## Contents

Note

<table>
<thead>
<tr>
<th>1 Classical Thermodynamics</th>
<th>1-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Systems and Properties</td>
<td>1-1</td>
</tr>
<tr>
<td>1.2 Thermodynamic Variables and Potentials</td>
<td>1-2</td>
</tr>
<tr>
<td>1.3 The Equations of State</td>
<td>1-3</td>
</tr>
<tr>
<td>1.4 Euler’s Equation</td>
<td>1-5</td>
</tr>
<tr>
<td>1.5 The Gibbs-Duhem Equation</td>
<td>1-5</td>
</tr>
<tr>
<td>1.6 Legendre Transformations</td>
<td>1-6</td>
</tr>
<tr>
<td>1.7 Massieu Transformations</td>
<td>1-7</td>
</tr>
<tr>
<td><strong>Appendix:</strong> Homogenous Functions and Euler’s Theorem</td>
<td>1-11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 Classical Mechanics</th>
<th>2-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Newton’s Laws</td>
<td>2-1</td>
</tr>
<tr>
<td>2.2 Conservative Systems</td>
<td>2-2</td>
</tr>
<tr>
<td>2.3 Generalized Coordinates</td>
<td>2-3</td>
</tr>
<tr>
<td>2.4 Hamilton’s Principle</td>
<td>2-5</td>
</tr>
<tr>
<td>2.5 Lagrange’s Equations</td>
<td>2-6</td>
</tr>
<tr>
<td>2.6 The Center of Mass</td>
<td>2-8</td>
</tr>
<tr>
<td><strong>Appendix:</strong> Calculus of Variations</td>
<td>2-11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3 Hamiltonian Mechanics and Phase Space</th>
<th>3-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Hamilton’s Equations</td>
<td>3-1</td>
</tr>
<tr>
<td>3.2 The Hamiltonian</td>
<td>3-3</td>
</tr>
<tr>
<td>3.3 Properties of the Hamiltonian</td>
<td>3-4</td>
</tr>
<tr>
<td>3.4 Phase Space</td>
<td>3-5</td>
</tr>
<tr>
<td>3.5 Examples</td>
<td>3-6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 The Ideas Behind Statistical Thermodynamics</th>
<th>4-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>4-1</td>
</tr>
<tr>
<td>4.2 Boltzmann’s Trajectory Method</td>
<td>4-2</td>
</tr>
<tr>
<td>4.3 The Gibbs Ensemble Method</td>
<td>4-3</td>
</tr>
<tr>
<td>4.4 The Equivalence of Methods and the Ergodic Hypothesis</td>
<td>4-4</td>
</tr>
<tr>
<td>4.5 The Goals of Statistical Thermodynamics</td>
<td>4-6</td>
</tr>
</tbody>
</table>
## CONTENTS

10.1 Introduction .................................................. 10-1
10.2 Vibration in Polyatomic Molecules ......................... 10-4
10.3 Rotation in Polyatomic Molecules ......................... 10-5
10.3.1 Principle Axes and the Moment of Inertia Tensor .... 10-5
10.3.2 The Rotational Partition Functions .................. 10-6
10.4 The Electronic Partition Function in Polyatomic Molecules 10-7
10.5 The Thermodynamics of Polyatomic Molecules ........... 10-8
10.5.1 Translation .............................................. 10-8
10.5.2 Vibration .............................................. 10-9
10.5.3 Rotation .............................................. 10-10

11 The Grand Canonical Ensemble ................................. 11-1
11.1 Introduction ............................................... 11-1
11.2 The Equations for the Ensemble ............................ 11-1
11.3 Solution of the Equations ................................. 11-3
11.4 Thermodynamics of the Grand Canonical Ensemble ....... 11-5
11.5 Entropy and Probability .................................... 11-7
11.6 The Relationship to the Canonical Ensemble ............. 11-7
11.7 A Direct Consequence of Having Independent Subsystems 11-8

12 The Equivalence of Ensembles .................................. 12-1
12.1 Expansion of the Grand Canonical Partition Function .... 12-2
12.2 Generalized Laplace Transform ............................ 12-2
12.3 Transformations Among Ensembles ........................ 12-4
12.3.1 The Maximum Term Method ............................ 12-4
12.3.2 A New Ensemble .................................... 12-6
12.3.3 A Cautionary Example: The Isobaric-Isothermal Ensemble 12-7
12.4 Summary: The Relationship Among Ensembles ........... 12-9

13 The Ideal Crystal ................................................. 13-1
13.1 Introduction ............................................... 13-1
13.2 The Models ................................................. 13-2
13.3 The Einstein Theory ...................................... 13-4
13.4 The Debye Theory ........................................ 13-6
13.4.1 The Frequency Distribution ........................... 13-6
13.4.2 The Debye Partition Function ....................... 13-7
13.5 The One-Dimensional Crystal .............................. 13-10
13.6 Appendix: Behavior of the Debye Function ............. 13-15
13.6.1 High Temperature (u small) .......................... 13-15
13.6.2 Low Temperature (u large) .......................... 13-17
13.7 Appendix: Differentiating Functions Defined by Integrals 13-20

14 Simple Quantum Statistics ....................................... 14-1
14.1 Quantum Statistics ........................................ 14-1
14.2 Simple Results ........................................... 14-6
14.3 The Ideal Gas Limit ...................................... 14-7

15 Ideal Quantum Gases ............................................. 15-1
15.1 Introduction .............................................. 15-1
15.2 Weakly Degenerate Ideal Fermi-Dirac Gas ............... 15-2
CONTENTS

B  Fundamental and Derived Physical Constants  4

C  List of Works Consulted  6
List of Tables

1.1 Natural Variables for Various Potentials ................. 1-3
1.2 Fundamental Definitions of the Potentials ................. 1-5
8.1 Derivatives for the Evaluation of the Translational Partition Function 8-3
8.2 Λ for Various Monatomic Gases at Their Boiling Point ........ 8-7
8.3 Θ_t for Various Monatomic Gases ......................... 8-8
8.4 Monatomic Gas Electronic States ......................... 8-13
8.5 The First Few Bernoulli Numbers ......................... 8-17
8.6 The First Few Euler-Maclaurin Coefficients ................. 8-17
9.1 Θ_v for Selected Diatomic Molecules ................. 9-5
9.2 Θ_r for Selected Diatomic Molecules ................. 9-9
9.3 Derivatives for the Evaluation of the Rotational Partition Function 9-10
10.1 Names for Rigid Rotators ................................. 10-6
13.1 The Debye Function and Heat Capacity for Reduced Temperatures 13-19
B.1 Fundamental and Derived Physical Constants ................. 4
List of Figures

2.1 An Atwood’s Machine .......................................................... 2-7
5.1 Spin entropy per spin as a function of fraction upspins ............ 5-10
5.2 The Gamma Function ......................................................... 5-11
6.1 Schematic Canonical Ensemble ........................................... 6-2
6.2 Energy of a Noninteracting Spin System ................................. 6-15
6.3 Heat Capacity of a Spin System ........................................... 6-16
9.1 Diatomic Molecule Vibrational Energy ................................. 9-6
9.2 Graph of $C_V/k$ vs $T/\Theta_V$ ........................................... 9-7
13.1 Graph of $C_V$ vs $T/\Theta$ for an Einstein Crystal .................. 13-5
13.2 Graph of the Debye Function vs $T/\Theta_D$ ............................ 13-9
13.3 The Debye Function $D(u)$ .............................................. 13-19
Statistical thermodynamics is a mathematical subject that uses a large number of symbols. It essentially exhausts the Latin alphabet, even considering ordinary, capital, and bold face letters as distinct. The Greek alphabet, capitals as well as lower case, is also pressed into service, but it is still not enough. The reason for this large set of symbols is that statistical thermodynamics uses the notations of quantum mechanics, classical mechanics, and classical thermodynamics in addition to its own symbols.

Even though standard symbols are always preferred here if possible, certain horrible choices often had to be made. The major ones are discussed here. Others are noted in passing in the text.

The standard symbol in chemical thermodynamics for the internal energy is $U$, while the standard symbol for energy in physics is $E$. I have decided to use $E$ for internal energy and energy in general, reserving $U$ for the potential energy. The standard for potential energy is $V$, but that is needed for the volume. Kinetic energy is denoted by $K$, which is an allowed symbol for it.

There is also a complication around the symbols for mechanical pressure and probability. For idiosyncratic reasons $p$ is used for pressure and $P$ for probability. IUPAC allows either $p$ or $P$ to be used for pressure, and many texts define them opposite to the way they are defined here.

The classical Hamiltonian is denoted here by $H$, as is the thermodynamic enthalpy. Which is meant should be clear from the context. The quantum mechanical operator is $\hat{H}$. Here $\hat{H}$ could have been used either for the classical Hamiltonian or the thermodynamic enthalpy, but the first conflicts both with custom and the use of the Latin letter $L$ for the Lagrangian and the second would be confusing to all chemists. On the other hand, $\hbar$ is used for the single particle Hamiltonian.

We differentiate the ordinary symbol for the entropy $S$ from the dimensionless entropy $S/k$ by using the symbol $\mathcal{S}$ for the latter.

The German fractur script $\mathfrak{A}$ is used twice. Mainly it is used to denote the number of members of a Gibbsian ensemble, but it is also used to denote the surface area of a hypersphere. No confusion should result from this.

\footnote{I have tried to follow the recommendations of the International Union of Pure and Applied Chemistry given in Pure and Applied Chemistry 51 (1), 1 - 41 (1979).}
Chapter 1

Classical Thermodynamics

1.1 Systems and Properties

Classical Thermodynamics divides the universe into two parts: the *system* and its *surroundings*. Neither of these has a precise definition. Perhaps the best definition of system is that given by Lewis and Randall:\(^1\)

> Whatever part of the objective world is the subject of thermodynamic discourse is customarily called a system.

All the rest of the universe may profitably be called surroundings.

Of course this must be somewhat modified. Thermodynamic systems must be large enough so that surface effects can be neglected.\(^2\) and they must be small enough so that relativistic effects can be neglected as well.

The system is what is important. Classical thermodynamics deals with systems at equilibrium. Equilibrium is a situation in which the properties of the system do not change in time, or at least in a time longer than the longest relaxation time of the system. A relaxation time is the typical time scale needed for a perturbation of the system to dissipate. For example, the time it takes for a pressure wave to fade away or the time it takes the temperature of the system to become uniform.\(^3\)

Classical thermodynamics is concerned both with the interrelationships among the properties of a system at equilibrium, and also with the changes in value of the


\(^2\)For an attempt to deal with small systems within the framework of classical thermodynamics, see Terrell L. Hill, *Thermodynamics of Small Systems, Parts I and II*, W. A. Benjamin, 1963 and 1964.

\(^3\)The first of these is governed by the speed of sound in the system; the second by the heat conductivity of the system.
properties of a system as it goes from an initial equilibrium state to a final equilibrium state.

Ordinarily the initial state of a system is maintained by the presence of internal constraints. If some or all of those constraints are removed or new ones imposed, the system usually changes to a new equilibrium state. Examples include forcefully decreasing the volume of a system or the removal of a barrier that previously prevented two subsystems from mixing.

1.2 Thermodynamic Variables and Potentials

Classical thermodynamics is involved with a small set of physical properties known as thermodynamic variables. Some of these variables are defined only for systems at equilibrium which is why classical thermodynamics is usually limited to such systems.

Thermodynamic variables are properties such as temperature, volume, pressure, and (internal) energy.

A frequently asked question is "why isn’t the internal energy just called the energy?" The answer is that we are normally not concerned with the gross motion of the system in its entirety or to a potential energy affecting the entire system uniformly. Thus we neglect the kinetic energy of the system due to the rotation of the earth and we neglect the potential energy due to the earth’s gravitational field if the strength of that potential energy does not vary appreciably through the volume of the system.

Each thermodynamic variable can be classed experimentally as being either extensive or intensive. Extensive properties are those that are directly proportional to the size or extent of a system, while intensive properties are independent of the size or extent of a system. As examples the internal energy is extensive while the temperature is intensive.

Most thermodynamic variables are state functions. That is, each of these has a value that depends only on the current state of a system and not on its past history. This in turn implies that the change in any state function during a process that takes it from one equilibrium state to another is independent of the path taken by that process. And it also implies that the state function is conserved during any cyclic process.

A cyclic process is a process that returns the system to its starting point at the end of each cycle. A given state function will have the same value at the end of such a cycle as it had at the start of the cycle; thus its change is zero, or, for short, it is conserved.

By virtue of their properties, several thermodynamic variables are especially impor-

---

4 Although the experiment can often be performed in one’s mind.
5 Exceptions include both heat and mechanical work.
tant. These are known as thermodynamic potentials. The ones most often used in classical thermodynamics are the internal energy $E$, the enthalpy $H$, the Gibbs free energy $G$, the Helmholtz free energy $A$, and the entropy $S$.

Each of these potentials is a state function. In addition, the thermodynamic potentials have another important property. Each is associated with its own set of natural variables. The natural variables for each of the thermodynamic potentials are:

$$
E = E(S, V, N) \\
H = H(S, p, N) \\
G = G(T, p, N) \\
A = A(T, V, N) \\
S = S(E, V, N)
$$

Table 1.1: Natural Variables for Various Potentials

When expressed in terms of their natural variables, the sign of the numerical change in a potential indicates whether or not the change is to an equilibrium state of greater or lesser stability. Stability here refers to the resistance of the system to change.

In the case of all the potentials except the entropy, increasing stability is indicated by a decrease in the potential. For the entropy, its value shows an increase for greater stability. Be cautioned that this only applied for the potential corresponding to the variables of the system.\(^6\) In one sense the potentials are not all equal. The internal energy $E$ and the entropy $S$ are defined in terms of extensive variables alone, while the other have intensive variables as part of their definition.\(^7\)

### 1.3 The Equations of State

If a thermodynamic potential is known in terms of its natural variables, then all other thermodynamic variables can be computed from that state function.

