c. As x also determines the volume of the gas, the free energy of the overall system is given by

$$
F(x) = E_s(x) + U(x) - T S(x)
$$

For an ideal gas, the energy does only depend on the temperature, which is constant

due to the heat bath. So the derivative of the free energy with respect to x is

$$
\begin{aligned}
\frac{dF}{dx} &= \frac{d}{dx}E_s(x) - T\frac{dS}{dx} \\
&= cx - T\frac{dS}{dV}\frac{dV}{dx} \\
&= cx - AT\frac{dS}{dV} = cx - AP
\end{aligned}
$$

where A is the area of the piston. We therefore find that a necessary condition for equilibrium is that

$$cx = AP$$

But the left hand side is simply the force exerted by the spring and the right hand side is the force exerted by the gas, so that this condition simply says that these two forces are equal. Thus applying the minimum principle for the Helmholtz energy yields the result that we expect.

Obviously we can apply Legendre transforms to the other extensive variables as well and to combinations of them. Functions obtained in this way are called *thermodynamic potentials*. We will not got into further details in this short introduction, but only mention the *enthalpy*, which is the Legendre transform with respect to the volume, and the *Gibbs free energy*, which is the Legendre transform with respect to volume and entropy at the same time.

We start with the enthalpy, which is widely used in chemistry as many processes in a laboratory take place at constant (often atmospheric) pressure. By definition, the enthalpy is

$$H(p, V, N) = U + PV$$

To interpret this quantity, imagine a system that has some internal energy U and takes some volume V inside a reservoir of constant pressure P. Imagine that you were somehow able to create this system from scratch. You would then have to make room inside the environment, i.e. you would have to create an empty space of volume V, which requires the work PV being done on the environment, and you would then have to create the system with its internal energy U. Thus the enthalpy represents the energy that you would have to invest (at least) to create a system from scratch in an environment of constant pressure. Conversely, suppose that the system would vanish. The energy you would then obtain is the internal energy of the system plus the work that now the environment does to fill up the volume left behind (this is nicely illustrated in section 1.6 of [5]). We will look at some applications of the enthalpy in our appendix.

Similarly, the Gibbs free energy is a natural potential to work with when considering processes with constant temperature and pressure and is given by

$$G(T, P, N) = U - TS + PV$$

Both of these potentials have properties similar to the Helmholtz energy. Let us consider this in more detail for the Gibbs energy. To derive a minimum principle for the Gibbs energy, let us now consider a system that is in touch with two reservoir - a heat reservoir HR and a pressure reservoir PR. Again we assume that the only way the heat reservoir can change its energy is to exchange heat with the system, and similarly that the only way for the pressure reservoir to change its energy is to exchange mechanical work with the system. And again we assume that the overall system is fully isolated, in particular that its energy and volume are constant. Then

we can argue as in the case of the Helmholtz free energy:

$$
\begin{aligned}
0 = dU_{tot} &= dU + dU_{HR} + dU_{PR} \\
&= dU + TdS_{HR} - PdV_{PR} \\
&= dU - TdS + TdS_{tot} + PdV \\
&= d(U - TS + PV) + TdS_{tot}
\end{aligned}
$$

so that

$$
dG = -TdS_{tot}
$$

As for the Helmholtz energy, this implies that the Gibbs energy has a minimum in the equilibrium state. More general, for systems that are in touch with a heat and a pressure reservoir, the Gibbs energy is minimized by the equilibrium state

As an example, let us calculate the Gibbs energy of an ideal gas. In a previous section, we have already derived the following expression for the chemical potential.

$$
\mu = -Ts_0 - TR\ln\left[\left(\frac{U}{U_0}\right)^c \frac{V}{V_0} N^{-c-1}\right] + (c+1)RT
$$

Let us express the chemical potential in terms of temperature and pressure instead of energy and volume. Using the ideal gas laws, we easily obtain

$$
\left(\frac{U}{U_0}\right)^c \frac{V}{V_0} N^{-c-1} = N_0^{-c-1}\left(\frac{T}{T_0}\right)^{c+1}\frac{P_0}{P}
$$

where N_0 is the particle number of our reference state. When we conveniently choose this to be N_A, so that our reference state contains one mole of substance, we therefore find that

$$
\mu = -Ts_0 - TR\ln\left[N_A^{-c-1}\left(\frac{T}{T_0}\right)^{c+1}\frac{P_0}{P}\right] + (c+1)RT
$$

showing that *the chemical potential of an ideal gas in isolation does not depend on the particle number but only on pressure and temperature.* Based on the results in the appendix on homogenous functions, this is what we expect, as the chemical potential is homogenous of degree 0 in N and therefore constant (along rays, but as there is only one particle involved this implies that is does not depend on N). This only becomes apparent if we express the chemical potential really in terms of T, P and N alone and unravel all hidden dependencies on N, for instance in S. As the Gibbs energy is homogenous in N, the Gibbs energy of a system with N particles is simply μN. Spelling this out explicitly and collecting terms, we find that

$$
G(T, P, N) = \frac{N}{N_A}\frac{T}{T_0}G_0 - NTR\ln\left[\left(\frac{T}{T_0}\right)^{c+1}\frac{P_0}{P}\right]
$$

where G_0 is the Gibbs energy of one mole of our gas at standard conditions. Be careful, however - we will later see an example where the chemical potential does depend on the number of other particles in the system. Also note that this is of course only true if the other variables are pressure and temperature, not if for instance we use volume and temperature, as the example of an ideal gas shows as well.

9 Isothermal and adiabatic processes

Among the many types of processes that typically occur in thermodynamics, there are two types that occur rather frequently and that we will briefly discuss in this section - isothermal and adiabatic processes.

A process is called *isothermal* if the temperature remains constant throughout the entire process, often because the process takes place in contact with a thermal reservoir. When a process is isothermal, the state of the system is completely characterized by two remaining variables, for instance pressure and volume. Typically, these two are further related by a *state law*. Let us take an ideal gas as an example. We know that

$$PV = NRT$$

which is constant for an isothermal process, so that $PV = $ const (known as the law of Boyle-Mariotte). We also know that the energy only depends on particle number and temperature, so that the energy of the gas remains constant during an isothermal process. This, however, is only true for an ideal gas - there are many isothermal processes that change the internal energy. An example is the vaporization of a liquid that we will study a bit later in more detail, where the internal energy changes due to the changing volume even though the temperature remains at the boiling point.

Note that when we draw a P-V diagram of an isothermal process, then the area under the curve is the work being exchanged between the system and the environment. For an ideal gas, for instance, which is compressed from volume V_1 to a volume $V_2 < V_1$, the work that needs to be done on the system is

$$W = -\int_{V_1}^{V_2} PdV = -NRT \int_{V_1}^{V_2} \frac{dV}{V} = NRT \ln \frac{V_1}{V_2}$$

Again for the special case of an ideal gas, this must go entirely into a change of the entropy as the internal energy U remains constant, which we can confirm by comparing this term to the part of the expression for the entropy of an ideal gas derived earlier that depends on the volume.

Let us now consider a different type of process - and *adiabatic* process, meaning that no heat is exchanged with the environment. Most of the time, this term is applied only to a process which is also quasi-static. As this implies that $Q = TdS$, we can conclude that such a process leaves the entropy constant (be careful, however - whether this is true depends on the exact definition of quasi-static, see the section on quasi-static and irreversible processes for a more detail discussion).

Let us try to understand what happens to an ideal gas that undergoes an adiabatic (and quasi-static) process, i.e. a process for which the entropy remains constant. We know that the entropy of an ideal gas is given by

$$S = Ns_0 + NR \ln \left[\left(\frac{U}{U_0} \right)^c \frac{V}{V_0} N^{-c-1} \right]$$

Assuming constant particle number, we can derive that a process leaves the entropy unchanged if

$$U^c V = \text{const}$$

Now by the ideal gas law, we know that

$$U = cNRT = cPV$$

so that our condition becomes

$$P^c V^{c+1} = \text{const}$$

where we have absorbed the factor c^c into the constant. Introducing the constant

$$\gamma = 1 + \frac{1}{c} = \frac{f+2}{f}$$

which is called the *adiabatic index*, this turns into

$$PV^\gamma = \text{const}$$

For a monatomic ideal gas, there are $f = 3$ degrees of freedom, so $\gamma = \frac{5}{2}$. In reality, an adiabatic process can at least approximately be achieved by a fast compression that does not allow the gas to continously consume heat from the environment, so that the heat transfer is zero or at least very small.

Note that this relation implies that the curve of an adiabatic process in P-V space is steeper than an isotherm. In [5], Schroeder gives a nice argument why we should expect this - if we do an adiabatic compression, we create heat that cannot be transferred to the environment, so that the temperature needs to increase. Thus we move between two isotherms in the P-V plane, and therefore this curve must be steeper than the isotherms.

Example 9.1. Let us look at isothermal and adiabatic processes in a bit more detail for a specific example. Suppose we are given a vessel that is separated into two chambers separated by a piston. Both chambers are filled with an ideal gas and initially occupy the same volume V and have the same number of particles N at the same pressure P.

Let us now assume that we can manipulate and move the piston from outside the vessel, so that the volume of both chambers changes. Let us first consider one of the chambers, say the left one, and denote its volume at time t by $V(t)$, so that $V(0) = V$. We will use the same notation for the other properties of this chamber, see diagram 5.

We will now consider two different cases. In the first case, we connect the vessel to a thermal reservoir so that during the entire process, the temperature in both chambers remains at its initial value. As we move the piston, the gas in one chamber is compressed, but the gas in the other chamber is expanded. The process is isothermic, so we know that

$$P(t)V(t) = PV$$

As the energy in both chambers depends only on the temperature and is therefore constant, the total energy is constant, so we expect that the net work being done on the system is zero. On the other hand heat is exchanged, so we should see a change in the entropy.

The other case that we consider is the case that the chambers are thermally isolated, so that the expansion and compression are adiabatic. In this case, we know that

$$P(t)V(t)^\gamma = PV^\gamma$$

Initially, the pressures in both chambers are equal, so that the net force on the piston is zero. However, once we have moved the piston a bit out of equilibrium, the pressure in one chamber will increase and that in the other chamber will decrease, so that there will be a net force acting on the piston that we need to overcome. This is mechanical work that we need to put into the system. As this energy cannot escape in the form of heat, we expect that the total energy is increased but the entropy remains the same.

To confirm our expectations for both cases, let us assume a more general process of moving the piston where the volume and the pressure of one of the chambers are related by

$$P(t)V(t)^\eta = PV^\eta$$

In the isothermic case, $\eta = 1$, in the adiabatic case, $\eta = \gamma$. Let us denote the degree by which the gas in the first chamber is compressed by
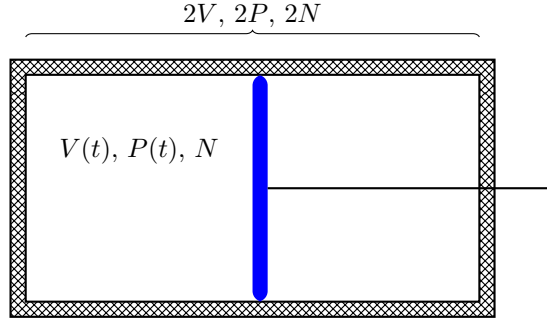
$$\lambda(t) = \frac{V(t)}{V}$$

Figure 5: A movable piston in a vessel

As the sum of the volumes of both chambers remains constant at $2V$, we see that the volume of the second chamber is given by

$$2V - \lambda(t)V = V(2 - \lambda(t))$$

Let us now establish formulas for both the entropy and the energy of the total system at time t. The energy of the first chamber is proportional to the temperature and thus to $P(t)V(t)$, so we find that

$$\begin{aligned} U(t) &= U\frac{P(t)V(t)}{PV} \\ &= U\frac{P(t)V(t)^\eta V(t)^{1-\eta}}{PV^\eta V^{1-\eta}} \\ &= U\frac{V(t)^{1-\eta}}{V^{1-\eta}} = U\lambda^{1-\eta} \end{aligned}$$

The same formula with λ replaced by $2 - \lambda$ holds for the second chamber, and we therefore find that the total energy is given by

$$U_{tot} = U[\lambda^{1-\eta} + (2-\lambda)^{1-\eta}]$$

Let us take the derivative with respect to λ to see how this changes over time. We have

$$\frac{dU_{tot}}{d\lambda} = (1-\eta)U[\lambda^{-\eta} - (2-\lambda)^{-\eta}]$$

For $\eta = 1$, this is zero. For $\eta > 1$ this is zero if and only if $\lambda = 1$, positive for $\lambda > 1$ and negative for $\lambda < 1$, so that we can conclude that U_{tot} has a minimum at $\lambda = 1$.

This confirms our expectations about the energy. In fact, in the isothermic case, $\eta = 1$ and the total energy is constant. In the adiabatic case, $\eta = \gamma > 1$ and we find that U_{tot} increases as we move the piston.

Let us now turn to the entropies. Using once more our formula for the entropy of an ideal gas, we find that the entropy of the first chamber is given by a constant plus

$$NR\ln U(t)^c + NR\ln V(t) = NR[\ln U\lambda^{c(1-\eta)} + \ln \lambda] = S_0 + NR\ln \lambda^{c+1-c\eta}$$

As the same expression holds for the second chamber, we find that the total entropy is given by

$$S_{tot} = S_{tot}(t=0) + NR(c+1-c\eta)\ln[\lambda(2-\lambda)]$$

Again we can differentiate this with respect to λ. We find that

$$\frac{d}{d\lambda}S_{tot} = (c + 1 - c\eta)\frac{2NR(1 - \lambda)}{\lambda(2 - \lambda)}$$

For $\eta = 1$, i.e. in the isothermic case, this shows that the entropy has a maximum at $\lambda = 1$, so that moving the piston increases the total entropy, as we have expected. In the adiabatic case,

$$\eta = \gamma = 1 + \frac{1}{c}$$

i.e.

$$c + 1 - c\eta = 0$$

and the total entropy is constant, as expected.

The upshot of the discussion is that to establish or change a constraint, we will in general have to apply an external force and therefore do work on the system. If the process by which we do this is adiabatic and quasi-static, then the entropy will remain constant, but the energy will be captured in the system. If we allow the energy to leave the system again to make sure that the total energy of the system remains constant, we will have to allow the flow of heat and this will come at the cost of an increase in entropy.

10   The Carnot cycle and heat engines

We have already briefly discussed heat engines and demonstrated that the second law of thermodynamics (in combination with the first law) implies that there is an upper bound on the efficiency of a heat engine. In fact, we have seen that for a heat engine in contact with a heat bath of temperature T_h and a second heat bath of temperature $T_l < T_h$, the efficiency with which the engine can turn heat drawn out of the hotter heat bath into work is at most

$$1 - \frac{T_l}{T_h}$$

We note that the pure fact that there is a bound is not too surprising - by the conservation of energy, the efficiency cannot exceed one. The more interesting part is that the bound is lower than one. In fact, the second law tells us that whenever we let heat flow between two bodies with different temperatures, there is a price we have to pay - an increase in entropy - and this limits the efficiency of a heat engine in addition to the limit given by the conservation of energy.

Let us now try to understand how a process that actually achieves this bound could look like. Thus we need to devise a cycle that our system undergoes and which takes it back to its original state while drawing heat out of the hot reservoir with temperature T_h, converting it into work and then letting excess heat flow into the cold reservoir with temperature T_l.

The first observation that we can make is that in order to avoid any unnecessary loss of entropy, our process needs to be quasi-static, and for any transfer of heat, the difference in temperature between the involved bodies should be as small as possible. So imagine we start with a state I in which our system has a temperature which is only a tiny bit smaller than T_h, i.e. $T \approx T_h$. Then a certain amount of heat $Q_h > 0$ will flow from the hot reservoir into the engine. During this time, the entropy of the engine changes by an amount

$$\Delta S_A = \frac{Q_h}{T_h}$$

Now we want to keep the temperature of our engine constant while the heat flows, which usually means that we have to let the engine do some work by expanding until it reaches a certain state II. Next, we let the engine undergo an adiabatic expansion. During this adiabatic expansion, the engine will do some additional work on the environment and its temperature will drop again. However, as no heat flows, the entropy will not change during that part of the cycle. We let the engine expand until its temperature has reached a value slightly above T_l. Denote this state by III.

In the third part of the cycle, we let the machine undergo another isothermal process during which it transfers some heat $Q_l < 0$ into the cold reservoir. Again, this will trigger a change in entropy given by

$$\Delta S_C = \frac{Q_l}{T_l}$$

Finally, we let the engine decrease its volume further in an adiabatic process during which it consumes some work from the environment. Thus the total change in entropy is

$$\Delta S_A + \Delta S_C$$

However, the entropy is a state variable and our process is a cycle, so that the total change in entropy is zero. Thus

$$\Delta S_A = -\Delta S_C$$

or

$$\frac{Q_h}{T_h} = -\frac{Q_l}{T_l}$$

so that

$$\frac{T_l}{T_h} = -\frac{Q_l}{Q_h}$$

Similarly, the total change in energy is zero. By the first law, this implies that

$$0 = Q_h + Q_l + W$$

where W is the total work done by the environment. Thus

$$
\begin{aligned}
W &= -Q_h - Q_l \\
&= -Q_h(1 + \frac{Q_l}{Q_h}) \\
&= -Q_h(1 - \frac{T_l}{T_h})
\end{aligned}
$$

We therefore find that

$$\frac{-W}{Q_h} = 1 - \frac{T_l}{T_h}$$

Now W is the work done on the system, so that $-W$ is the work that is done on the environment. The left hand side is therefore exactly the efficiency as we have defined it earlier, and we find that the efficiency of a Carnot engine is in fact equal to the theoretical limit.

To visualize the process, it is useful to display the process as a curve in our configuration space. The upper part of diagram (6) displays this process in the state space spanned by pressure P and volume V. We start at state I with an isothermal expansion at temperature T_h, then switch at some point to an adiabatic expansion, followed by an isothermal compression at temperature T_l and finally use an adiabatic process to return to our starting point.
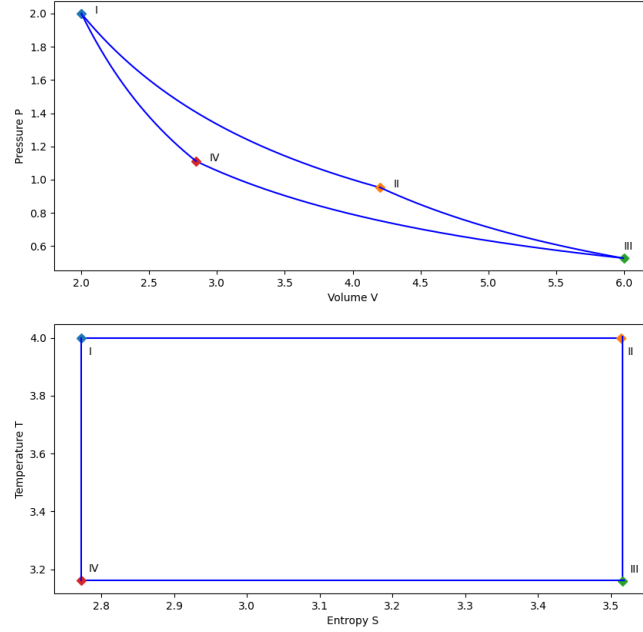
Figure 6: Carnot cycle

If we describe the process in terms of the conjugated variables S and T, we obtain the lower part of diagram 6. Here we see that during the first and third phase of the process, the temperature remains constant, while the entropy changes, as heat is transferred in an isothermal process. In phase two and four, no heat is transferred, and as we assume the process to be quasi-static, the entropy does not change. Again, the cycle returns to its original state as S and T are state variables.

Also note that the Carnot cycle demonstrates nicely that the amount of heat emitted or absorbed is actually path-dependent. In diagram 6, the system can move from state I to state III by either following the first part of the cycle or the (reversed) second part of the cycle. In the first case, the amount of heat absorbed is Q_h, in the second case, this is Q_l, and our calculation shows that these are not equal unless $T_h = T_l$.

