

An introduction to Thermodynamics

December 6, 2024

Contents

0.1	Microstates and macroscopic quantities	1
0.2	Entropy	2
0.3	The energy representation	6
0.4	The principal of minimal internal energy	8
0.5	The maximal work theorem	11
0.6	Legendre transform and free energy	13
0.7	The van der Waals model	17
0.8	The microcanonical ensemble	22
0.9	The canonical ensemble	27
0.10	Ideal gas and blackbody radiation	33

0.1 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 to 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.

0.2 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 with the following property: the values assumed by the extensive parameters in the equilibrium state 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.

To see this principle 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. 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.

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.

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

More generally, given an arbitrary composite system, one usually chooses 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 $U = \text{const}$.

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 where 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

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

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

What happens to the entropy in the process of reaching equilibrium? Initially, the entropy is the sum $S^1 + S^2$ of the initial entropies of the individual systems. Our postulate states that after equilibrium has been reached, the entropy will be maximal among all possible states that are compatible with the still existing external

constraints. As the initial distribution of the energies is clearly compatible with these constraints, we can conclude that the new entropy is at least $S_1 + S_2$, i.e. the total entropy can change in the course of this process, but can never decrease. This is known as the *second law of thermodynamics*. Be careful, however, as this is only true for the total entropy - the entropy of an individual subsystem can actually decrease over time.

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

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

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.

$T = \frac{1}{\beta} = \left(\frac{\partial S}{\partial U}\right)^{-1}$ $P = -\frac{\partial U}{\partial V} = T \frac{\partial S}{\partial V}$ $\mu = \frac{\partial U}{\partial N} = -T \frac{\partial S}{\partial N}$

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

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

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

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

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 processes 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 0.2.1 (see 4.1 and 4.4 in [4]). Let us now make our example more concrete by assuming a specific relation between entropy and the extensive parameters which holds for both subsystems. Specifically, let us assume that the entropy is given by

$$S = \bar{S}(V, N) + NC \ln U$$

with a positive constant C , where \bar{S} is a function of volume and particle number alone (we will see in example 0.3.1) that this is true for an ideal gas). Then, by taking partial derivatives, we find that

$$\frac{1}{T} = \frac{\partial S}{\partial U} = CN \frac{\partial \ln U}{\partial U} = \frac{CN}{U}$$

so that we obtain

$$U = NCT$$

Note that this holds for both systems, and we also assume that the constant C is the same for both systems. 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. We want to compare the entropy before allowing heat to flow

and after we have removed the constraint and reached equilibrium again. The initial entropy is the sum of the entropies of both subsystems. The same is true for the final entropy. So the change of the total entropy is the sum of the changes of the individual entropies. In any case, the energy will be preserved, so that, if T_f is the final temperature in which the composite system has settled, we have

$$U = (N_1 + N_2)CT_f = N_1CT_1 + N_2CT_2$$

In other words, 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

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

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.

0.3 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

$$dU = TdS - PdV + \mu dN$$

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*.

Note that the entropy does of course depend on the functional relations between U , V and N . These relations can either be derived from statistical considerations, which we will do in a later chapter, or can be assumed to be given, based on empirical considerations. To illustrate this, let us consider an example.

Example 0.3.1. An *ideal gas* is a thermodynamical system which is given by the following fundamental relations:

$$PV = NRT$$

$$U = cNRT$$

Here R is a constant called the *gas constant* and c is another constant called the *heat capacity* of the gas. 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

$$S(U, V, N) = NS\left(\frac{U}{N}, \frac{V}{N}, 1\right) = s\left(\frac{U}{N}, \frac{V}{N}\right)$$

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

$$ds = cR d \ln U - R d \ln V$$

This allows us to find an expression for s , namely

$$s = \text{const} + cR \ln U - R \ln V$$

and by multiplying by N again, we find that

$$S = N \cdot \text{const} + cRN \ln U - RN \ln V$$

If we want, we can now define a reference macrostate and express the unknown constant in terms of the entropy of this reference state. 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. We note that the fact that the chemical potential can be eliminated from the consideration is formally expressed in the *Gibbs-Duhem equation*, see [4], section 3.2).