To see this consider the total derivative of the Helmholtz free energy $A$ with respect to its natural values $T, V,$ and $N$:

$$
dA(T, V, N) = \left(\frac{\partial A}{\partial T}\right)_{V, N} dT + \left(\frac{\partial A}{\partial V}\right)_{T, N} dV + \left(\frac{\partial A}{\partial N}\right)_{T, V} dN \quad (1.3.1)
$$

The values of all of these derivatives are known:

$$
\left(\frac{\partial A}{\partial T}\right)_{V, N} = -S \quad \left(\frac{\partial A}{\partial V}\right)_{T, N} = -p \quad \left(\frac{\partial A}{\partial N}\right)_{T, V} = \mu \quad (1.3.2)
$$

\(^6\)Thus, for example, the conversion of water vapor to solid particles (known as snow) is spontaneous at $-1^\circ C$ and constant pressure even though the entropy of the system decreases

\(^7\)It looks as if the entropy is different in another respect as well, since it has units of joules per kelvin while the other four have units of joules alone. This difference stems only from historical accident. Later we will introduce equations for dimensionless thermodynamic potentials and this seeming difference will disappear.
where $\mu$ is the chemical potential. Thus we end up knowing $S$, $p$, and $\mu$ in terms of $T$, $V$, and $N$. Further, because of the formal definition of $A$:

$$A = E - TS$$  \hspace{1cm} (1.3.3)

we also know $E = A + TS$, $C_V$, the constant volume heat capacity since $C_V = (\partial E/\partial T)_{V,N}$ and all the other thermodynamic functions as well.

Equation (1.3.2) on page 1-3 does something more. The three relations involving the thermodynamic variables $S$, $p$, and $\mu$ are collectively known as equations of state. These can be written symbolically as:

$$S = S(T,V,N), \quad p = p(T,V,N), \quad \text{and} \quad \mu = \mu(T,V,N),$$  \hspace{1cm} (1.3.4)

Here only the middle relation in Equation (1.3.4) would be familiar, at least in the case of an ideal gas where

$$p = \frac{NkT}{V},$$  \hspace{1cm} (1.3.5)

where $k$ is Boltzmann’s constant and $N$ is the number of particles in the ideal gas. The others are also equations of state since they allow the computation of a new thermodynamic variable from the old ones.

---

**Example 1.1**

As an example assume that we have the fundamental equation

$$E(S,V,N) = \frac{K S^4}{N V^2}$$

where $K$ is a constant. We are to find the three equations of state.

The first thing to do is to ensure that the given equation is a homogeneous equation of the first degree in $N, V,$ and $S$. This is clearly true.

The answers are simple:

$$T = \left(\frac{\partial E}{\partial S}\right) = \frac{4K S^3}{N V^2},$$

$$p = -\left(\frac{\partial E}{\partial V}\right) = \frac{2K S^4}{N V^3},$$

$$\mu = \left(\frac{\partial E}{\partial N}\right) = -\frac{K S^4}{N^2 V^2}$$

---

*In classical thermodynamics, you are more used to seeing these as $R$ and $n$, the gas constant and the number of moles, respectively. Here one must get used to thinking small... .*
1.4 Euler’s Equation

As mentioned above, the internal energy is somewhat special in that it is defined in terms of extensive variables only. Indeed we know it as:

\[ dE(S, V, N) = TdS - pdV + \mu dN. \quad \text{(1.4.1)} \]

Because \( S, V, \) and \( N \) are extensive, Equation (1.4.1) is homogeneous of degree 1 in those variables and so \( E \) is extensive in those variables as well, though we already knew that. As a result Euler’s Theorem\(^9\) holds and we can write:

\[ E(S, V, N) = S \left( \frac{\partial E}{\partial S} \right)_{N,V} + V \left( \frac{\partial E}{\partial V} \right)_{S,N} + N \left( \frac{\partial E}{\partial N} \right)_{S,V}, \quad \text{(1.4.2)} \]

and since we know the thermodynamic variables given by the derivatives we have the very important equation:

\[ E = TS - pV + \mu N. \quad \text{(1.4.3)} \]

The other thermodynamic potentials are generated from this. For example the enthalpy \( H \) is formally defined as:

\[ H = E + pV. \quad \text{(1.4.4)} \]

Differentiating this and using Equation (1.4.1) results in

\[ dH = TdS + Vdp + \mu dN. \quad \text{(1.4.5)} \]

Similarly we have \( A = E - TS \) and \( G = H - TS \). In each case the newly added or subtracted quantity consists of an extensive-intensive variable pair whose overall units are those of energy. These are sometimes called conjugate variables.

Going one step more, since \( A = E - TS \) and \( E = TS - pV + \mu N \), we have that \( A = -pV + \mu N \). Doing this for all the potentials results in a set of fundamental definitions:

\[
\begin{align*}
E &= TS - pV + \mu N \\
H &= TS + \mu N \\
A &= -pV + \mu N \\
G &= \mu N
\end{align*}
\]

Table 1.2: Fundamental Definitions of the Potentials

1.5 The Gibbs-Duhem Equation

If we take the total derivative of Equation (1.4.3) we get:

\[ dE = TdS + SdT - pdV - Vdp + \mu dN + Nd\mu, \quad \text{(1.5.1)} \]

\(^9\)For a full discussion of Euler’s theorem see the Appendix to this chapter on page 1-11.
and if we then subtract (1.4.1) on page 1-5 from this, we obtain
\[ SdT - Vdp + N\mu = 0. \] (1.5.2)

There is nothing special about starting with \( dE \). We might just as well start with \( dG \):
\[ dG = -SdT + Vdp + \mu dN, \] (1.5.3)
and since we know from 1.2 on page 1-5 that \( G = \mu N \), we take the total derivative to get:
\[ dG = \mu dN + N\mu, \] (1.5.4)
and subtracting, we end up with
\[ SdT - Vdp + N\mu = 0, \]
which is Equation (1.5.2) again.

Equation (1.5.2) is known as the Gibbs-Duhem equation. It shows that the three intensive variables are not independent and that in fact given any two, the third can in principle be found.\(^{10}\)

### 1.6 Legendre Transformations

Legendre transformations are common in classical thermodynamics. They are simply not called that in most texts.\(^{11}\) A Legendre transformation is a transformation that allows the swapping of one variable in an equation for its conjugate in such a way that all information available from that equation is maintained.

As an example the second thermodynamic function introduced in most thermodynamics course is the enthalpy \( H \).\(^{12}\) The enthalpy is derived from the internal energy \( E \), which is a function of its natural variables \( S, V, \) and \( N \). What is wanted is a function that is dependent on \( p \) instead of \( V \). Note that \( p \) and \( V \) are conjugate variables. The transformation is simple, \( H \) is defined as:
\[ H = E + pV, \] (1.6.1)
so that
\[ dH = dE + pdV + Vdp. \] (1.6.2)
Now \( dE \) is given by \( dE = TdS - pdV + \mu dN \), and placing this in Equation (1.6.2) yeilds:
\[ dH = TdS + Vdp + \mu dN, \] (1.6.3)
with the additional effect that since \( E = TS - pV + \mu N \),
\[ H = -pV + \mu N. \] (1.6.4)

---

\(^{10}\)If there is more than one component, then both \( \mu \) and \( N \) must have an index and there will be one set of these conjugate variables for each component. If there are then \( m \) components, there will be only \( m + 1 \) independent components instead of the \( m + 2 \) terms that might be expected.

\(^{11}\)Probably to avoid terrifying most undergraduate students.

\(^{12}\)The first thermodynamic function introduced is, of course, the internal energy \( E \).
The process is quite general. Given the conjugate pairs $TS$, $pV$, and $\mu N$ a number of thermodynamic potentials can be generated as functions of their natural variables. Swapping $\mu$ for $N$ is quite simple. We define

$$J = E - \mu N,$$

so that

$$dJ = TdS - pdV - N\mu,$$  \hspace{1cm} (1.6.6)

which isn’t often used in thermodynamics and so has neither a name nor a symbol of its own.$^{13}$

One quantity that will come up later in our discussion of statistical thermodynamics is obtained by starting with $dE$ and swapping $T$ for $S$ and $\mu$ for $N$. Again, since this does not come up often in thermodynamics it too has no standard name or symbol. We shall call it $I$, also for no good reason.

Formally, $I$ is defined by:

$$I = E - TS - \mu N,$$  \hspace{1cm} (1.6.7)

and is obtained by doing two Legendre transformations. The result is:

$$dI = -SdT - pdV - N\mu,$$  \hspace{1cm} (1.6.8)

with

$$I = -pV.$$  \hspace{1cm} (1.6.9)

### 1.7 Massieu Transformations

Unfortunately, while the thermodynamic functions discussed so far are quite useful in classical thermodynamics, they are in the wrong form for statistical thermodynamics. Why? Because the equations developed for statistical thermodynamics are all dimensionless! Why? That will emerge as we study the subject.

So we need to develop an entire set of dimensionless thermodynamic functions. This is not a daunting task and is, in fact, quite easy.

We begin with Equation (1.4.1) on page 1-5:

$$dE = TdS - pdV + \mu dN,$$

which is already familiar. We convert this to:

$$TdS = dE + pdV - \mu dN.$$  \hspace{1cm} (1.7.1)

Each term here has the units of energy. We convert this to dimensionless form by dividing each term by $kT$, where $k$ is Boltzmann’s constant. We get:

$$\frac{dS}{k} = \left( \frac{1}{kT} \right) dE + \left( \frac{p}{kT} \right) dV - \left( \frac{\mu}{kT} \right) dN.$$  \hspace{1cm} (1.7.2)

$^{13}$I’ve used $J$ for it for no very good reason.
The variables are the natural variables for the entropy, $E$, $V$, and $N$. We use Euler’s theorem to discover that:

\[
\frac{S}{k} = \left( \frac{1}{kT} \right) E + \left( \frac{p}{kT} \right) V - \left( \frac{\mu}{kT} \right) N, \tag{1.7.3}
\]

and the equivalent Gibbs-Duhem equation\(^{15}\)

\[
Ed \left( \frac{1}{kT} \right) + Vd \left( \frac{p}{kT} \right) - Nd \left( \frac{\mu}{kT} \right) = 0. \tag{1.7.4}
\]

It looks like there is nothing new here, but there is. Look at Equation (1.7.2) on page 1-7. The intensive variables are not $p$, $T$, and $\mu$, but $p/kT$, $1/kT$, and $\mu/kT$ but three entirely new variables!

We are not used to dealing with single variables that are combinations of old variables. To ease the pain we introduce three definitions that simplify things:\(^{16}\)

\[
\beta = \frac{1}{kT}, \tag{1.7.5}
\]
\[
\pi = \frac{p}{kT}, \quad \text{and} \tag{1.7.6}
\]
\[
\gamma = -\frac{\mu}{kT}. \tag{1.7.7}
\]

For $S/k$ we use the the unfamiliar symbol $\mathfrak{S}$.\(^{17}\)

With this notation Equation (1.7.2) on page 1-7 is:

\[
d\mathfrak{S} = \beta dE + \pi dV + \gamma dN \tag{1.7.8}
\]

which looks (and is) far more simple.

Dimensionless equations are manipulated in the same manner that ordinary thermodynamic equations are handled. But the transformations in this case are called Massieu Transformations,\(^{18}\) introduced by Massieu in 1869 for just this purpose. The transformations are really Legendre transformations, but done on conjugate variables that are, together, dimensionless.

The formal definition of $\mathfrak{S}$ is Equation (1.7.3) rewritten here using the new notation:

\[
\mathfrak{S}(E, V, N) = \beta E + \pi V - \gamma N. \tag{1.7.9}
\]

It is now possible to define eight new dimensionless thermodynamic functions, the

---

\(^{14}\)We could have started with $E = TS - pV + \mu N$, solved for $S$ and made that dimensionless. As one realizes, in thermodynamics there are many ways to get to the same place.

\(^{15}\)Which is only a dimensionless form of the Gibbs-Duhem equation we already have.

\(^{16}\)Students always object to this since the new variables are totally unfamiliar. But if I don’t do this people start differentiating parts of the new variables and get into horrible trouble.

\(^{17}\)Because there is no familiar symbol. The $\mathfrak{S}$ is an “$S$” in the old Germanic Fraktur typeface that dates back to the 1500’s and remained in use until the mid 20th century. It can still be found in some places.

first being Equation (1.7.9) on page 1-8:

\[ \Phi(\beta, V, N) = \mathcal{S} - \beta E, \] \hspace{1cm} (1.7.10)
\[ \Lambda(E, \pi, N) = \mathcal{S} - \pi V, \] \hspace{1cm} (1.7.11)
\[ \Gamma(E, V, \gamma) = \mathcal{S} - \gamma N, \] \hspace{1cm} (1.7.12)
\[ \Theta(\beta, \pi, N) = \mathcal{S} - \beta E - \pi V, \] \hspace{1cm} (1.7.13)
\[ \Upsilon(E, \pi, \gamma) = \mathcal{S} - \pi V - \gamma N, \] \hspace{1cm} (1.7.14)
\[ \Psi(\beta, V, \gamma) = \mathcal{S} - \beta E - \gamma N, \] \hspace{1cm} (1.7.15)
\[ \Pi(\beta, \pi, \gamma) = \mathcal{S} - \beta E - \pi V - \gamma N. \] \hspace{1cm} (1.7.16)

The notation used here is the notation of Planes and Vives,\textsuperscript{19} with some modification.\textsuperscript{20}

Differential relationships are obtained by using one of the equations above, taking its total derivative, and then substituting Equation (1.7.8) on page 1-8 into the result and simplifying. As an example start with (1.7.10) and take the total derivative to get:

\[ d\Phi = d\mathcal{S} - \beta dE - Ed\beta, \] \hspace{1cm} (1.7.17)

and now insert Equation (1.7.8) on page 1-8 into this to get:

\[ d\Phi = -Ed\beta + \pi dV + \gamma dN. \] \hspace{1cm} (1.7.18)

All of the dimensionless potentials can be converted to differentials in the same way. Here is a listing of them:

\[ d\mathcal{S} = \beta dE + \pi dV - \gamma dN, \] \hspace{1cm} (1.7.19)
\[ d\Phi = -Ed\beta + \pi dV + \gamma dN, \] \hspace{1cm} (1.7.20)
\[ d\Lambda = \beta dE - V d\pi + \gamma dN, \] \hspace{1cm} (1.7.21)
\[ d\Gamma = \beta dE + \pi dV - Nd\gamma, \] \hspace{1cm} (1.7.22)
\[ d\Theta = -Ed\beta - V d\pi + \gamma dN, \] \hspace{1cm} (1.7.23)
\[ d\Upsilon = -\beta dE - V d\pi - Nd\gamma, \] \hspace{1cm} (1.7.24)
\[ d\Psi = -Ed\beta + \pi dV - N d\gamma, \] \hspace{1cm} (1.7.25)
\[ d\Pi = -Ed\beta - V d\pi + N d\gamma \] \hspace{1cm} (1.7.26)

It turns out that \(d\Pi = 0\), which can easily be proven.