11 The van der Waals model

In this section, we put the principles discussed so far into action, using the van der Waals model to describe an actual gas. In contrast to an ideal gas, that ignores interactions between the particles, the van der Waals model describes a substance with interaction. In this model, the fundamental equation is

$$F = F(T, N, V) = -Nk_BT[\ln(\frac{V - bN}{N}) + \frac{3}{2}\ln(k_BT)] - a\frac{N^2}{V}$$

where a and b are positive constants that account for interactions between the particles (a) and the volume of the particles (b). As F is minus the Legendre

transform of $U = U(S, V, N)$, we know that

$$S = -\frac{\partial F}{\partial T} = \frac{3}{2}Nk_B + Nk_B[\ln \frac{V - bN}{N} + \frac{3}{2}\ln k_B T]$$

Using the relation

$$U = TS + F$$

we can therefore find an expression for U as

$$U(T, N, V) = \frac{3}{2}Nk_B T - a\frac{N^2}{V}$$

We can also calculate the pressure and the chemical potential easily in this representation. For instance, the pressure is given by

$$P = -\frac{\partial F}{\partial V} = \frac{Nk_B T}{V - bN} - a\frac{N^2}{V^2}$$

which yields an expression for the Gibbs free energy

$$G = F + PV = -Nk_B T[\ln \frac{V - bN}{N} + \frac{3}{2}\ln k_B T] + \frac{Nk_B TV}{V - bN} - \frac{2aN^2}{V}$$

Let us look at some quantities that are typically used to describe the behavior of our system. Then first quantity is called the *isothermal compressibility*. It measures to what extent a gas or fluid can be compressed when some pressure is exercised, and is defined as

$$\kappa_T = -\frac{1}{V}(\frac{\partial V}{\partial P})_{T,N}$$

Here we use the usual notation that the variables that are to be fixed are added as a subscript to the partial derivative. So we need to express V in terms of P, T and N and then take the partial derivative with respect to P. Clearly, the inverse of this quantity is

$$\frac{1}{\kappa_t} = -V(\frac{\partial P}{\partial V})_{T,N}$$

That can easily be computed. We have

$$(\frac{\partial P}{\partial V})_{T,N} = -\frac{Nk_B T}{(V - bN)^2} + \frac{2aN^2}{V^3}$$

For $a = 0$, corresponding to the case of an ideal gas, this is clearly negative. This corresponds to a positive value of the compressibility and is what we would expect - when we increase the pressure, the volume will decrease. However, for positive a, this derivative can have zeros. This can be a simple zero, or a zero of higher order. The point (T_c, V_c) where a multiple zero occurs, i.e. for which

$$(\frac{\partial P}{\partial V})_{T,N} = 0$$

and

$$(\frac{\partial^2 P}{\partial V^2})_{T,N} = 0$$

is called the *critical point*. This corresponds to the equations

$$\frac{Nk_b T}{(V - bN)^2} = \frac{2aN^2}{V^3}$$

$$\frac{Nk_b T}{(V - bN)^3} = \frac{3aN^2}{V^4}$$

Dividing the first by the second condition yields

$$V - bN = \frac{2}{3}V$$

so that

$$V_c = 3bN$$

Plugging this into the first equation gives

$$k_B T_c = \frac{8a}{27b}$$

The pressure at the critical point is then easily calculated to be

$$P_c = \frac{a}{27b^2}$$

It is useful and simplifies a few calculations to express the temperature and pressure in a dimensionless way relative to the critical point. Thus we introduce the notation

$$\bar{T} = \frac{T}{T_c}$$
$$\bar{V} = \frac{V}{V_c}$$
$$\bar{P} = \frac{P}{P_c}$$

Then, for instance, the expression for the normalized pressure becomes

$$\bar{P} = \frac{8\bar{T}}{3\bar{V} - 1} - \frac{3}{\bar{V}^2}$$

(see [4], example 1 in section 9.4 for a reassuring cross-check). We also see that to solve this for the volume given temperature and pressure, we have to find the roots of a third-order polynomial in the volume. Thus there will in general be three roots. We will soon see that one of these roots is in the instable region, while the other two correspond to the volumes of the gas and liquid phase.

In figure 7, we have displayed some of the relations between pressure, temperature and volume.

The diagram on the left displays the dependency of P on V for four different values of T between 2.9 (lowest line) and 3.3 (highest line). The second line from the top has $T = T_c$. We clearly see that the nature of the lines changes as expected and we introduce zeros of the derivative, i.e. extremal points.

An example for such a curve (for $T = 2.9 = 0.93T_c$) is displayed in detail in the picture at the right, this time with the axis reverted. We see that, as we expect, the relation between V and P can no longer be inverted, i.e. for one value of P, there is more than one matching value of V. The zeros of the derivative are marked with red diamonds. The same markers have been added in the diagram at the bottom, which displays the partial derivative

$$\frac{\partial P}{\partial V}$$

for $T = 2.9$ and $N = 1$. In the area between these points (that correspond to an infinitely large value for the isothermal compressibility), the derivative is positive. This means that exercising pressure would in fact cause the material to expand, which is clearly unphysical and would lead to an unstable behavior (if we place this material in a small container, an expansion would in turn increase the pressure
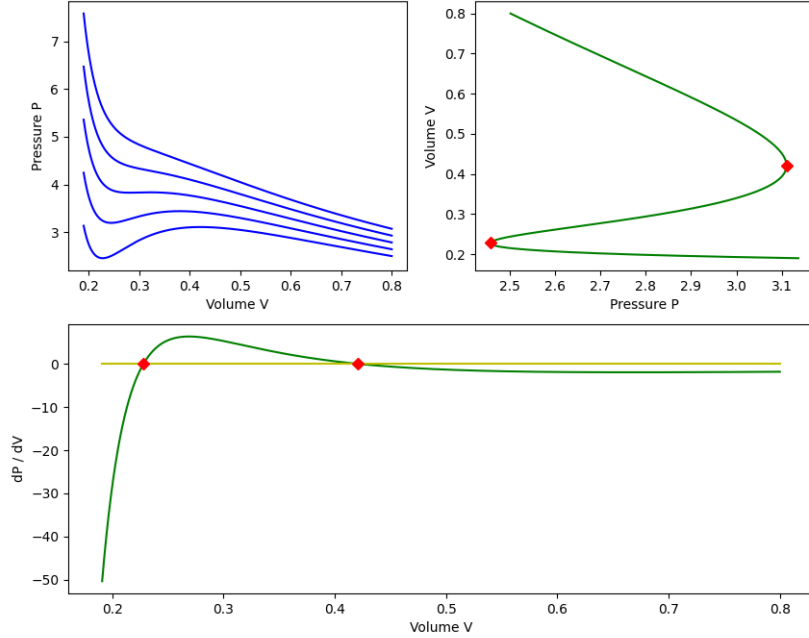
Figure 7: P-V diagrams for a Van der Waals fluid (a=1.05, N=1, b=0.1, k=1)

exercised by the air in the container and the process would, once initiated, continue).
Thus there is a *region of the phase space that does not represent stable systems*.
Intuitively, when the state moves along the isothermal, it will proceed directly from
a point on the right of the instable region to a point on the left of the region, with
a discontinuous behavior of the volume. This is an indication of a *phase transition*.

This might sound a bit strange, but recall that all the quantities that we are
dealing with are only defined in equilibrium, and we have essentially argued that
certain regions of the phase space do not allow for an equilibrium. Thus when a
system reaches this region, it will no longer be able to maintain an equilibrium. It
will continue to evolve its microstate, and at some point, when the phase transition is
complete, it will reappear in the configuration space, in alignment with our previous
discussion of quasi-static processes. Thus it is not the system that somehow evolves
in an unpredictable way during a phase transition, it is our description in terms of
equilibria that breaks down at this point.

Let us take a closer look at the role that the Gibbs energy plays in a phase
transition. For that purpose, let us assume that a portion of the gas described by
the van der Waals equations is contained in a vessel and let us consider a system
that consists of a comparatively small, but fixed number of particles within the
vessel, but can change its volume and its internal energy - this could just be a
small bubble within the system that can absorb or emit heat and shrink or grow.
Thinking of the surrounding part as a heat and pressure reservoir we want to apply
the fact that under these conditions, the Gibbs energy is minimized.

So let us now get back to the van der Waals equation and express the Gibbs
energy in terms of the normalized variables used earlier. We are only interested in
the terms depending on the volume $\bar{V}$, and we will use the normalized molar Gibbs

energy, i.e.

$$g = \frac{G}{Nk_B T_c}$$

The first term that depends on V can be written as follows

$$-\frac{1}{Nk_B T_c} Nk_B T \ln \frac{V - bN}{N} = -\frac{T}{T_c} \ln \frac{V - bN}{N}$$

$$= -\bar{T} \ln \frac{3\bar{V}bN - bN}{N}$$

$$= -\bar{T} \ln b(3\bar{V} - 1) = -\bar{T} \ln(3\bar{V} - 1) - \bar{T} \ln b$$

where we have used the relation

$$V = V_c \bar{V} = 3\bar{V}bN$$

We will only keep the first logarithm and move the second term into the part that does not depend on V. Similarly, the second term in the Gibbs energy turns into

$$\frac{1}{Nk_B T_c} \frac{Nk_B TV}{V - bN} = \bar{T} \frac{V}{V - bN}$$

$$= \bar{T} \frac{3\bar{V}bN}{3\bar{V}bN - bN}$$

$$= \bar{T} \frac{3\bar{V}}{3\bar{V} - 1}$$

$$= \bar{T} \frac{3\bar{V} - 1}{3\bar{V} - 1} + \bar{T} \frac{1}{3\bar{V} - 1} = \bar{T} + \bar{T} \frac{1}{3\bar{V} - 1}$$

Again, we move the first term into the part of the Gibbs energy that only depends on the temperature. Finally, the third term of the Gibbs energy that depends on V turns into

$$-\frac{1}{Nk_B T_c} \frac{2aN^2}{V} = -\frac{1}{k_B T_c} \frac{2aN}{3bN\bar{V}}$$

$$= -\frac{27b}{8a} \frac{2a}{3b\bar{V}} = -\frac{9}{4\bar{V}}$$

where we have used that $k_B T_c = \frac{8a}{27b}$. Collecting all terms depending on V, we thus arrive at

$$g = -\bar{T} \ln(3\bar{V} - 1) + \frac{\bar{T}}{3\bar{V} - 1} - \frac{9}{4\bar{V}} + f(\bar{T})$$

with a function depending on $\bar{T}$ only (usually one would write down the Gibbs energy in terms of temperature and pressure, but in this case, we cannot do this as the volume is not a function of pressure as it is the case in the absence of phase transitions). The part depending on V and the corresponding P-V curve for a fixed temperature $T = 0.85 * T_c$ are plotted in figure 8.

When a state moves along the isotherm starting below H, it will, at some point, experience a phase transition. It will then disappear from the configuration space and appear again at the left hand side of the diagram. It could do so at point H and reappear at point B (which, as we can see in the lower diagram, are two points with the same pressure but different values) or it could transition into the second phase upon reaching point G and reappear at point A (another pair of points with the same pressure and different volumes). How can we determine at which point the transition really happens?

To understand this, consider a little subsystem as before, for which we now know that the Gibbs energy has a minimum, but this time suppose that this system
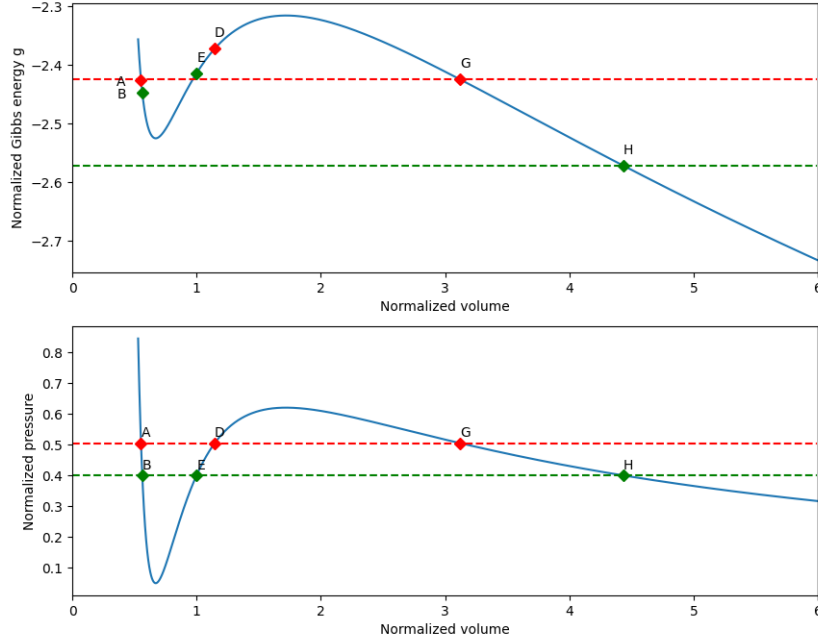
Figure 8: Gibbs energy along a van der Waals isotherm

is by itself a composition of two subsystems, which are both in contact with the surrounding reservoirs. The first subsystem, with N_1 particles and volume V_1, is still in the phase I on the left hand side of the unstable region, whereas the second subsystem, with N_2 particles and occupying volume V_2, is already in phase II. For each of the two phases, we have a well pressure and can do a Legendre transform to arrive at a Gibbs energy that depends only on temperature, pressure and particle number and is extensive in the particle number. As pressure and temperature are the same for both subsystems, we know that

$$G = G_I(T, p, N_1) + G_{II}(T, P, N_2) = g_I(T, P)N_1 + g_{II}(T, V_2)(N - N_1)$$

where N is the number of particles in our subsystems and g is the respective molar Gibbs energy. Now this is a minimum if and only if

$$0 = \frac{G}{\partial N_1} = g_I(T, P) - g_{II}(T, P)$$

i.e. if the Gibbs energies at both sides of the instable region are identical. If the Gibbs energies are different, say if $g_I(T, P) > g_{II}(T, P)$, then the system is not stable as transferring additional substance from phase I to phase II will increase the Gibbs energy. Thus we can conclude that we can have a stable state if and only if the two molar Gibbs energies are equal.

We remark that this and similar arguments that we will apply always assume that the regions of liquid and gas are nicely separated, i.e. that the liquid part of the substance and the vaporized part each occupy their own well defined volume, but are in contact at the boundary of that volume. Thus there is no real mixing and we do not need to consider the entropy of mixing, even though we will use the term mixture for such a state in which gas and liquid coexist.

Assuming stability, let us now take a look at the plot of the Gibbs energy. We see that for the pair of states H and B, the Gibbs energy on the left hand side of the unstable region is much higher than the energy on the right hand side. As the system is striving for a minimum of the Gibbs energy, no transition from H to B will take place. If, however, we increase the pressure further and the state moves further along the isotherm and eventually reaches G, the Gibbs energies on both sides (points G and A) will be equal. So at this point the transition will take place. We note that this will make the Gibbs energy continuous as a function of the pressure p.

Put differently, the points G and A mark two states in two different phases with the same pressure and temperatures, but different volumes and hence densities, that have the same Gibbs free energy and are therefore able to coexist in a stable state. This is called a *coexistence state.* If we think of the two phases as gas and liquid phase, this is just the boiling point.

The equilbrium condition is an additional relation between pressure and temperature that determines the pressure as a function of temperature. Thus, there is one pressure for which coexistence is possible for each given value of the temperature less than T_c. When plotted into a P-T diagram, as in figure 9 (where we have again plotted the normalized pressure and temperature), we find a curve that ends at the critical point. A diagram like this is called a *phase diagram.*
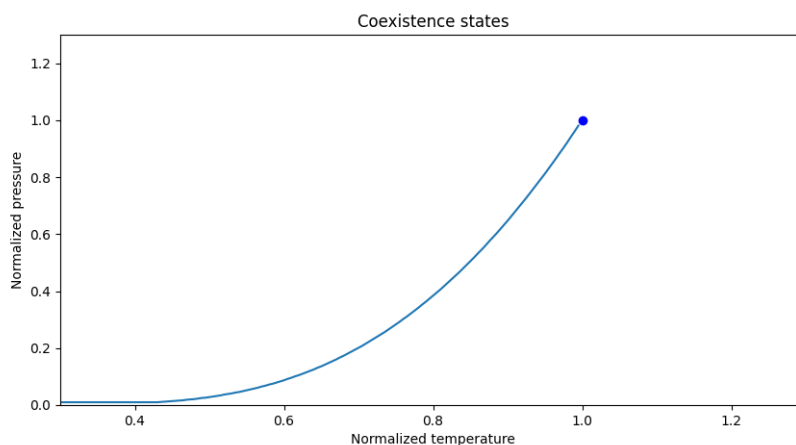

Figure 9: Phase diagram for a van der Waals system

Thus the pressure at which a phase transition occurs is a function of the temperature alone which has an interesting consequence that is sometimes know as the *Gibbs phase rule.* Away from the phase boundary, we can choose pressure and temperature independently - we can, for instance, prepare a given mass of water at atmospheric pressure at any temperature below the boiling point. Thus P and T are independent intensive variables that - at fixed particle number - determine the state of the system completely.

At the phase boundary, this changes. For a given temperature, the Gibbs equilibrium condition is an additional constraint that determines the pressure as a function of the temperature. Thus pressure and temperature are no longer independent, and the number of intensive variables required to describe the system is reduced by one. On the other hand, there are now many possible states for a fixed temperature (and thus pressure), namely mixtures with a different composition. Thus we need one additional parameter, for instance the fraction of substance in the liquid phase, to completely determine the state of the system. In a certain way, we have exchanged

the pressure for a new thermodynamic variable.

12 More on phase transitions

Some textbooks take a different approach to phase transitions that is based on the properties of the Helmholtz free energy. As a few important details are typically left out when this is presented, it might be worth to take a closer look at this argument. We will again use the van der Waals gas as an example, but the structure of the argument does not depend on any specifics of this system.

To simplify some of the calculations, let us íntroduce a normalized Helmholtz energy

$$f = \frac{F}{k_B N T_c}$$

as we have done it for the Gibbs free energy and find an expression for it in terms of the reduced quantities. Reusing some of the calculations that we have done for the normalized Gibbs energy, we find that

$$f = -\bar{T}\ln(3\bar{V} - 1) - \frac{9}{8\bar{V}} + f_0(T)$$

with a function f_0 depending on the temperature alone. To get an idea for the shape of this curve, let us take a look at diagram 10
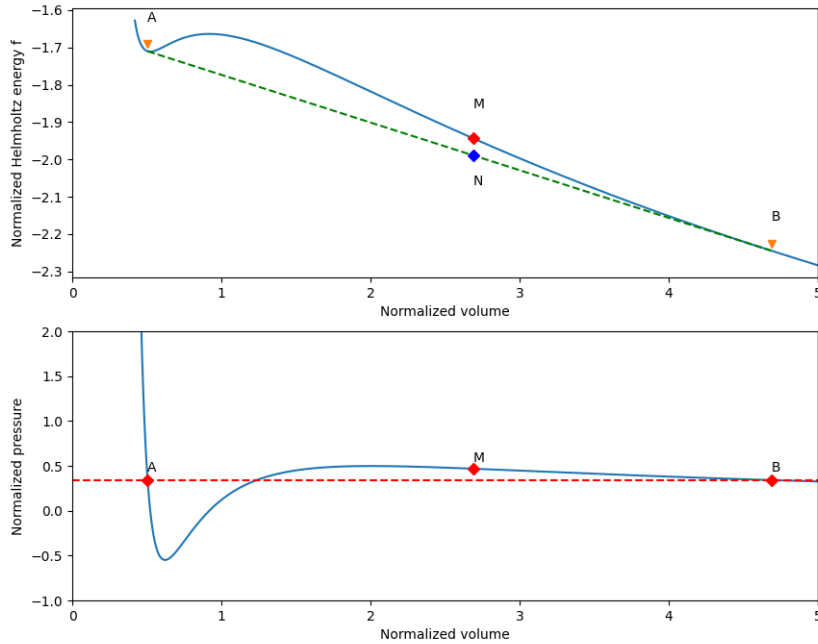

Figure 10: Helmholtz energy of a van der Waals gas

The upper part of this diagram displays the free Helmholtz energy as a function of the reduced volume while the lower part of the diagram shows the corresponding values of the reduced pressure which (up to a constant due to normalization) is the derivative of the Helmholtz energy. We already know that we will find two volumes

V_A and V_B for which the pressure is the same, which we can interpret as the volumes of the liquid phase and the gas phase exposed to the same pressure.

We also know that the pressure is - up to a normalization factor that is due to our usage of reduced quantities - minus the derivative of the Helmholtz energy with respect to the reduced volume. In addition, the derivative of the pressure with respect to the volume changes its sign and is positive in a small region. In this region, the second derivative of the Helmholtz energy is negative, meaning that the Helmholtz energy fails to be a convex function of the volume in this range.

Let us try to understand the physical implications of this lack of convexity. Again we consider two states A and B with different volumes, but the same pressure P, at a fixed temperature T. We also assume that (P, T) is on the coexistence curve, i.e. P is the boiling pressure at temperature T. Suppose we have a gas that has a volume slightly above V_B and we very slowly and cautiously continue to compress gas further, all this in the presence of a heat reservoir so that the temperature stays constant. When we hit V_B, in most cases, a phase transition will take place, i.e. the gas will liquify and we re-enter the configuration space at volume V_A. However, it is very well possible that we can move the system a bit beyond V_B without a phase transition taking place. This will produce what is called a *metastable state* that we denote by M.

Now let us try to produce a state N with the same volume by mixing the liquid and gas phases. Thus we consider a system that consists of N_A particles of liquid and N_B particles of gas, subject to the condition that $N_A + N_B = N$. Using the notation

$$n_A = \frac{N_A}{N}$$

and similarly for N_B, this condition turns into

$$n_A + n_B = 1$$

Within the mixture, the part that is still in the gas phase lives at the temperature (due to the heat bath) and pressure (because for a coexistence state, the Gibbs equilibrium condition determines the pressure for a given temperature) as that in state A, so that its specific density (the volume per particle) is the same. The volume of the gas part of the mixture is therefore given by

$$N_A \frac{V_A}{N} = n_a V_A$$

where we still use the symbol V_A to denote the volume of N particles in the gas phase. A similar relation holds for the liquid. Therefore the volume of the mixture is given by

$$n_A V_A + n_B V_B = n_A V_A + (1 - n_A) V_B$$

Clearly, we can choose n_A to realize any volume between V_A and V_B that we want in this way, including the volume V_M of the state M.

Let us now calculate the Helmholtz energy of the combined system. This is simply the sum of the Helmholtz energies of the individual systems, i.e. (recall our notation for the volumes)

$$F = F(T, \frac{N_A}{N} V_A, N_A) + F(T, \frac{N_B}{N} V_B, N_B)$$
$$= N_A F(T, \frac{V_A}{N}, 1) + N_B F(T, \frac{V_B}{N}, 1)$$

where we have used the fact that the Helmholtz energy is an extensive quantity. Dividing both sides by N and using the extensivity once more thus yields

$$\frac{F}{N} = n_A F(T, \frac{V_A}{N}, 1) + (1 - n_A)F(T, \frac{V_B}{N}, 1)$$
$$= n_A \frac{1}{N} F(T, V_A, N) + (1 - n_A)\frac{1}{N} F(T, V_B, N)$$

and dividing this by $k_B T_c$ therefore shows that the normalized Helmholtz energy of the mixture is given by

$$f = n_A f(\bar{T}, \bar{V}_A, N) + (1 - n_A)f(\bar{T}, \bar{V}_B, N)$$

This just the convex combination of the free Helmholtz energies of the liquid phase and gas phase. In the $f-V$ diagram (the upper part of diagram 10), this is therefore the point N located at the same volume as M on the dashed straight line connecting the two values of the Helmholtz energy at V_A and V_B.

