4

The Classical Ideal Gas: Configurational Entropy

The single, all-encompassing problem of thermodynamics is the determination of the equilibrium state that eventually results after the removal of internal constraints in a closed, composite system.

Herbert B. Callen

This chapter begins the derivation of the entropy of the classical ideal gas, as outlined in Chapter 2. Our goal is to derive the entropy function for each member of a large set of isolated thermodynamic systems, with the total entropy given by the sum over the individual entropies. The essential property of these entropy functions is that when two or more systems are allowed to exchange particles, the sum of the total entropy is a maximum at the equilibrium value of the quantity which can be exchanged. We will use this property as a basis for the definition of entropy, following the ideas of Boltzmann. The exchange of energy or volume will be considered in Chapter 6.

The first step is to separate the calculation of the entropy into two parts: One for the contributions of the positions of the particles, and one for the contributions of their momenta. In an ideal gas, these are assumed to be independent, although this is not true for all interacting particles. As we will see, the total entropy is then just the sum of the two contributions.

The contributions of the probability distribution for the positions of the particles, which we will call the configurational entropy, will be calculated in this chapter. After probabilities are generalized to continuous variables in Chapter 5, the contributions of the momenta to the entropy will be given in Chapter 6.

4.1 Separation of Entropy into Two Parts

The assumption that the positions and momenta of the particles in an ideal gas are independent allows us to consider each separately.

As we saw in Section 3.3, the independence of the positions and momenta means that their joint probability can be expressed as a product of functions of q and p alone,

$$P(q,p) = P_q(q)P_p(p). \tag{4.1}$$

According to Boltzmann's 1877 definition, the entropy *of a composite system* is proportional to the logarithm of the probability of the extensive variables to within additive and multiplicative constants. Since eq. (4.1) shows that the probability distribution in phase space can be expressed as a product, the total entropy will be expressed as a sum of the contributions of the positions and the momenta.

The probability distribution in configuration space, $P_q(q)$, depends only on the volume V and the number of particles, N. Consequently, the configurational entropy, S_q, depends only on V and N; that is, $S_q = S_q(V,N)$.

The probability distribution in momentum space, $P_p(p)$, depends only on the total energy, E, and the number of particles, N. Consequently, the energy-dependent contribution to the entropy from the momenta, S_p, depends only on E and N; that is, $S_p = S_p(E,N)$.

The total entropy of the ideal gas is given by the sum of the configurational and the energy-dependent terms:

$$S_{total}(E,V,N) = S_q(V,N) + S_p(E,N). \tag{4.2}$$

The thermodynamic quantities E, V, and N are referred to as 'extensive' parameters (or observables, or variables) because they measure the amount, or extent, of something. They are to be contrasted with 'intensive' parameters, such as temperature or pressure (to be defined later), which do not automatically become bigger for bigger systems.

4.2   Probability Distribution of Particles

Before presenting the solution for a large number of systems in Section 4.8, we will first restrict the discussion to the distribution of particles between two systems, and see how this leads to an expression for the configurational entropy.

Let us consider a composite system consisting of two boxes (or subsystems) containing a total of $N_{T,jk}$ distinguishable, non-interacting particles. We will label the boxes j and k, with volumes V_j and V_k. The total volume of the two boxes is $V_{T,jk} = V_j + V_k$. The number of particles in subsystem j is N_j, with $N_k = N_{T,jk} - N_j$ being the number of particles in subsystem k.

We can either constrain the number of particles in each box to be fixed, or allow the numbers to fluctuate by removing the wall that separates the boxes. The total number of particles $N_{T,jk}$ is constant in either case. We will first consider the probability distribution for the number of particles in each subsystem when the constraint (impenetrable wall) is released (by removing the wall or making a hole in it). We will consider the case of isolated systems in Section 4.3.