0.4 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, \dots, 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

$$U \times \text{id}: (s, \{x_i\}) \mapsto (U(s, \{x_i\}), x_i)$$

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. Moreover, 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

$$0 = \dot{s}(t) = \langle dS, (\dot{u}(t), \{\dot{x}_i(t)\}) \rangle \quad (0.1)$$

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 (0.1) 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

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

Applying the chain rule once more, we have that

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

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.*

It is instructive to plot a typical entropy and visualize how this works ¹. In figure 1, we have displayed the entropy of the composite system considered in example 0.2.1. 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 X of the first subsystem. The blue solid lines represent values of the entropy for four different, but fixed values of the energy. The red 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.

¹This is in fact a Python generated version of a by now classical and often copied illustration in [4], which has the advantage that in an interactive Python, one can rotate the surface and zoom into it.

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.

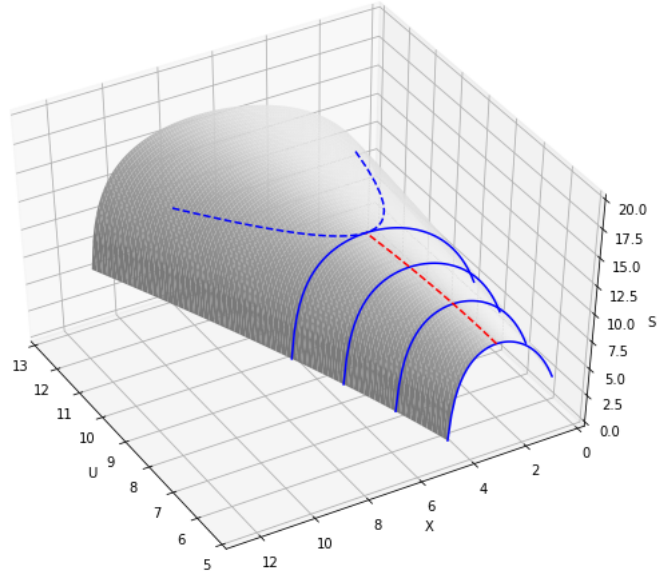


Figure 1: Plot of an entropy function

To be able to work directly with a minimum principle in the energy representation often has advantages. Suppose, for instance, we study systems that are not completely isolated, but permit the exchange of energy with the environment in the form of mechanical work. Then the energy will not be constant, whereas the entropy will be preserved, as no heat flows into the system or out of the system. It is then much more natural to work with a constraint reflecting this, i.e. to consider states with a fixed entropy, and the energy minimum principle then tells us how the system moves through the state space as other constraints are removed.

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.

0.5 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) 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 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* 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.

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 cylinder. 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.

0.6 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

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

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

$$p = f'(x) \tag{0.2}$$

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

$$\begin{aligned} \frac{df^*}{dp} &= \frac{d}{dp}(px(p)) - \frac{d}{dp}f(x(p)) \\ &= p \frac{dx}{dp} + x(p) - f'(x(p)) \frac{dx}{dp} = x \end{aligned}$$

where we have used equation 0.2 in the last line to see that the first and third term cancel. Thus we obtain the fundamental relations

$$\begin{aligned} p &= \frac{df}{dx} \\ x &= \frac{df^*}{dp} \end{aligned}$$

Example 0.6.1. 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

$$\begin{aligned} f'(x_1(p)) &= p \\ g'(x_2(p)) &= p \end{aligned}$$

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

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

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

$$df = p dx + v dy$$

then

$$df^* = x dp - v dy$$

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

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

denote the Legendre transform with respect to the coordinate x_i as

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

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

$$F(T, V, N) = -U[S] = U - TS \quad (0.3)$$

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

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

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 . Consider some process which is quasi-static and in particular keeps the temperature of our system constant and equal to the temperature of the reservoir. Differentiating equation 0.3, interpreted as defining a function of (U, V, N) , and restricting to tangent vectors along the hypersurfaces of constant temperature so that $dT = 0$, we obtain