Some, but not all of these are recognizable combinations of familiar thermodynamic functions. It isn’t hard to prove the following:

\[ \Phi = -\beta A, \quad \Lambda = \beta(TS - pV), \quad \Gamma = \beta H, \] \hspace{1cm} (1.7.27)
\[ \Theta = -\beta G, \quad \Upsilon = \beta E, \quad \Psi = \beta pV. \]

Note that \(\Lambda\) and \(\Psi\) have no simple counterparts, but while \(\Lambda\) won’t be used, \(\Psi\) certainly will be. It should be noted that \(\Phi\), or more accurately \(-A/T\) is sometimes


\textsuperscript{20}There is no standard notation and there really ought to be.
called the **Massieu function** while \( \Theta \), or again more precisely, \(-G/T\) is sometimes called the **Gibbs function**.

We have used a slightly modified version of the notation of Planes and Vives here. Other notations for these functions exist. The most important of these is that of Callen,\(^{21}\) in which he uses \( S[1/T] \) to stand for the Massieu transformation in which \( \beta \) replaces \( E \)

\[
S[1/T] = \frac{S}{k} - \left( \frac{1}{kT} \right) E. \tag{1.7.28}
\]

We’ve used \( \Phi \) for this function.

The reader should note that \( S[1/T] \) is *not* the entropy. \( S[1/T] \) is a symbol for the entropy transformed as:

\[
dS[1/T] = \left( \frac{1}{k} \right) dS - \left( \frac{1}{kT} \right) dE - Ed \left( \frac{1}{kT} \right), \tag{1.7.29}
\]

which, after substitution of Equation (1.7.2) on page 1-7 for \( S/k \) leads to:

\[
dS[1/T] = -Ed \left( \frac{1}{kT} \right) + \left( \frac{p}{kT} \right) dV - \left( \frac{\mu}{kT} \right) dN. \tag{1.7.30}
\]

This seems to be unnecessarily complicated. We’ve not adopted the Callen notation here because even though it shows what is being used in the transformation, it is still too easy to confuse it with the entropy.

Appendix: Homogenous Functions and Euler’s Theorem

There are many Euler’s theorems, but only one beloved by students of thermodynamics.22 Before Euler’s theorem can be discussed in detail some terminology needs to be defined:

First, look at a function of $m$ variables

$$z = f(x_1, x_2, \ldots, x_m).$$

(1.8.1)

We replace each of the $m$ variables $x_i$ with $\lambda x_i$ where $\lambda$ is an arbitrary quantity.23 We now have:

$$z = f(\lambda x_1, \lambda x_2, \ldots, \lambda x_m).$$

(1.8.2)

Here is the basic point: if we can factor $\lambda$ to some power (say $n$) out of Equation (1.8.2) we then say that the function $f(\ldots)$ is homogeneous of degree $n$ in the variables $x_1 \ldots x_m$.

Here’s an example: Consider the function of three variables, $x$, $y$, and $z$\textcolor{red}{w = x^2 y^2 + x y z}$

(1.8.3)

Now replace each $x$, $y$, and $z$ by $\lambda x$, $\lambda y$, and $\lambda z$ to get:

$$w = (\lambda x)(\lambda y)(\lambda z) + \frac{(\lambda x)^2(\lambda y)^2}{(\lambda z)}.$$\textcolor{red}{w = \lambda^3 \left( x^2 y^2 + \frac{x y z}{z} \right).}$$

(1.8.4)

Now $\lambda^3$ factors out of each of these terms leaving

$$w = \lambda^3 \left( x^2 y^2 + \frac{x y z}{z} \right).$$

(1.8.5)

behind. Thus, according to the definition, the function in Equation (1.8.3) is homogeneous of degree 3.

A function can even be homogenous of degree zero, for example

$$w = \frac{x}{y}.$$\textcolor{red}{w = \lambda \left( \frac{x}{y} \right).}$$

(1.8.6)

22Leonhard Euler (pronounced "oiler") lived from 1707 to 1783. The son of a clergyman, he showed an early aptitude for mathematics. He was sent to the University of Basel where he became the pupil of John Bernoulli, one of the famous Bernoulli brothers. In 1727 (the year Newton died) Euler set off for St. Petersburg where the Bernoullis had settled. In 1733 he took the Chair of Natural Philosophy there, later moving to Berlin at the invitation of Frederick the Great (yes, in those days rulers worried about mathematicians) and lived there until 1766, when he returned to St. Petersburg where he died in 1783. He became blind in his later years, working mathematics in his head and dictating the results to his pupils. His output was prodigious. Hundreds of his theorems and countless lesser results are still quoted in the literature. Perhaps one of his most famous results is: $e^{i\pi} + 1 = 0.$

23Note that $\lambda$ is an arbitrary parameter. The best way to think about arbitrary parameters is to think of them as variables standing for some quantity to be specified later. For instance $\lambda$ could later be 1 or 2 or whatever. If it is 1, it effectively disappears and we are left with the original formula.
And, of course, most functions aren’t homogeneous at all, such as
\[ w = x + y^2 - 7, \]  
(1.8.7)  
since replacing \( x \) and \( y \) with \( \lambda x \) and \( \lambda y \) gives:
\[ w = \lambda x + \lambda^2 y^2 - 7, \]  
(1.8.8)  
which doesn’t factor at all.\(^{24}\)

To make things slightly more complicated, a function does not have to be homogeneous in all of its variables. For example the function of three variables:
\[ w = g(x, y, z) = 3xyz + \frac{5z^3}{y} \]  
(1.8.9)  
is homogeneous of degree 2 in \( y \) and \( z \) and isn’t homogeneous at all in \( x \).

Just to have it written down clearly, let us define homogeniety more mathematically:

\textbf{Definition 1.1 (Homogeneous Functions)} Given a function of \( m \) variables \( F(x_1, x_2, \ldots, x_m) \), that function is said to be homogeneous of degree \( n \) in those variables if and only if
\[ F(\lambda x_1, \lambda x_2, \ldots, \lambda x_m) = \lambda^n F(x_1, x_2, \ldots, x_m) \]  
(1.8.10)

With homogeneous functions defined, Euler’s Theorem can be stated:

\textbf{Theorem 1.1 (Euler’s Theorem)} Given a function of \( m \) variables that is homogeneous of degree \( n \), \( F(x_1, x_2, \ldots, x_m) \), also having continuous partial derivatives with respect to those \( m \) variables, then:
\[ x_1 \left( \frac{\partial F}{\partial x_1} \right) + x_2 \left( \frac{\partial F}{\partial x_2} \right) + \ldots, x_m \left( \frac{\partial F}{\partial x_m} \right) = n F(x_1, x_2, \ldots, x_m) \]  
(1.8.11)

The proof is easy enough, and quite beautiful.\(^ {25}\)

Look at Equation (1.8.11). There is nothing at all obvious about it. It \textit{does} look almost familiar. If the left-hand side had \( dx \)'s instead of \( x \)'s it would equal \( df \), the total derivative. But instead it \textit{doesn't} have \( dx \)'s and it isn’t equal to a total derivative, it’s equal to the original function!

So how does one prove such a thing? Well, the clue is that it applies to homogeneous functions. So we will use the definition of homogeneous functions in Equation (1.8.10) in a slightly modified form. To do this let’s introduce a set of new

\(^{24}\)Of course, \textit{most} functions aren’t homogeneous at all. But then, we aren’t interested in those functions.
\(^{25}\)Yes, I know, mathematicians and thermodynamicists have weird notions of what the word \textit{beautiful} means.
variables: $x_1', x_2', \ldots$, where $x_1' = \lambda x_1$, and so on. Then we can write Eq. 14.9 as:

$$f(x_1', x_2', \ldots, x_m') = \lambda^n f(x_1, x_2, \ldots, x_m),$$  \hspace{1cm} (1.8.12)

To prove the theorem, all that needs to be done is to differentiate Equation (1.8.12) with respect to $\lambda$ to get:

$$\left( \frac{\partial f}{\partial x_1'} \right) \left( \frac{\partial x_1'}{\partial \lambda} \right) + \left( \frac{\partial f}{\partial x_2'} \right) \left( \frac{\partial x_2'}{\partial \lambda} \right) + \ldots + \left( \frac{\partial f}{\partial x_m'} \right) \left( \frac{\partial x_m'}{\partial \lambda} \right) = n\lambda^{n-1} f(x_1, x_2, \ldots, x_m),$$  \hspace{1cm} (1.8.13)

where the chain rule has been used to differentiate the $x'$ with respect to $\lambda$.

Two more simple steps and we are done. First, since $x_i' = \lambda x_i$, $(\partial x_i'/\partial \lambda)$ is simply $x_i$. Using this gives:

$$\left( \frac{\partial f}{\partial x_1'} \right) x_1 + \left( \frac{\partial f}{\partial x_2'} \right) x_2 + \ldots + \left( \frac{\partial f}{\partial x_m'} \right) x_m = n\lambda^{n-1} f(x_1, x_2, \ldots, x_m).$$  \hspace{1cm} (1.8.14)

Now we have only to set $\lambda$ equal to 1. This converts all remaining $x_i'$ to $x_i$:

$$\left( \frac{\partial f}{\partial x_1'} \right) x_1 + \left( \frac{\partial f}{\partial x_2'} \right) x_2 + \ldots + \left( \frac{\partial f}{\partial x_m'} \right) x_m = nf(x_1, x_2, \ldots, x_m),$$  \hspace{1cm} (1.8.15)

which was to be proven.

---

26 You will see why later. Treat this as a mystery story and try to figure out how this is going to be done.

27 Whoa!, you say. You can’t do that! Lambda is a constant! To which I reply, “why not?” We will make it a constant in the end. This is an old and very worthwhile trick in mathematics. As long as I set $\lambda$ equal to 1 in any final formula, this is perfectly ok.
Chapter 2

Classical Mechanics

2.1 Newton’s Laws

Newton’s Laws provide one basis for classical mechanics as we know it today. These laws are not derived, they are assumed as axioms. The laws are:

Definition 2.1 (First Law) In the absence of external forces, the velocity of the center of mass of a body is constant in time.

Definition 2.2 (Second Law) A body with constant mass $m$ moving under the influence of a force $F$ undergoes an acceleration $a$ where $F = ma$.

Definition 2.3 (Third Law) If a given body exerts a force $F$ on a second body, the second body also exerts a force on the given body. This force is equal to $F$ but opposite in direction.

These laws are, in principle, sufficient to solve all problems in classical mechanics. In particular, since the acceleration is the second derivative of position with respect to time, the Second Law gives rise to a second order differential equation for each particle in a system of particles. Thus for each of $N$ particles we have an equation of the form:

$$m \frac{d^2r_i}{dt^2} = F_i(r_i), \quad (2.1.1)$$

1. We phrase things differently than Newton and have the advantage of many years of clarification of the principles involved.
2. Bold face symbols indicate vectors.
3. In general $F = dp/dt$ where $p$ is the momentum and $t$ is the time. In this work the mass will always be constant, hence the simpler (and often more usual) definition.
where \( \mathbf{r}_i \) is the (vector) position of the \( i \)'th particle and \( \mathbf{F}_i \) is the force acting on particle \( i \). Note that the force \( \mathbf{F}_i \) is generally a function of the positions of all particles, but not a function of either the velocities of those particles or of the time \( t \).\(^4\)

Given \( N \) particles, the Equations (2.1.1) on page 2-1 will form a set of \( N \) simultaneous second order differential equations that can be solved (often numerically) if one has the \( 2N \) constants of integration that will be needed. These \( 2N \) constants can be taken to be \( N \) positions \( \mathbf{r}_i \) and \( N \) velocities \( \dot{\mathbf{r}}_i \), where the dot indicates differentiation with respect to time.\(^5\) Here it will frequently be the case that the integration constants will be the starting positions and velocities at time \( t = 0 \).

### 2.2 Conservative Systems

Newtonian mechanics deals with all sorts of systems, including systems with friction, velocity dependent potential energies, etc. Here we are much more restricted. First, we will almost always be dealing with systems of particles, be they atoms, molecules, or other entities that can be regarded as particles. Second, we will be dealing with forces that are derivable from a potential energy that is a function of coordinates alone. Thus the force \( \mathbf{F}_i \) on the \( i \)'th particle is given by:

\[
\mathbf{F}_i = -\nabla_i U(\mathbf{r}_1, \ldots, \mathbf{r}_N),
\]

where \( \mathbf{r}_j \) is the vector coordinates of the \( j \)'th particle and \( N \) is the total number of particles.

Systems in which this is true are called conservative systems because, as will be shown, this is the condition needed in order for the total energy \( E \) to be conserved.

The work involved in moving particle \( i \) from position \( A \) to position \( B \) is given by:

\[
W_{i,BA} = \int_A^B \mathbf{F}_i \cdot d\mathbf{s}_1,
\]

where \( \mathbf{s}_i \) is the path taken by particle \( i \) in going from \( A \) to \( B \). This is the same definition of work that is used in thermodynamics, except for the sign.

With the mass constant, we can use the definition of force to introduce the kinetic energy.\(^6\)

\[
\int_A^B \mathbf{F}_i \cdot d\mathbf{s}_1 = m_i \int_A^B \frac{d\mathbf{v}_i}{dt} \cdot \mathbf{v}_i dt = m_i \int_A^B \frac{d}{dt}(\mathbf{v}_i^2) dt,
\]

where, in the middle equation \( \mathbf{F} \) is replaced by \( m \frac{d\mathbf{v}_i}{dt} \) and \( \mathbf{s} \) is replaced by \( \mathbf{v}_i dt \).

---

\(^4\)These conditions are specific to our study, and while generally true for atomic and molecular systems, they are not true in other situations.