In keeping with our intention of making the simplest reasonable assumptions about the probability distributions of the configurations, we will assume that the positions of the particles are not only independent of the momenta, but are also mutually independent of each other. The probability density $P_q(q)$ can then be written as a product,

$$P_q(q) = P_{N_{T,jk}}(\{\vec{r}_i\}) = \prod_{i=1}^{N_{T,jk}} P_i(\vec{r}_i), \tag{4.3}$$

where $P_i(\vec{r}_i)$ is the probability distribution of the single particle i. If we further assume that a given particle is equally likely to be anywhere in the composite system, the probability of it being in subsystem j is $V_j/V_{T,jk}$.[1]

Remember that this is an assumption that could, in principle, be tested with repeated experiments. It is not necessarily true. However, we are strongly prejudiced in this matter. If we were to carry out repeated experiments and find that a particular particle was almost always in subsystem j, we would probably conclude that there is something about the system that breaks the symmetry.

The assumption that everything we do not know is equally probable—subject to the constraints—is the simplest assumption we can make. It is the starting point for all statistical mechanics. Fortunately, it provides extremely good predictions.

If the $N_{T,jk}$ particles are free to go back and forth between the two subsystems, the probability distribution of N_j is given by the binomial distribution, eq. (3.47), as discussed in Section 3.8:

$$
\begin{aligned}
P(N_j|N_{T,jk}) &= \frac{N_{T,jk}!}{N_j!(N_{T,jk}-N_j)!} \left(\frac{V_j}{V_{T,jk}}\right)^{N_j} \left(1-\left(\frac{V_j}{V_{T,jk}}\right)\right)^{N_{T,jk}-N_j} \\
&= \binom{N_{T,jk}}{N_j} \left(\frac{V_j}{V_{T,jk}}\right)^{N_j} \left(1-\left(\frac{V_j}{V_{T,jk}}\right)\right)^{N_{T,jk}-N_j}.
\end{aligned}
\tag{4.4}
$$

To emphasize the equal standing of the two subsystems, it is often useful to write this equation in the symmetric form

$$P(N_j,N_k) = \frac{N_{T,jk}!}{N_j!N_k!} \left(\frac{V_j}{V_{T,jk}}\right)^{N_j} \left(\frac{V_k}{V_{T,jk}}\right)^{N_k}, \tag{4.5}$$

with the implicit constraint that $N_j + N_k = N_{T,jk}$.

4.3 Distribution of Particles between Two Isolated Systems that were Previously in Equilibrium

If the two subsystems are first in equilibrium with each other and then separated by inserting a partition or closing the hole, the probability of finding N_j particles in system j and N_k particles in system k is exactly the same as if the subsystems remained in

---

[1] This can be shown more formally after the discussion of continuous random variables in Section 5.2 of Chapter 5.

equilibrium with each other. It is still given by eq. (4.5). This probability is defined for all values of N_j and $N_k = N_{T,jk} - N_j$ in two subsystems that are isolated from each other.

Therefore, eq. (4.4) gives the probabilities of the variables N_j and N_k, even if the two systems, j and k, are isolated from each other. Since the initial state of the composite system in this example is specified by N_j and N_k, eq. (4.4) gives the probability of the initial state that Boltzmann was referring to in his 1877 paper, which was quoted at the beginning of Chapter 3.

4.4   Consequences of the Binomial Distribution

As shown in Section 3.7, the average value of N_j from the binomial distribution is

$$\langle N_j \rangle = N_{T,jk} \left(\frac{V_j}{V_{T,jk}} \right). \tag{4.6}$$

By symmetry, we also have

$$\langle N_k \rangle = N_{T,jk} \left(\frac{V_k}{V_{T,jk}} \right), \tag{4.7}$$

so that

$$\frac{\langle N_j \rangle}{V_j} = \frac{\langle N_k \rangle}{V_k} = \frac{N_{T,jk}}{V_{T,jk}}. \tag{4.8}$$

The width of the probability distribution for N_j is given by the standard deviation, as given in eq. (3.43),

$$\delta N_j = \left[N \left(\frac{V_j}{V_{T,jk}} \right) \left(1 - \frac{V_j}{V_{T,jk}} \right) \right]^{1/2} \tag{4.9}$$

$$= \left[N \left(\frac{V_j}{V_{T,jk}} \right) \left(\frac{V_k}{V_{T,jk}} \right) \right]^{1/2}$$

$$= \left[\langle N_j \rangle \left(\frac{V_k}{V_{T,jk}} \right) \right]^{1/2}.$$

In general, the mode, or location of the maximum of the probability distribution, differs from the location of the average value $\langle N_j \rangle$ by something of order $1/\langle N_j \rangle$. If $\langle N_j \rangle \approx 10^{20}$, this error is unmeasurable. By a fortunate quirk in the mathematics, the average value has exactly the same value as the mode of *approximate* probability density using Stirling's approximation. This turns out to be very convenient, although not necessary.