This is now where the lack of convexity comes into play - this point is located *below* the point M, i.e. has a lower Helmholtz energy than M. Now recall that the Helmholtz energy is minimized in the presence of a heat bath and at constant volume. More precisely, we can think of the state M as living in a vessel enclosing the volume V_M that is kept at constant temperature by a heat bath but is mechanically isolated. We can then think of a process that slowly converts a part of the gas in the metastable state into the mixture of stable gas and liquid represented by N, all this inside the vessel and thus subject to the condition of fixed volume V_M. Eventually, the full volume, i.e. the vessel, will be in state N - our mixed state. As the entire process took place at constant temperature we can appeal to the principle of minimal energy for the Helmholtz energy to conclude that M is not stable.

Moreover, the new state N is stable, as we have assumed that our states A and B are on the coexistence curve, i.e. equalities of pressure and molar Gibbs energy hold. We can conclude that, if we perturb M a bit, it will move to state N and remain there.

It is interesting to relate the equilibrium conditions to the graphical representation of the Helmholtz energy in diagram 10. The first condition - equality of the pressures - simply means that the slope of the graph is the same at the two points A and B. To interpret the second condition, let us express the Helmholtz energies of the two states A and B in terms of the Gibbs energy at the (common) pressure P. We have

$$F(T, V_A, N) = G_A(T, P, N) - PV_A$$
$$F(T, V_B, N) = G_B(T, P, N) - PV_B$$

Thus if the two Gibbs energies are equal, then

$$F(T, V_A, N) - F(T, V_B, N) = -P(V_A - V_B)$$

As the slope is $-P$, this condition now simply states that the slopes are not only equal at the points A and B, but that the dashed connecting the two points A and B on which N is located has the exact same slope! Thus this line is tangent at both points, which is the reason why this construction is sometimes called the *double tangent construction*. As we know that the dashed line is the Helmholtz energy of the mixed state, this is also the condition that we require to make sure that the pressure remains continuous across the entire phase transition.

There is a much shorter version of this argument that works as follows. We know that the Gibbs energies of both phases are equal at the boiling pressure. Thus the Gibbs energy of a mixture does not change when we change its composition.

Therefore when we transition from state B through a mixed state with both phases to state B, the Gibbs energy remains constant. As

$$G = F + PV$$

the Helmholtz energie therefore moves along a line with slope P, which is the line that we have indicated. This is below the graph of F, so a mixture will the be energetically preferred state.

There is, however, a reason why I have decided to discuss the role of the Helmholtz energy in a bit more detail. When a student hits upon the above illustration in a textbook and learns that the Helmholtz free energy is not convex and therefore the system tends to move into a mixture of two phases, this is a bit against what that student might have learned about the Helmholtz energy. In most cases, a phase transition like vaporization does *not* take place at constant volume - in fact, a vapor usually occupies a much higher volume than the liquid - but the Helmholtz free energy is what you usually minimize for (combined) systems that do not do any work on the remaining environment, i.e. have constant volume. However, as the above discussion hopefully demonstrate, we do *not* use the Helmholtz energy to find the equilibrium volume of the system (which would be futile, as the curve shows that the Helmholtz energy does not have a minimum at any of the states that we consider). Instead, we only use it for one very specific argument - we compare a metastable state M at a given volume and a mixed state N at the *same volume*, and then conclude that, as one of them has a smaller Helmholtz energy than the other, the metastable state is not stable. This is a completely legitimate usage of the minimum principle for the Helmholtz energy.

Let us now turn to another quantity that we have not yet discussed so far - the entropy. As a starting point, let us again take the van der Waals gas as an example whose entropy is given by

$$S = \frac{3}{2}Nk_B + Nk_B[\ln \frac{V - bN}{N} + \frac{3}{2} \ln k_B T]$$

Now suppose we consider a van der Waals liquid at constant atmospheric pressure, maybe in an open vessel, and in contact with a heat bath, and we slowly increase the temperature of the heat bath and therefore the gas. This process corresponds to an isobar in the phase diagram. At some point, we will reach the coexistence curve. The liquid will vapor and thus increase its volume. However, the temperature of the system will remain the same, as we are still in contact with the heat bath. Thus the above formula shows us that the entropy will increase by an amount

$$\Delta S = Nk_B \ln \frac{V_g - bN}{V_l - bN} = Nk_B \ln \frac{3\bar{V}_g - 1}{3\bar{V}_l - 1}$$

The *latent heat L* is defined as the heat corresponding to this change in entropy, i.e.

$$L = T\Delta S$$

In the case of the van der Waals model, we find that

$$\frac{L}{Nk_B T_c} = \bar{T} \ln \frac{3\bar{V}_l - 1}{3\bar{V}_g - 1}$$

During this process, all potentials of the gas change, and it is instructive to relate them to see what is happening. First, the volume of the gas increases, so there is some work

$$P\Delta V = P(V_g - V_l)$$

being done on the environment. Using the van der Waals relation between pressure and volume, we can write

$$PV_g = \frac{Nk_BTV_g}{V_g - bN} - \frac{aN^2}{V}$$

and similarly for PV_l, so that

$$P\Delta V = Nk_BT(\frac{V_g}{V_g - bN} - \frac{V_l}{V_l - bN}) - aN^2(\frac{1}{V_g} - \frac{1}{V_l})$$

The latent heat is given by

$$L = T\Delta S = Nk_BT\ln(V_g - bN) - Nk_BT\ln(V_l - bN)$$

The total energy in the gas phase is given by

$$U_g = \frac{3}{2}Nk_BT - a\frac{N^2}{V_g}$$

so that

$$\Delta U = -aN^2(\frac{1}{V_g} - \frac{1}{V_l})$$

From the first law of thermodynamics, we expect that

$$\Delta U = T\Delta S - P\Delta V$$

i.e. we expect that

$$-aN^2(\frac{1}{V_g} - \frac{1}{V_l}) = Nk_BT\ln(V_g - bN) - Nk_BT\ln(V_l - bN)$$
$$- Nk_BT(\frac{1}{V_g - bN} - \frac{1}{V_l - bN}) + aN^2(\frac{1}{V_g} - \frac{1}{V_l})$$

which we can rearrange as

$$0 = Nk_BT\frac{1}{V_l - bN} - Nk_BT\ln(V_l - bN) - \frac{2aN^2}{V_l}$$
$$- \big[Nk_BT\frac{1}{V_g - bN} - Nk_BT\ln(V_g - bN) - \frac{2aN^2}{V_g}\big]$$

Comparing this with our expression for the Gibbs energy

$$G = -Nk_BT[\ln\frac{V - bN}{N} + \frac{3}{2}\ln k_BT] + \frac{Nk_BTV}{V - bN} - \frac{2aN^2}{V}$$

we find that this is just the change in the Gibbs energy. But we have already seen that the condition for the coexistence curve is exactly that this change is zero. Thus we find that the first law holds (as it should be) or, put differently, we once more see that the equality of the Gibbs energies is the correct condition for the phase transition to take place.

Thus during the phase transition, the system will draw some energy in the form of heat from the heat bath, this is the latent heat $T\Delta S$. A part of this energy is used to increase the internal energy and remains in the system. Another part is used to do work on the environment by increasing the volume against the constant atmospheric pressure. The Gibbs energy will remain constant (this is our coexistence condition). Due to the relation
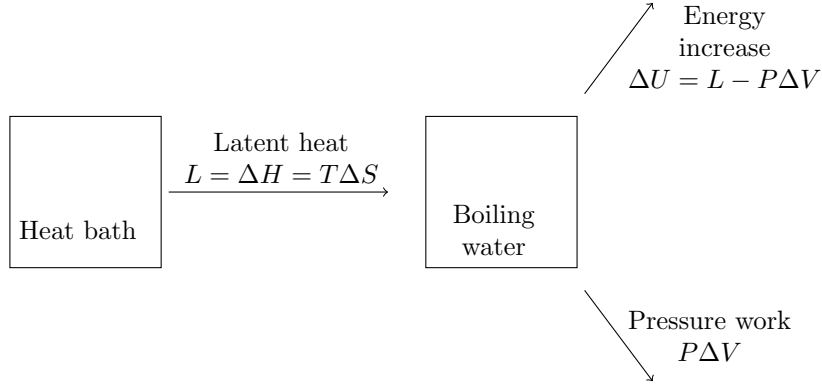
$$G = F + PV$$

Figure 11: Latent heat

we see that the change in the free Helmholtz energy is equal to the work being done on the environment. And finally, the change in the free enthalpy $H = G - TS$ is equal to the latent heat $T\Delta S$, again because $\Delta G = 0$. This is summarized in diagram 11.

Let us do a numerical example to get a feeling for the involved quantities. From [9], we can obtain the molar formation enthalpies of liquid and vaporized water which are (at atmospheric pressure and standard temperature - to be more precise we would of course have to use the values at the boiling points 373 K of water)

$$\Delta_f H_2O(l) = -285\text{kJ/mol}$$
$$\Delta_f H_2O(g) = -241\text{kJ/mol}$$

The difference is the latent heat which is therefore

$$L = 44 \cdot 10^3 \text{J/mol}$$

The true value is roughly 41kJ/mol, the difference being due to the values that we have used being valid at room temperature, not at the boiling point. To approximate the work due to the change of volume when water vaporizes, we ignore the volume of the water and approximate the liquid of one mol of vapor by

$$V = \frac{RT}{P}$$

so that

$$P\Delta V \approx RT = 8.3 \frac{\text{J}}{\text{mol} \cdot \text{K}} \cdot 373\text{K} = 3.1\text{kJ}$$

Comparing this to the molar latent heat, we find that only a small part of the heat drawn from the reservoir is used to perform volume expansion work. The larger part is absorbed by the increase of internal energy - this is the energy that is required to break the bonds between the water molecules to turn the water into vapor.

In the example of the van der Waals liquid, we were able to find a closed expression for the latent heat. In general, this is more difficult. However, there is a beautiful equation that allows us to determine the latent heat from the slope of the phase boundary. To find this equation, consider a process that moves a system along the phase boundary in the *P-T* plane (recall diagram 9). Let γ denote the path on the phase boundary corresponding to this process, and let us analyze the derivative of the Gibbs energies for each of the phases during this process. More precisely, we

look at $G = G(T, P, N)$ for both phases. For the gas phase, the derivative of this with respect to the curve parameter is, by the chain rule

$$dG_g \cdot \dot{\gamma}(0) = -S_g dT \cdot \dot{\gamma}(0) + V_g dP \cdot \dot{\gamma}(0)$$

and the same holds for the liquid phase. Now we know that while we move along the phase boundary, the equilibrium condition is maintained, i.e. the two Gibbs energies are equal. Thus the derivatives along the curve need to be equal as well, and we find that

$$-S_g dT \cdot \dot{\gamma}(0) + V_g dP \cdot \dot{\gamma}(0) = -S_l dT \cdot \dot{\gamma}(0) + V_l dP \cdot \dot{\gamma}(0)$$

or, collecting terms

$$(V_g - V_l)dP \cdot \dot{\gamma}(0) = (S_g - S_l)dT \cdot \dot{\gamma}(0)$$

Dividing by the change of the temperature on both sides and returning to a slightly more sloppy notation, we find that

$$\frac{dP}{dT} = \frac{S_g - S_l}{V_g - V_l}$$

or

$$\frac{dP}{dT} = \frac{L}{T \Delta V}$$

In other words, the latent heat $T \Delta S = T(S_g - S_l)$ is equal to the slope of the phase boundary times the change in volume $V_g - V_l$ times the temperature. This is the famous *Clausius-Clapeyron equation*.

Recalling that the change in the enthalpy H is equal to the latent heat, we could also write this as

$$\frac{dP}{dT} = \frac{\Delta H}{T \Delta V}$$

Finally, we can approximate the change in volume by the volume of the gas (which is usually much higher as that of the solid) and treat the gas as an ideal gas, so that

$$\frac{dP}{dT} = \frac{\Delta H}{TV} = \frac{P \Delta H}{T^2 N k_B} = \frac{P \Delta H^0}{T^2 R}$$

where now ΔH^0 is the molar latent heat. Assuming that L is constant and doing a separation of variables to solve the resulting partial differential equation shows that the vaporization pressure is approximatel

$$P \approx \text{const} \cdot e^{\frac{-L}{RT}}$$

Of course, this will only hold in a small region of the curve where the assumption that L is constant is approximately valid, so use this with care.

13 The microcanonical ensemble

Up to now, we have taken the existence of the entropy for granted respectively formulated it as a postulate, and focused our attention on macroscopic states and their development over time. We will now explore microstates and find that the entropy can in fact be defined as a statistical quantity.

To describe microstates, let us consider a system with a finite state space (we consider discrete state spaces for the sake of simplicity, but for most of what follows, the generalization to a continuous state space is obvious). Thus we have a set $\mathcal{S}$ of

states and assume that each of the states describes the system completely from a microscopical point of view. We also assume that each microstate has a well defined total energy E, i.e. that we have an energy function

$$E \colon \mathcal{S} \to \mathbb{R}$$

Note that other macroscopic quantities like the temperature and the pressure are only defined in equilibrium, so we should not expect them to be given by a function on the state space - we will get back to this point later.

The state space can be very general, and it is useful to introduce some examples that illustrate the concepts that we will discuss.

Example 13.1 (The two state system). Let us assume that we have a substance which is composed of particles that only have two possible discrete states. One of these states has energy 0, the other state has energy ϵ. If we consider a system of N such particles, then we can describe the microstate of the system by specifying, for each particle, whether it is in the ground state or in the excited state. Thus we can represent each state as a vector in $\{0,1\}^N$ and therefore the state space has 2^N elements. The energy function is given by

$$E(x_1, x_2, \ldots, x_n) = \epsilon \sum_i x_i$$

An example could be a paramagnetic substance in a magnetic field which we model as being composed of microscopic dipoles and the state reflects the orientation of the dipole relative to an external magnetic field (parallel or antiparallel).

Example 13.2 (The Einstein solid). Let us now consider a very simple model for a solid, developed by A. Einstein. We consider a collection of N very small microscopic systems, for instance particles in a crystalline solid that we envision as being bound to a specific equilibrium location by a harmonic force. That is, each particle is an independent system described by a harmonic oscillator (so the model ignores interactions between the particles) in three dimensions, or, equivalently, by three harmonic oscillators in one dimension. Thus the entire system consists of $3N$ independent harmonic oscillators. We assume that for each oscillator, the energy levels are given by $n\hbar\omega$ with the same ω for all oscillators. We therefore can describe each oscillator by its occupation number n, and consequently the state space is given by $\mathbb{N}^{3N}$, and the energy of a microstate is given by

$$E(n_1, n_2, \ldots,) = \hbar\omega \sum_i n_i$$

Let us now again consider some system, described by a state space $\mathcal{S}$ and macroscopic variables, say U, V, N and maybe some other variables $\{X_i\}$. We seek to relate the entropy of the system to the microstates. Obviously, different microstates will give us different values for the energy. In the example of the Einstein model, we expect that the total energy will be equal to the sum of the energies of all oscillators and therefore depend on the state. There will, however, be some subset of the state space that is compatible with the macrostate. The number of microstates that are compatible with a given macrostate is called the *multiplicity* of the microstate and is conventially denoted as

$$\Omega(U, V, N, \{X_i\})$$

even though this collides a bit with the usual usage of Ω in probability theory and statistics (this is the reason why we use the letter $\mathcal{S}$ for the state space).

If a system is composed of two subsystems, then there will be a number of compatible microstates for the first system and a number of compatible microstates

for the second system, and if we assume the systems to be initially isolated, the total multiplicity is

$$\Omega = \Omega_1 \Omega_2$$

Now let us suppose that we remove a barrier between the systems, and allow some particles to pass from one system to the other system or allow the particles in the two systems to interact. As we have removed a constraint, it is reasonable to assume that the number of microstates which are compatible with the remaining constraints will increase. Thus we now have

$$\Omega \geq \Omega_1 \Omega_2$$

Thus we have found a quantity which is multiplicative with respect to subsystems and tends to increase as constraints are removed. This is already very similar to the entropy, with the difference that the entropy is additive, not multiplicative. This can be fixed with a logarithm, and we arrive at the following definition which goes back to S. Boltzmann.

DEFINITION 13.1. *The* Boltzmann entropy *of a system with a given macrostate* $(U, V, N, \{X_i\})$ *is the logarithm of the number of microstates compatible with the values of the macroscopic variables, i.e.*

$$S(U, V, \ X_i) = k_B \ln \Omega(U, V, \{X_i\})$$

The constant k_B is called the Boltzmann constant.

There is a little subtlety concerning this definition that is worth mentioning. In many cases, we will want to apply this to system of many particles which have discrete energy levels. The actual energy of a system is then the sum of the individual energies and can therefore also take only discrete values. Thus $\Omega(U)$ would be zero for almost all values of U and the entropy would be minus infinity. To avoid this, Callen proposes in [4] section 15.5 that we technically have to consider all states with energies in a range $(U - \Delta U, U + \Delta U)$ and makes a nice argument why for a sufficiently high-dimensional state space, the choice of ΔU does not matter. However, this creates a certain conflict with the idea that the entropy is a function of the macroscopic variables and thus depends on the average energy of a system, not on the exact value of the energy in a certain state. This can be resolved by thinking of both, the entropy and the average energy, as properties of the probability distribution, not as that of an individual state.

Let us therefore try to relate Boltzmann's definition of entropy to the Shannon entropy that we know from information and probability theory. So far, we have not yet said anything about the probability to be in a specific microstate, i.e. we have not yet defined a probability distribution on the space of states $\mathcal{S}$. We now close this gap and make an assumption which is central to statistical mechanics:

All microstates that are compatible with the imposed constraints and values of the macroscopic variables have equal possibility.

Put differently, the probability distribution we are looking for is the uniform distribution on the part of the state space that represents states compatible with the constraints. Thus, for given values of U, V, N and the $\{X_i\}$, we assume the uniform distribution on the space

$$\mathcal{S}_{(U,V,\{X_i\})} = \{s | (E(s), V(s), N(s), X_i(s)) = (U, V, N, X_i)\}$$

What is the Shannon entropy of this probability distribution? By definition, the

Shannon entropy is given by

$$
\begin{aligned}
S_{Shannon} &= - \sum_{s \in \mathcal{S}(U,V,N,\{X_i\})} p(s) \ln p(s) \\
&= - \sum_{s \in \mathcal{S}(U,V,N,\{X_i\})} \frac{1}{\Omega(U,V,N,\{X_i\})} \ln \frac{1}{\Omega(U,V,N,\{X_i\})} \\
&= \sum_{s \in \mathcal{S}(U,V,N,\{X_i\})} \frac{1}{\Omega(U,V,N,\{X_i\})} \ln \Omega(U,V,N,\{X_i\}) \\
&= \ln \Omega(U,V,N,\{X_i\})
\end{aligned}
$$

and we find that the Shannon entropy is exactly the Boltzmann entropy, up to the constant k_B. As we know that the uniform distribution maximizes the Shannon entropy, we could as well have started with the postulate that the probability distribution of microstates maximizes the entropy for the given macroscopic state and then derived the uniform distribution from that principle, which is done in a few textbooks.

Example 13.3 (Two state system). Let us try to apply what we have learned to the case of a two-state system. Let N_+ denote the number of particles which are in the excited state. This is related to the total energy U of the system by

$$
N_+ \epsilon = U
$$

where again ϵ is the energy of the excited state. Defining a microstate which is compatible with a given energy U thus amounts to designating N_+ out of the N particles as excited, i.e. the number of possible microstates is equal to the number of possibilities to pick N_+ out of N particles, which is

$$
\Omega(U) = \binom{N}{N_+} = \frac{N!}{N_+!(N - N_+)!}
$$

The logarithm is

$$
\ln N! - \ln N_+! - \ln(N - N_+)!
$$

which, by applying Stirling's formula

$$
\ln N! \approx N \ln N - N
$$

and collecting terms, can be approximated by

$$
N \ln N - N_+ \ln N_+ - (N - N_+) \ln(N - N_+)
$$

which is the same as

$$
(N - N_+) \ln \frac{N}{N - N_+} - N_+ \ln \frac{N_+}{N}
$$

so that we eventually obtain

$$
S = k_B \left(\frac{U}{\epsilon} - N \right) \ln \left(1 - \frac{U}{\epsilon N} \right) - k_B \frac{U}{\epsilon} \ln \frac{U}{\epsilon N}
$$

Note that our model only allows for discrete values of the energy, namely multiples of ϵ, so that it does not make sense to formally define the temperature in terms of a partial derivative. However, we can of course consider a discrete version of the derivative by looking at the change in the entropy if the energy changes by one unit, i.e. if one particle changes its state so that N_+ goes up by one.
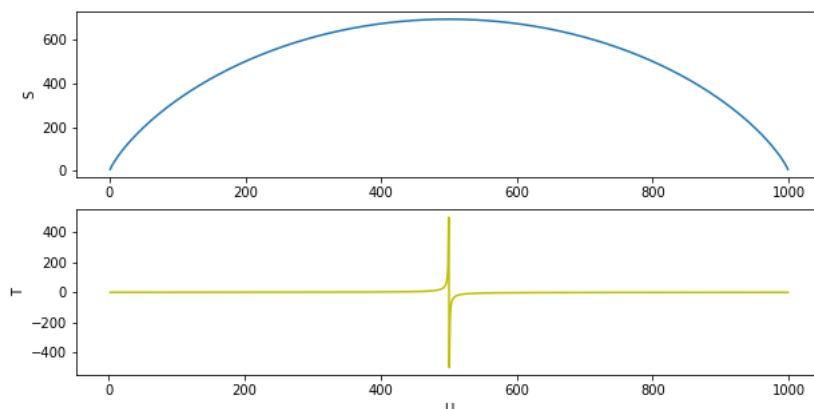
Figure 12: Entropy and temperature of a two-state system

In diagram 12, we have plotted the entropy and the temperature of a two-state system with $N = 1000$ particles and $\epsilon = 1$. We see that the entropy reaches a maximum when half of the particles are in an excited state. As the energy increases from zero to this state, the temperature goes up slowly, but then diverges when this state is reached and becomes negative if we move into areas where more than half of the particles are in an excited state. If the system is in this state and part of a composite system, the entropy of the composite system will actually go up if the two-state system transfers energy to the other system. Thus the two-state system will readily transfer heat to any other system, so that in a certain sense, the negative temperature should be considered as infinite, but positive temperature. States of this kind can actually be produced and observed for very short periods of time, we refer the reader to [5], section 3.3 for a discussion.