\(^5\)This is Newton's notation for a derivative and is basically all that survives of Newton's notation in calculus. The notation we most often use, such as \( dt/dt \), is that of Leibnitz.

\(^6\)Much of this and following material was drawn primarily from H. Goldstein, Classical Mechanics, Addison-Wesley, 1957.
The integral is now easily done with the result that $W_{i,BA}$ is given by:

$$W_{i,BA} = \frac{m_i}{2} (v_{i,B}^2 - v_{i,A}^2). \quad (2.2.4)$$

Here $v_{i,A}^2$ is the square of the speed of particle $i$ when it is at position $A$ and similarly for $v_{i,B}^2$. In spite of the notation, both speeds are really scalars because $v^2 = v \cdot v$, and the dot product produces a scalar.

The quantity

$$K_i = \frac{1}{2} m_i v_i^2 \quad (2.2.5)$$

is known as the kinetic energy of particle $i$. As a result the work done in moving particle $i$ from position $A$ to position $B$ is

$$W_{i,BA} = K_{i,B} - K_{i,A}, \quad (2.2.6)$$

or, in words, the difference in the kinetic energy at the two points.

This is not the only result we can get from Equation (2.2.2) on page 2-2. In general the work in particle $i$ going from $A$ to $B$ will depend on the details of the path. But for our systems $F_i = -\nabla U_i$ and so the work in Equation (2.2.2) on page 2-2 is independent of path, just as it would be in thermodynamics. Thus the work done is just the change in the potential energy between states $A$ and $B$ or:

$$W_{i,BA} = -\Delta U_i = -(U_{i,B} - U_{i,A}). \quad (2.2.7)$$

There is a further, very important conclusion. The work in moving particle $i$ from $A$ to $B$ has been written two ways, one involving the kinetic energy and one involving the potential energy:

$$W_{i,BA} = K_{i,B} - K_{i,A} = -(U_{i,B} - U_{i,A}), \quad (2.2.8)$$

and combining these two gives:

$$K_{i,A} + U_{i,A} = K_{i,B} + U_{i,B}, \quad (2.2.9)$$

each of these is equal to the total energy $E$ of particle $i$, and so we have the fact that $E$ is constant for particle $i$ during any change. But there is nothing special about particle $i$. What is true for it is true for all the other particles in the system. The result is The Law of Conservation of Energy. And it comes from the fact that the force is derivable from a potential energy function. Hence the name conservative system for systems with such potential energy functions. In fact, here in this work we will always deal only with conservative systems unless otherwise specified.

### 2.3 Generalized Coordinates

Problems can often be drastically simplified by a change in coordinates. For example think of two particles bonded together by some potential but otherwise free to move in space.
We can let \( \mathbf{r}_1 \) and \( \mathbf{r}_2 \) be two coordinate vectors of three components each, together representing the six coordinates in the problem.

To completely specify the motion of these two particles we’d need 12 initial conditions: six initial positions (three for each particle) and six velocities (three for each particle).

That would give us six differential equations of the second order to solve. And if the particles are bonded by a potential that depends on the distance between them, the forces involved would depend on all six coordinates. That means that the six differential equations are coupled and must be solved simultaneously.

This is a dreary task in cartesian coordinates.

But you’ve doubtless already thought of a better way. We know from elementary physics that the motion of the center of mass of a group of particles is independent of the motion of the particles about the center of mass.\(^7\)

So we do one of those awful transformation of coordinates so that we have our six new position coordinates, three representing the motion of the center of mass, \( X \), \( Y \), and \( Z \), and three representing the motion of the particles about their center of mass. A convenient choice there would be spherical coordinates \( r \), \( \theta \), and \( \phi \).

Since there is no force acting on the center of mass, we know from Newton’s First Law that it is moving with a constant velocity (which may be zero) in some fixed direction.

We further assume that the potential energy depends only on \( r \), and not on the angles \( \theta \) and \( \phi \). Now the potential energy occurs in only one differential equation, the one for \( r \). So instead of six coupled second order differential equations, We have three that we’ve already integrated (the center of mass coordinates), two which are trivial since there is no potential involved, and only one uncoupled equation that could cause us any difficulty.

That’s a big gain.

But can we simply apply Newton’s laws to these new coordinates?

The three coordinates of the center of mass are cartesian, so it would seem that there’s no problem there. They should (and do) obey Newton’s laws in the form we’d talked about above.

But what about the spherical coordinates? What equations of motion do they observe? I suspect that you already know that spherical coordinates are not quite as simple as cartesian ones.

There is another "difficulty". What if the two particles were stuck together with an iron rod so that their distance apart could not change? Now we have what is known as a constraint, an unchanging relationship among the variables.

Here the constraint is best expressed as \( r = \text{constant} \) But obviously there are

---

\(^7\)We’ll take that as axiomatic right now since we’ve not discussed this yet. But we shortly will. See Section 2.6 on page 2-8.
situations where constraints are not so simple such as a bead frictionlessly sliding on a wire of complex shape.

These are complicated equations and hard to solve. Can this situation be rectified?

2.4 Hamilton’s Principle

The answer is yes. To do it we need a reformulation of the laws of mechanics that will do two things: The first is that it will let us use **generalized coordinates**. These are any quantities that could be used to specify a position in some way. The second is a corresponding set of equations to solve that use generalized coordinates without having to produce new equations for every different set of coordinates used.

As has been already pointed out, Newton’s Laws are axioms. We accept them because they work. But Newton’s Laws are not the only possible formulation of classical mechanics. There are others. One of these formulations is based upon **Hamilton’s Principle**.

The statement of Hamilton’s principle is somewhat obscure because it deals with something that we will call the **action**. Action has been variously defined over the years, but here we shall define it as the time integral of the **difference** between the kinetic and potential energies of a system. We denote this difference by $L$:

$$L = K - U$$

(2.4.1)

and so the action can be written as:

$$A = \int_{t_1}^{t_2} L dt.$$  

(2.4.2)

Here the kinetic energy $K$ and the potential energy $U$ are considered functions of generalized coordinates $q$ and velocities $\dot{q}$. The integral then takes place between the fixed points $(q_1, \dot{q}_1, t_1)$ and $(q_2, \dot{q}_2, t_2)$.

Now the system can get from its initial state to its final state in many ways. Hamilton’s principle states that the path actually taken is the one that makes the action $A$ an extremum.$^9$

Hamilton’s Principle can be deduced from Newton’s Laws,$^{10}$ and, as will be shown, Newton’s Laws can be deduced from Hamilton’s principle. This makes them equivalent in one sense. But they differ in another. Newton’s Laws are differential laws and thus give you the next set of values for $(q$ and $\dot{q})$ to be found from the current set, given the forces on the system. The Hamiltonian formulation however, looks at all the paths as a whole and chooses that path that extremizes the action.

The remaining problem is to discover exactly which path is being specified by Hamilton’s Principle. This, in turn, requires that we use the **calculus of variations**, a

---

$^8$Of course all formulations give the same answers. The world does not change because we use different mathematical formulations.

$^9$An extremum is a maximum or a minimum. In mechanics the path often makes the action a minimum and hence Hamilton’s principle is often called The Principle of Least Action.

$^{10}$See any text on classical mechanics.
mathematical technique that will allow the integrand in Equation (2.4.2) on page 2-5 to be varied so as to extremize the action.\footnote{This is discussed in some detail in Section 2.6 on page 2-11.}

Hamilton’s Principle dictates that the appropriate equations to solve are those known as Lagrange’s Equations. They involve a special function called a Lagrangian. These equations were first worked out in 1788 by Joseph-Louis Lagrange.\footnote{Joseph-Louis Lagrange (1736-1813) was born in Turin in the Italian Piedmont as Giuseppe Lodovico Lagrangia and became a world-famous mathematician. He succeeded Euler in a major position in Berlin before emigrating to France where he spent the rest of his life. He is buried in the Pantheon in Paris.} In other words the path that extremizes the action is the one that satisfies Lagrange’s equations.

The big advantage here is that these equations are formulated in terms of the kinetic and potential energies and not in terms of the cartesian coordinates of the system. The equations do not care what coordinates are used as long as they are sufficient to specify the system.

\section{2.5 Lagrange’s Equations}

We can now define the Lagrangian as:

\[ L = K(q, \dot{q}, t) - U(q), \quad (2.5.1) \]

where \( K \) is the kinetic energy, assumed to be a function of the generalized coordinates \( q \), the corresponding velocities \( \dot{q} \), and \( t \), the time. The potential energy \( U \) is assumed to be a function of the generalized coordinates alone. The system is assumed to be conservative.

Lagrange’s Equation is given by:

\[ \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) - \left( \frac{\partial L}{\partial q_i} \right) = 0, \quad i = 1, 2, \ldots, N, \quad (2.5.2) \]

where \( N \) is the number of degrees of freedom in the system. We speak of degrees of freedom because a generalized coordinate can involve more than one particle, or, as in the motion of the center of mass, all of the particles.

Taking a somewhat broader view, if we have a system with \( N \) degrees of freedom, then the \( q_k \) are the coordinates, however stated, of the \( k \)’th degree of freedom and \( \dot{q}_k \) is the corresponding velocity. Also associated with the generalized velocity \( \dot{q}_k \) is the \textit{generalized momentum} \( p_k \) given by:

\[ p_k = \left( \frac{\partial L}{\partial \dot{q}_k} \right). \quad (2.5.3) \]

If there are no restrictions on the particles, then a system of \( N \) particles can be described by \( 3N \) scalar equations. We choose cartesian coordinates and note that the kinetic energy \( K \) is given by:

\[ K = \frac{1}{2} \sum_j m_j (\dot{x}_j^2 + \dot{y}_j^2 + \dot{z}_j^2), \quad (2.5.4) \]
and if we also assume that the potential energy $U$ is a function of position alone, then

$$L = \frac{1}{2} \sum_j m_j (\dot{x}_j^2 + \dot{y}_j^2 + \dot{z}_j^2) - U(x_1, y_1, z_1, \ldots, x_N, y_N, z_N),$$

(2.5.5)

then the generalized momenta are given by:

$$p_k = \left( \frac{\partial L}{\partial \dot{x}_j} \right) = m_j \dot{x}_j \text{ etc.,}$$

(2.5.6)

so that

$$m \ddot{x}_k = - \left( \frac{\partial U}{\partial x_k} \right), \quad k = 1, 2, \ldots, N$$

(2.5.7)

and similarly for $y$ and $z$. Thus we see that we recover Newton’s Equations from the Lagrangian.

As an example of the use of generalized coordinates, consider the Atwood’s Machine. This is a device that involves two masses, $m_1$ and $m_2$ connected by a flexible unstretchable cord of length $A$. The cord goes over a pulley and the masses hang from the pulley, the first at a height $x_1$, the second at a height of $x_2$.

Figure 2.1: An Atwood’s Machine

This is a situation with two degrees of freedom, the lengths $x_1$ and $x_2$. But there is a *constraint* in that we must have

$$A = x_1 + x_2,$$

(2.5.8)

As a result of this there is really only one degree of freedom.
We let \( q \) be the height of the first mass so that
\[
x_1 = q \quad \text{and} \quad x_2 = A - q.
\] (2.5.9)

The kinetic energy is
\[
K = \frac{1}{2}m_1 \dot{x}_1^2 + \frac{1}{2}m_2 \dot{x}_2^2 = \frac{1}{2}(m_1 + m_2)\dot{q}^2,
\] (2.5.10)

and the potential energy is
\[
U = [-m_1 g q - m_2 g (A - q)]
\] (2.5.11)

so that the Lagrangian is:
\[
L = \frac{1}{2}(m_1 + m_2)\dot{q}^2 - [-m_1 g q - m_2 g (A - q)].
\] (2.5.12)

Using this in Lagrange’s Equations leads to:
\[
(m_1 + m_2)\ddot{q} = (m_1 - m_2)g,
\] (2.5.13)

and we never had to deal with the constraint in any specific way.

While this is a simple example, it can be seen how constraints can be removed by a good choice of coordinates.

## 2.6 The Center of Mass

As a final example we show that any system of particles subject not only to internal forces but to external ones as well, can be broken into two types of motion: the motion of the center of mass and the motion of the particles about the center of mass.

Let us be as general as possible and consider a group of \( n \) particles, each of mass \( m_i \), and vector coordinate \( \mathbf{r}_i \), \((i = 1, \ldots, n)\).

Each of these particles \( i \) is acted on by two types of force. One, caused by some external source, is denoted by \( \mathbf{f}_i \). The other is caused by pairwise interactions between the particles themselves. That is, the total force on particle \( i \) due to all other particles is the simple sum of the individual forces between each pair. Such forces are called pairwise additive.\(^\text{14}\) The force due to the interaction between particles \( i \) and \( j \) is denoted by \( \mathbf{f}_{ij} \). In all cases here and below the indices run from 1 to \( n \).

\(^{13}\)The material in this section follows closely the development of Slater and Frank, *Mechanics*, McGraw-Hill, 1947, p. 90ff, and that of Whittaker, *Analytical Dynamics*, Dover, 1944. This is a reprint of the Fourth Edition of 1937 published by the Cambridge University Press. Yes, sometimes the oldest books are the clearest.

\(^{14}\)This is not true in detail for atoms and molecules since the interaction between one pair can cause shifts in the electron density such as to affect their interactions with other particles. An extreme example of this occurs when some of the particles have chemical bonds between them.
The total force on particle $i$ is $F_i$. Newton's Second Law, Definition 2.2 on page 2-1 gives us:

$$m_i \ddot{\mathbf{r}}_i = F_i = f_i + \sum_{j \neq i} f_{ij} \quad (2.6.1)$$

If we sum up Equation (2.6.1) for all values of $i$, we get:

$$\sum_i m_i \ddot{\mathbf{r}}_i = \sum_i f_i + \sum_i \sum_{j \neq i} f_{ij} \quad (2.6.2)$$

Now Newton’s Third Law, Definition 2.3 on page 2-1 causes the double sum to vanish. This is because for every term $f_{ij}$ in the last summation in Equation (2.6.2) there is also a term $f_{ji}$ which is equal and opposite to $f_{ij}$. Thus we have:

$$\sum_i m_i \ddot{\mathbf{r}}_i = \frac{d^2}{dt^2} \sum_i m_i \mathbf{r}_i = \sum_i f_i = F \quad (2.6.3)$$

where $F$ is the vector sum of all the external forces $f_i$ acting on the system.