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

Now consider the total entropy of the composite system that consists of the system of interest and the thermal reservoir. As the volume and particle number of the thermal reservoir are constant, we can calculate the total differential of the entropy of the reservoir and obtain

$$dS_R = \frac{1}{T} dU_R = -\frac{1}{T} dU + \frac{1}{T} dU_{tot}$$

so that the differential of the total entropy is

$$\begin{aligned} dS_{tot} &= dS - \frac{1}{T} dU + \frac{1}{T} dU_{tot} \\ &= \frac{1}{T} (PdV - \mu dN) + \frac{1}{T} dU_{tot} \\ &= -\frac{1}{T} dF + \frac{1}{T} dU_{tot} \end{aligned}$$

or

$$TdS_{tot} = -dF + dU_{tot}$$

Let us now suppose that we let the composite system somehow interact with the environment. If this interaction is reversible, so that the total entropy of the system S_{tot} is constant, the left hand side of this equation is zero. This implies that a state minimizes the total energy if and only if it minimizes the Helmholtz potential of the primary subsystem. Thus the results from section 0.4 immediately yield 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*.

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 go 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 and given by

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

These potentials have properties similar to the Helmholtz energy, for instance we can deduce in a similar way that the Gibbs free energy is minimized in equilibrium if a process is constrained to a fixed temperature and pressure - we will actually demonstrate an application of this in the next section.

0.7 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

$$F = F(T, N, V) = -Nk_B T \left[\ln \left(\frac{V - bN}{N} \right) + \frac{3}{2} \ln(k_B T) \right] - a \frac{N^2}{V}$$

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

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

Using the relation

$$U = TS + F$$

we can therefore find an expression for U as

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

and for the Gibbs free energy

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

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

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

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

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

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

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

That can easily be computed. We have

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

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

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

and

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

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

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

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

$$\begin{aligned}\bar{T} &= \frac{T}{T_c} \\ \bar{V} &= \frac{V}{V_c} \\ \bar{P} &= \frac{P}{P_c}\end{aligned}$$

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).

In figure 2, we have displayed some of the relations between pressure, temperature and volume.

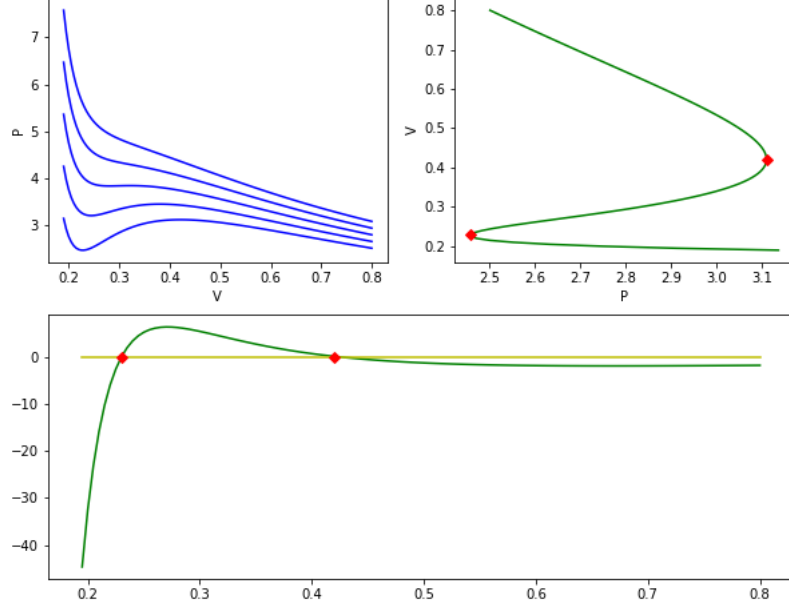


Figure 2: P-V diagrams for a Van der Waals fluid ($a=1.05$, $N=1$, $b=0.1$, $k=1$)

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 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.