So far we have not been very clear on what exactly we mean by the probability to be in a microstate. To make this more precise, the usual assumption is that we are in fact looking at a very large number of identical systems, called the *ensemble*. The constellation that we have been studying so far, i.e. a collection of system that all have the same energy, is called the *microcanonical ensemble*. The probability for a specific microstate is then defined to be the fraction of systems in the ensemble that we find in that state if we make an observation at some point in time.

Of course we could ask for a different kind of probability: if we make a large number of measurements, at different points in time, of only one system, what is the fraction of times we would find that system in a given microstate? It it in fact one of the central assumptions that is usually taken that this gives the result, i.e. that time average equal average across the ensemble, which is usually expressed by saying that the system is *ergodic*.

It is instructive to use the theory of Markov chains to see why that is a reasonable assumption. So let us assume that we are given a finite state space (for the sake of simplicity) and that we observe the development of the microstate over time. Thus, we take measurements at discrete points in time, say at time $1, 2, \ldots$. Let us denote the state of the system at time n by X_n. Between two subsequent observations, the system can move from a state i into a different state j. Now let us make a couple of assumptions for the transition probabilities.

First, let us assume that the probability to move into a state j while being at i is independent of the time and the previous states. In the example of a two-state system above, a transition could mean that an excited particle spontaneously emits

a photon and falls back into its ground state, and the photon is absorbed by some other particle that moves into its excited state. It appears reasonable that the probability that this happens for a given target state depends on the current state, but not on previous states and not on the time. Thus the collection of the X_n forms a discrete time Markov chain.

Next, let us also assume that every state can be reached from any other state with positive probability in a finite time, and that there is a non-zero probability to stay in a given state. This corresponds to the assumptions that the Markov chain is irreducible and aperiodic. Both assumptions are not really required, as we know that we could split the state space of a reducible chain into smaller pieces and similarly, if the chain were not aperiodic, we could consider the subchain at times $d, 2d, \ldots$. So a more general system should be reducible to a system for which these two assumptions holds.

The third, and central, assumption is that the transitions between the microstates are reversible, i.e. the probability to move from state i to state j is the same as the probability to move from state j to state i. If this happens, then the transition matrix is symmetric, and it follows immediately that the chain has the uniform distribution as an invariant distribution. Thus the chain is finite, positive recurrent and therefore Harris recurrent, irreducible and aperiodic. Thus it will converge to the uniform distribution, and the law of large numbers holds. In particular, for any random variable f, i.e. any observable quantity on the state space, we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \frac{1}{\#\mathcal{S}} \sum_{s \in \mathcal{S}} f(s)$$

Put differently, for large times, the time average of any observable quantity is equal to the average across the state space (which in turn, by our defintion of the probability as being the fraction across the ensemble, is the average across the ensemble).

Taking as f the characteristic function of any subset A of the state space, we find in particular that the average time that the system spends in that part of the state space is equal to $p(A)$ which is in turn - as we are assuming a uniform distribution - proportional to the multiplicity $\Omega(A)$, i.e. the number of states in A. Thus, if we have an area for which the multiplicity is very small compared to the total number of states, the system will rarely enter that region, and we will need a very large number of observations to catch the system in one of these states.

Example 13.4 (Einstein solid). Let us now try to calculate the entropy for the Einstein solid. We consider an Einstein solid that has $N/3$ particles and hence N oscillators. If the total energy of the system is $U = q\hbar\omega$ for some integer q, then exactly those microstates are compatible where the sum of all occupation numbers is q. It is not difficult to show that the number of states is given by

$$\Omega(N, q) = \binom{q + N - 1}{q} = \frac{(q + N - 1)!}{(q!)(N - 1)!}$$

(see for example [5], section 2.2 for a derivation). We can again apply the Stirling formula and ignore the difference between $N-1$ and N to approximate the entropy:

$$S(N, q) = k_B(q + N) \ln(q + N) - k_B q \ln q - k_B N \ln N$$

Using the rule $\frac{d}{dx} x \ln x = \ln x + 1$ twice, we obtain that derivative of this with respect to q is

$$\frac{dS}{dq} = k_B \ln(q + N) - k_B \ln q = k_B \ln(1 + \frac{N}{q})$$

and observing $dU = \hbar\omega dq$, we find that

$$\frac{\hbar\omega}{T} = \hbar\omega \frac{dS}{dU} = \frac{dS}{dq} = k_B \ln(1 + \frac{N}{q})$$

as an expression for the temperature. For large values of q, we can approximate the logarithm by

$$\ln(1 + \frac{N}{q}) \approx \frac{N}{q}$$

and we see that the temperature will be almost proportional to q and therefore the energy, so that the heat capacity tends to a constant value. In the diagram below, we have plotted the values of entropy and temperature for a system with $N = 500$ and $\hbar\omega = 1$.
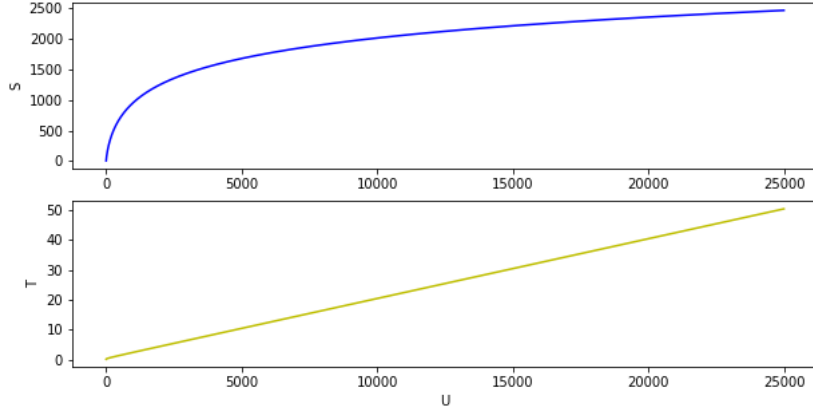
Figure 13: Entropy and temperate of an Einstein solid with N = 500

Also note that we can rewrite the entropy in terms of the temperature and the energy as follows. First, we summarize the terms in the expression for the entropy with a factor $k_B q$ and write them as

$$k_B q \ln(q + N) - k_B q \ln q = k_B q \ln \frac{q + N}{q}$$

$$= q k_B \ln(1 + \frac{N}{q}) = \frac{q\hbar\omega}{T} = \frac{U}{T}$$

Similarly

$$k_B N \ln(q + N) - k_B N \ln N = k_B N \ln \frac{q + N}{N}$$

$$= k_B N \ln \frac{q + N}{q} = k_B N \ln(1 + \frac{q}{N})$$

so that we obtain

$$S = \frac{U}{T} + k_B N \ln(1 + \frac{q}{N})$$

as our final expression for the entropy of an Einstein solid.

14 The canonical ensemble

So far, we have worked with the microcanonical ensemble which describes the entropy of a system with fixed energy. In many cases, however, it is more convenient to consider systems for which the energy can vary, but the temperature is constant. As in our derivation of the Helmholtz potential, we can do this by considering a composite system that consists of the system of interest - that we again call the primary system - and a thermal reservoir. We want to understand how the probability distribution of the states of our system changes due to the contact with the reservoir.

Example 14.1 (Einstein solid). To get an idea for what is going on, let us start with an example. We consider a system that consists of two Einstein solids with N_A and N_B oscillators. We consider system B to be the reservoir and system A to be the primary system. Thus we assume that $N_B >> N_A$, but that N_A and N_B are both sufficiently large so that the Stirling approximation holds. Using our formula for the temperature of an Einstein solid that we have derived earlier, we find that in the equilibrium state

$$\frac{q_A}{q_B} = \frac{N_A}{N_B}$$

For simplicity, we also set $\hbar\omega = 1$ in what follows. Let $q_t = q_A + q_B$ denote the total number of quanta of the system.

Let us now look at a specific state s of the smaller system and try to understand under what conditions we will find the overall system in this state. A state of the overall system is a combination of s with a state of the reservoir. If we denote by $q(s)$ the number of quanta of state s, then $q_t - q$ quanta will be left for the reservoir. Thus we have

$$\Omega_B(N_B, q_t - q(s))$$

states of system B that we can combine with s to obtain a valid state for the overall system, and this is the number of possible states of the overall system for which system A is in state s. To obtain the probability of this to happen, we can again use our assumption that for the states of the overall system, all state is equally likely. So the probability of this event is the number of states for which it occurs divided by the total number of states, i.e.

$$p(s) = \frac{\Omega_B(N_B, q_t - q)}{\Omega(N_A + N_B, q_t)}$$

Taking the logarithm of both sides and using our identification of the entropy with the Boltzmann constant times the logarithm of the multiplicity, we obtain

$$\ln p(s) = \frac{1}{k_B}[S_B(q_t - q(s), N_B) - S(q_t, N_A + N_B)]$$

Using the expression for the entropy of an Einstein solid in terms of temperature and energy that we have derived before and keeping in mind that the temperatures of all systems are equal to T, we find that

$$\ln p(s) = -\frac{q(s)}{k_B T} + N_B \ln(1 + \frac{q_t - q(s)}{N_B}) - (N_A + N_B)\ln(1 + \frac{q_t}{N_A + N_B})$$

The third term does not depend on q at all, only on the total energy and the total number of particles that we assume to be fixed. The second term does depend on q, but as long as q is much smaller than q_t, we can safely ignore that dependency. Thus we obtain that

$$\ln p(s) \approx -\frac{q(s)}{k_B T} + const.$$

and therefore

$$p(s) \simeq e^{-\frac{q(s)}{k_B T}}$$

We find that the probability for a state s of the smaller system depends on the exponential of the energy that this state has. This might be a bit surprising, so let us look at an extreme case to understand why this happens. Suppose our small system consists of exactly one oscillator, and we consider a state where all the energy is concentrated in our small system. Then all quanta are placed in this single oscillator, and none of the quanta are in any off the oscillators in system B. Thus there is exactly one state for which the small system has energy q_t, and the probability to be in this state is extremely low. If we reduce the energy in system A, more and more states will be possible for system B and the probability will increase.

We will soon see that this is no coincidence, but a general phenomena - when we consider a system in thermal equilibrium with a reservoir, the probability distribution of states of the small system will be of a similar form. Thus when we allow the smaller system to exchange heat with the reservoir, the probability distribution changes from an initially uniform distribution to an exponential distribution, and with it change all macroscopic variables that are in fact average values and thus depend on this probability distribution. If you want, you can image the actual configuration space as a space of probability distributions which determines all macroscopically observable variables, like the average energy or the (Shannon) entropy, and you could in fact rephrase the second law by stating that the distribution will tend to an equilibrium distribution which maximizes the Shannon entropy, we will come back to this point a bit later.

The approximations that we have made for our example do in fact work reasonably well also for comparatively low numbers of N_A and N_B. In diagram 14, we have plotted the exact values of the probability distribution (blue bars) for $N_A = 1$, $N_B = 12$ and $q_{tot} = 20$, computed using binomial coefficients, along with the prediction made by the Boltzmann distribution (yellow). We see that the actual distribution is reasonably close to a Boltzmann distribution even in this case. The average energy U is $U \approx 1.54$, and the most likely energy level is $E = 0$, with a probability of ≈ 0.38 (so there is one energy level below the most likely level and the average, a situation which is untypical and disappears if we increase the number of particles).
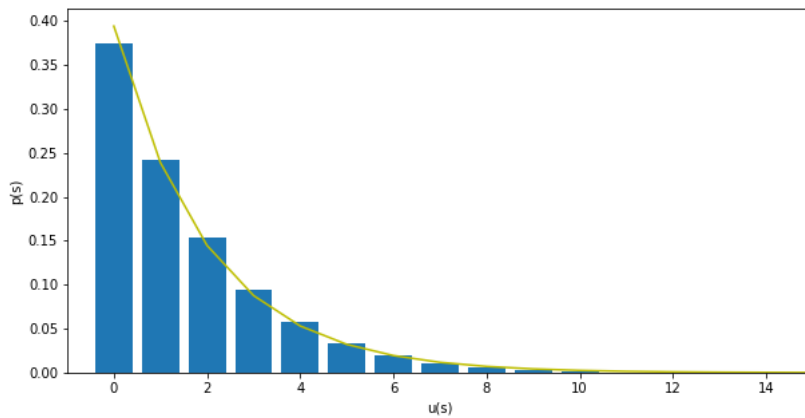

Figure 14: Boltzmann distribution for an Einstein solid

Let us now try to mimic the chain of arguments that we have used in our example to obtain a similar result in a more general setting. So we consider a system with energy U, volume V and particle number N, which is in contact with a thermal reservoir, with energy U_R, fixed volume V_R and fixed particle number N_R. Let T denote the temperature of the reservoir, which, by the very definition of a reservoir, is a constant. We want to derive an expression for the probability of a microstate with given energy.

We consider the composite system consisting of the system of interest and the reservoir as an isolated system with fixed total energy U_{tot}, fixed total volume V_{tot} and fixed particle number N_{tot}. Thus we can apply the microcanonical formalism to the composite system and assume that all microstates of the composite system compatible with U_{tot}, V_{tot} and N_{tot} have equal probability.

Now let us pick a specific microstate s of the system of interest, and let $E(s)$ denote the energy of that microstate (we continue to denote the macroscopic variable total energy by U and the microscopic energy of a state by E). This state can be combined with a state of the reservoir to a state of the composite system compatible with the constraints if and only if the state of the reservoir has energy $U_{tot} - E(s)$. Thus the number of microstates of the composite system for which the system of interest is in state s and which are compatible with the constraints is equal to the number of reservoir states with energy $U_{tot} - E(s)$, i.e. to

$$\Omega_R(U_{tot} - E(s))$$

As all admissible states of the composite system are supposed to be equally likely, the probability to find the system in state s is therefore

$$p(s) = \frac{\Omega_R(U_{tot} - E(s))}{\Omega_{tot}} \tag{4}$$

Now we can express numerator and denominator in terms of the entropy. The numerator is

$$\Omega_R(U_{tot} - E(s)) = \exp(\frac{1}{k_B} S_R(U_{tot} - E(s)))$$

and the denominator is

$$\Omega_{tot} = \exp \frac{1}{k_B} S_{tot}(U_{tot})$$

If we wait until the system has exchanged heat with the reservoir to adjust its temperature to be equal to the temperature T of the reservoir, the energy of the system will have reached a certain value U and the energy of the reservoir in equilibrium will therefore be

$$U_R = U_{tot} - U$$

Let us see how both entropy terms can be related to U. For the denominator, we can write, using the additivity of the entropy

$$\Omega_{tot} = \exp \frac{1}{k_B} S_{tot}(U_{tot}) = \exp(\frac{1}{k_B} S(U)) \exp(\frac{1}{k_B} S_R(U_{tot} - U))$$

For the numerator, we write

$$U_{tot} - E(s) = (U_{tot} - U) + (U - E(s)) = U_R + (U - E(s))$$

Thus we can try to express $S_R(U_{tot} - E(s))$ by using a Taylor expansion of the entropy of the reservoir around the point U_R. As we have assumed that the particle number and volume of the reservoir are fixed, the only relevant first derivative is

$$\frac{\partial S_R}{\partial U_R} = \frac{1}{T_R} = \frac{1}{T}$$

and the second derivative is

$$\frac{\partial^2 S_R}{\partial U_R^2} = \frac{\partial}{\partial U_R}\frac{1}{T} = -\frac{1}{T^2}\left(\frac{\partial T}{\partial U_R}\right)_{V,N}$$

where the last term is the inverse of the heat capacity. Now, the characteristic property of a reservoir is that we can add heat without raising its temperature (with other words, the heat capacity of the reservoir is infinite), meaning that the derivative of the temperature with respect to the energy is zero so that the second and all higher derivatives are zero. We thus obtain the expansion

$$S_R(U_{tot} - E(s)) = S_R(U_{tot} - U) + \frac{1}{T}(U - E(s))$$

so that

$$\Omega_R(U_{tot} - E(s)) = \exp\left(\frac{1}{k_B}S_R(U_{tot} - U)\right)\exp\left(\frac{1}{Tk_B}(U - E(s))\right)$$

Let us now introduce the abbreviation

$$\beta = \frac{1}{k_B T}$$

Then we obtain by putting all of this together that

$$p(s) = \exp(\beta(U - E(s)))\exp(-\beta T S(U)) = \exp(\beta(U - TS))\exp(-\beta E(s))$$

Now recall that U is the average, i.e. the macroscopically observed energy of our primary system (an assumption that we have in fact not even used so far). Therefore the quantity $U - TS$ is nothing but the Helmholtz potential of the primary system, and we obtain the fundamental equation

$$\boxed{p(s) = e^{\beta(F - E(s))}}$$

If we sum this over all possible states s, we find that

$$1 = \sum_s p(s) = e^{\beta F}\sum_s e^{-\beta E(s)}$$

Thus we recognize the exponential of minus the Helmholtz potential as a normalization factor which is conventionally denoted by Z and called the *partition function*:

$$Z = \sum_s e^{-\beta E(s)} = e^{-\beta F}$$

Using this, we can express our result in the form

$$\boxed{p(s) = \frac{1}{Z}e^{-\beta E(s)}}$$

This probability distribution is called the *Boltzmann distribution* or *canonical distribution*. Note that the partition function depends - as the Helmholtz energy - on T, V, N and potentially on other macroscopic variables.

Let us now calculate the average energy of the system - or more precisely its expectation value. Of course we expect that this is equal to U. The computation that we will do involves a general pattern that is used very frequently when working with Boltzmann distributions, so we will do this in a bit more generality.

The quantity βE will, of course, depend on the state, but it obviously also depends on the temperature (which is part of β). In a general model, however, the Energy could depend on other parameters as well. Suppose for instance that our system interacts with a magnetic field. Then the energy will have a term depending on that field as well. So let us assume in full generality that βE depends on some parameter X (which could be the temperature or a magnetic field) in addition to the dependendy on the state. Then, typically the derivative of βE with respect to X has a certain physical meaning. For $X = \beta$, for instance,

$$\frac{\partial}{\partial X}(\beta E) = E$$

so that the partial derivative is the energy. If E depends on some magnetic field, then the partial derivative of βE with respect to that field will be the magnetic moment and so forth. Of course, Z will then also depend on this parameter. Let us try to compute the derivative of Z or more precisely $\ln Z$ with respect to this parameter.

$$\begin{aligned}
\frac{\partial}{\partial X}\ln Z &= \frac{1}{Z}\frac{\partial}{\partial X}\sum_s \exp(-\beta E) \\
&= \frac{1}{Z}\sum_s \frac{\partial}{\partial X}(\exp(-\beta E)) \\
&= -\frac{1}{Z}\sum_s \exp(-\beta E)\frac{\partial}{\partial X}(\beta E) \\
&= -\sum_s p(s)\frac{\partial}{\partial X}(\beta E) = -\langle \frac{\partial}{\partial X}(\beta E)\rangle
\end{aligned}$$

In other words *the mean value of a derivative of the energy is minus the corresponding derivative of* $\ln Z$. This is a very useful relation. We can immediately apply it to the energy which is the derivative of βE with respect to β and find that

$$\langle E\rangle = -\frac{\partial}{\partial\beta}\ln Z$$

But having our explicit expression of $\ln Z$ in terms of the Helmholtz energy, we can compute that derivative:

$$\begin{aligned}
-\frac{\partial}{\partial\beta}\ln Z &= \frac{\partial}{\partial\beta}(\beta F) \\
&= F + \beta\frac{\partial F}{\partial\beta} \\
&= F - \frac{k_B}{\beta}\frac{\partial F}{\partial T} \\
&= F - T\frac{\partial F}{\partial T} = F + TS = U
\end{aligned}$$

so that we eventually obtain

$$\langle E\rangle = -\frac{\partial}{\partial\beta}\ln Z = U$$

as expected. We can also calculate the probability to find the system in a specific

energy E. In fact,

$$P(E) = \sum_{s|E(s)=E} p(s)$$

$$= \frac{1}{Z} \sum_{s|E(s)=E} e^{-\beta E}$$

$$= \frac{1}{Z} \Omega(E) e^{-\beta E}$$

$$= \frac{1}{Z} e^{\frac{1}{k_B} S(E)} e^{-\beta E} = \frac{1}{Z} e^{\beta(TS(E)-E)}$$

Motivated by the macroscopic minimum principle for the Helmholtz energy derived earlier, it is tempting to interpret the exponent as the Helmholtz energy. However, this is not exactly true because to get the Helmholtz energy for an energy level E, we would have to use the temperature that corresponds to this energy, not the temperature T which corresponds to the equilibrium energy U.

One additional remark on that formula is in order. If we use this formula with $U = E$ and recall that Z is $\exp(-\beta F)$, we seem to obtain the result that the system is in a state with energy U with certainty, i.e. with probability one. This is of course not true. The reason for this seemingly paradox situation is that the above derivation is only valid if E is one of the discrete energy levels that the system can attain! If E is none of these values, then $\Omega(E) = 0$ and the entropy is formally minus infinity, so that the derivation breaks down. Now, for a large number of particles, the average energy U will be slightly above the lowest energy, but still extremely close to it. Therefore the average energy U is not one of the allowed energy levels and the probability for the system to have that energy is actually zero (which would also be the result if we tried to fix the problem by passing to a continuous model). Thus the only conclusion that we can draw is that among the allowed, discrete energy levels, the one whose Helmholtz energy (computed with the temperature T) is closest to the minimum is the most likely energy level.