If we now denote the total mass of the system by $M$, then the location of the center of mass of the system $\mathbf{R}$ is given by:

$$\mathbf{R} = \frac{\sum m_i \mathbf{r}_i}{M} \quad (2.6.4)$$

and we then use this in Equation (2.6.3) we get:

$$\frac{d^2}{dt^2} (M \mathbf{R}) = M \ddot{\mathbf{R}} = F \quad (2.6.5)$$

The meaning of this is simple. A system of particles acted upon by an external force behaves as if the force acts upon a single particle at the center of mass where that particle has a mass equal to the total mass of the system.

Now we introduce generalized coordinates. We will have one set for the location of the center of mass of the system, and another set for the location of each of the $n$ particles relative to that center of mass.

To do this we define new coordinates for the system of particles with the center at the center of gravity of the system and the axes parallel to the old axes. The new coordinates will be denoted by $\mathbf{r}^i$:

$$\mathbf{r}^i = \mathbf{r}_i - \mathbf{R} \quad (2.6.6)$$

The kinetic energy $K$ for such a system is given by:

$$K = \frac{1}{2} \sum_i m_i \dot{\mathbf{r}}^2_i \quad (2.6.7)$$

We now rearrange Equation (2.6.6) slightly and differentiate it once with respect to time to get:

$$\dot{\mathbf{r}}^i = \dot{\mathbf{r}}^i + \dot{\mathbf{R}} \quad (2.6.8)$$
and place that into Equation (2.6.7) on page 2-9. The result after squaring Equation (2.6.8) on page 2-9 is:

\[ K = \frac{1}{2} \sum_i m_i \left( \dot{r}_i' + \dot{R} \right)^2 + 2\dot{R} \sum_i m_i \dot{r}_i' \]  
(2.6.9)

The last term is zero as can be seen by resubstituting \( \dot{r}_i + \dot{R} \) for \( \dot{r}_i' \) in Equation (2.6.9):

\[ \dot{R} \sum_i m_i \dot{r}_i' = \dot{R} \sum_i m_i (\dot{r}_i + \dot{R}) = \dot{R} \sum_i m_i \dot{r}_i - \dot{R}^2 \sum_i m_i = M\dot{R}^2 - M\dot{R}^2 = 0 \]  
(2.6.10)

Thus we have an important result:

\[ K = \frac{1}{2} \sum_i m_i \left[ \dot{R}^2 + \dot{r}_i'^2 \right] \]  
(2.6.11)

which, in words, means that the kinetic energy consists of two additive parts, the kinetic energy of the center of mass and the kinetic energy around the center of mass.

If the external forces \( f_i \) and the internal forces \( f_{ij} \) are both conservative\(^{15}\) then the potential energy \( U \) of the system is given by:

\[ U(R, r_1 \ldots r_n) = U(R) + U(r_1 \ldots r_n) \]  
(2.6.12)

Then Lagrangian for the system, \( K - U \) splits into two Lagrangians, one for the motion of the center of mass and the other for the relative motion about the center of mass:

\[ L(R, r_1 \ldots r_n) = L(R) + L(r_1 \ldots r_n) \]  
(2.6.13)

The result is that under very general conditions we can split off the motion of the center of mass of a system\(^{16}\) and treat only the motion about the center of mass. The energy involved in the motion about the center of mass is what is called in thermodynamics, the internal energy.

\(^{15}\) And hence derivable from a potential \( U \) by differentiation

\(^{16}\) We are actually rarely interested in that.
Appendix: Calculus of Variations

One of the standard problems of elementary calculus is the determination of the maximum or minimum of a function $f(x)$. Trivially, an extremum occurs at all values of $x$ such that $df/dx = 0$ and $d^2f/dx^2 \neq 0$. The corresponding value of $x$ is a maximum if the second derivative is less than zero and a minimum if it is greater than zero. If the second derivative is zero, $x$ is an inflection point.

The calculus of variations is concerned with a similar problem. We are here interested in the integral

$$I = \int_a^b F(y, dy/dx, x) \, dx, \quad (2.7.1)$$

where $a$ and $b$ are constants and $y$ is a function of $x$. Here $F(y, dy/dx, x)$ is some function, different for different problems, of the variable $y$, an implicit function of $x$, the derivative of $y$ with respect to $x$, and $x$ itself. The problem is to find the function $y(x)$ such that the integral $I$ in Equation (2.7.1) is an extremum, that is, a maximum, or (most often), a minimum. The integral is called a functional, a function $I$ of a function $y(x)$.

This seems like a strange kind of thing, but surprisingly it turns up in a number of situations. And more importantly, variational problems of this sort seem to underly a number of branches of physics including mechanics, optics, and quantum mechanics.

As an example of the kind of problem that arises in the calculus of variations, consider the following question: What is the shortest distance between two points in a plane?\(^{17}\) The answer will be some function $y(x)$ that goes through the two points $(x_1, y_1)$ and $(x_2, y_2)$. The element of arc length is

$$ds = \left[ \left( \frac{dx}{dx} \right)^2 + \left( \frac{dy}{dx} \right)^2 \right]^{1/2}, \quad (2.7.2)$$

and what we want is to minimize the value of $I$ in

$$I = \int_{x_1}^{x_2} ds = \int_{x_1}^{x_2} \left( \frac{dx}{dx} \right)^2 + \left( \frac{dy}{dx} \right)^2 \right]^{1/2} dx = \int_{x_1}^{x_2} \left[ 1 + \left( \frac{dy}{dx} \right)^2 \right]^{1/2} \, dx. \quad (2.7.3)$$

Clearly the value of $I$ will vary with the choice of the particular function $y(x)$. The question is: how do we find $y(x)$? That’s the question we answer below.

Each different choice for the function $y(x)$ leads to a different path from $(x_1, y_1)$ to $(x_2, y_2)$. Let us assume that $y(x)$ denotes the correct answer and that $Y(x)$ is some other function lying close to $y$.\(^{18}\) This makes sense since if we vary the path, $I$ in Equation (2.7.1) varies in such a way that any nearby path will be slightly longer if $I$ is a minimum, or slightly shorter if $I$ is a maximum. So if we use function $Y(x)$ instead of $y(x)$ in Equation (2.7.1), we expect $I$ to change slightly. This suggests the following notation:

$$Y(x) - y(x) = \delta y, \quad (2.7.4)$$

\(^{17}\)I suspect that you may already know the answer to this.

\(^{18}\)Close here means that $Y(x)$ lies completely in a small domain surrounding all the points in the path $y(x)$.
Chapter 2. Classical Mechanics

and

\[ F(Y, dY/dx, x) - F(y, dy/dx, x) = \delta F, \quad (2.7.5) \]

where \( \delta \) has the usual meaning of "a small change in...", or better: "a variation in...". Note that there is no variation in \( x \) as \( x \) remains constant as we go from \( y(x) \) to \( Y(x) \).

If we examine \( \delta(dy/dx) \) we see that:

\[ \frac{\delta dy}{dx} = \frac{dY}{dx} - \frac{dy}{dx} = \frac{d}{dx}(Y - y) = \frac{d}{dx}\delta y, \quad (2.7.6) \]

so that we see that the operations \( \delta \) and \( d/dx \) commute in the sense that the order of these operations can be interchanged without any change in the result.

This looks confusing because of the \( dy/dx \) terms all over the place. Let us make the following substitutions:

\[ z = \left( \frac{dy}{dx} \right) \quad \text{and} \quad Z = \left( \frac{dY}{dx} \right). \quad (2.7.7) \]

Now we have \( F = F(y, z, x) \), which looks much better.

We now take the total derivative of \( F \) with respect to \( y, z \), and \( x \):

\[ \delta F = F(y + \delta y, z + \delta z, x) - F(y, z, x) \]

\[ = \left( \frac{\partial F}{\partial y} \right) \delta y + \left( \frac{\partial F}{\partial z} \right) \delta z, \quad (2.7.8) \]

where there is no term in \( \delta x \) because \( x \) does not vary between \( F(y + \delta y, z + \delta z, x) \) and \( F(y, z, x) \).

In terms of this notation,\(^\text{19}\) in order to have \( \int_a^b Fdx \) be an extremum, we need to have

\[ \delta I = \int_a^b (\delta F)dx = 0. \quad (2.7.9) \]

We rewrite this using Equation (2.7.8)

\[ \int_a^b \left[ \left( \frac{\partial F}{\partial y} \right) \delta y + \left( \frac{\partial F}{\partial z} \right) \frac{d}{dx}(\delta y) \right] dx = 0, \quad (2.7.10) \]

and then since \( dz = d(dy/dx) = d/dx(\delta y) \):

\[ \int_a^b \left[ \left( \frac{\partial F}{\partial y} \right) \delta y + \left( \frac{\partial F}{\partial z} \right) \frac{d}{dx}(\delta y) \right] dx = 0. \quad (2.7.11) \]

The second term in Equation (2.7.11) can be integrated by parts:

\[ \int_a^b \left( \frac{\partial F}{\partial z} \right) \frac{d}{dx}(\delta y)dx = \left[ \left( \frac{\partial F}{\partial z} \right) \delta y \right]^b_a - \int_a^b \left( \frac{d}{dx} \left( \frac{\partial F}{\partial z} \right) \right) \delta ydx. \quad (2.7.12) \]

\(^{19}\)Don’t worry, we will put \( dy/dx \) back for \( z \) in the end.
The first term after the equals sign is identically zero since at each end point \( \delta y = 0 \). So we are left with:

\[
\int_a^b \left[ \left( \frac{\partial F}{\partial y} \right) - \frac{d}{dx} \left( \frac{\partial F}{\partial z} \right) \right] \delta y \, dx = 0, \tag{2.7.13}
\]

which is the condition that \( I \) be an extremum.

Since the change \( \delta y \) is arbitrary, and since Equation (2.7.13) must be true for all choices of \( \delta y \) suitably small, then the only way Equation (2.7.13) can be true is if the integrand is identically zero:

\[
\left( \frac{\partial F}{\partial y} \right) - \frac{d}{dx} \left( \frac{\partial F}{\partial (dy/dx)} \right) = 0, \tag{2.7.14}
\]

where, as promised \( dy/dx \) has been written for \( z \).

This equation was first derived by Euler.

Returning to the problem of the shortest distance between two points in a plane, we recall that we have:

\[
I = \int_{x_1}^{x_2} \left[ 1 + \left( \frac{dy}{dx} \right)^2 \right]^{1/2} \, dx, \tag{2.7.15}
\]

where clearly

\[
F = \left[ 1 + \left( \frac{dy}{dx} \right)^2 \right]^{1/2}. \tag{2.7.16}
\]

Resorting once again to \( z \)-notation, this is:

\[
F = (1 + z^2)^{1/2}. \tag{2.7.17}
\]

Using Equation (2.7.14) the term \( \partial F/\partial y = 0 \) because \( F \) is not a function of \( y \). This leaves only the second term:

\[
\frac{d}{dx} \left( \frac{\partial F}{\partial z} \right) = \frac{d}{dx} \left( \frac{\partial}{\partial z} \right) \left[ (1 + z^2)^{1/2} \right] = 0, \tag{2.7.18}
\]

\[
-\frac{d}{dx} \frac{z}{(1 + z^2)^{1/2}} = -\frac{dz/dx}{(1 + z^2)^{3/2}} = 0, \tag{2.7.19}
\]

and while there may be particular values of \( z \) and \( dz/dx \) that make Equation (2.7.19) zero, the only way it can be zero for all values of \( z \) will be if \( dz/dx = 0 \). That in turn means that \( z = \) constant and since \( z = dy/dx \), then \( dy/dx = \) constant and so \( y = c_1 x + c_2 \) where \( c_1 \) and \( c_2 \) are constants. This is the equation of a straight line. So we conclude that the shortest distance between two points in a plane is a straight line.\[^{20}\]

\[^{20}\]I trust that you are all amazed by this!
Chapter 3

Hamiltonian Mechanics and Phase Space

3.1 Hamilton’s Equations

The Lagrangian is an exceptionally useful concept. In many branches of physics one starts with the Lagrangian and proceeds onward from there.

However there is another formulation of physics that is more natural for our purposes. Indeed you have already run into it in many different guises – particularly in quantum mechanics.

This formulation uses what is called the Hamiltonian in a set of equations called Hamilton’s Equations.

We have previously talked about the Lagrangian for a system of $N$ particles. It can be written symbolically as:

$$L = L(q_j, \dot{q}_j, t),$$

where the $q_j$ and $\dot{q}_j$ stand for the entire set of generalized coordinates and velocities for the problem.\(^1\)

We’ve also talked (briefly) about the generalized momentum $p_j$ that goes with a particular velocity $\dot{q}_j$

$$p_j = \frac{\partial L}{\partial \dot{q}_j}.$$

\(^1\)There can be fewer than $6N$ total variables because there may well be constraints between some of them.
CHAPTER 3. HAMILTONIAN MECHANICS AND PHASE SPACE 3-2

Using this we can take Lagrange’s equation

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = 0,$$

(3.1.3)

and rewrite it as:

$$\dot{p}_j = \frac{\partial L}{\partial q_j}.$$

(3.1.4)

Employing the Lagrangian in Lagrange’s equation leads to a set of second order partial differential equations that describe the motion of the system in time. The variables are the positions and the velocities.

However, sometimes it is inconvenient to use the velocities $\dot{q}_j$ and more useful to use momenta instead. In fact doing this leads to a very symmetric formulation of classical mechanics.

Doing the switch from velocities to momenta as independent variables isn’t just a matter of multiplying all the velocities by $m$ and changing the resulting $mv$’s to momenta. We are using generalized coordinates and the momentum that goes with a coordinate $q$ is not necessarily $m\dot{q}$. The appropriate momenta are those given by (3.1.2) on page 3-1.

To do this more complex change in variable we need to use a Legendre transformation. We last ran into these in Section 1.6 on page 1-6.