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.

We can then model the part of the vessel outside this small subsystem as a heat reservoir that keeps the temperature T stable and, at the same time, as a pressure reservoir, so that the pressure P is constant as well. The subsystem itself could again be a composite system and therefore be described by its total energy, volume and other, internal parameters. If we use the index R to indicate that a quantity refers to the reservoir and quantity without subscript to refer to our small subsystem, we then have

$$\begin{aligned} dU_{tot} &= dU + dU_R \\ &= dU - PdV_R + TdS_R \\ &= dU + PdV - TdS + TdS_{tot} \\ &= dU + d(PV) - d(TS) + TdS_{tot} = dG + TdS_{tot} \end{aligned}$$

where we have used that the number of particles within the subsystem and therefore within the reservoir is constant to get from the first line to the second line, the fact that the total volume is constant to get from there to the third line and the fact that temperature and pressure are constant in the last line.

When we now assume in addition that the vessel is isolated so that no heat can enter or leave and therefore S_{tot} is constant, we find that

$$dG = dU_{tot}$$

But according to the principle of minimal energy, the total energy is minimized in the equilibrium, and therefore we can conclude (for instance by integrating this identity of differentials along a small curve representing a variation) 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 subject to the condition of constant total entropy - this is the equivalent of the respective condition for the Helmholtz energy if merely the temperature is kept stable by a heat reservoir.

Armed with this information, let us now get back to the van der Waals equation and study in more detail the phase transition along an isotherm. Using the dimensionless variables introduced earlier, we can express the normalized Gibbs energy as

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

with a function depending on T only. The part depending on T and the corresponding P-V curve for a fixed temperature $T = 0.85 * T_c$ are plotted in figure 3.

When a state moves along the isotherm starting below H , it will, at some point, experience a phase transition. It will then cross the instable region and appear again at the left hand side of the diagram. It could do so at point H and reappear at point B, or it could transition into the second phase upon reaching point G and reappear at point A. 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 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.

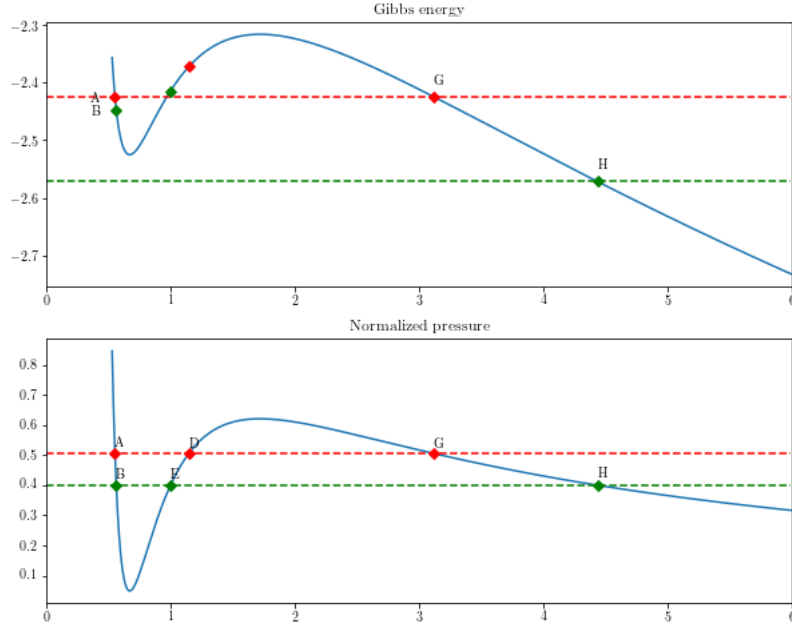


Figure 3: Gibbs energy along a van der Waals isotherm

In each of the phases, the Gibbs energy is proportional to $g(T, V_i)N_i$ with g as above. As the Gibbs energy is additive in the particle number (and pressure and temperature are the same for both subsystems), we know that