Be careful at this point - the most likely energy is not the energy of the most likely state. The most likely energy level is the one that mininizes the Helmholtz energy as discussed above, while the most likely state is the one with lowest energy. To see why this is not a contradiction, just consider a hypothetical system with three states, one with probability 0.4 and two others with probability 0.3. If the two other states have the same energy, than the probability of that energy of .6 which is higher than the probability of the energy of the most likely state. This is because to calculate probabilities for an energy, you have to take the number of states that realize this energy into account, i.e. the entropy. Also note that the statement that the most likely energy level is the one with the lowest Helmholtz energy is only true when we fix the probability distribution and with it the partition function and the average energy.

Let us now differentiate the relation between energy and partition function once more to find the heat capacity. We have

$$\frac{\partial U}{\partial \beta} = \frac{\partial}{\partial \beta} \frac{1}{Z} \sum_s E(s) e^{-\beta E(s)}$$

$$= -\frac{1}{Z} \frac{\partial \ln Z}{\partial \beta} \sum_s E(s) e^{-\beta E(s)} - \frac{1}{Z} \sum_s E(s)^2 e^{-\beta E(s)}$$

$$= \langle E \rangle^2 - \langle E^2 \rangle = -var(E)$$

from which we obtain that

$$C_V = \frac{\partial U}{\partial T} = \frac{var(E)}{k_B T^2}$$

Let us try to interpret this result. First, the heat capacity of a system is typically proportional to the particle number, i.e. it is given by a constant (the molar heat capacity) times the particle number. Thus if we scale a system up by the factor λ, this goes up by a factor λ as well. What really controls how sharp the peak of the energy distribution is, however, is not really the variance, but the variance in relation to the absolute value of the energy, or, more precisely, the quotient of the standard deviation (the square root of the variance) and the mean energy (sometimes called the coefficient of variation). By what we have just said, the standard deviation scales with $\sqrt{\lambda}$, while the energy, being an extensive quantity, should scale with λ. Thus the quotient scales with $\lambda^{-\frac{1}{2}}$ and therefore goes to zero as $\lambda \to \infty$. Thus, in the limit of a system with an infinite number of particles, the *thermodynamic limit*, this quantity becomes zero and the energy becomes sharp.

Finally, we can express the entropy in terms of the partition function. We know of course that the entropy is minus the partial derivative of the Helmholtz energy by the temperature. So we find immediately

$$S = k_B \frac{\partial}{\partial T} T \ln Z$$

It is reassuring that up to the factor k_B, this is the same as the Shannon entropy of the distribution. In fact, by defnition

$$\ln p(s) = - \ln Z - \beta E(s)$$

and thus

$$
\begin{aligned}
S_{Shannon} &= - \sum_s p(s) \ln p(s) \\
&= \sum_s p(s) \ln Z + \sum_s \beta p(s) E(s) \\
&= \ln Z \sum_s p(s) + \beta \langle E \rangle \\
&= \ln Z + \beta U = \beta(U - F)
\end{aligned}
$$

Now using again the definition of the Helmholtz energy as $F = U - TS$, we have $U - F = TS$, so that

$$S_{Shannon} = \beta TS = \frac{1}{k_B T} TS = \frac{1}{k_B} S$$

This is not just pure coincidence - [11] has a nice argument in section 1.3.3 that Shannon and Boltzmann entropy are identical for all physical systems.

The equations that we have derived are the fundamental equations describing the Boltzmann distribution and are so important that they deserve being highlighted once more.

$$p(s) = \frac{1}{Z} e^{-\beta E(s)}$$

$$Z = \sum_s e^{-\beta E(s)} = e^{-\beta F}$$

$$F = -k_B T \ln Z$$

$$S = k_B \frac{\partial}{\partial T} T \ln Z$$

$$\langle E \rangle = -\frac{\partial}{\partial \beta} \ln Z$$

$$var(E) = -\frac{\partial U}{\partial \beta} = k_B T^2 C_V$$

We note that given the partition function, we can use the second or third relation in this list to obtain the Helmholtz potential from which - as we know - we can either obtain all thermodynamical quantities directly or to which we can apply an inverse Legendre transform to obtain the energy. Thus knowledge of the partition function will enable us to derive all classical thermodynamical properties of the system.

It is interesting to compare the various energy related quantities that we have discussed so far. So pick a certain macrostate, which, as we know, fixes the average energy $U = \langle E \rangle$ as well as the entropy S. We know that the probability of an individual state s is given

$$p(s) = \exp(\beta(F - E(s)))$$

Being a probability, this cannot exceed one, so that we immediately find

$$E(s) \geq F$$

for every microstate s. If there is equality for a microstate, the probability to be in that state is one, i.e. the system only has one possible state, a case that we ignore. Thus if we denote the smallest possible energy by E_0, we find that $E_0 > F$.

There is a similar relation between the average energy U and the smallest possible energy. In fact

$$U = \langle E \rangle = \sum_s p(s) E(s) \geq \sum_s p(s) E_0 = E_0$$

and if we have equality, then

$$E_0 = \sum_s p(s) E(s) = \sum_s p(s)(E - E_0) + \sum_s p(s) E_0 = \sum_s p(s)(E - E_0) + E_0$$

from which we could again conclude that only states with energy E_0 have a probability different from zero. Excluding this pathological case, we therefore find that

$$F < E_0 < \langle E \rangle$$

Due to the large factor $k_B T$ in the exponent, the probability will go down rapidly if the energy is significantly (i.e. in the order of $k_B T$ times a small number) higher than E_0 and therefore F. Thus typically, the states that have a relevant probability of occuring have energies that are rather close to the ground state energy E_0, and states with much higher energy are rather unlikely. Diagram 15 summarizes our findings.

So far we have treated the particle number as constant. In some situations, however, fluctuations of the particle number can no longer be ignored. In that
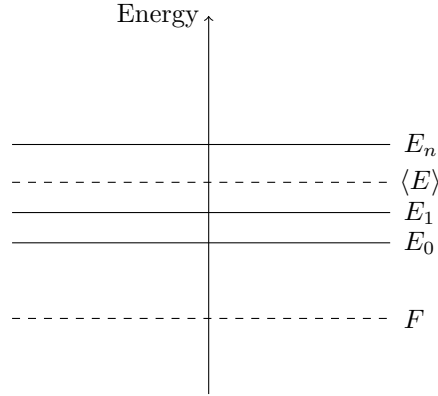
Figure 15: Energy levels in the Boltzmann distribution

case, we can replace the Boltzmann distribution by a distribution called the *Gibbs distribution*. The derivation is very similar. In addition to assuming that our system is in contact with a heat reservoir, we also have to assume that the reservoir acts as a chemical reservoir, so that in equilibrium, the chemical potential is constant. In equilibrium, the primary system will then settle down at a certain energy U and a certain particle number N. In equation 4, we can then expand the logarithm of the numerator again into a Taylor series, where the terms of higher order vanish, but this time we need a contribution of the derivatives with respect to U and N. If we then collect terms and the smoke settles, we find that

$$p(s) = e^{\beta \Phi(T, V, \mu)} e^{-\beta(E(s) - \mu N(s))}$$

where

$$\Phi = U - TS - \mu N$$

is the *grand free energy* or the *grand canonical potential*. Obviously, Ψ is minus the Legendre transform of the energy with respect to S and N. This distribution is called the *Gibbs distribution*[2]. It is now clear how other combinations of constraints lead to distributions in which other Legendre transforms of the energy appear.

Later, we will often consider systems which are combinations of non-interacting subsystems. In this case, the partition function is multiplicative. To make this more precise, assume that we are given two systems S_1, S_2 and a combined system, so that the state of the combined system is described by a pair $s = (s_1, s_2)$ where s_i is the state of system S_i. If there is no interaction term in the energy, then we can write the energy as

$$E(s) = E_1(s_1) + E_2(s_2)$$

The partition function then becomes

$$Z = \sum_{s_1, s_2} \exp(-\beta E(s))$$
$$= \sum_{s_1} \exp(-\beta E_1(s_1)) \sum_{s_2} \exp(-\beta E_2(s_2)) = Z_1 Z_2$$

As we will soon see, this can simplify the calculation of partition functions greatly.

---

[2]The terminology is a bit confusing at this point. If we allow the volume to vary at constant pressure and do the same derivation, the Gibbs energy appears in the exponent, but it seems that this distribution is NOT called the Gibbs distribution in most textbooks

Let me add a few closing remarks. First, it is worth noting that both, the microcanonical and the canonical distribution, can be derived from a more abstract principle - they are the distributions that maximize the Shannon entropy under the constraints specific to the model. In this sense, they are the distributions that forget everything except the information provided by the constraints, a point of view which is for instance discussed in [4], chapter 17.

Finally, recall that we have obtained the derivation of the Boltzmann distribution using a Taylor expansion of the entropy of the reservoir, which is an approach found in many textbooks, including [4], and (though in a slightly disguised form) [5]. We remark, however, that there is a different approach to arrive at the Boltzmann distribution, starting with Shannon's definition of entropy and then using Lagrange multipliers to maximize the entropy under the constraints of a given average energy and a normalization condition, we refer to section 17.2 in [4] for details on this approach.

15   A toy example - a two-state system

To illustrate the concepts discussed in the previous section, let us work out a simple toy example in all details - a two-state system with only one degree of freedom. Specifically, we consider a single particle that can be in one of two states - $s = 1$ corresponding to spin up, and $s = -1$ corresponding to spin down. We think of our particle as being influenced by an external magnetic field B, and the energy is

$$E(s) = -Bs$$

so that the only two allowed energy levels are

$$E_0 = E(s = 1) = -B$$

and

$$E_1 = E(s = -1) = B$$

We assume that $B > 0$ so that $E_0 < E_1$. In this simple case, we can of course immediately write down the partition function. We have

$$Z = \sum_s \exp(-\beta E(s)) = \exp(-\beta B) + \exp(\beta B) = 2\cosh(\beta B)$$

We can also easily obtain down the full Helmholtz energy as

$$F = -k_B T \ln Z = -k_B T \ln 2\cosh(\beta B)$$

Let us now calculate the average energy - either using our formula in terms of derivatives of Z or directly. Let us do both to see that the outcome is the same. First, the direct calculation is

$$\langle E \rangle = p(s = 1)E_0 + p(s = -1)E_1$$
$$= \frac{1}{Z}(-B\exp(\beta B) + B\exp(-\beta B))$$
$$= \frac{B}{Z}(\exp(-\beta B) - \exp(\beta B)) = -2\frac{B}{Z}\sinh\beta B = -B\tanh\beta B$$

We already see that this is not exactly $-B$, but a bit higher, but converges to $-B$ for large values of B. Let us do the same calculation using the partition function.

$$\langle E \rangle = -\frac{\partial}{\partial \beta}\ln Z = -\frac{1}{Z}\frac{\partial}{\partial \beta}Z$$
$$= -\frac{1}{Z}\frac{\partial}{\partial \beta}2\cosh(\beta B) = -B\frac{1}{Z}2\sinh\beta B$$

and we get the same result as before.

Let us now move on to the entropy. There are several ways to get this, and again we will use two of them to show that the outcome is the same. The first approach is to use the relation

$$F = U - TS$$

so that

$$S = \frac{1}{T}(U - F)$$

This involves quantities that we have already computed, so we obtain

$$S = -\frac{B}{T}\tanh\beta B + k_B \ln 2\cosh(\beta B)$$

The (harder) way is to calculate this is by taking the derivative of the Helmholtz potential with respect to the temperature. Thus

$$
\begin{aligned}
S &= -\frac{\partial F}{\partial T} = -\frac{\partial F}{\partial \beta}\frac{d\beta}{dT} \\
&= -\frac{1}{T^2}\frac{\partial}{\partial \beta}T \ln 2\cosh(\beta B) \\
&= -\frac{1}{T}\frac{\partial}{\partial \beta}\ln 2\cosh(\beta B) - \frac{1}{T^2}\ln 2\cosh(\beta B)\frac{dT}{d\beta} \\
&= -\frac{1}{T}\frac{1}{2\cosh(\beta B)}\frac{\partial}{\partial \beta}2\cosh(\beta B) + \frac{1}{T^2}k_B T^2 \ln 2\cosh(\beta B) \\
&= -\frac{B}{T}\tanh\beta B + k_B \ln 2\cosh(\beta B)
\end{aligned}
$$

giving again the same result.

So far we have considered a system with only one particle. Now suppose that we have N particles which, however, do not interact, i.e. a state is given by a vector $(s_1, s_2, \dots)$ in $\{-1, 1\}^N$. The energy of a state is then

$$E(s) = -B\sum_i s_i = -Bm(s)$$

where we define

$$m(s) = \sum_i s_i$$

Now we know that the partition function is multiplicative, so that we can immediately conclude that the partition function is given by

$$Z = 2^N \cosh(\beta B)^N$$

This factor N then appears in front of all quantities that we obtain from $\ln Z$, like the Helmholtz free energy or the entropy, which is what we expect as these are extensive quantities. Let us now take a closer look at the energy levels of this system. Again, the lowest energy is given by

$$E_0 = -NB$$

corresponding to a state where all spins are aligned with the magnetic field. In general, if N_+ sites have state $+1$ and N_- sites have state -1, the energy is given by

$$E = -B(N_+ - N_-) = B(N_- - (N - N_-)) = 2BN_- + E_0$$

so that

$$N_- = \frac{1}{2B}(E - E_0)$$

and

$$N_+ = N - N_- = N - \frac{1}{2B}(E - E_0)$$
$$= -\frac{2E_0}{2B} - \frac{1}{2B}(E - E_0) = \frac{-E_0 - E}{2B}$$

Let us now calculate the multiplicity of a given energy level E, i.e. we want to many how many states s we can find that have $E(s) = E$. Now the energy determines N_-, so this is equivalent to calculating the number of states that share the same value of N_-. But this is just the number of ways in which we can designate N_- sites as pointing downwards, i.e. this is

$$\Omega(N_-) = \binom{N}{N_-} = \frac{N!}{N_-!(N - N_-)!} = \frac{N!}{N_-!N_+!}$$

For large numbers of N, N_+ and N_-, we can again use Stirling's formula to approximate this, and find after a short calculation that this is

$$\ln \Omega = N \ln N - N_- \ln N_- - N - + \ln N_+$$

which we can rewrite as follows

$$\ln \Omega = (N_- + N_+) \ln N - N_- \ln N_- - N_+ \ln N_+$$
$$= -N_- \ln \frac{N_-}{N} - N_+ \ln \frac{N_+}{N}$$
$$= \ln \Omega = -N[n_+ \ln n_+ + n_- \ln n_-]$$

with the abbreviation $n_\pm = \frac{N_\pm}{N}$. This gives us a formula for $\Omega(E)$ as

$$\ln \Omega(E) = N\left[\frac{E - E_0}{2E_0} \ln \frac{E_0 - E}{2E0} - \frac{E + E_0}{2E_0} \ln \frac{E + E_0}{2E_0}\right]$$

Of course this formula breaks down if E approaches the maximum or minimum energy that the system can have, as in this case N_- and N_+ become very small and our approximations do no longer work.

At the first glance, this looks very different from S - and keep in mind that these are two different things. $\Omega(E)$ is zero for almost all E and only defined for the sharp energy levels that the energy of a state can take on. On the other hand, $S(U)$ is a property of the probability distribution, not a single state, so we should not expect this to be equal. However, if an energy is close to the average energy U, we would expect these two quantities to take similar values.

To see that this holds in this case, let us consider the quantity

$$q = \frac{E}{E_0}$$

and rewrite our expression for $\Omega(E)$ in terms of q. First

$$\frac{E_0 - E}{2E_0} = \frac{1}{2}\frac{E_0 - E}{E_0} = \frac{1}{2}\left(1 - \frac{E}{E_0}\right) = \frac{1}{2}(1 - q)$$

and

$$\frac{E_0 + E}{2E_0} = \frac{1}{2}\frac{E_0 + E}{E_0} = \frac{1}{2}\left(1 + \frac{E}{E_0}\right) = \frac{1}{2}(1 + q)$$

Thus if q is small, we can approximate

$$\frac{E - E_0}{2E_0} \ln \frac{E_0 - E}{2E0} = -\frac{1}{2}(1 - q) \ln \frac{1}{2}(1 - q) \approx \frac{1}{2}(1 - q)(-\ln \frac{1}{2} + q)$$

and similarly

$$\frac{E + E_0}{2E_0} \ln \frac{E + E_0}{2E_0} = \frac{1}{2}(1 + q) \ln \frac{1}{2}(1 + q) \approx \frac{1}{2}(1 + q)(\ln \frac{1}{2} + q)$$

so that

$$\ln \Omega(E) = \frac{N}{2}[(1 - q)(-\ln \frac{1}{2} + q) - (1 + q)(\ln \frac{1}{2} + q)] = N(\ln 2 - q^2)$$

Similarly, let us define

$$u = \frac{U}{E_0} = \frac{U}{-BN}$$

Our relation

$$U = -NB \tanh \beta B$$

shows that

$$\tanh \beta B = -\frac{U}{NB} = u$$

so that, for small values of u, we can again use a Taylor approximation for $\tanh^{-1}$ to find that

$$\beta B = u$$

i.e.

$$\beta = \frac{1}{k_B T} = \frac{u}{B}$$

so that

$$\frac{B}{T} = k_B u$$

The entropy is then given by

$$S = -k_B N u^2 + k_B N \ln 2 = k_B N(\ln 2 - u^2)$$

Thus we find that, as expected, the macroscopic entropy of the average energy is approximately equal to the entropy of the closest energy level.

16 Ideal gas and blackbody radiation

In the next sections, we will discuss two famous classical applications - the derivation of the laws of the ideal gas and the calculation of the energy spectrum of a black body.

Let us start by looking, more generally, at a gas which is made of quantum mechanical particles, following the exposition in [5], section 6.7. We assume that the particles do not interact, i.e. we ignore potential energy, and we also ignore internal kinetic energy like rotational energy. So the energy of a particle with momentum p is given by

$$\frac{p^2}{2m}$$

If we describe the state of each of these particles by a wave function and impose the boundary condition that the wave function is zero at the walls of the box, we can compute the number of states which are available to a particle.

Let us start with a one dimensional box. The treatment of a particle with mass m in a one-dimensional box of length L with infinitely high potential wall can be found in any introductory textbooks on quantum mechanics. The energy eigenvalues are

$$E_n = \frac{n^2\hbar^2\pi^2}{2mL^2}$$

and all eigenstates have multiplicity one, so the state is fully specified by the energy. We can now use this information to write down the partition function for an individual particle and obtain

$$Z_1 = \sum_s e^{-\beta E(s)} = \sum_n e^{-n^2\frac{\beta\pi^2\hbar^2}{2mL^2}}$$

If the energy levels are very close together (which is the case for a macroscopic L), we can approximate the sum by an integral and find that

$$Z_1 = \int_0^\infty e^{-x^2\frac{\beta\pi^2\hbar^2}{2mL^2}}\, dx = \frac{1}{2}\int_{-\infty}^\infty e^{-ax^2}$$

where we have used the fact that the integral is symmetric and introduced the notation

$$a = \frac{\beta\pi^2\hbar^2}{2mL^2}$$

But the integral is a Gaussian integral, and we therefore obtain

$$Z_1 = \frac{1}{2}\sqrt{\frac{\pi}{a}} = \frac{L}{h}\sqrt{2m\pi k_B T} = \frac{L}{\lambda}$$

where we have introduced the quantity

$$\lambda(T) = \sqrt{\frac{h^2}{2\pi m k_B T}} = \sqrt{\frac{2\pi\hbar^2}{m k_B T}}$$

which is called the *thermal wavelength* of the particle at temperature T. This argument is easily generalized to the case of three dimensions as the partition function factorizes by dimension, and we obtain that the partition function of a particle constrained in a cubic box of sidelength L and volume $V = L^3$

$$Z_1 = \frac{V}{\lambda(T)^3}$$

This is a very appealing result - the partition function is supposed to reflect the number of states available to the system, and we find that this given by the number of ways how we can split the volume of the box into smaller volumes related to the wavelength of a particle.

Having the partition function for a single particle, we can now write down the partition function for the entire gas, as the partition function in the canonical formalism is multiplicative if we assume that the individual constituents are independent. However, there is one subtlety that we need to consider: quantum mechanical particles are indistinguishable. Thus when simply multiply the partition functions of the individual particles, we overcount. A simple fix is to divide by the number of permutations $N!$, but this is not quite correct because it might be that some particles are in the same state, even though this will be rather unlikely if the number of available states is large. However, let us ignore this problem - which would lead into

the realm of quantum statistics - for the moment and simply divide by $N!$. Thus we find that

$$Z(T, V, N) = \frac{1}{N!} \frac{V^N}{\lambda(T)^{3N}}$$

for the partition function of the ideal gas.

We can now obtain all relevant thermodynamical quantities from the partition function. Let us work out the average energy as an example. We have

$$\langle E \rangle = -\frac{\partial}{\partial \beta} \ln Z$$

$$= -\frac{\partial}{\partial \beta}(-3N \ln \lambda)$$

$$= 3N \frac{1}{\lambda} \frac{\partial}{\partial \beta} \lambda(T)$$

Now a short calculation shows that

$$\frac{\partial}{\partial \beta} \lambda(T) = \frac{\lambda}{2\beta}$$

and we therefore find that

$$\langle E \rangle = \frac{3}{2} N \frac{1}{\beta} = \frac{3}{2} N k_B T$$

which is the classical expression for the energy of an ideal gas. The entropy can be similarly calculated, but requires a bit more work. We start with

$$\frac{\partial T}{\ln Z} = \frac{\partial \ln Z}{\partial \beta} \frac{\partial \beta}{\partial T} = \frac{3}{2} \frac{N}{T}$$

using our result above for the derivative of $\ln Z$. Applying our master formula for the entropy in terms of the partition function, we now have

$$S = k_B \frac{\partial}{\partial T} T \ln Z$$

$$= k_B \ln Z + k_B T \frac{\partial}{\partial T} \ln Z = k_B \ln Z + \frac{3}{2} N k_B$$

To simplify the first term, we will have to use the Stirling approximation. Taking the logarithm, we find that

$$\ln Z = -\ln N! + N \ln V - 3N \ln \lambda$$

$$= -\ln N! + N \ln \frac{V}{\lambda^3}$$

$$= -N \ln N + N + N \ln \frac{V}{\lambda^3} = N + N \ln \frac{V}{N\lambda^3}$$

Combining this, we find that

$$S = N k_B \left[\ln \frac{V}{N\lambda^3} + \frac{5}{2} \right]$$

This formula has a nice physical interpretation. The numerator in the logarithm is the total volume available to the particles, and the denominator is the space taken by N particles and their typical wavelength. If we increase this ratio, we give the particles more available states, thus the entropy increases. We can do this be either increasing the volume, so the gas expands, or decreasing the wavelength $\lambda(T)$ which

is the same as increasing the temperature, so that an individual particle has a higher energy.