4.5   Actual Number versus Average Number

It is important to make a clear distinction between the actual number of particles N_j at any given time and the average number of particles $\langle N_j \rangle$.

The actual number of particles, N_j, is a property of the system. It is an integer, and it fluctuates with time if the system can exchange particles with another system.

The average number of particles, $\langle N_j \rangle$, is part of a description of the system and not a property of the system itself. It is not an integer, and is time-independent in equilibrium.

The magnitude of the fluctuations of the actual number of particles is given by the standard deviation, δN_j, from eq. (4.9), which is of the order of $\sqrt{\langle N_j \rangle}$. If there are about 10^{20} particles in subsystem j, the actual number of particles N_j will fluctuate around the value $\langle N_j \rangle$ by about $\delta N_j \approx 10^{10}$ particles. The numerical difference between N_j and $\langle N_j \rangle$ is very big—and it becomes even bigger for bigger systems!

Given the large difference between N_j and $\langle N_j \rangle$, why is $\langle N_j \rangle$ at all useful? The answer lies in the fact that macroscopic measurements do not count individual molecules. The typical method used to measure the number of molecules is to weigh the sample and divide by the weight of a molecule. The weight is measured experimentally with some relative error—usually between 1% and one part in 10^5. Consequently, using $\langle N_j \rangle$ as a description of the system is good whenever the relative width of the probability distribution, $\delta N_j / \langle N_j \rangle$, is small.

From eq. (4.6), the relative width of the probability distribution is given by

$$\frac{\delta N_j}{\langle N_j \rangle} = \frac{1}{\langle N_j \rangle}\left[\langle N_j \rangle \left(\frac{V_k}{V_{T,jk}} \right) \right]^{1/2} = \sqrt{\frac{1}{\langle N_j \rangle}} \sqrt{\frac{V_k}{V_{T,jk}}}. \tag{4.10}$$

The relative width is proportional to $1/\sqrt{\langle N_j \rangle}$, which becomes very small for a macroscopic system. For 10^{20} particles, the relative uncertainty in the probability distribution is about 10^{-10}, which is much smaller than the sensitivity of thermodynamic experiments.

In the nineteenth century, as thermodynamics was being developed, the atomic hypothesis was far from being accepted by all scientists. Thermodynamics was formulated in terms of the mass of a sample, rather than the number of molecules—which in any case, many scientists did not believe in. Fluctuations were not seen experimentally, so scientists did not make a distinction between the average mass in a subsystem and the actual mass.

Maintaining the distinction between N_j and $\langle N_j \rangle$ is made more difficult by the common tendency to use a notation that obscures the difference. In thermodynamics, it would be more exact to use $\langle N_j \rangle$ and similar expressions for the energy and other observable quantities. However, it would be tiresome to continually include the brackets in all equations, and the brackets are invariably dropped. We will also follow this practice in the chapters on thermodynamics, although we will maintain the distinction in discussing statistical mechanics.

Fortunately, it is fairly easy to remember the distinction between the actual energy or number of particles in a subsystem and the average values, once the distinction has been recognized.

4.6   The 'Thermodynamic Limit'

It is sometimes said that thermodynamics is only valid in the limit of infinite system size. This is implied by the standard terminology, which defines the 'thermodynamic limit' as the limit of infinite size while holding the ratio of the number of particles to the volume fixed.

An obvious difficulty with such a point of view is that we only carry out experiments on finite systems, which would imply that thermodynamics could never apply to the real world.

Another difficulty with restricting thermodynamics to infinite systems is that the finite-size effects due to containers, surfaces, interfaces, and phase transitions are lost.

The point of view taken in this book is that thermodynamics is valid in the real world when the uncertainties due to statistical fluctuations are much smaller than the experimental errors in measured quantities. For macroscopic systems containing 10^{20} or more molecules, this means that the statistical uncertainties are of the order 10^{-10} or less, which is several orders of magnitude smaller than typical experimental errors. Even if we consider colloids with about 10^{12} particles, the statistical uncertainties are about 10^{-6}, which is still smaller than most measurements.