$$G = g(T, V_1)N_1 + g(T, V_2)N_2 = g(T, V_1)N_1 + g(T, V_2)(N - N_1)$$

where N is the number of particles in our subsystems. Now this is a minimum if and only if

$$0 = \frac{\partial G}{\partial N_1} = g(T, V_1) - g(T, V_2)$$

i.e. if the Gibbs energies at both sides of the unstable region are identical. If the Gibbs energies are different, say if $g(T, V_1) > g(T, V_2)$, 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 Gibbs energies are equal.

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, at least not in a quasi-stable process. 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 function 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. 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 4 (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*.

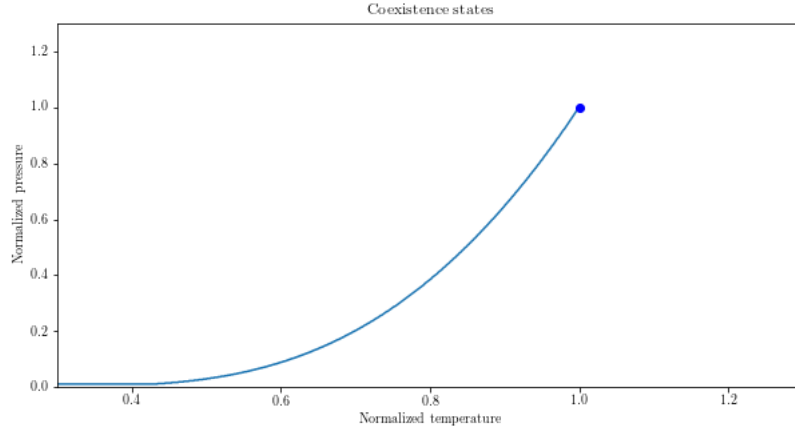


Figure 4: Phase diagram for a van der Waals system

0.8 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

$$E: \mathcal{S} \rightarrow \mathbb{R}$$

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 0.8.1 (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

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

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 0.8.2 (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

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

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 conventionally 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 0.8.1. *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.

Let us try to relate this 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

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

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 0.8.3 (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.

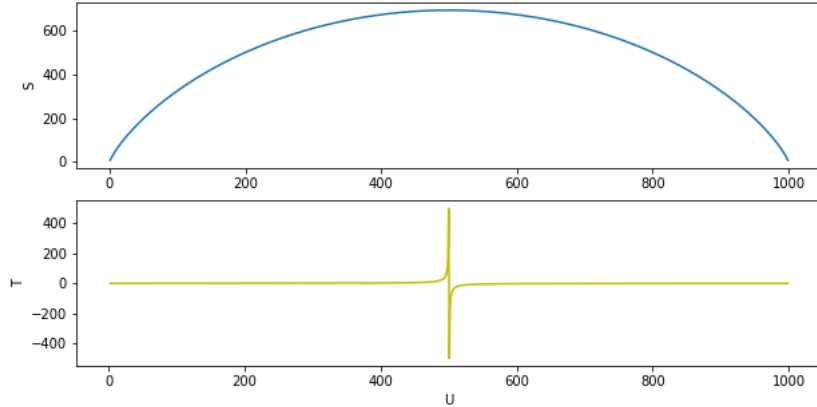


Figure 5: Entropy and temperature of a two-state system

In diagram 5, 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 is 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 $1, 2, \dots$. 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 $d, 2d, \dots$. 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

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

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 definition 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 0.8.4 (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:

$$S(N, q) = k_B(q + N) \ln\left(1 + \frac{q}{N}\right) - k_B q \ln \frac{q}{N}$$

Taking the derivative with respect to q , we find that

$$\frac{\hbar\omega}{T} = \frac{\partial S}{\partial q} = k_B \ln\left(1 + \frac{q}{N}\right) - k_B \ln \frac{q}{N} = k_B \ln\left(1 + \frac{N}{q}\right)$$

so that we can rewrite our expression for S to obtain

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

We see in particular that for large values of q , 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 $\hbar\omega = 1$.