If we want to put this into its usual form, expressed in terms of the energy, we have to eliminate λ from the expression. Now

$$\ln \lambda(T)^3 = \frac{3}{2} \ln \lambda^2 = \frac{3}{2} \ln \frac{\hbar^2}{2\pi m k_B T} = \frac{3}{2} \ln \frac{3N\hbar^2}{4\pi m} - \frac{3}{2} \ln \frac{3}{2} N k_B T$$

and we obtain

$$\ln \lambda(T)^3 = -\frac{3}{2} \ln U + \frac{3}{2} \ln N + \frac{3}{2} \ln \frac{3\hbar^2}{4\pi m}$$

We therefore can write the logarithm in our expression for the entropy as follows

$$\ln \frac{V}{N\lambda^3} = \ln \left[(\frac{V}{N})(\frac{U}{N})^{\frac{3}{2}} \right] + \frac{3}{2} \ln \frac{4\pi m}{3\hbar^2}$$

from which obtain the formula

$$S(U, V, N) = N k_B \ln \left[(\frac{V}{N})(\frac{U}{N})^{\frac{3}{2}} \right] + \frac{5}{2} N k_B + \frac{3}{2} N k_B \ln \frac{4\pi m}{3\hbar^2}$$

which is known as the *Sackur-Tetrode equation*. We also find that the volume appears only in one term, so that we can easily calculate the derivative with respect to the volume and therefore the pressure:

$$\frac{P}{T} = \frac{\partial S}{\partial V} = N k_B \frac{\partial}{\partial V} \ln V = \frac{N k_B}{V}$$

We therefore immediately obtain the characteristical equation

$$PV = N k_B T$$

that relates volume, pressure, particle number and temperature of an ideal gas.

Now let us turn to our second example - the so-called *black body*. A standard model for such a system is an empty cavity with conducting walls that is in a thermal equilibrium with its environment. The atoms in the wall will then constantly emit radiation into the interior of the cavity, so in equilibrium, we expect to see a stable radiation field inside the cavity. We want to calculate the energy of this field.

So let us suppose that we consider a cube with sides of length L. Without getting too much into the details of electrodynamics, let us suppose that the electromagnetic waves in the interior, i.e. the solutions to Maxwell's equations, are superpositions of standing waves with wave vectors of the form

$$k = \frac{\pi}{L} n \tag{5}$$

where

$$n = (n_x, n_y, n_z) \in \mathbb{N}^3$$

and angular frequency $\omega(k)$, subject to the condition

$$\omega(k)^2 = k^2 c^2 \tag{6}$$

We model each standing wave of the field as an independent harmonic oscillator. As a consequence of Maxwell's equations applied to waves, the magnetic field will be completely determined once we know the electric field, and the electric field will be perpendicular to the wave vector. Thus, for each wave vector we have two degrees of freedom, corresponding to two independent vectors in the plane parallel to the

wave vector. To summarize: we model the radiation in the cavity as a collection of independent harmonic oscillators, two for each allowed wave vector.

The dispersion relation (6) together with the requirement (5) will only allow a finite number of wave vectors for a region $d\omega$ around a given angular frequency ω. Let us try to calculate this number, which is called the *density of states*.

For that purpose, we will first try to calculate the number of wave vectors with angular frequence less than ω. Thus we need to calculate the number of vectors n with integer coordinates that fulfill

$$n^2 \leq \frac{\omega^2 L^2}{c^2 \pi^2}$$

Roughly, this is equal to the volume of a ball with radius

$$\frac{\omega L}{c \pi}$$

divided by 8, as we only take the vectors for which all coordinates are positive, i.e.

$$\frac{1}{8} \frac{4}{3} \pi \frac{\omega^3 V}{c^3 \pi^3} = \frac{1}{6} \frac{V}{c^3 \pi^2} \omega^3$$

Differentiating with respect to ω and taking into account that each state vector gives rise to two oscillators, we find that

$$D(\omega) = \frac{V}{c^3 \pi^2} \omega^2$$

Now let us calculate the average energy of each oscillator. We consider each oscillator as being a system in a heat bath, where the heat bath is provided by the walls of the cavity. We can then apply the canonical formalism to each oscillator. To calculate its energy, we treat each oscillator as a quantum mechanical oscillator with energy levels

$$E_n = \hbar\omega \tag{7}$$

The state of the oscillator is then given by the number n of quanta of energy that it currently contains, called the *occupation number* (thus our vision of the radiation field in the cavity is that of a collection of oscillators that can exchange energy in quanta of $\hbar\omega$ - you could call these quanta photons, but it does in fact not matter for this derivation). According to the Boltzmann distribution, the probability of being in state n is given by

$$P(n) = \frac{1}{Z} e^{-n\beta\hbar\omega}$$

The partition function is then a geometric series

$$Z = \sum_{n=0}^{\infty} e^{-n\hbar\omega\beta} = \frac{1}{1 - e^{-\hbar\omega\beta}}$$

We can now easily calculate the average energy of each oscillator as a function of ω:

$$\langle E \rangle = -\frac{\partial}{\partial\beta} \ln Z$$

$$= -\frac{\partial}{\partial\beta} \frac{1}{1 - e^{-\hbar\omega\beta}}$$

$$= \frac{1}{1 - e^{-\hbar\omega\beta}} \frac{\partial}{\partial\beta} (1 - e^{-\hbar\omega\beta}) = \frac{\hbar\omega}{e^{\hbar\omega\beta} - 1}$$

which corresponds to an average occupancy number of

$$\langle n \rangle = \frac{1}{e^{\hbar\omega\beta} - 1}$$

Combining this with the state density that we have calculated above, we therefore find that the *energy density* of the radiation, i.e. the energy carried by the frequency in a range $(\omega, \omega + d\omega)$ is given by

$$\frac{V}{c^3\pi^2} \frac{\hbar\omega^3}{e^{\hbar\omega\beta} - 1} d\omega$$

or, dividing by the volume, that the energy density per frequency and volume is given by

$$\frac{1}{c^3\pi^2} \frac{\hbar\omega^3}{e^{\hbar\omega\beta} - 1} d\omega$$

To evaluate this further, it is useful to introduce the dimensionless variable

$$x = \frac{\hbar\omega}{k_B T} = \hbar\omega\beta$$

so that we can rewrite the density as

$$\frac{(k_B T)^4}{(\hbar c)^3 \pi^2} \frac{x^3}{e^x - 1} dx$$

Consequently, the total energy of the radiation field per unit area is given by

$$\frac{U}{V} = \frac{(k_B T)^4}{(\hbar c)^3 \pi^2} \int_0^\infty \frac{x^3}{e^x - 1} dx$$

Now there are a number of interesting observations that we can make. First, the integral is finite, as the numerator grows faster than any power of x. In fact (see [4], section 16.8, or [5], Appendix B) the integral turns out to be

$$\frac{\pi^4}{15}$$

so that we finally obtain

$$\frac{U}{V} = \frac{\pi^2 (k_B T)^4}{15(\hbar c)^3}$$

In particular, we recover the classical *Stefan-Boltzmann law* that predicts that the energy is proportional to $(k_B T)^4 V$. The second observation is that the distribution of the total energy to the frequencies of the radiation is given by the integrand, which is plotted below.

We see that the curve has a maximum, which can be numerically determined to be at $x = 2.821$, i.e. at $\hbar\omega = 2.821 k_B T$.

Now let us assume for a moment we had used a classical oscillator instead of a quantum oscillator - or, equivalently, let us calculate the classical limit, i.e. the limit as $\hbar \to 0$. Of course the state density does not change. What changes is the average energy per oscillator. In fact, we can do a Taylor expansion of the inverse of the energy to find that

$$\langle E \rangle^{-1} = \frac{e^{\hbar\omega\beta} - 1}{\hbar\omega} = \frac{\hbar\omega\beta + \frac{1}{2}\hbar^2\omega^2\beta^2 + \dots}{\hbar\omega} = \beta + \dots$$

so that

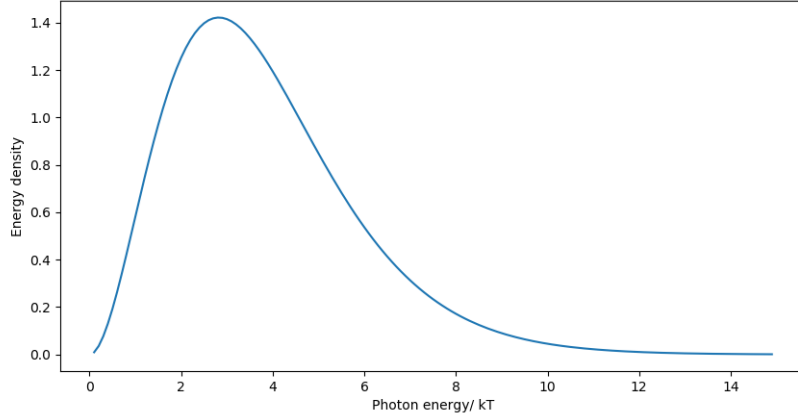$$\lim_{\hbar \to 0} \langle E \rangle = \beta^{-1} = k_B T$$

Figure 16: Planck spectrum for a black body

This is in perfect agreement with our expectation - a classical oscillator has two kinds of energy, kinetic energy and potential energy, and it is a general rule (known as the equipartition theorem) that each type of energy and each degree of freedom contribute $k_B T$ to the average energy of a classical system. Unfortunately, an average energy that does not depend on ω is a disaster because the energy density is then proportional to ω^2 and therefore the integral over ω and thus the total energy diverges. Thus the prediction of a classical treatment would be that the radiation in a cavity at thermal equilibrium has infinite energy, which is of course absurd. This issue was known as the *ultraviolet catastrophe*. Historically, this lead M. Planck in 1900 to the postulate that the energy of a mode of the radiation field is quantized according to equation (7) which can be considered the birth of quantum mechanics.

17   The Ising model

The Ising model is probably one of the most influential models in modern physics. It was originally conceived by the physicist E. Ising in his doctoral thesis in 1924 in an attempt to explain ferromagnetism. In the Ising model, we consider a d-dimensional lattice and, at each site of the lattice, a hypothetical particle which has a spin that can either be +1 (up) or -1 (down). A state is then described by recording for each site i whether the particle at this site has spin up or spin down, i.e. by a vector

$$s = (s_1, s_2, \ldots, s_N) \in \{-1, 1\}^N$$

where N is the number of sites in the lattice. Each particle can interact with the other particles on the lattice, but we assume that the interaction drops off quickly with increasing distance, so that only the interactions with the nearest neighbors are relevant. In addition, each particle can interact with an external magnetic field B. The energy of a state is then given by

$$E(s) = -\frac{1}{2}J\sum_{i,j} d(i,j)s_i s_j - B\sum_i s_i$$

with a distance function

$$d(i,j) = \begin{cases} 1 & \text{i and j are direct neighbors} \\ 0 & \text{otherwise} \end{cases}$$

and a constant J that describes the size of the interaction. It is common to use the notation $\langle i, j \rangle$ for a pair of lattice points that are direct neighbors. With this convention, our formula reads

$$E(s) = -J \sum_{\langle i,j \rangle} s_i s_j - B \sum_i s_i$$

as we count every pair only once. The model is called *ferromagnetic* if $J > 0$ which we will assume. For a given state s, we define the *magnetization* to be

$$m(s) = \frac{1}{N} \sum_i s_i$$

and the quantity that we are primarily interested in is the average magnetization

$$M = \langle m \rangle = \sum_s p(s) m(s)$$

The idea behind this model is as follows. Given our previous discussion of the Boltzmann distribution, we should expect that for a given temperature T, those states are most likely that minimize the energy. This can be achieved in two ways - a spin can align with the external field B or it can align with the neighboring spins to minimize the first part of the energy (this is why we have chosen $J > 0$). For a strong external magnetic field B, the first contribution will probably dominate, so we should expect that the states of lowest energies are those were all spins are aligned with the external field. In this state, all the spins are aligned with the field and consequently among each other. This creates a magnetic moment, i.e. the overall system becomes magnetic, a fact which is well known under the name *paramagnetism.*

If we now slowly turn off the external field, it might happen that the system initially remains in this state. If now one of the spins flips spontaneously and is now anti-aligned, the energy will increase, so that our state remains the state with the lowest energy and the body remains magnetic even though the external field has been switched off. This is called *ferromagnetism* and this is what Ising wanted to explain with his model.

To prepare for our discussion of the Ising model, let us first consider the case that $J = 0$, i.e. there is no interaction, and let us try to compute the partition function for this simple case. As now

$$E(s) = -B \sum_i s_i$$

we have

$$Z = \sum_s \exp(-\beta E(s))$$
$$= \sum_s \exp(\beta B \sum_i s(i)) = \sum_s \exp(\beta B N m(s))$$

This looks daunting, but recall that there is no interaction - the energy that we consider is the sum of the individual contributions $-B s_i$ for each site i, so we can try to exploit the fact that the partition function is multiplicative. To do this, we first need to calculate the partition function Z_1 for the case $N = 1$. This is easy, as there are only two possible states:

$$Z_1 = \sum_{s=-1,1} \exp(\beta B s)) = \exp(-\beta B) + \exp(\beta B) = 2 \cosh(\beta B)$$

With this and the fact that the partition function is multiplicative, we immediately obtain the expression

$$Z = Z_1^N = 2^N \cosh(\beta B)^N$$

for the case $J = 0$.

Of course, the situation is much more complicated in the case $J \neq 0$, and it appears hopeless to find a closed expression for the general model (there are, however, closed expressions for the case $d = 1$ due to Ising himself and for the case $B = 0$ and $d = 2$ due to Onsager). However, there is an approximation method that already displays some of the real behaviour of the model. To motivate this method, let us spell out the interaction in terms of the average magnetization M, or, more precisely, on the deviations $s_i - M$ that reflect the fluctuations of the spins. Now trivially

$$
\begin{aligned}
s_i s_j &= ((s_i - M) + M)((s_j - M) + M) \\
&= (s_i - M)(s_j - M) + (s_i - M)M + (s_j - M)M + M^2
\end{aligned}
$$

Let us take a closer look at the contribution that the second term makes to the interaction. For that purpose, let q denote the number of neighbors that each lattice point has (i.e. $q = 2$ in one dimension, $q = 4$ in two dimensions and so on).

$$
\begin{aligned}
-\frac{1}{2}J \sum_{i,j} d(i,j)(s_i - M)M &= -MJ \sum_{i,j} d(i,j)(s_i - M) \\
&= -\frac{1}{2}MJ \sum_i \sum_j d(i,j)(s_i - M) \\
&= -\frac{1}{2}MJ \sum_i (s_i - M) \sum_j d(i,j) \\
&= -\frac{1}{2}MJq \sum_i (s_i - M) = -\frac{1}{2}MJq \sum_i s_i + \frac{1}{2}NM^2 Jq
\end{aligned}
$$

A similar contribution comes from the third term. The last term finally makes the contribution

$$-\frac{1}{2}J \sum_{i,j} d(i,j)M^2 = -\frac{1}{2}JqNM^2$$

Thus we can write the total energy of a state s as follows,

$$E(s) = -\frac{1}{2}J \sum_{i,j} (s_i - M)(s_j - M) + \frac{1}{2}JqNM^2 - (JMq + B)\sum_i s_i$$

The first term is the energy due to the fluctuations, think of this as a second order term. The second term represents a constant contribution from the average magnetization. The third term is a term that contains the external field and an additional field that each spin experiences which is created by its q nearest neighbors, this is the contribution from the derivations $s_i - M$ in first order. The *mean field approximation* now drops the first term and uses the energy term

$$E(s) = \frac{1}{2}JqNM^2 - (JMq + B)\sum_i s_i$$

Let us now write down the partition function arising from this energy term. We find that

$$Z = \sum_s \exp(-\beta E(s)) = \sum_s \exp(\beta - \frac{1}{2}JqNM^2) \exp(\beta(JMq + B)\sum_i s_i)$$

The first term is a constant factor that can be pulled out of the sum so that

$$Z = \exp(-\beta \frac{1}{2} JqNM^2) \sum_s \exp(\beta(JMq + B) \sum_i s_i)$$

The remaining term, however, is formally the same as that in the case without interaction that we have already studied, and we can therefore write down the partition function immediately.

$$Z = \exp(-\frac{1}{2} \beta JqNM^2) 2^N \cosh(\beta(JMq + B))^N$$

This looks nice. However, we have cheated a bit. Formally, you could do these manipulations for every value of M. Physically, however, we know that this only has a chance to be a good approximation if M is actually the average magnetisation of our probability distribution, i.e $M = \langle m \rangle$. So let us calculate this magnetization. We know that

$$\sum_i s_i = -\frac{\partial}{\partial B} E(s)$$

so that by our general statements on average values of derivatives for Boltzmann distributions,

$$\langle m \rangle = \frac{1}{N} \langle \sum_i s_i \rangle = -\frac{1}{N\beta} \langle \frac{\partial}{\partial B} \beta E(s) \rangle = \frac{1}{N\beta} \frac{\partial}{\partial B} \ln Z$$

Going back to the mean field approximation, we can compute the derivative and find that

$$\frac{\partial Z}{\partial B} = \exp(-\frac{1}{2} \beta JqNM^2) 2^N \cosh(\beta(JMq + B))^{N-1} N\beta \sinh(\beta(JMq + B))$$

which we can also write as

$$\exp(-\frac{1}{2} \beta JqNM^2) 2^N \cosh(\beta(JMq + B))^N N\beta \tanh(\beta(JMq + B))$$

so that we arrive at

$$\frac{\partial Z}{\partial B} = \tanh(\beta(JMq + B)) N\beta Z$$

i.e.

$$\langle m \rangle = \tanh(\beta(JMq + B))$$

Thus for the mean field approximation to hold, we need to impose the additional *consistency condition*

$$M = \tanh(\beta(JMq + B))$$

For the sake of completeness, let us also record an explicit expression for the Helmholtz energy:

$$F = -\frac{1}{\beta} \ln Z = \frac{1}{2} JqNM^2 - \frac{N}{\beta} \ln 2 \cosh(\beta(JMq + B))$$

and of course

$$M = \langle m \rangle = \frac{1}{N\beta} \frac{\partial}{\partial B} \ln Z = -\frac{1}{N} \frac{\partial F}{\partial B}$$

showing that M is, up to a constant, conjugate to B.

Let us now pause for a moment and reflect on what we have done. We have approximated the true behaviour of an Ising model by a thermodynamical system

which is described by the temperature, the average magnetization M and the external field B. We have then found an expression for the Helmholtz free energy of this system in terms of these quantities. We have seen that two of these quantities M and B are conjugate variables, similar to pressure and volume in system like the ideal gas. And, like for the ideal gas and other systems like the van der Waals model, we have found an additional relation - the consistency relation above - between two conjugated variables, specifically M and B.

For the van der Waals model, we have found that the relation between pressure and volume is multivalued - for a given pressure, we have found several possible values of the volume. Let us now analyze the consistency relation further to see whether the situation is similar. For that purpose, we need to determine the shape of the function

$$f(M) = M - \tanh(\beta(JMq + B))$$

and see how many zeroes is has for fixed B, i.e. how many values of M we find for a given B. This will tell us the possible magnetizations in equilibrium. The derivative is

$$\frac{df}{dM} = 1 - \frac{\beta Jq}{\cosh^2(\beta(JMq + B))}$$

which is zero if and only if

$$\cosh^2(\beta(JMq + B)) = \beta Jq$$

The cosh function takes all values between $+1$ and $+\infty$. Thus we immediately find that if $\beta Jq < 1$, or, equivalently,

$$T > \frac{Jq}{k_B} = T_c$$

the derivative has no zeros and is everywhere positive. As the range of the tanh term is between -1 and 1, we also know that f tends to $-\infty$ at $M = -\infty$ and to $+\infty$ at $M = +\infty$, so that there is exactly one zero.

This changes if the temperature is below the critical point T_c. Then the derivative has a zero, in fact is has two zeroes as the *cosh* function itself is already symmetric. As f has limits $\pm\infty$, we can conclude that there are in fact three zeroes, i.e. we might find several possible magnetizations M for a given value of the external field B. We again interpret this as a *phase transition*.

The situation is especially interesting if $B = 0$. Then our consistency condition reads

$$M = \tanh(\beta JMq) = \tanh(\frac{M}{\bar{T}})$$

where we have again introduced the reduced temperature

$$\bar{T} = \frac{T}{T_c} = \frac{Tk_B}{Jq} = \frac{1}{Jq\beta}$$

which is symmetric in M. Thus we find that below the critical temperature, there are three solutions for M: a magnetization of zero and two additional solutions with the same magnitude but different signs. These represent ferromagnetic states in the absence of an external field.