To see how this works here let’s take the total derivative of the Lagrangian:

$$dL = \sum_j \frac{\partial L}{\partial q_j} dq_j + \sum_j \frac{\partial L}{\partial \dot{q}_j} d\dot{q}_j + \frac{\partial L}{\partial t} dt.$$

(3.1.5)

Using Equations (3.1.2) on page 3-1 and (3.1.4) we can write this as:

$$dL = \sum_j \dot{p}_j dq_j + \sum_j p_j d\dot{q}_j + \frac{\partial L}{\partial t} dt,$$

(3.1.6)

and this can be rewritten as:

$$dL = \sum_j d(p_j \dot{q}_j) + \sum_j \dot{p}_j dq_j - \sum_j \dot{q}_j dp_j + \frac{\partial L}{\partial t} dt,$$

(3.1.7)

which is better seen by taking Equation (3.1.7) and working backwards.

What we have now can be written in the form:

$$d \left( \sum_j p_j \dot{q}_j - L \right) = -\sum_j \dot{p}_j dq_j + \sum_j \dot{q}_j dp_j - \frac{\partial L}{\partial t} dt,$$

(3.1.8)

which, to the annoyance of the reader, will now simply be tucked away for a moment.

Let’s invent a new function called the Hamiltonian after Sir W. R. Hamilton who actually invented it back in the first part of the 19th century.\(^\text{2}\) In his honor we’ll

\(^2\)Sir William Rowan Hamilton (4 August 1805 – 2 September 1865) was born in Dublin and educated at Trinity College, Dublin, where he spent the rest of his life. He was clearly a mathematical genius and was appointed to a Professorship at Trinity prior to his graduation. He was 22 at the time. Hamilton’s work included not only mechanics but optics, and mathematics as well where he is celebrated for having invented quaternions. He was still working on his book Elements of Quaternions when he died in 1865.
call it $H$.

$$H = \sum_j p_j \dot{q}_j - L.$$  \hfill (3.1.9)

Curiously, this is exactly what is on the left hand side of Equation (3.1.8) on page 3-2

There’s one catch. While the Lagrangian is written in terms of the generalized coordinates, velocities, and time, we want to write the Hamiltonian as a function of the generalized coordinates, momenta, and time. If we take the Hamiltonian to be a function of those variables we can then take the total derivative of Equation (3.1.9). This is:

$$dH = \sum_j \frac{\partial H}{\partial q_j} dq_j + \sum_j \frac{\partial H}{\partial p_j} dp_j + \frac{\partial H}{\partial t} dt.$$  \hfill (3.1.10)

If we now compare this with Equation (3.1.8) on page 3-2 we can identify the derivatives term by term:

$$\dot{q}_j = \frac{\partial H}{\partial p_j},$$  \hfill (3.1.11)

$$\dot{p}_j = -\frac{\partial H}{\partial q_j}, \quad \text{and}$$  \hfill (3.1.12)

$$\left(\frac{\partial H}{\partial t}\right) = -\left(\frac{\partial L}{\partial t}\right).$$  \hfill (3.1.13)

These are called Hamilton’s Equations and make up a set of $2N$ first order differential equations for the motion of the system. Further, the generalized coordinate $q$ and the generalized momentum $p$ are said to be canonical conjugates of each other. They individually may have any units, but their product must have the units of action (which are the same as those of angular momentum).

### 3.2 The Hamiltonian

So far we have defined the Hamiltonian by Equation (3.1.9), but it is quite a bother to first construct the Lagrangian and then do an actual Legendre transformation on it to get that Hamiltonian. So that is not the most useful definition.

We can obtain a much more useful one from Equation (3.1.9). That equation contains the term $\sum_i p_i \dot{q}_i$. Using Equation (3.1.4) on page 3-2 (which defines the momentum $p_i$) and if we assume that the potential energy is a function of the generalized coordinates alone, we can write:

$$\sum_i p_i \dot{q}_i = \sum_i \left(\frac{\partial L}{\partial \dot{q}_i}\right) = \sum_i \left(\frac{\partial K}{\partial \dot{q}_i}\right).$$  \hfill (3.2.1)

Now $K$ is the kinetic energy, a scalar. As a result it does not depend on the coordinate system used to compute it, so we can assume that it is computed in a

\[This is because Equation (3.1.8) on page 3-2 is also the total derivative of the Hamiltonian and a function can have only one total derivative with the same independent variables.\]
cartesian coordinate system. Then the kinetic energy is the sum of terms of the form $m_i \dot{r}_i^2 / 2$ and

$$p_i = \left( \frac{\partial L}{\partial \dot{r}_i} \right) = \frac{1}{2} \frac{\partial (m_i \dot{r}_i^2)}{\partial \dot{r}_i} = m_i \dot{r}_i$$ (3.2.2)

In Equation (3.2.1) on page 3-3 this term is multiplied by $\dot{q}_i$ which is here $\dot{r}_i$, so the result is:

$$\sum_i p_i \dot{q}_i = \sum_i m_i \dot{r}_i^2 = 2K.$$ (3.2.3)

If we now look back at Equation (3.1.9) on page 3-3 we find that the Hamiltonian can be written as:

$$H = \sum_i p_i \dot{q}_i - L = 2K - (K - U) = K + U.$$ (3.2.4)

The importance of this cannot be overestimated. The Hamiltonian, under very general conditions, can be written as the sum of the kinetic and potential energies. All that need be done is to pick a system of generalized coordinates and write the scalars $K$ and $U$ in terms of these.  

### 3.3 Properties of the Hamiltonian

We can find how the Hamiltonian changes in time by looking at its time derivative. Since the Hamiltonian is most generally a function of coordinates, momenta, and the time, i.e $H(p, q, t)$, then we can get the time derivative by first writing the total derivative as:

$$dH = \sum_j \left[ \left( \frac{\partial H}{\partial q_j} \right) dq_j + \left( \frac{\partial H}{\partial p_j} \right) dp_j \right] + \left( \frac{\partial H}{\partial t} \right) dt,$$ (3.3.1)

(which is the same as Equation (3.1.10) on page 3-3) and then “dividing” by $dt$ to get:

$$\frac{dH}{dt} = \dot{H} = \sum_j \left[ \left( \frac{\partial H}{\partial q_j} \right) \dot{q}_j + \left( \frac{\partial H}{\partial p_j} \right) \dot{p}_j \right] + \left( \frac{\partial H}{\partial t} \right).$$ (3.3.2)

Looking back at Hamilton’s Equations (3.1.11) on page 3-3 and (3.1.12) on page 3-3, Equation (3.3.2) becomes:

$$\frac{dH}{dt} = \dot{H} = \sum_j \left( -\dot{p}_j \dot{q}_j + \dot{p}_j \dot{q}_j \right) + \left( \frac{\partial H}{\partial t} \right),$$ (3.3.3)

or

$$\dot{H} = \left( \frac{\partial H}{\partial t} \right) = -\left( \frac{\partial L}{\partial t} \right),$$ (3.3.4)

where the last is from Equation (3.1.13) on page 3-3.
Now if the Lagrangian $L$ is not an explicit function of the time, then $\frac{\partial L}{\partial t} = 0$ and

$$\dot{H} = 0,$$  \hspace{1cm} (3.3.5)

so that the Hamiltonian for a system without explicit time dependence is simply a constant.\footnote{Note that, for instance, the Lagrangian can be a function of the position which in turn will be an implicit function of time. Nevertheless, the partial derivative of the Lagrangian with respect to time is taken with that variable being held constant. Hence if the Lagrangian is not an explicit function of time, Equation (3.3.5) holds.}

Given that the Hamiltonian does not change in time and given that the Hamiltonian is the total energy, we see that for these systems, the total energy of the system is constant in time. In other words the system is \textit{conservative}.

### 3.4 Phase Space

We are used to \textit{coordinate space} which for one particle is a three dimensional space in which the particle is located. If we have two particles, we need three dimensions each for the two particles.

We can represent the positions of the two particles then either as two points on a three-dimensional graph or as \textit{one single point on a six-dimensional graph}.

For $N$ particles we can have $N$ points on a three dimensional graph, or one point on a $3N$-dimensional graph.

Each representation is sometimes useful. For example if our $N$ particles are contained in a box, the cloud of points representing those particles will be confined to a single region of coordinate space. Or if we use a $3N$-dimensional graph, our single point is likewise contained in a corresponding region of coordinate space.

We could, if we wanted, similarly graph the momentum of a particle in a three-dimensional \textit{momentum space}.

And again, if we wished, we could graph the momenta of $N$ particles either by $3N$ points on a three-dimensional graph or a single point in a $3N$-dimensional graph.

If we look at Hamilton’s Equations (3.1.11) on page 3-3 and (3.1.12) on page 3-3 we see that $p$ and $q$ are treated almost symmetrically (the difference is the minus sign in Equation (3.1.12) on page 3-3.) That tempts us to consider graphing both the position and the momentum of our $N$ particles on a single graph. It turns out that this is a good thing to do.

The space defined by such a graph is known as \textbf{phase space}. And again, it can be either a 6-dimensional graph with a cloud of $N$ points in it or a $6N$-dimensional graph with a single point on it. The first is called a $\mu$-space or \textit{micro phase space}. The second is called a $\Gamma$-space or \textit{grand phase space}.

The first or $\mu$-phase space plots the individual particles in a system. The second or $\Gamma$-phase space plots the system itself. As time goes on the points on both graphs...
move because the positions and the momenta of the particles change in time. So the cloud of points in the 6-dimensional graph is in constant motion, or the single point in the 6N-dimensional graph traces out a path as it moves in time. This last path is called a **trajectory**.

A moment’s thought should show that through any point in system phase space there can be one and only one trajectory.

Why? Because given the $6N$ coordinates and momenta of all the particles in the system at some time $t$, Hamilton’s equations give us the the “next” point that the system will occupy at time $t + \delta t$. And if two trajectories have the same initial conditions (that is, give the same point in phase space) they will always have exactly the same “next” point. If trajectories crossed, there would have to be two next points at the point of intersection. We’ve just seen that there is only one. Thus trajectories in phase space cannot cross themselves.

This works going backwards in time also. The equations of mechanics are independent of the direction of time. So being at the same point at time $t$ means that two trajectories would have to have been at the same point at time $t - \delta t$. Trajectories “crossing” each other at a point implies that the previous and next points must be different. We’ve seen that that can’t happen. Thus trajectories can never cross each other.

### 3.5 Examples

The purpose of these examples is to show how the Hamiltonian formulation of mechanics works in practice.

---

**Example 3.1**

Again let us consider a free particle of mass $m$ moving without any forces acting on it whatsoever. The particle is initially at $x_0$, $y_0$, and $z_0$ just as before. But this time instead of initial velocities we specify initial momenta $p_{x,0}$, $p_{y,0}$, and $p_{z,0}$.

The Hamiltonian for this problem is easily written:

$$H = \left[ \frac{p_x^2}{2m} + \frac{p_y^2}{2m} + \frac{p_z^2}{2m} \right].$$  \hspace{1cm} (3.5.1)

Hamilton’s equations are:

$$\dot{x} = \left( \frac{\partial H}{\partial p_x} \right) = \frac{p_x}{m} \quad \dot{p}_x = - \left( \frac{\partial H}{\partial x} \right) = 0, \hspace{1cm} (3.5.2)$$

$$\dot{y} = \left( \frac{\partial H}{\partial p_y} \right) = \frac{p_y}{m} \quad \dot{p}_y = - \left( \frac{\partial H}{\partial y} \right) = 0, \hspace{1cm} (3.5.3)$$

$$\dot{z} = \left( \frac{\partial H}{\partial p_z} \right) = \frac{p_z}{m} \quad \dot{p}_z = - \left( \frac{\partial H}{\partial z} \right) = 0. \hspace{1cm} (3.5.4)$$
There is no potential energy. Thus all the equations for the momenta equal zero. This means that the momenta are all constants.

There is a general point here. When a coordinate does not appear in the potential, then the corresponding momentum equation is equal to zero and that momentum is constant. This can often be used to make the solution of the remaining equations more simple.

So now we have:

\[ p_x = p_{x,0} \quad p_y = p_{y,0} \quad p_z = p_{z,0}. \]  

(3.5.5)

Substitution of these into the other Hamilton’s Equations gives:

\[ \dot{x} = \frac{p_{x,0}}{m} \quad \dot{y} = \frac{p_{y,0}}{m} \quad \dot{z} = \frac{p_{z,0}}{m}, \]  

(3.5.6)

which immediately integrate to

\[ x = \frac{p_{x,0}}{m} t + x_o \quad y = \frac{p_{y,0}}{m} t + y_o \quad z = \frac{p_{z,0}}{m} t + z_o, \]  

(3.5.7)

which is the final result. The meaning of this is that a free particle (one with no potential energy) moves in a straight line with constant linear momentum.\(^6\)

---

Here’s another example:

---

**Example 3.2**

Consider a pendulum swinging from the origin and hanging down in the direction of the negative \( y \)-axis. It swings back and forth in the \( x \) direction. The pendulum rod is of length \( R \) and is, of course, weightless. The pendulum bob is of mass \( m \). The potential energy is:

\[ U(x, y) = -mgR \cos \theta, \]  

(3.5.8)

where \( g \) is a constant (the acceleration of gravity) and \( \theta \) is the angle between the pendulum rod and the \( y \)-axis.

Convenient coordinates for this problem are \( r \), the distance from the origin to the pendulum bob, and \( \theta \), the pendulum angle.

With these polar coordinates the kinetic energy is:

\[ K = \frac{mR^2 \dot{\theta}^2}{2}, \]  

(3.5.9)

and the Lagrangian

\[ L(\theta, \dot{\theta}) = K - U = \frac{m}{2} R^2 \dot{\theta}^2 + mgR \cos \theta. \]  

(3.5.10)

\(^6\)This is a satisfying answer because we already knew it was true. It is good when our theory correctly predicts what is known. Imagine if that were not the case!
CHAPTER 3. HAMILTONIAN MECHANICS AND PHASE SPACE

Remember that we need the Lagrangian in order to find the proper momentum conjugate to $\theta$. Of course, with experience one can omit this step and simply write down the conjugate momentum. That momentum is given by:

$$p_\theta = \left( \frac{\partial L}{\partial \dot{\theta}} \right) = R^2 \dot{\theta}. \quad (3.5.11)$$

Solving for theta and squaring gives

$$\dot{\theta}^2 = \frac{p_\theta^2}{m^2 R^4}, \quad (3.5.12)$$

which now gets inserted into the kinetic energy above to make it a function of the angular momentum and not the angular velocity. We can then write the Hamiltonian:

$$H(\theta, p_\theta) = \frac{1}{2} \frac{p_\theta^2}{m R^2} - mgR \cos \theta. \quad (3.5.13)$$

From Hamilton’s equations we get:

$$\dot{\theta} = \left( \frac{\partial H}{\partial p_\theta} \right) = \frac{p_\theta}{m R^2} \quad (3.5.14)$$

$$\dot{p}_\theta = - \left( \frac{\partial H}{\partial \theta} \right) = -mgR \sin \theta. \quad (3.5.15)$$

Sadly, these equations cannot be solved in terms of simple functions.\(^7\) The equations can be simplified by making the standard small angle assumption. We assume that $\theta$ is small. Then the potential energy becomes $U = -mgR\theta$ and the last equation above is then:

$$\dot{p}_\theta = - \left( \frac{\partial H}{\partial \theta} \right) = -mgR \theta. \quad (3.5.16)$$

We can assume a solution of the form:

$$\theta = A \sin(\omega t + B) \quad (3.5.17)$$

and get the standard results.