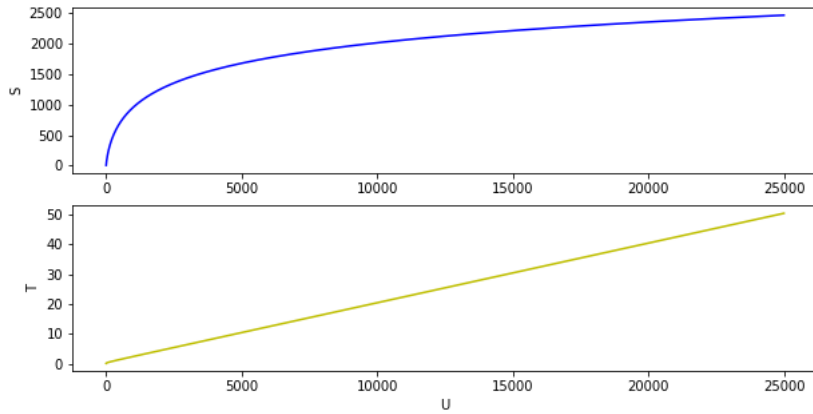


Figure 6: Entropy and temperature of an Einstein solid with $N = 500$

0.9 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.

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 , volume V_R and 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 U .

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. For simplicity, let us assume that the particle number and the volume of our subsystem are fixed as well.

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

$$p(s) = \frac{\Omega_R(U_{tot} - E(s))}{\Omega_{tot}} \quad (0.4)$$

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

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

and the denominator is

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

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 terms can be related to U . For the denominator, we can write, using the additivity of the entropy

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

For the numerator, we will do a Taylor expansion of the entropy of the reservoir around the point U_R . The first derivative is

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

The second derivative is related to the heat capacity of the reservoir at constant volume:

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

Now, the characteristic property of a reservoir is that we can add heat without significantly raising its temperature. Thus the heat capacity of a reservoir is essentially infinite, and therefore the inverse of the heat capacity is essentially zero. Therefore we should be able to obtain a very good approximation if we ignore the second derivative and all higher derivatives. As we assume that the volume and particle number are constant, we therefore obtain the expansion

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

so that

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

Let us now introduce the abbreviation

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

Then we obtain by putting all of this together that

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

Now formally this works for every choice of U , but we have chosen U to be the average, i.e. the macroscopically observed energy of our primary system. Therefore the quantity $U - T S$ is nothing but the Helmholtz potential of the primary system, and we obtain the fundamental equation

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

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

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

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 expectation is given by

$$\langle E \rangle = \sum_s E(s) p(s)$$

Using our expression for $p(s)$ we find that

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

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

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

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,

$$\begin{aligned}
 P(E) &= \sum_{s|U(s)=E} p(s) \\
 &= \frac{1}{Z} \sum_{s|U(s)=E} e^{-\beta E} \\
 &= \frac{1}{Z} \Omega(E) e^{-\beta E} \\
 &= \frac{1}{Z} e^{\frac{1}{k_B} S(E)} e^{-\beta E} = \frac{1}{Z} e^{\beta(TS(E)-E)}
 \end{aligned}$$

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 .