For finite B, our consistency condition is - again expressed in terms of the reduced temperature - given by

$$M = \tanh(\frac{1}{\bar{T}}(M + \frac{B}{Jq}))$$

Clearly we can only hope for solutions in the range $[-1, 1]$ as this is the range of the *tanh* function. For small values of B, this has again three solutions. However,

as B increases, two of the zeroes move closer together and eventually disappear, and for larger B, there will be only one zero and there is no phase transition any more. Diagram 17 displays the consistency condition as a function for $T = 0.8 T_c$ and $J = 1$ as a function of M, so that the zeroes correspond to the values of M that are consistent with the given field strength B.
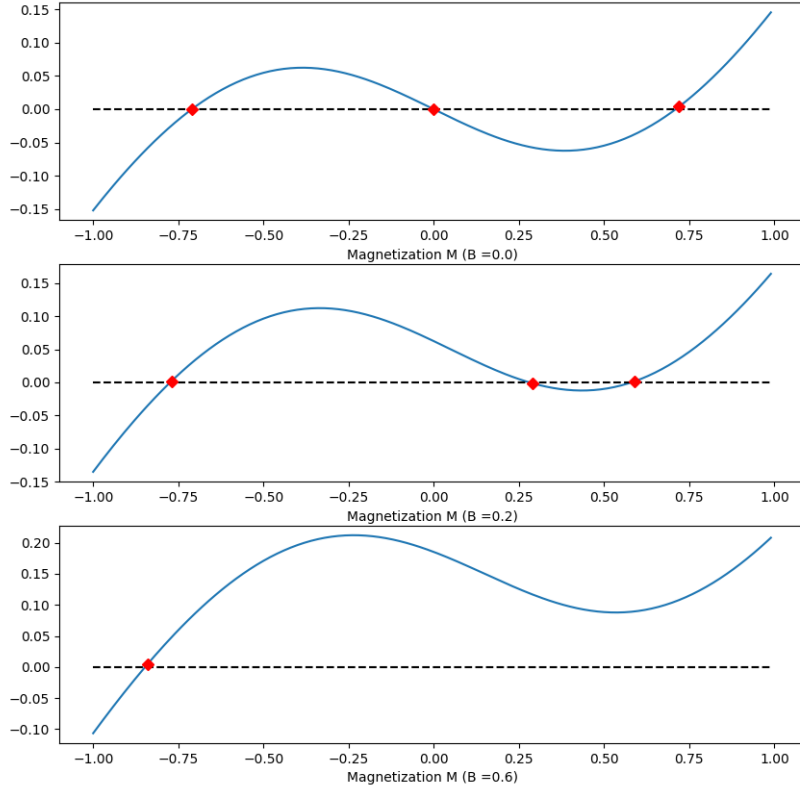

Figure 17: Graphical representation of the consistency condition

At this point, let us summarize the analogies with the van der Waals model that we have found so far. We describe the system in terms of temperature, magnetization and external field strength. For a given external field strength, we find up to three values of the magnetization, very similar to the different values of the volume that we found for a given pressure in the van der Waals model. In an experimental setup, we would control and slowly change the pressure respectively the external field and then observe a phase transition, expressed by a sudden jump in the volume respectively the magnetization. The first of these parameters is therefore called the *control parameter*, while the parameter that indicates a change in phase is called the *order parameter*. We can express the control parameter as a proper function of the order parameter, but the order parameter is a multi-valued function of the control parameter.

van der Waals model	Ising model
Volume	Magnetization M
Pressure	External field
Order parameter: V	Order parameter: M
Control parameter: P	Control parameter: B
$P = P(V)$	$B = B(M)$
$F = F(T,V)$	$F = F(T,M)$

18   More on mean field theory

In the previous section, we have made a heuristic argument to obtain an approximation for the true Helmholtz energy of an Ising model, and our argument that this model is only a good approximation if the self consistency condition holds contained a fair bit of handwaving. Let us now try to put this onto slightly more solid ground, following [4], section 20.

Suppose we are considering a system in which the energy E of a microstate is given by some complex relation between properties of the state and maybe external forces like a magnetic field. Our general approach to deriving thermodynamic properties would then be to sum up the partition function and derive the Helmholtz free energy from it, whose partial derivatives are the quantities of interest.

Of course, this is in general hopeless, but it might be easier for a simplified energy term E_0. In the case of the Ising model, for instance, the energy is given by an interaction term and a term depending only on one s_i. Without the interaction term, the energy splits and correspondingly we can obtain the partition function as a product, with one factor for each lattice point. In the presence of interaction terms, this does no longer work.

In this situation, we can try to apply perturbation theory to our problem. Thus we split our energy as

$$E = E_0 + E_1$$

into an "easy" term that we can hope to solve and a complex term. We then introduce a coupling parameter λ, i.e. we consider the energy

$$E(\lambda) = E_0 + \lambda E_1$$

and try to understand how our system changes if we vary λ. For $\lambda = 0$, we can obtain the partition function and an expression for the Helmholtz energy and we hope to learn something from this for $\lambda = 1$ as well.

Thus let us try to understand how the Helmholtz energy $F(\lambda)$ changes if we vary λ. By the general pattern of how derivatives of energy terms are related to derivatives of $\ln Z$, we have

$$\frac{\partial}{\partial \lambda} F = -\frac{1}{\beta} \frac{\partial}{\partial \lambda} \ln Z$$

$$= \frac{1}{\beta} \langle \frac{\partial}{\partial \lambda} (\beta E(\lambda)) \rangle = \langle E_1 \rangle$$

Note that this still depends on λ, as it should, as we take the average with respect to the Boltzmann distribution corresponding to the energy $E(\lambda)$. However, this is already encouraging - if the term E_1 is somehow small in the average, F changes only slightly with λ, so that we can hope that our approach works.

Let us now calculate the second derivative of F with respect to λ, which is a bit

more complicated. First, we use our previous result to write

$$
\begin{aligned}
\frac{\partial^2}{\partial \lambda^2} F &= \frac{\partial}{\partial \lambda} \langle E_1 \rangle \\
&= \frac{\partial}{\partial \lambda} \sum_s p(s) E_1(s) \\
&= \sum_s \frac{\partial}{\partial \lambda} \frac{1}{Z} \exp(-\beta E(\lambda, s)) E_1(s) \\
&= \sum_s E_1(s) \frac{\partial}{\partial \lambda} \frac{1}{Z} \exp(-\beta E(\lambda, s))
\end{aligned}
$$

When calculating the partial derivative within the sum, we need to be careful, as Z depends on the parameter λ as well. To be able to apply the product rule, let us first calculate

$$
\frac{\partial}{\partial \lambda} \frac{1}{Z} = -\frac{1}{Z^2} \frac{\partial Z}{\partial \lambda} = -\frac{1}{Z} \frac{\partial}{\partial \lambda} \ln Z = \frac{\beta}{Z} \langle E_1 \rangle
$$

We can now use the product rule and find that

$$
\begin{aligned}
\frac{\partial}{\partial \lambda} \frac{1}{Z} \exp(-\beta E(\lambda, s)) &= \exp(-\beta E(\lambda, s)) \frac{\partial}{\partial \lambda} \frac{1}{Z} + \frac{1}{Z} \frac{\partial}{\partial \lambda} \exp(-\beta E(\lambda, s)) \\
&= \beta \langle E_1 \rangle \frac{1}{Z} \exp(-\beta E(\lambda, s)) + p(s) \frac{\partial}{\partial \lambda} (-\beta E(\lambda, s)) \\
&= \beta \langle E_1 \rangle p(s) - \beta p(s) E_1 = \beta p(s) (\langle E_1 \rangle - E_1)
\end{aligned}
$$

Let us now plug this result into our formula for the second derivative. We find that

$$
\begin{aligned}
\frac{\partial^2}{\partial \lambda^2} F &= \sum_s E_1(s) \frac{\partial}{\partial \lambda} \frac{1}{Z} \exp(-\beta E(\lambda, s)) \\
&= \beta \sum_s E_1(s) p(s) (\langle E_1 \rangle - E_1) \\
&= \beta \langle E_1 \rangle \sum_s E_1(s) p(s) - \beta \sum_s E_1(s)^2 p(s) = \beta (\langle E_1 \rangle^2 - \langle E_1^2 \rangle) = -\beta \operatorname{var}(E_1)
\end{aligned}
$$

This is another interesting result - apparently our approximatio is most useful if E_1 is not only small in the mean value, but if its fluctuations are small. In any case, however, this tells us something about the sign of the second derivative. The variance of any probability distribution is a non-negative number, so that we immediately find that

$$
\frac{\partial^2}{\partial \lambda^2} F \leq 0
$$

This implies that, as a function of λ, F is concave, and therefore we can conclude that for all $\lambda \in [0, 1]$, the true value $F(\lambda)$ is equal to or below the tangent at $F(0)$, in other words

$$
F(\lambda) \leq F(0) + \lambda F'(0) = F_0 + \lambda \langle E_1 \rangle
$$

If we now let $\lambda = 1$, then $F(\lambda)$ is the true Helmholtz potential F we are interested in, and we obtain that

$$
F \leq F_0 + \langle E_1 \rangle
$$

This inequality - that is called the *Bogoliubov inequality* - looks trivial but is actually rather powerful. To understand why this is so useful, let us consider how this is typically used in practice.

The approximation scheme that is typically used is based on the observation that right hand side is the approximation of the true value of F by a tangent, i.e. this is a *first order approximation*, while F_0 itself is only an approximation to order zero. Thus you will typically choose not F_0 as an approximation for the true value of F, but the right hand side of the Bogoliubov inequality. This quantity is often called the *variational free energy* and denoted by F_v, so

$$F_v = F_0 + \langle E_1 \rangle$$

Typically, however, you have a choice when choosing your reference system F_0, and different choices of this model parameter will provide different values for F_0 and then of course $\langle E_1 \rangle$ and the approximation $F_0 + \langle E_1 \rangle$. To find out which one is the best approximation, you would like to compare them to the true value - but this is unfortunately not an option, as the true value is unknown. The Bogoliubov inequality, however, tells you that whatever choice you make, the true value will be than the variational free energy. Thus if you change your choice and find that the variational free energy gets smaller, it will be closer to the true value than before. Thus, you *choose the model system that makes the approximated Helmholtz potential as small as possible.*

In many cases, the choice of the reference model is expressed by some parameter or set of parameters Λ. Then $F_0 = F_0(\Lambda, X)$ where X stands for all other parameters, including the temperature, of your system, and consequently the variational free energy depends on these parameters as well. Then the Bogoliubov inequality tells you that in order to obtain the best approximation, you should first "fit" the model by minimizing F_v with respect to Λ, i.e. to solve

$$\frac{\partial}{\partial \Lambda} F_v(\Lambda, X) = 0$$

This will give you a preferred choice for Λ that can, however, depend on X. You can then use this relation to express Λ in terms of X and then obtain a variational free energy that depends on X alone and is the approximation for $F(X)$ you are looking for.

19 Applications in chemistry - the law of mass action

In this and the next section, we take a closer look at some applications of thermodynamical concepts to chemistry. In a typical chemical reaction, all thermodynamical properties that we have discussed so far tend to change. The number of particles of a given species changes (this is the very nature of a reaction) as reactants are converted into products of the reaction, and chemical energy is converted into other forms of energy. Many reactions involve a change in volume, for instance because a gas is produced or the densities of the involved reactants differ, and this change in volume in turn results in work being done on the environment. Often, excess energy is transferred to the environment in the form of heat. And finally, the entropy changes as well - if for instance two molecules which were initially separated are combined into a new molecule, the number of degrees of freedom and thus the entropy are reduced. We will see that a strategy to deal with this complexity is to fix certain thermodynamic variables like temperature or pressure and to use Legendre transforms to effectively focus on some of the changing quantities in isolation.

Our first application will be the derivation of the *law of mass actions* that describes the equilibrium state of chemical reactions. As a starting point, let us

take a closer look at the Gibbs energy that we have introduced earlier. Recall that the Gibbs energy is given by

$$G(T, P, N) = U - TS + PV$$

which is obtained from the internal U energy by applying two Legendre transformations, once with respect to the entropy and once with respect to the volume. Again, our general theory gives us an expression for the total differential of the Gibbs energy:
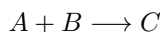
$$dG = -SdT + VdP + \mu dN$$

We could also derive this identity directly from the definition:

$$
\begin{aligned}
dG &= d(U - TS + PV) \\
&= dU - d(TS) + d(PV) \\
&= (TdS - PdV + \mu dN) - d(TS) + d(PV) \\
&= TdS - PdV + \mu dN - TdS - SdT + VdP + PdV \\
&= -SdT + VdP + \mu dN
\end{aligned}
$$

We have also already seen (see the discussion and derivation in the section on phase transitions) that the Gibbs free energy is minimized for systems that are in contact with a reservoir that keeps the system at constant pressure and constant temperature.

The Gibbs free energy is tailored to describe processes that take place at constant pressure and temperature, which is often the case for chemical reactions. As an application, we will now derive the so-called law of mass actions that governs the conditions under which a chemical reaction achieves equilibrium.

To get started, consider a simple chemical reactions involving three substances A, B, C of the form

$$A + B \longrightarrow C$$

With the obvious notation, the differential of the Gibbs free energy is then

$$dG = -SdT + VdP + \mu_a dN_A + \mu_B dN_B + \mu_C dN_C$$

In equilibrium, this needs to be zero. If the process takes place under the condition of constant pressure and temperature, this implies that for an infinitesimal change in particle numbers

$$0 = \mu_A \Delta N_A + \mu_B \Delta N_B + \mu_C \Delta N_B$$

Now an infinitesimal change in this sense is one step of the reaction which creates one molecule of type C while destroying one molecule of type A and one of type B so that

$$
\begin{aligned}
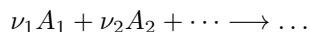\Delta N_A = \Delta N_B &= -1 \\
\Delta N_C &= 1
\end{aligned}
$$

With that, our condition becomes

$$0 = -\mu_A - \mu_B + \mu_C$$

or

$$\mu_C = \mu_A + \mu_B$$

In general, for a reaction of the form

$$\nu_1 A_1 + \nu_2 A_2 + \cdots \longrightarrow \ldots$$

the resulting equation would be

$$\sum_{reactants} \nu_i \mu_i = \sum_{products} \nu_i \mu_i$$

i.e. the sum of the chemical potentials counted with the respective stoichiometric multiplicities is the same for the reactants as well as for the products. Now the point of this equation is that the chemical potentials μ_i are (in general) functions of temperature, pressure and concentration (or partial pressure for gases). If we know these expressions, we can hope to obtain an equilibrium conditions that involves the concentrations so that we can determine the concentrations at equilibrium.

Before we do this, it is worth pointing out another consequence of our discussion. Suppose we start with a system that is not in equilibrium. If one infinitesimal step of our reaction takes place (for instance, the hypothetical step of converting one molecule of one of the reactants), the Gibbs energy of the system will change by some amount ΔG which - as we have seen above - can be calculated from the chemical potentials and the stoichiometric coefficients (and are tabulated for many chemical reactions). Now we know that the system strives to decrease the Gibbs energy on its way to equilibrium. Thus a negative value of ΔG is a necessary condition for a reaction to take place spontaneously.

Let us now try to derive (or at least guess) such a relation. This is not an exact derivation but somehow heuristic and motivated by statistical mechanics. We start by deriving an expression for the Gibbs energy of a system containing only one type of substance. We know that the Gibbs free energy is extensive with respect to the particle number N, i.e.

$$G(T, P, \lambda N) = \lambda G(T, P, N)$$

In particular

$$G(T, P, N) = G(T, P, N \cdot 1) = NG(T, P, 1)$$

If we use the notation

$$\mu^0(T, P) = G(T, P, 1)$$

we obtain

$$G(T, P, N) = G(T, P, N \cdot 1) = N\mu^0(T, P)$$

and also see that

$$\frac{\partial G}{\partial N} = \mu^0(T, P)$$

so that we recognize μ^0 as the chemical potential at temperature T and P. Let us now assume that we are considering a *dilute solution* of constant pressure and temperature, i.e. a substance B (the solute) dissolved in some substance A (the solvent). We assume that the concentration of B is so small that every B-molecule only interacts with the A-molecules that are direct neighbors, but never with another B-molecule.

Let us now suppose that we are given a pure solvent. By the discussion above, the Gibbs energy is then

$$G = N_A \mu^0(T, P)$$

If we now add a molecule of type B, then the Gibbs energy changes by an amount given by

$$dG = d(U - TS + PV)) = dU - TdS + PdV$$

In general, both the volume and the internal energy will increase by a certain amount. However, under the assumption that the newly added molecule does only interact with its immediate neighbors, these increases will not depend on the overall number N_A (which we assume to be so large that the actual number of immediate

neighbors of the newly added molecule does not depend on it). So these components will be functions of temperature and pressure. The increase of entropy, however, has a part that does depend on N_A, as we have a choice where to put the B molecule. Using the N_A molecules of type A as a coordinate system, we can decide to put the B molecule near any of those, so that we increase the number of possible microstates by a factor N_A. By our general formula for the Boltzmann entropy, we therefore have a contribution of the form $k_B \ln N_A$ in addition to any possible contributions depending only on temperature and pressure, so that we obtain

$$dG = f(T, P) - Tk_B \ln N_A$$

If we add two molecules, we can apply the same pattern of reasoning, with the additional observation that - as the molecules are indistinguishable - we increase the number of microstates by a factor $N_A^2/2$ and not N_A^2, so that

$$dG = 2f(T, P) - 2Tk_B \ln N_A + Tk_B \ln 2$$

In general, our expression will be

$$dG = N_B f(T, P) - N_B Tk_B \ln N_A + Tk_B \ln N_B!$$

which (using Stirlings formula) we can approximate to be

$$dG = N_B f(T, P) - N_B Tk_B \ln N_A + Tk_B N_B(\ln N_B - 1)$$

Thus we obtain the following expression for the Gibbs free energy:

$$G = N_A \mu^0(T, P) + N_B f(T, P) - N_B Tk_B \ln N_A + Tk_B N_B \ln N_B - Tk_B N_B$$

In textbooks on chemistry, it is common to express the Gibbs energy not in terms of absolute particle numbers, but in terms of molarities. The molarities of A is related to the particle number by

$$m_A = \frac{N_A}{\mathcal{N}_A}$$

where $\mathcal{N}_A$ is Avogrado's constant, and similarly

$$m_B = \frac{N_B}{\mathcal{N}_A}$$

Using this, observing that $\mathcal{N}_A k_B = R$, absorbing the additional factor $\mathcal{N}_A$ into the definition of f and collecting terms not depending on m_B, we find that

$$G = g(T, P, m_A) + m_B f(T, P) + m_B RT \ln \frac{m_B}{m_A} - m_B TR$$

Taking the partial derivative with respect to m_B, we therefore obtain

$$\mu_B(T, P, m_A, m_B) = f(T, P) + RT \ln \frac{m_B}{m_A}$$

with a new function g depending only on T and P. As $m_B << m_A$, the quotient is a constant times the *concentration* $c(B)$ of B in the solution. We therefore find that the chemical potential in a dilute solution is a term depending only on temperature and pressure plus a term depending on the concentration. Denoting the first term by $\mu_{0,B}(T, P)$ we arrive at

$$\mu_B(T, P, m_A, m_B) = \mu_{0,B}(T, P) + RT \ln c(B)$$

Note that this example also nicely demonstrates that the chemical potential μ_i can in general depend all particle numbers, even on N_i itself. Also note that this is not a contradiction to the fact that the Gibbs energy should be homogenous, as the potential depends on the relation of N_A and N_B, so that if we multiply both by the same scaling factor λ, the chemical potential remains unchanged (it is constant on rays or homogenous of degree 0, see also Theorem 22.1).

Let us now plug this equation for the chemical potential of a solute into our previous condition for chemical equilibrium to arrive at a condition for the equilibrium of a reaction taking place in a dilute solution. We find

$$\sum_{reactants} \nu_i \mu_{0,i}(T,P) + RT \sum_{reactants} \nu_i \ln c_i = \sum_{products} \nu_i \mu_{0,i}(T,P) + RT \sum_{products} \nu_i \ln c_i$$

Collecting terms, we can rewrite this as

$$RT \ln \frac{\prod_{products} c_i^{\nu_i}}{\prod_{reactants} c_i^{\nu_i}} = \sum_{reactants} \nu_i \mu_{0,i}(T,P) - \sum_{products} \nu_i \mu_{0,i}(T,P)$$

Note that the right hand side of this equation is a number that only depends on temperature and pressure and of course on the involved substances, but not on the concentrations. Dividing both sides by RT and exponentiating gives us the usual formulation of the *law of mass actions*: in equilibrium, the product of the concentrations of products divided by the product of the concentrations of the reactants is a constant

$$\frac{\prod_{products} c_i^{\nu_i}}{\prod_{reactants} c_i^{\nu_i}} = K$$

that depends only on temperature and pressure and on the reaction itself. To relate this constant K to Gibbs energies, let us define

$$\Delta_R G^0 = \sum_{products} \nu_i \mu_{0,i}(T,P) - \sum_{reactants} \nu_i \mu_{0,i}(T,P)$$

Then we find that

$$RT \ln \frac{\prod_{products} c_i^{\nu_i}}{\prod_{reactants} c_i^{\nu_i}} = -\Delta_r G^0$$

or

$$K = \exp(-\frac{\Delta_r G^0}{RT})$$

in aligment with section chapter 5 of [5].

How can we interpret the quantity $\Delta_r G^0$? To see this, let us consider a chemical reaction as above. When a reaction step takes place, all the molarities m_i change, but their changes are related by the stoichiometric coefficients. In fact, it is not difficult to see that if we know how the molarity of one reactant or product changes, we know all the other changes. Thus the state of the system moves on a one-dimensional hypersurface, i.e. a graph, as the reaction proceeds.