\(^7\)They can be solved in terms of elliptical integrals, but let’s not do that.
Chapter 4

The Ideas Behind Statistical Thermodynamics

4.1 Introduction

To be very approximate about things, by the end of the first third of the nineteenth century classical mechanics had pretty much assumed the form that we see now. By the end of the middle third of the nineteenth century classical thermodynamics had also pretty much been worked into its modern form.

So it should come as no surprise that in the last third of the nineteenth century saw the start of finding the connection between classical mechanics and classical thermodynamics.

Clearly there had to be a connection. Classical mechanics was assumed to apply to the atomic world and in principle one should be able to compute mechanical properties such as the pressure exerted by a container filled with gas, the several moles of differential equations that would need to be solved notwithstanding. And if, as was commonly assumed, the temperature represented some sort of average energy of a container of gas, that ought to be calculable as well. Indeed, even relationships such as the change in energy with temperature should be, at least in principle, computable from classical mechanics.

There was even a good hint as to how to do it. Specifying the state of a single phase, single component molecular system in classical mechanics required several moles of initial positions and momenta. Specifying the state of the same system in classical thermodynamics required only three variables.1

---

1This number is calculated from the Gibbs Phase Rule, not discovered until the last third of the nineteenth century, but the number was known experimentally for many decades before that.
Somehow, the moles of microscopic variables boiled down to three macroscopic ones. That strongly hinted that averages were involved.\footnote{One does not have to be an expert to see that the thermodynamic internal energy depends on the internal energy of all the molecules in the system.}

The question was: what kind of averages?

The first theories that could be classed as "statistical" were those developed in the study of the kinetic theory of gases, developed by Maxwell and Boltzmann, culminating in the Maxwell-Boltzmann Distribution Law well-known today.\footnote{See Jeans' Kinetic Theory of Gases} But applying that method to anything but the simplest cases was exceptionally difficult.

The largest problem was that in the first two-thirds of the 19th century most (but not all) physicists believed that atoms and molecules were only "convenient fictions". It took the work of Einstein and Perrin to convince them.\footnote{Chemists, on the other hand, had no problem with atoms. They were more than convenient in chemical theory. And they explained experiments very well. Even so, there were a few "convenient fiction" chemists even as late as the 1920's in the United States and elsewhere.} And that was much later.

\section*{4.2 Boltzmann's Trajectory Method}

Ludwig Boltzmann was one of the many prodigies of the 19th century.\footnote{He was born on 20 February, 1844 in Vienna and died on 5 September 1906 at a vacation town near Trieste. His life spanned the last half of the 19th century. He earned his PhD in 1866. His thesis was on the kinetic theory of gases. In his lifetime he was named Professor at a number of German and Austrian universities. Boltzmann was subject to severe fits of depression and evidently had what is called today bipolar disorder. He several times attempted suicide and, in 1906, was successful.}

One way to compute macroscopic properties from microscopic ones would be to plot the trajectory of a system in phase space (Γ-space with $6N$ coordinates). All one needed do is compute the desired properties at each point along the trajectory and then average the values. Thus one would be using something like

\begin{equation}
\langle X \rangle = \lim_{t \to \infty} \frac{1}{t} \sum_{i} X_i t_i ,
\end{equation}

where $X_i$ is the value of the property at point $i$ on the trajectory and $t_i$ is the time the system spent at point $i$. In principle the trajectory is followed for all time.

Now there is a serious problem with that equation. How much time does a system spend at a point? That is, for how long does a system have exactly one set of exact positions and momenta? The answer is the obvious one: the system spends zero time at any of the infinite number of points on the trajectory. This makes $t_i$ zero and makes evaluation of Equation (4.2.1) impossible.

This problem was solved by Boltzmann in a revolutionary way. Boltzmann’s breakthrough was to break phase space up into little cells of finite volume

\begin{equation}
\delta V = \delta q_1 \delta q_2 \ldots \delta q_{3N} \delta p_1 \delta p_2 \ldots \delta p_{3N} ,
\end{equation}
small enough so that the value of the mechanical quantity \( X \) we are interested in will not change in any appreciable way during the now finite time the system spends in \( \delta V \).

This is called **coarse graining** because it breaks up phase space into a huge collection of little but finite cells.\(^6\) Now Equation (4.2.1) on page 4-2 makes sense, since the system will spend a small but finite time \( t_i \) in cell \( \delta V_i \).

Of course doing the computation is still hard. But Boltzmann was able to compute the physical properties of an ideal gas using this technique.

Today it is much easier to compute things using this method. While a huge number of particles can still not be handled, modern computers can compute the phase space trajectory from Hamilton’s equations with varying degrees of precision from a few dozen to hundreds of thousands.\(^7\) This technique is today known as **molecular dynamics** and is a very important theoretical technique.

### 4.3 The Gibbs Ensemble Method

Boltzmann’s technique was very hard to use prior to the invention of computers. Gibbs\(^8\) had two insights into this situation. Instead of following a single system in time along a trajectory, Gibbs developed a time-free technique. He proposed that one mentally create a huge collections of macroscopically identical systems all having the same set of external values. To be concrete, here we will take those to be \( N, V, \) and \( E \), but they don’t have to be those.

Now the systems are assumed to be macroscopically identical. But there is no reason for them to be microscopically identical. Every single point along a Boltzmann trajectory corresponds to a different microscopic state.

So Gibbs imagined an **ensemble** or collection of systems, each frozen at some point in phase space. The number of such systems, \( \mathfrak{A} \), is assumed to be much greater than the number of different microstates \( \Omega \) in the phase space. The important point being that even though \( \Omega \) is huge, \( \mathfrak{A} \) is even more huge, but both are **finite** numbers.

The next insight was to **assume** that the trajectory in phase space of any equilibrium system would eventually go through all the coarse-grained cells \( \delta V \) in the phase space.\(^9\) So all he had to do was to select his \( \mathfrak{A} \) systems randomly from the entire phase space. Since the ensemble is assumed to contain many more systems than there are available microstates \( \Omega \), we can assume that the number of Gibbsian ensemble members \( a_i \) that are found to be in \( \delta V_i \) is always one or more. Then the

---

\(^6\)In effect, though the terminology was not known in Boltzmann’s day, he quantized space. And we shall see later that quantum mechanics gives us a reasonable way to determine exactly what the volume of these little cells should be.

\(^7\)The more particles, the less precise the computation. Entire galaxies of stars have been simulated using this technique. And quite complex molecules have been done as well.

\(^8\)Yes, that Gibbs...

\(^9\)This is known as the **ergodic hypothesis** and will be discussed further below. It has never been proven. Indeed, systems exist for which it is not at all true. Nevertheless, it seems indeed to be true for the huge majority of real systems.
probability that a random system will be found in microstate $i$ is given by:

$$P_i = \frac{a_i}{\sum_i a_i} = \frac{a_i}{A}.$$ \hfill (4.3.1)

Then the value of any property $X$ can be found by computing the value of $X$ in the $i$'th microstate, $X_i$ and then

$$\langle X \rangle = \sum_i P_i X_i.$$ \hfill (4.3.2)

This turns out to be a very convenient method to use with paper and pencil, and was for a long time the predominant way to do computations in statistical thermodynamics. The invention of computers made the picking of random systems even more convenient and such methods are in wide use today. They are collectively known as Monte Carlo methods.

### 4.4 The Equivalence of Methods and the Ergodic Hypothesis

To begin, note that the phase space associated with a given problem contains all points consistent with the given values of the independent macroscopic variables. These are sets such as $E, V$, and $N$ or perhaps $T, V$, and $N$, and so on.

Every such point is a possible state of a member of the Gibbs Ensemble for this system. But every such point does not necessarily lie on a Boltzmann trajectory.

If each point in the phase space corresponds to an à priori possible point on the Boltzmann trajectory, then the Boltzmann trajectory method and the Gibbs ensemble method clearly give the same results.

So a major question is: does the Boltzmann trajectory in fact go through every single point in the phase space?

This is known as the **Ergodic Hypothesis**.

Curiously, the answer is no. We can construct model systems where this isn’t true at all. Imagine a system of $N$ point particles in a cubical box of volume $V$ with total energy $E$. Now if all of the molecules in the system have $y$- and $z$-components of velocity exactly equal to zero the molecules will then simply bounce back and forth between the walls perpendicular to the $x$-axis and never hit the other four sides of the box – even though there is no energetic reason why they should do this. It is all a matter of the initial conditions.

Obviously the Gibbsian approach will have molecules in all possible initial states, most of which will not have zero $y$- and $z$-velocity components. The Boltzmann approach will not.

Of course this example is a bit contrived but the point is clear. The initial conditions on a system may very well restrict that system to move in only a portion of the available phase space.
Thus the Ergodic Hypothesis can not be true.

What is true is a modified version of the Ergodic hypothesis called the weak ergodic theorem which says:

**Theorem 4.1 (The Weak Ergodic Theorem)** *The trajectory of a system obeying Hamilton’s equations will, in time, almost always come arbitrarily close to any given point in the phase space.*

This is known as the **Poincaré Recurrence Theorem** which, in more detail shows that the closer you want the trajectory to come to a given point, the longer you may have to wait, but you will return to the vicinity of the given point over and over again.

We still have the problem of strange situations such as the one described above where some regions of phase space are simply inaccessible to the trajectory. We can’t wish those away and it is likely that such systems make up a vanishingly small subset of all possible systems.\(^\text{10}\) So we shall have to live with that.

In practice, the problem essentially never comes up. So some authors sidestep the problem by making the equivalence of the space average (Gibbs) and the time average (Boltzmann) a postulate.\(^\text{11}\)

A valid question is ”Why does the Ergodic Theorem work even if it isn’t true?” The answer seems to lie in the coarse graining introduced by Boltzmann. Because of it, an evolving mechanical system never occupies a point in phase space; it occupies a very small but finite volume. So the trajectory isn’t a line, but instead is a tube of some sort snaking through space.

Now the tube is not allowed to intersect itself since every microstate in phase space is succeeded by one and only one new microstate. So since the phase space also has a finite volume\(^\text{12}\) then the trajectory tube must fill all of phase space in a finite time, in the end reconnecting to its origin,\(^\text{13}\) or the trajectory must rejoin itself before filling all of phase space. This last is unlikely and becomes less likely the coarser the graining. Even in our artificial model presented above, this works. The particles are points and hence never collide with each other, but coarse graining gives them a non-zero \(y\) - and \(z\)-component of both momentum and position.

This is not the only deep conceptual problem in statistical thermodynamics. This and several more are discussed in S. K. Ma’s book\(^\text{14}\) which the interested reader should see.

---

\(^{10}\)What the mathematicians love to call a set of measure zero.

\(^{11}\)See, for example, Friedman, Harold L., *A Course in Statistical Mechanics*, Prentice-Hall, 1985, page 5, First Postulate.

\(^{12}\)Both the space coordinates and the momenta have upper and lower limits.

\(^{13}\)You can see in this the basis for the **Poincaré Recurrence Theorem**

4.5 The Goals of Statistical Thermodynamics

The first goal of statistical thermodynamics is to lay down a theoretical formulation of the methods used to compute macroscopic properties from microscopic ones.

The second is to show by using models of various complexity, how these calculations can be done.

The simplest theoretical formulations use the Gibbs ensemble technique coupled with a very few simple ideas from quantum mechanics. For instance, the frequently discussed notion of a microstate corresponds quite naturally to regions mapped out by the Heisenberg Uncertainty Principle, \( \Delta p \Delta q \geq h \). To see this consider an \( N \)-particle phase space with \( 6N \) dimensions. There will be one momentum \( p_i \) conjugate to each coordinate \( q_i \) and there will be \( 3N \) pairs of such conjugate variables. Then a microstate can be defined as a region of phase space of dimension \( h^{3N} \), where \( h \) is Planck’s constant. This works because we cannot specify both the \( p \)'s and the \( q \)'s for a system any more accurately than that.

Another idea taken from quantum mechanics is that the energy of a system is frequently not continuous, but instead is limited to a countable infinity of discrete values \( E_i \). The advantage here is not so easily seen, but it lies in the fact that we can sum over these energies rather than integrate, as we’d have to for continuous energies.

And there are times when the symmetry rules governing wave functions will be useful, but that will come up in its own time.

Last, we will discuss quantum statistics to the extent of finding the properties of ideal quantum gases.

The question of models is more complex. Our theoretical equations will be formal equations. That is, they will be relationships between thermodynamic quantities that we want to compute, and quantities calculated from microscopic data through a model. As an example if we wish to deal with an ideal gas, we can consider a gas made up of point particles of mass \( m \). Or we can consider it made up of tiny hard spheres of diameter \( d \) and mass \( m \). This last model is more realistic, but much harder to use in computations.

So models will play a major role. We will frequently find that the more realistic the model, the more complex the the calculations.\(^ {15} \) So we will often start with a simple model and make it more complicated as we go along.

In general we deal with two kinds of systems. The first contain particles that do not interact with each other, though they may be affected by an external field. Such systems are called ideal systems for reasons that will quickly become visible. Later we will deal with systems of interacting particles cleverly called non-ideal systems.