Finally, let us differentiate the relation between energy and partition function

once more to find the heat capacity. We have

$$\begin{aligned}\frac{\partial U}{\partial \beta} &= \frac{\partial}{\partial \beta} \frac{1}{Z} \sum_s E(s) e^{-\beta E(s)} \\ &= -\frac{1}{Z} \frac{\partial \ln Z}{\partial \beta} \sum_s E(s) e^{-\beta E(s)} - \frac{1}{Z} \sum_s E(s)^2 e^{-\beta E(s)} \\ &= \langle E \rangle^2 - \langle E^2 \rangle = -\text{var}(E)\end{aligned}$$

from which we obtain that

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

One additional remark on that formula is in order. If we use this formula with $U = E$, the temperature of the system is in fact equal to T as this is the equilibrium state, 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.

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

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

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.

$$\begin{aligned}p(s) &= \frac{1}{Z} e^{-\beta E(s)} \\ Z &= \sum_s e^{-\beta E(s)} = e^{-\beta F} \\ F &= -k_B T \ln Z \\ S &= k_B \frac{\partial}{\partial T} T \ln Z \\ \langle E \rangle &= -\frac{\partial}{\partial \beta} \ln Z \\ \text{var}(E) &= -\frac{\partial U}{\partial \beta} = k_B T^2 C_V\end{aligned}$$

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.

Example 0.9.1 (Einstein solid). Let us 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 $N_B \gg N_A$, 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 us now look at a specific state s of the smaller system with energy $q = q(s)$. Taking the logarithm of equation (0.4), we see that

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

where q_t is the total energy of the system. Using the expression for the entropy of an Einstein solid in terms of temperature and energy that we have derived before, we find that

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

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 first term is $-\beta q(s)$. 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

$$\ln p(s) \approx -\beta T + \text{const.}$$

and therefore

$$p(s) \simeq e^{-\beta q(s)}$$

as expected. This approximation does in fact work reasonably well also for comparatively low numbers of N_A and N_B . In diagram 7, 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).

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 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 0.4, 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

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

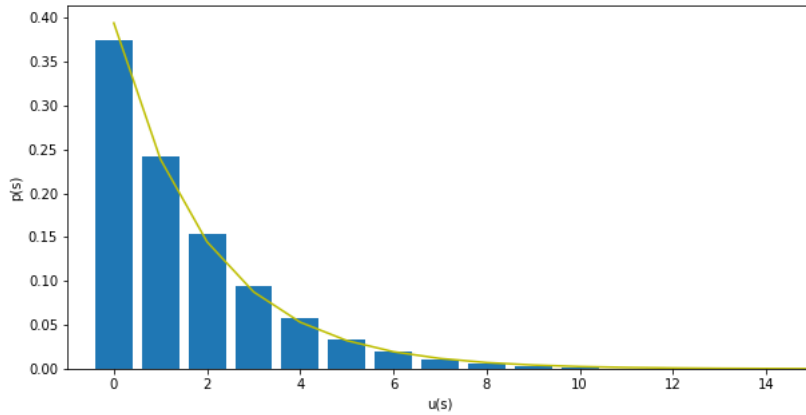


Figure 7: Boltzmann distribution for an Einstein solid

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.

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.

The second remark is that in the last example, we have applied the canonical formalism to a system that consists of one particle only. This appears to be a contradiction, as that system by itself is not a thermodynamic system that is sufficiently large to have a well defined externally observable macroscopic state. However, there are good justifications why the canonical formalism can nevertheless be applied, based on the idea of an ensemble and the fact that the fundamental quantities are additive - see [4], section 16.5.