Taking the limit of infinite system size can be a very useful mathematical approximation, especially when studying phase transitions, but it is not essential to either understand or apply thermodynamics.

4.7   Probability and Entropy

We have seen that the equilibrium value of N_j is determined by the maximum of the probability distribution (or mode) given by

$$P\left(N_j, N_k\right) = \frac{N_{T,jk}!}{N_j! N_k!} \left(\frac{V_j}{V_{T,jk}}\right)^{N_j} \left(\frac{V_k}{V_{T,jk}}\right)^{N_k}, \tag{4.11}$$

with the constraint that $N_j + N_k = N_{T,jk}$. We can show the dependence of eq. (4.11) on the composite nature of the total system by introducing a new function,

$$\Omega_q(N, V) = \frac{V^N}{N!}, \tag{4.12}$$

where V and N are generic variables that stand for the relevant volume and particle number (V_j, V_k or $V_{T,jk}$), and (N_j, N_k or $N_{T,jk}$). This allows us to rewrite eq. (4.11) as

$$P(N_j, N_k) = \frac{\Omega_q(N_j, V_j)\,\Omega_q(N_k, V_k)}{\Omega_q(N_{T,jk}, V_{T,jk})}. \tag{4.13}$$

At this point, we are ready to make an extremely important observation. Since the logarithm is a monotonic function of its argument, the maximum of $P(N_j, N_k)$ and the maximum of $\ln[P(N_j, N_k)]$ occur at the same values of N_j and N_k; they both represent the mode of the probability distribution. Using $\ln[P(N_j, N_k)]$ turns out to be much more convenient than using $P(N_j, N_k)$, partly because we can divide it naturally into three distinct terms:

$$\ln[P(N_j, N_k)] = \ln\left[\frac{V_j^{N_j}}{N_j!}\right] + \ln\left[\frac{V_k^{N_k}}{N_k!}\right] - \ln\left[\frac{V_{T,jk}^{N_{T,jk}}}{N_{T,jk}!}\right] \tag{4.14}$$

$$= \ln\Omega_q(N_j, V_j) + \ln\Omega_q(N_k, V_k) - \ln\Omega_q(N_{T,jk}, V_{T,jk}).$$

The first term on the right (in both forms of this equation) depends only on the variables for subsystem j, the second depends only on the variables for subsystem k, and the third term depends only on the variables for the total composite system. Note that since $N_{T,jk}$ and $V_{T,jk}$ are assumed to be constants, $\ln\Omega_q(N_{T,jk}, V_{T,jk})$ is also a constant.

It will be convenient to define a function

$$S_q(N, V) \equiv k\ln\left(\frac{V^N}{N!}\right) + kXN \tag{4.15}$$

$$\equiv k\ln\Omega_q(N, V) + kXN, \tag{4.16}$$

where N and V are again generic variables, and k and X are both (at this point) arbitrary constants. The maximum of the function

$$S_{q,jk}(N_j, V_j, N_k, V_k) = k\ln[P(N_j, N_k)] + S_q(N_{T,jk}, V_{T,jk}) \tag{4.17}$$

$$= S_q(N_j, V_j) + S_q(N_k, V_k),$$

with the usual constraint that $N_j + N_k = N_{T,jk}$, then gives us the location of the mode of the distribution of the number of particles, which is an excellent approximation to $\langle N_j \rangle$.

We have seen in Section 3.8 that the width of the probability distribution is proportional to $\sqrt{\langle N_j \rangle}$. Since the probability distribution is normalized, the value of its peak must be proportional to $1/\sqrt{\langle N_j \rangle}$. This can also be seen from the gaussian approximation in eq. (3.57), with $pN \to \langle N_j \rangle$, and $1-p \to V_k/V_{T,jk}$. At the equilibrium values of N_j and N_k, this gives

$$\ln P\left(N_j, N_k\right)\big|_{equil} \approx -\frac{1}{2}\ln\left(2\pi\left\langle N_j\right\rangle\left(V_k/V_{T,jk}\right)\right). \tag{4.18}$$