To parametrize this hypersurface, let us introduce a parameter ξ that is called the *extent* of the reaction (see for instance section 4.6 in [7]) and is related to the changes in molarity by

$$\Delta m_i = \nu_i \xi$$

for products and

$$\Delta m_i = -\nu_i \xi$$

for reactants. Put differently

$$\frac{d}{d\xi} m_i(\xi) = \pm \nu_i$$

Now the Gibbs energy at any point of the reaction depends on ξ as well, and we can use the chain rule to calculate the derivative:

$$\frac{d}{d\xi}G(\xi) = \sum_i \frac{\partial G}{\partial m_i}\frac{dm_i}{d\xi} = \sum_{products} \nu_i\mu_i - \sum_{reactants} \nu_i\mu_i$$

In particular, we find that at standard conditions, we have

$$\frac{d}{d\xi}G(\xi)|_{\xi=0} = \sum_{products} \nu_i\mu_i^0(P,T) - \sum_{reactants} \nu_i\mu_i^0(P,T)i = \Delta_r G^0$$

Thus the quantity $\Delta_r G^0$ is the rate of change of the Gibbs energy during a reaction at standard conditions. For a finite but small value of ξ, we can approximate the change in Gibbs energy by taking the Taylor expansion to the first order, i.e. as

$$\Delta G \approx \xi \cdot \Delta_r G^0$$

or

$$\Delta_r G^0 \approx \frac{\Delta G}{\xi}$$

As ξ has units of mole, we find that $\Delta_r G^0$ describes the *change in Gibbs free energy per mole at standard conditions*. Therefore this quantity is often called molar free Gibbs energy in chemistry books, and in fact the same name is often used for the chemical potentials. As, by definition, the molar free Gibbs energy is the weighted sum of the chemical potentials, it can be determined by looking up these values in a table like those in [9], and adding them up, taking stoichiometric coefficients into account.

For non-infinitesimal extents ξ, the above relation between the change in Gibbs free energy and the molar Gibbs energy is not exact. To see why, note that the second derivative of $G(\xi)$ with respect to ξ is given by

$$\frac{d^2}{d\xi^2}G(\xi) = \frac{d}{d\xi}\sum_{products}\nu_i\mu_i - \frac{d}{d\xi}\sum_{reactants}\nu_i\mu_i$$

$$= \sum_{products}\nu_i\frac{d\mu_i}{d\xi} - \sum_{reactants}\nu_i\frac{d\mu_i}{d\xi}$$

which is in general not zero, as the chemical potentials depend on the concentrations and therefore on ξ.

Let us try to estimate the error that we make by assuming a simple linear relation between ΔG and ξ. We know by the above relation between the chemical potential and the concentration that at standard conditions ($c = 1$), the derivative of the chemical potential with respect to the concentration is RT which is in the order of $2.5 kJ \cdot \text{mol}^{-1}$. Now the dependency of the concentration on ξ is simple, namely

$$c_i = 1 \pm \nu_i\xi$$

so that the derivatives of the potential with respect to ξ are of the same order of magnitude. For many reactions, however, the chemical potentials are in the order of a few hundred $kJ \cdot \text{mol}^{-1}$. Therefore the change of the chemical potential with a changing concentration can be neglected, and we can roughly determine the change in Gibbs free energy by simply multiplying the molar Gibbs energy $\Delta_r G^0$ with the extent.

Note, however, that we still used a little approximation in this derivation. We have calculated the chemical potential of a solute in a dilute solution and found

that it depends on the environment, i.e. on the amount of solvent present. We have then, however, applied this individually to each reactant and product of the reaction, ignoring that the chemical potential of one substance could be influenced by the presence of another substance. Thus we have ignored the interactions between the different solutes. This appears to be justified as these cross-terms can probably be neglected compared to the terms involving the solvent. If we wanted to be exact, we would have to use the *activities* a_i instead of the concentrations, which are defined by

$$\mu_i = \mu_{i,0} + RT \ln a_i$$

For an *ideal solution*, the activity can then be taken to be the concentration, but for a general solution, this might no longer be true and the *activity coefficient* which relates the activity to the concentration might be different from one.

     Let us close this section with a numerical example. We consider the autoprotolysis of water, i.e. the reaction

$$2H_2O \longrightarrow OH^- + H_3O^+$$

To be able to better look up the chemical potentials, let us write this as

$$H_2O \longrightarrow OH^- + H^+$$
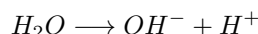
as the Gibbs energy of the remaining H_2O appears on both sides. Looking up the values for the individual Gibbs energies $\Delta_r G^0$ (i.e. the chemical potentials) in [9] gives us

$$\Delta_r G^0(H_2O) = -237.129\text{kJ/mol}^{-1}$$
$$\Delta_r G^0(OH^-) = -157.244\text{kJ/mol}^{-1}$$
$$\Delta_r G^0(H^+) = 0\text{kJ/mol}^{-1}$$

Therefore

$$\Delta_r G^0 = \Delta_r G^0(H^+) + \Delta_r G^0(OH^-) - \Delta_r G^0(H_2O)$$
$$= (-157.244 + 237.129)\text{kJ/mol}^{-1} = 79.885\text{kJ/mol}^{-1}$$

Thus the value of the constant K is

$$K = \exp(\frac{-\Delta_r G^0}{RT})$$
$$= \exp(-\frac{79.885\text{kJ/mol}^{-1}}{8.314\frac{J}{K} \times 298.15K}) = \exp(-32.227) \approx 1 \times 10^{-14}$$

Assuming that the concentration of water is essentially constant and equal to one, we find that

$$c(H_3O^+)c(OH^-) = 1 \times 10^{-14}$$

If this reaction is the only source of these ions, we also know that these two concentrations must be equal, so that

$$c(H_3O^+) = 1 \times 10^{-7}$$

Minus the logarithm of the concentration of oxonium H_3O^+ iones is by definition the pH-value, and we recover the well-known fact that the pH-value of pure water is 7.

20   Applications in chemistry - the reaction quotient

Let us now see how the Gibbs free energy comes into play if we consider reactions away from equilibrium. For that purpose, let us consider a reaction that takes place at standard temperature and pressure, but at non-standard concentrations c_i. As in the previous section, we can then describe the extent to which the reaction takes place by a parameter ξ, and, using the same arguments as above, we find that

$$\frac{dG}{d\xi} = \sum_{products} \nu_i \mu_i(T, P, c_i) - \sum_{reactants} \nu_i \mu_i(T, P, c_i)$$

Let us call this quantity $\Delta_r G$:

$$\Delta_r G = \sum_{products} \nu_i \mu_i(T, P, c_i) - \sum_{reactants} \nu_i \mu_i(T, P, c_i)$$

If the concentrations are the standard concentrations, i.e. $c_i = 1$, then this is the standard molar Gibbs free energy $\Delta_r G^0$. By the same reasoning as above, we also find that

$$\Delta G \approx \xi \Delta_r G$$

for small values of ξ, so this number again approximates the change in Gibbs free energy during a reaction per mole, but now at non-standard concentrations.

Let us now try to determine the value $\Delta_r G$ for general values of the concentrations. Again, we make use of the relation

$$\mu_i = \mu_i^0 + RT \ln c_i$$

Plugging this into our definition of $\Delta_r G$ and collecting terms, we find that

$$\Delta_r G = \Delta_r G^0 + \ln \frac{\prod_{products} c_i^{\nu_i}}{\prod_{reactants} c_i^{\nu_i}}$$

The fraction on the right hand side is called the *reaction quotient* and is usually denoted by Q. Thus we have shown that

$$\Delta_r G = \Delta_r G^0 + RT \ln Q$$

Again, the first term can be looked up or calculated easily from tabulated values. Thus this formula can be used to calculate the (approximate) change in the Gibbs energy during a reaction which does not take place at standard concenctrations.

It is interesting to relate this quantity (which is not a constant, but changes as the reaction proceeds) to the equilibrium constant K. The condition for equilibrium is

$$0 = \frac{dG}{d\xi} = \Delta_r G$$

or

$$RT \ln Q = -\Delta_r G^0$$

Dividing by RT and exponentiating on both sides yields

$$Q = \exp(\frac{-\Delta_r G^0}{Rt}) = K$$

Thus (not surprisingly) the reaction is in equilibrium if and only if the reaction quotient (a number which changes during the reaction) is equal to the equilibrium

constant K (a constant that only depends on the conditions and the reaction in question and can be computed from the molar Gibbs free energy).

As the Gibbs energy tends to a minimum, we also see that the reaction runs in the forward direction if and only if

$$\frac{dG}{d\xi} < 0$$

i.e. if and only if

$$RT \ln Q < -\Delta_r G^0$$

or

$$Q < K$$

and conversely, the reaction runs backwards if $Q > K$. Thus the reaction moves towards equilibrium by increasing or decreasing Q as needed.

21 Applications in chemistry - the reaction enthalpy

In previous sections, we have used the Helmholtz free energy and the Gibbs free energy, both of which happen to be Legendre transforms of the internal energy U. The third Legendre transform that we have already briefly touched on is the enthalpy.

To get a better intuition for what a *change* in enthalpy actually describes, let us consider a physical system in which a reaction takes place (think of a vessel filled with a dilute solution or with gas). Let us suppose that one wall of this system contains a movable piston that the system can adjust to transfer energy in the form of mechanical work into a pressure reservoir PR, which is otherwise closed and isolated from the environment. Suppose further that there is also a heat sink HS to which our system is connected and to which it can transfer heat, but which is again isolated towards the environment. We do not assume that this is a heat reservoir in the classical sense with unlimited heat capacity, but simply that it is connected to the primary system by the diathermic wall (thus we do only require constant pressure, not constant temperature).

To make this more tangible, let us envision a *coffee cup calorimeter* as described in chapter 7 of [7]. This calorimeter is simply a well insulated coffee cup, maybe out of styrofoam, and a cover with two holes in it. The cup contains water as a solvent, and one hole is used to insert a temperature sensor, while the other hole is used to insert a stirrer. Small amounts of the reactants of the reaction in question are dissolved in the water in the cup. Then the system is given some time to achieve equilibrium (implying that the temperature of the solvent and that of the solute will be the same). Next the stirrer is used to initiate the reaction. When the reaction takes place, the volume of the system typically changes, but as the cup is not fully airtight, this results in the change of volume of the water at constant pressure. Thus the water in the cup serves as both the pressure reservoir (in combination with the surrounding atmosphere) and heat sink, while the system under consideration is composed of reactants and products of the reaction that we want to study. The heat transferred into the surrounding water can be measured using the temperature sensor. Picture 18, (taken from [8]) displays a sketch of such a device.

Let us now calculate the changes in the inner energies of the various subsystems. Our primary system - the system in which the reaction takes place - undergoes a change in energy given by

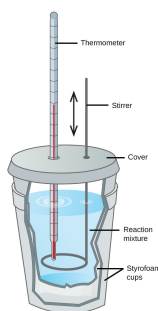$$dU = d(H - PV) = dH - d(PV) = dH - PdV$$

Figure 18: Coffee Cup Calorimeter

where we have used the definition of the enthalpy and the fact that the process takes place at constant pressure, so that $dP = 0$. Let now Q denote the amount of energy transferred as heat from our system into the heat sink. Then the inner energy of the heat sink increases by Q, i.e.

$$dU_{HS} = Q$$

as the heat sink is otherwise isolated so that no mechanical or chemical work is done on it. Similarly, the energy of the pressure reservoir increases by PdV

$$dU_{PR} = PdV$$

Thus the overall energy changes by

$$\begin{aligned} dU_{tot} &= dU + dU_{HS} + dU_{PR} \\ &= (dH - PdV) + Q + PdV = dH + Q \end{aligned}$$

However, as we assume that the overall system is isolated towards the environment, the total change of energy is zero. Thus

$$0 = dH + Q$$

or

$$Q = -dH$$

This shows that the *change in enthalpy is (minus) the heat released by a process at constant pressure.* If the change of enthalpy is negative, the amount of heat released is positive, so we have an *exothermic reaction.* If the change of enthalpy is positive, the reaction consumes heat and is *endothermic.*

One of the reasons why the enthalpy is so useful is that it is - by construction - a function on the configuration space. Thus the change in enthalpy depends only on the final and the initial state of the system (which is in general not true for heat as the heat transfer is given by the one-form TdS which is not exact and therefore depends on the process, i.e. on the path in configuration space representing the process). To see why this is useful in practice, let us consider an example. Say we wanted to calculate the change in enthalpy of the reaction

$$A \longrightarrow B$$
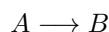
and suppose further that we already knew the enthalpy changes for the reactions

$$A \longrightarrow C$$

and

$$C \longrightarrow B$$

Now each of the reaction reactants and products that we have abbreviated with A, B, C are states of our system - a certain amount of molecules of the respective type dissolved in water (as long as we stick to some standard conditions, which are conventially chosen to be a pressure of 1 bar and a temperature of 298 K). Thus each of them has a certain value of the enthalpy, and the change in enthalpy is simply the difference of the enthalpy after and before the reaction took place. This implies, however, that we can simply calculate enthalpy changes of new reactions by adding up enthalpy changes of existing reactions in the obvious way, specifically

$$\Delta H_{A \longrightarrow B} = H_B - H_A$$
$$\Delta H_{A \longrightarrow C} = H_C - H_A$$
$$\Delta H_{C \longrightarrow B} = H_B - H_C$$

and therefore

$$\Delta H_{A \longrightarrow B} = \Delta H_{A \longrightarrow C} + \Delta H_{C \longrightarrow B}$$

This fact is known as Hess's law of enthalpies and allows us to calculate the changes of enthalpies for complex reactions by piercing together the enthalpy changes for simpler reactions that we might already have measured and tabulated.

Finally, let us relate the enthalpy with the Gibbs free energy that we have discussed in the previous section. By definition

$$G = U + pV - TS = H - TS$$

so that

$$dG = dH - d(TS) = dH - TdS$$

at constant temperature. Thus if the enthalpy change of a reaction is negative, i.e. the reaction is exothermic, and the entropy change of the reaction is positive, the reaction will take place spontaneously (reducing the enthalpy and increasing the entropy at the same time). If a reaction is endothermic and the entropy change is negative, the Gibbs energy is positive, so that the reaction will not take place. If, however, the enthalpy change is negative but the entropy change is negative as well, it depends on the temperature whether the reaction takes place spontaneously (note, however, that both enthalpy change and Gibbs energy depend on the temperature as well, so the condition for a negative Gibbs energy change is not as simple as our notation might indicate).

We close this section by discussing another relation between the Gibbs free energy and the enthalpy known as the van't Hoff equation. Recall that with the convention typically used in chemistry textbooks, the relation between the equilibrium constant and the Gibbs free energy change for the reaction at constant pressure and temperature of one mole of substance reads

$$K = \exp(-\frac{\Delta G}{RT})$$

or

$$\ln K = -\frac{\Delta G}{RT}$$

As $G = H - TS$, this is the same as

$$\ln K = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}$$

where ΔS is the entropy change of the reaction and ΔH is the change of enthalpy. Of course both quantities depend on the temperature at which the reaction takes

place. However, let us ignore this dependency for a moment and assume that these quantities are - for small changes in the temperature - independent of the temperature. Then we can take the derivative with respect to T and obtain

$$\frac{d}{dT}\ln K = \frac{\Delta H}{RT^2}$$

This equation that approximates how the equilibrium constant depends on the temperature is known as van't Hoff equation. Alternatively, we could write this as

$$\ln\frac{K(T_2)}{K(T_1)} = \frac{\Delta H}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)$$

Note that for an endothermic reaction, this predicts that when we increase the temperature, the equilibrium is moved to the right, and for an exothermic reaction the equilibrium is moved to the left.

The van't Hoff equation has an interesting graphical interpretation. Suppose that we could measure the equilibrium constant for different temperatures (close enough to the standard temperature that our approximation does not break down) and plot the natural logarithm of K versus the inverse temperature T^{-1}. Then the van't Hoff equation tells us that the result will be a line. The slope of this line multiplied by R is minus the enthalpy change of the reaction, while the point value at which it intersects the y-axis multiplied by R gives the entropy change of the reaction. This plot is known as *van't Hoff plot*, and can be used to determine enthalpy change and entropy change without using a calorimeter.

22 Appendix - the Gibbs-Duhem equation

One of the postulates of thermodynamics is that the entropy is an extensive quantity: when we double the size of a system (i.e. its volume, its number of particles and its internal energy) the entropy will double. Technically, this comes down to the requirements that

$$S(\lambda U, \lambda N, \lambda V) = \lambda S(U, N, V)$$

for all scaling factors λ. Functions with this property are called homogenous and it turns out that these functions have a very special form.

DEFINITION 22.1. *A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is called* homogenous *when the relation*

$$f(\lambda x) = \lambda f(x)$$

holds for all $\lambda \geq 0$ and all $x \in \mathbb{R}^n$.

THEOREM 22.1 (Eulers theorem on homogenous functions). *If $f\colon \mathbb{R}^n \to \mathbb{R}$ is smooth and homogenous, then for every point x, the relation*

$$f(x) = \sum_{i=1}^{n}(\partial_i f)(x)x_i$$

holds. Moreover, the partial derivatives $\partial_i f$ have the property that $(\partial_i f)(\lambda x) = (\partial_i f)(x)$, i.e. they are constant along rays.

Proof. Suppose we are given an arbitrary but fixed point x. Consider the function g defined on $\mathbb{R}$ by

$$g(\lambda) = f(\lambda x) - \lambda f(x)$$

As f is homogenous, this function is zero on $[0, \infty)$, so its derivative on $(0, \infty)$ is as well. On the other hand, we can use the chain rule to see that

$$0 = g'(\lambda) = \sum_i (\partial_i f)(\lambda x) x_i - f(x)$$

Setting $\lambda = 1$ yields our desired result. To prove our second statement, note that, by definition,

$$(\partial_i f)(\lambda x) = \frac{d}{dt}|_{t=0} f(\lambda x + t e_i)$$

which we can write as

$$\frac{d}{dt}|_{t=0} f(\lambda(x + u e_i)) = \lambda \frac{d}{dt}|_{t=0} f(x + u e_i)$$

with the substitution $u = \lambda^{-1} t$. As $dt = \lambda du$, the λ term cancels and we find that

$$(\partial_i f)(\lambda x) = \frac{d}{du}|_{t=0} f(x + u e_i) = (\partial_i f)(x)$$

as claimed. $\square$

Let us now apply this theorem to the entropy, which is an extensive quantity (under the assumption that our system is homogenous). We obtain

$$S(U, V, N) = \frac{\partial S}{\partial U} U + \frac{\partial S}{\partial V} V + \frac{\partial S}{\partial N} N$$

Using the identifications of the various partial derivatives with temperature, pressure and chemical potential, this turns into

$$S(U, V, N) = \frac{1}{T} U + \frac{1}{T} PV - \frac{1}{T} \mu N$$

Multiplying this by the temperature and rearranging yields

$$U(S, V, N) = TS - PV + \mu N$$

With more than one type of particles involved, this equation becomes

$$U(S, V, N) = TS - PV + \sum_i \mu_i N_i$$

Using this, we can now easily find an expression for the Gibbs free energy. As

$$G = U + PV - TS$$

we see that

$$G(T, P, N_i) = \sum_i \mu_i N_i$$

which is an amazingly simple expression (which, however, has some hidden complexities as the μ_i will in general depend on parameters, i.e. μ_i might even depend on N_j as the example of the formula for the chemical potential in a dilute solution nicely illustrates).

There is one case, however, where this relation is especially nice, which is the case of only one particle. In general, the chemical potential will then depend on T, P and N. Our result on the derivatives of homogeneous functions, however, implies that μ is homogeneous of degree zero in N. Thus

$$\mu(T, P, N) = \mu(T, P, 1)$$

i.e. μ does not depend on the particle number any more. The Gibbs energy is then simply

$$G(T, P, N) = \mu(T, P)N$$

and we see that

$$\mu(T, P) = G(T, P, 1) = \frac{G(T, P, N)}{N}$$

is the same as the specific (or molar, depending on units that we use) Gibbs energy.

To see yet another implication of the homogenous function theorem, let us take the derivative of our expression for U. We obtain

$$dU = d(TS - PV) + \sum_i d(\mu_i N_i)$$

$$= SdT + TdS - PdV - VdP + \sum_i \mu_i dN_i + \sum_i N_i d\mu_i$$

On the other hand, we know by the fundamental relation of thermodynamics that

$$dU = TdS - PdV + \sum_i \mu_i dN_i$$

Collecting the remaining terms yields

$$0 = SdT - VdP + \sum_i N_i d\mu_i$$

or, rearranged

$$-SdT + VdP = \sum_i N_i d\mu_i$$

This is the famous *Gibbs-Duhem equation* that states that the intrinsic thermodynamical variables are not fully independent. It implies in particular that for a quasi-static process at constant pressure and temperature, the derivatives of the chemical potentials are related by

$$0 = \sum_i N_i d\mu_i$$

References

[1] H. Bauer, *Wahrscheinlichkeitstheorie*, de Gruyter, Berlin, New York 1991

[2] A. Klenke, *Probability theory* Springer, London 2008

[3] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal *Vol. 27*, pp. 379–423, pp. 623–656, July, October 1948

[4] H.B. Callen, *Thermodynamics and an introduction to Thermostatistics*, Wiley, New York 1985

[5] D.V. Schroeder, *An introduction to thermal physics*, Addison-Wesley, San Francisco 2000

[6] Robert H. Sweden, *An Introduction to Statistical Mechanics and Thermodynamics*, Oxford University Press 2012

[7] R.H. Petrucci et al., *General chemistry - principles and modern applications*, 11th edition, Pearson, Toronto 2017

[8] OpenStax, *Chemistry 2e*, available at `https://openstax.org/details/books/chemistry-2e`

[9] National Bureau of Standards, *The NBS table of chemical thermodynamical properties*, Journal of Physical and Chemical Reference Data, Volume 11 (1982), Supplement No. 2

[10] M. Schwartz, *Statistical mechanics lecture notes*, available at `https://scholar.harvard.edu/schwartz/teaching`

[11] D. Tong, *Statistical physics - lecture notes*, available at `https://www.damtp.cam.ac.uk/user/tong/statphys.html`