0.10 Ideal gas and blackbody radiation

We close these notes on thermodynamics with 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

$$\frac{p^2}{2m}$$

²The 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

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

$$\begin{aligned} \langle E \rangle &= -\frac{\partial}{\partial \beta} \ln Z \\ &= -\frac{\partial}{\partial \beta} (-3N \ln \lambda) \\ &= 3N \frac{1}{\lambda} \frac{\partial}{\partial \beta} \lambda(T) \end{aligned}$$

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

$$\begin{aligned} S &= k_B \frac{\partial}{\partial T} T \ln Z \\ &= k_B \ln Z + k_B T \frac{\partial}{\partial T} \ln Z = k_B \ln Z + \frac{3}{2} N k_B \end{aligned}$$

To simplify the first term, we will have to use the Stirling approximation. Taking the logarithm, we find that

$$\begin{aligned} \ln Z &= -\ln N! + N \ln V - 3N \ln \lambda \\ &= -\ln N! + N \ln \frac{V}{\lambda^3} \\ &= -N \ln N + N + N \ln \frac{V}{\lambda^3} = N + N \ln \frac{V}{N \lambda^3} \end{aligned}$$

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 by 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

$$\ln \frac{V}{N\lambda^3} = \ln \left[\left(\frac{V}{N} \right) \left(\frac{U}{N} \right)^{\frac{3}{2}} \right] + \frac{3}{2} \ln \frac{4\pi m}{3\hbar^2}$$

from which obtain the formula

$$S(U, V, N) = N k_B \ln \left[\left(\frac{V}{N} \right) \left(\frac{U}{N} \right)^{\frac{3}{2}} \right] + \frac{5}{2} N k_B + \frac{3}{2} N k_B \ln \frac{4\pi m}{3\hbar^2}$$

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 characteristic 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

$$k = \frac{\pi}{L} n \tag{0.5}$$

where

$$n = (n_x, n_y, n_z) \in \mathbb{N}^3$$

and angular frequency $\omega(k)$, subject to the condition

$$\omega(k)^2 = k^2 c^2 \tag{0.6}$$

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 (0.6) together with the requirement (0.5) 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 frequency 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

$$E_n = \hbar\omega \quad (0.7)$$

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 :

$$\begin{aligned} \langle E \rangle &= -\frac{\partial}{\partial\beta} \ln Z \\ &= -\frac{\partial}{\partial\beta} \frac{1}{1 - e^{-\hbar\omega\beta}} \\ &= \frac{1}{1 - e^{-\hbar\omega\beta}} \frac{\partial}{\partial\beta} (1 - e^{-\hbar\omega\beta}) = \frac{\hbar\omega}{e^{\hbar\omega\beta} - 1} \end{aligned}$$

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 $\hbar \rightarrow 0$. 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

$$\lim_{\hbar \rightarrow 0} \langle E \rangle = \beta^{-1} = k_B T$$

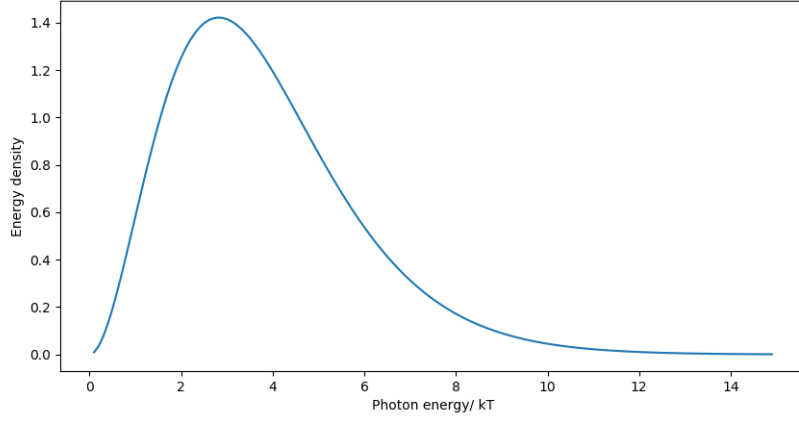


Figure 8: Planck spectrum for a black body

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 led M. Planck in 1900 to the postulate that the energy of a mode of the radiation field is quantized according to equation (0.7) which can be considered the birth of quantum mechanics.

Bibliography

- [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