Since the function $S_q\left(N_{T,jk}, V_{T,jk}\right)$ is of order $N_{T,jk}$, and $N_{T,jk} > \langle N_j\rangle >> \ln\langle N_j\rangle$, the term $k\ln\left[P\left(N_j, N_k\right)\right]$ in eq. (4.17) is completely negligible at the equilibrium values of N_j and N_k. Therefore, *in equilibrium*, we have

$$S_{q,jk}\left(N_j, V_j, N_k, V_k\right) = S_q\left(N_j, V_j\right) + S_q\left(N_k, V_k\right) = S_q\left(N_{T,jk}, V_{T,jk}\right). \tag{4.19}$$

Note that in eq. (4.19) we have used the numbers N_j and N_k instead of the averages $\langle N_j\rangle$ and $\langle N_k\rangle$. This represents the approximation that the width of the particle-number distribution vanishes. It corresponds to the microcanonical ensemble, which assumes that width of the energy distribution vanishes. It should be remembered that they are both approximations, although very good ones. We will discuss their consequences in some detail in Chapter 21.

We can identify $S_{q,jk}\left(N_j, V_j, N_k, V_k\right)$ as the part of the entropy of the composite system that is associated with the configurations; that is, with the positions of the particles. We will call $S_{q,jk}\left(N_j, V_j, N_k, V_k\right)$ the total configurational entropy of the composite system. The functions $S_q\left(N_j, V_j\right)$ and $S_q\left(N_k, V_k\right)$ are called the configurational entropies of subsystems j and k.

We have therefore found functions of the variables of each subsystem, such that the maximum of their sum yields the location of the equilibrium values

$$S_{q,jk}\left(N_j, V_j, N_k, V_k\right) = S_q\left(N_j, V_j\right) + S_q\left(N_k, V_k\right) \tag{4.20}$$

subject to the usual constraint that $N_j + N_k = N_{T,jk}$.

The fact that the contributions to the configurational entropy from each subsystem are simply added to find the configurational entropy of the composite system is a very important property. It is traditionally known, reasonably enough, as 'additivity'. On the other hand, the property might also be called 'separability', since we first derived the entropy of the composite system and were then able to separate the expression into the sum of individual contributions from each subsystem.

The entropy of a composite system has the property of being additive whether or not the subsystems are in equilibrium with each other.

The function that we have identified as the configurational entropy follows Boltzmann's definition (see Section 2.5) in being the logarithm of the probability *of a composite system* (within additive and multiplicative constants). The fact that it is maximized at equilibrium (takes on the value of its mode) agrees with Boltzmann's intuition. This turns out to be the most important property of the entropy in thermodynamics, as we will see in Chapter 9.

Boltzmann's definition of the entropy always produces a function that takes on its maximum, or mode, in equilibrium. Anticipating the discussion of the laws of thermodynamics in Part II, this property is equivalent to the Second Law of Thermodynamics,

which we have therefore derived from statistical principles. (There is still the matter of irreversibility, which will be addressed in Chapter 22.)

4.8   The Generalization to $M \geq 2$ Systems

In thermodynamics, we are usually interested in the interactions between more than just two (sub)systems. We will need to consider the effects of putting two systems into equilibrium with each other, separating them, and then letting one of the systems equilibrate with a different system. We will also need to consider equilibrium conditions involving three or more systems.

The equations we have derived are easily generalized to an arbitrarily large number of systems using the results for the multinomial probability distribution in eq. (3.82), as derived in Section 3.13. In fact, the systems can be taken to include all systems in the world that might possibly be brought into contact with each other.

If we have $M \geq 2$ systems, whether exchanging particles or isolated from each other, the probability distribution for $\{N_j | j = 1, \ldots, M\}$ is

$$P(\{N_j | j = 1, \ldots, M\}) = \left(\frac{N_T!}{\prod_{j=1}^{M} N_j!} \right) \prod_{k=1}^{M} \left(\frac{V_k}{V_T} \right)^{N_k}. \tag{4.21}$$

In this equation, $N_T = \sum_{j=1}^{M} N_j$ and $V_T = \sum_{j=1}^{M} V_j$ are constants. The individual system volumes, V_j are also regarded as constants for now. Eq. (4.21) was derived in Chapter 3 as eq. (3.83).

Using the definition of $\Omega_q(N, V)$ in eq. (4.12), the multinomial probability can be written as

$$P(\{N_j, V_j | j = 1, \ldots, M\}) = \frac{\prod_{k=1}^{M} \Omega_Q(N_k, V_k)}{\Omega_Q(N_T, V_T)}. \tag{4.22}$$

This equation is a generalization of eq. (4.13). It is valid whether some or all of the individual systems are in equilibrium with each other, or if all M systems are isolated from each other. If every system is isolated from every other system, eq. (4.22) is the probability of the initial state $\{N_j, V_j | j = 1, \ldots, M\}$.

Finally, using the definition of entropy in eq. (4.15), we can write

$$S_q(\{N_j, V_j | j = 1, \ldots, M\}) = \sum_{k=1}^{M} S_q(N_k, V_k) - S_q(N_T, V_T) + C, \tag{4.23}$$

where C is an arbitrary constant. N_T, V_T, and C are constants, whose values have no physical consequences.

Eq. (4.23) shows that the entropy of an individual system, j, $S_q(N_j, V_j)$, is exactly the same as derived from considering just two systems, and, in fact, the expression for $S_q(N_j, V_j)$, is independent of the total number of systems. The validity of this equation is not dependent on whether the systems are isolated or in equilibrium.

> The fact that the expression for the entropy of system j contains a factor of $1/N_j!$, is due to the use of interacting systems in the definition of entropy, following Boltzmann and the form of the multinomial coefficient. This derivation has not been universally accepted, and a wide variety of other explanations have been proposed, some claiming that it is impossible to understand the factor of $1/N_j!$ without invoking quantum mechanics. The controversies have run for over 140 years without any sign of slowing down. I recommend the literature on Gibbs' paradox to any student interested in pursuing the matter.

4.9   An Analytic Approximation for the Configurational Entropy

The expression for the configurational entropy becomes much more practical if we introduce Stirling's approximation from Section 3.11. The entropy can then be differentiated and integrated with standard methods of calculus, as well as being much easier to deal with numerically.

Because we are interested in large numbers of particles, only the simplest version of Stirling's approximation is needed: $\ln N! \approx N \ln N - N$. The relative magnitude of the correction terms is only of the order of $\ln N/N$, which is completely negligible. It is rare to find an approximation quite this good, and we should enjoy it thoroughly.

Using Stirling's approximation, our expression for the configurational entropy becomes

$$S_q(N, V) \approx kN \left[\ln \left(\frac{V}{N} \right) + X \right], \tag{4.24}$$

where N and V are generic variables that take the values N_T and V_T, as well as $\{N_k, V_k | k = 1, \ldots, M\}$. X and k are still arbitrary constants. In Chapter 8 we will discuss further constraints that give them specific values.

Eq. (4.24) is our final result for the configurational entropy of the classical ideal gas, under the assumption that the number of particles is known exactly. This definition will be supplemented in Chapter 6 with contributions from the momentum terms in the Hamiltonian. But first, we need to develop mathematical methods to deal with continuous random variables, which we will do in Chapter 5.

4.10 Problems

....

PROBLEM 4.1
Using the entropy to find the equilibrium equations

This object of this problem is to find the equilibrium conditions for ideal-gas systems that are allowed to exchange particles. Use Stirling's approximation and eqs. (4.15), (6.40), and (4.24).

1. Consider two ideal-gas systems with volumes V_j and V_k. They are initially isolated and contain N_j and N_k, respectively. After they are brought together and allowed to come to equilibrium, find the equation that expresses the condition for the equilibrium values, N_j^* and N_k^*, and solve that equation.
2. Now consider three ideal-gas systems with volumes V_j, V_k, and V_3. The systems are initially isolated and contain N_j, N_k, and N_3 particles. Now let them all exchange particles and come to equilibrium. Find two equations to express the equilibrium conditions, and solve them to find the equilibrium values N_j^*, N_k^*, and N_3^*.
3. Again consider the same three ideal-gas systems with volumes V_j, V_k, and V_3. The systems are again initially isolated and contain N_j, N_k, and N_3 particles.
   First, bring systems j and k together and let them come to equilibrium. Find the equilibrium values N_j^* and N_k^*.
   Now separate systems j and k. Bring systems k and 3 together and let them come to equilibrium. Find the number of particles in each of the three systems.

....