What we will not deal with in any detail are modern computer methods of doing realistic computations. For that the student is asked to see Mark Tuckerman’s Sta-\(^ {15} \)Surprise! Surprise! \(^ {15} \)
Chapter 5

The Microcanonical Ensemble

5.1 Introduction

As remarked in Section 4.5 on page 4-6, the preferred method for doing theoretical calculations is the Gibbsian Ensemble method. We illustrate this here with a simple but not fantastically useful example.\(^1\) In this model we assume a single-phase, single-component system with fixed number of particles \(N\), a fixed volume \(V\), and a fixed energy \(U\). Those are the standard independent variables for the what is called the microcanonical ensemble.

It goes without saying that these values are slightly restricted. After all, neither \(N\) nor \(V\) can be negative and their ratio, the density, cannot be so great as to leave the range of chemical processes and enter the realm of nuclear physics.

We assume that the system in the volume \(V\) has energy \(E\) which may be degenerate. The system is then in one of the \(\Omega\) degenerate levels of the energy \(U\) of the system.

We will apply the Gibbsian approach to calculate the probability that the system is in a particular degenerate state.

To apply the Gibbsian approach we mentally construct a huge number, \(A\), of macroscopic replicas of our system.\(^2\) Each will have the same identical values of \(N\), \(V\), and \(E\), and each will be in one of the one or another of the \(\Omega\) degenerate energy levels corresponding to that energy. Note that this system is totally isolated. It interchanges neither heat, work, nor matter with its surroundings.

---

\(^1\)So why do we bother? Because this is the simplest ensemble and in looking at it we introduce quantities useful later on.

\(^2\)This is not to be confused with the \(A\) used for action in another chapter.
We then mentally freeze the systems in whatever degenerate level they happen to be in at that moment. We now have a static collection of \( \mathfrak{A} \) systems, each in a definite microstate.

This collection of systems with the independent variables \( N, V, \) and \( U \) is called a \textit{microcanonical ensemble}.

## 5.2 Occupation Numbers

We chose the number \( \mathfrak{A} \) of ensemble members to satisfy the condition:

\[
\mathfrak{A} \gg \Omega ,
\]

where \( \Omega \) is the degeneracy of the quantum mechanical state corresponding to the imposed external conditions.

This is done to ensure that there is a far large number of systems than degenerate levels. This way we can expect that usually there will be at least several systems in any given microstate. Indeed, what we hope is that (1) all possible microstates of the system are sampled by this process and (2) that there will be many members of the ensemble in each of those possible microstates.

We let \( a_j \) be the number of systems in the ensemble that are in microstate \( j \). These \( a \)'s are called the \textbf{occupation numbers} of the microstates of the ensemble. It is immediately clear that:

\[
\sum_{j=1}^{\Omega} a_j = \mathfrak{A} .
\]

If we knew the \( a_j \)'s, we could determine the macroscopic properties \( \langle X \rangle \) of the system because we assume that, as stated before (see Chapter 4 on page 4-1), that we can associate the property \( X_j \) with each microstate \( j \). And, most importantly, it is clear that the probability of finding any given ensemble member in state \( j \) is just \( a_j/\mathfrak{A} \). So given the \( a_j \)'s we can work out the thermodynamics. It is worth stopping for a moment and considering this statement. It is both momentous and obvious. Obvious, because the systems are in random states and that’s how probability is defined, and momentous because that leads directly to thermodynamics.

However, we don’t yet have enough information to find the \( a_j \)'s. There are many \textit{different sets of} \( a_j \)'s that will satisfy Equation \( (5.2.2) \). For instance, if there were three systems in an ensemble that had two possible microstates, then \( \mathfrak{A} = 3 \) and \( \Omega = 2 \). We can describe this toy ensemble by three digits, the first being the microstate of the first system, the second being the microstate of the second, etc.

The possible arrangements of the systems among the microstates are then:

\[
111, \ 112, \ 121, \ 122, \ 211, \ 212, \ 221, \ 222
\]

All of these have \( a_1 + a_2 = 3 \).
There is only one way to have $a_1 = 3$ while there are three ways to have $a_1 = 2$. So even though we’ve placed the systems evenly among the possible microstates, the various sets of occupation numbers $a_j$ are not equally likely.

Of course, we’ve assumed that all microstates are equally likely to be occupied. But is that even true?

### 5.3 The Principle of Democratic Ignorance

The problem before us is, with the assumption that all systems are in the same system macroscopic energy level having degeneracy $\Omega$, then

*Are all the degenerate microstates equally likely to be occupied?*

We do not know which, if any, of the $\Omega$ degenerate microstates the system prefers to be in. Worse, we have no obvious way of judging if any group of microstates is more or less likely to be occupied than any other.

So we are at a loss as to how to proceed.

What we need is some sort of rule that will help us. And there is such a general rule in science that covers situations like this. We’ll call it the *rule of democratic ignorance*. It is essentially this:

**Definition 5.1 (The Principle of Democratic Ignorance)**  
When there is a situation with a number of possibilities, and when there is absolutely no reason to prefer one possibility over any other, then all possibilities have to considered to be equally likely.

Of course this rule can not be proven. But it is instructive to consider several possible alternatives:

1. **The Rule of Primogeniture**: Whatever state is numbered 1 is the most likely; number 2 is second most likely, and so on.

2. **The Rule of Authority**: The most probable state is the one the author says is the most probable.

3. **The Rule of Mystery**: There is a most probable state but we can never know what it is.

The first alternative is clearly silly. Numbering the states is up to the person who does the numbering. It is a pure accident which state is numbered first and cannot reflect any physical reality. Thus we can safely ignore Rule 1.

Alternative 2 is equally silly. Simple moral authority only substitutes someone else’s ignorance for yours.
Alternative 3 is an abdication of responsibility. It is throwing up our hands and saying that we cannot solve this problem and should go on to do something else as a career, like perhaps being a movie star.  

It is doubtful that anyone can suggest an acceptable rule other than the Rule of Democratic Ignorance. Certainly, nobody yet has been able to do that. So in the absence of anything better, we are left with the Rule of Democratic Ignorance as an axiom of statistical thermodynamics.

Applying the Rule of Democratic Ignorance to the situation at hand, we must conclude that: All members of the microcanonical ensemble have an equal chance of being in any of the $\Omega$ different microstates. This is equivalent to saying that all $\Omega^A$ different ways of arranging the $A$ systems among the $\Omega$ states are equally likely.

This does not mean that each of the sets of $a$’s are equally likely. For instance in the example with three systems and two states above there are eight equally likely arrangements, three of them lead to the set $\{2,1\}$ but only one way leads to the set $\{3,0\}$.

It turns out that of all possible arrangements, some few are far more likely than any other. That can’t be guessed beforehand, but it is an interesting consequence of the fact that $N$ is huge.

Accepting this for now (we will retroactively prove it later) our question: What values can the $a$’s be expected to have can now be changed to:

What set of $a$’s has the greatest number of ways of occurring?

### 5.4 A Subproblem: The Multinomial Coefficient

Listing the number of ways in which three systems can be arranged so that, for example, there are two of them in the first state and one in the second could be simplified if we had a formula for computing how many such ways there are.

In particular we are going to be very interested in knowing how many ways there are to get any particular given set of $a$’s. Put more formally:

How many ways are there of arranging $A$ systems so that there are $a_1$ in the first quantum state, $a_2$ in the second, etc.

Or, to fit our interest better:

How many ways are there of arranging $A$ objects into $\Omega$ piles such that there are $a_1$ in the first pile, $a_2$ in the second pile, etc.

\(^3\text{Actually, one might like that...}\)
If we can find a formula for this, it is likely that we can find which of these arrangements has the most ways of being attained and hence the highest probability of occurring.

And the answer to this is already known. If we let the number of ways of making such piles be $W$, then it can be shown that

$$W(a_1, a_2, ..., a_\Omega) = \frac{\Omega!}{a_1!a_2!...a_\Omega!} = \frac{\Omega!}{\prod_j a_j!}. \quad (5.4.1)$$

Here’s an example:

**Example 5.1**

How many arrangements are there of three systems that result in the occupation numbers $a_1 = 2$ and $a_2 = 1$?

$$W(2, 1) = \frac{3!}{2!1!} = \frac{6}{2} = 3. \quad (5.4.2)$$

The quantity $W$ is known as the multinomial coefficient. It is a generalization of the binomial coefficient to multinomials. It arises naturally in algebra in the expansion of multinomials. In particular:

$$(x_1 + x_2 + ... + x_\Omega)^\mathfrak{A} = \sum_{a_1=0}^{\mathfrak{A}} \sum_{a_2=0}^{\mathfrak{A}} \cdots \sum_{a_\Omega=0}^{\mathfrak{A}} W(a_1, a_2, ..., a_\Omega)x_1^{a_1}x_2^{a_2}...x_\Omega^{a_\Omega}. \quad (5.4.3)$$

This gives us a quick and useful side result. If all of the $x$’s are set to 1 we get:

$$\Omega^\mathfrak{A} = \sum_{[a_1=0,...,\mathfrak{A}],[a_2=0,...,\mathfrak{A}],[...]} W(a_1, a_2, ..., a_\Omega). \quad (5.4.4)$$

Thus the total number of distinct ways of arranging $\mathfrak{A}$ objects into $\Omega$ piles is $\Omega^\mathfrak{A}$.

This is a very very very large number.

If, for instance, $\Omega$ were only 100 and $\mathfrak{A}$ were only 1000 (both much smaller than the numbers we’d expect to run into), there would be $10^{2000}$ different ways to arrange the 1000 objects among the 100 piles.

With realistic values of $\mathfrak{A}$ and $\Omega$, the value of $W$ is astronomically large. This is important because large numbers can sometimes have strange properties, as we shall see.
5.5 The Result

So we have the result in Equation (5.4.1) on page 5-5 that

$W(a_1, a_2, ..., a_\Omega) = \frac{\mathfrak{A}!}{a_1! a_2! ... a_\Omega!} = \mathfrak{A}! \prod_j a_j!$, \tag{5.5.1}

and we seek the set of $a$’s that will make $W$ a maximum. And we already, via the Principle of Democratic Ignorance, have the answer. We have no reason to prefer any state $i$ over any other, so it is clear that:

$a_1 = a_2 = \ldots = a_j = \ldots = a_\Omega$, \tag{5.5.2}

and so since the $a$’s must add up to the number of members of the ensemble $\mathfrak{A}$,

$a_i = \frac{\mathfrak{A}}{\Omega}$. \tag{5.5.3}

Now this obviously isn’t the only possibility. We don’t prefer any one state over any other, but that does not keep chance from allowing a few more to be in state $i$ and a few less in state $j$. But we shall see that Equation (5.5.3) in fact will maximize $W$.

For now we will assume that these values maximize $W$ so $W$ becomes $W_m$, the maximum value of $W$:

$W_m = \frac{\mathfrak{A}!}{\Omega (\mathfrak{A}/\Omega)!}$ \tag{5.5.4}

This can be evaluated using **Stirling’s Approximation**\(^4\) given by Equation (5.8.1) on page 5-11:

$\ln n! \approx n \ln n - n$. \tag{5.5.5}

The simplest way to do this is to work with the logarithm of $W$. This works because the maximum of $\ln W$ occurs at the same point as the maximum of $W$ since $d \ln W = (1/W) dW$ and here $W$ is always positive and greater than zero.

Thus

$\ln \mathfrak{A}! = \mathfrak{A} \ln \mathfrak{A} - \mathfrak{A}$ \tag{5.5.6}

and

$\ln \left( \frac{\mathfrak{A}}{\Omega} \right) = \frac{\mathfrak{A}}{\Omega} \ln \left( \frac{\mathfrak{A}}{\Omega} \right) \frac{\mathfrak{A}}{\Omega}$. \tag{5.5.7}

Given that

$\ln W_m = \ln \mathfrak{A}! - \sum_{j=1}^{\Omega} \ln \left( \frac{\mathfrak{A}}{\Omega} \right)! = \ln \mathfrak{A}! - \Omega \ln \left( \frac{\mathfrak{A}}{\Omega} \right)!$, \tag{5.5.8}

then

$\ln W_m = \mathfrak{A} \ln \mathfrak{A} - \mathfrak{A} - \Omega \left( \frac{\ln \mathfrak{A}}{\Omega} \ln \frac{\mathfrak{A}}{\Omega} - \frac{\mathfrak{A}}{\Omega} \right)$, \tag{5.5.9}

\(^4\)See Section 5.8 on page 5-11 for details if you are not familiar with Stirling’s Approximation.
and
\[
\ln W_m = A \ln A - A \ln \frac{A}{\Omega} + A
\]  \hspace{1cm} (5.5.10)

The result is then
\[
\ln W_m = A \ln \Omega \quad \text{or} \quad W_m = \Omega^A. \tag{5.5.11}
\]

This brings up a curious situation. If we add up the values of \(W\) for all possible arrangements of the \(a\)'s, we get Equation (5.4.4) on page 5-5
\[
\Omega^A = \sum_{[a_1=0,\ldots,A],[a_2=0,\ldots,A],\ldots} W(a_1,a_2,\ldots,a_\Omega), \tag{5.5.12}
\]
and yet Equation (5.5.11) says that the value of \(W\) for the set of \(a\)'s that produce the maximum is exactly the same value! How can this be?

The answer is that the largest value of \(W\) is so much larger than the other values of \(W\) that to all intents and purposes the largest value is the value of the sum of them all.\(^5\) Of course the sums are not really exactly equal. Stirling’s Approximation is just that, and small terms have been left out. Had they been included, the two sums would have differed by an extremely small amount.

So we have shown that the maximum value of \(W\) is attained when the \(a_i\)'s are all equal. And we have shown that the maximum value of \(W\) is \(\Omega^A\).

## 5.6 Thermodynamics

Properties in the microcanonical ensemble are computed as we’d expect. We assume that the mechanical property we are interested in can be computed for each microstate \(j\). And since all the \(a_j\)'s are equal, the probability of any particular state, normalized to 1 is \(P_j\):
\[
P_j = \frac{1}{\Omega}. \tag{5.6.1}
\]

Thus the thermodynamic properties of a system in the microcanonical ensemble depend on the degeneracy of the energy state, that degeneracy is very important.\(^6\)

The independent variables in our microcanonical systems are \(N\), \(V\), and \(E\). Thus \(\Omega\) here is a function of \(N\), \(V\), and \(E\), and is more properly written as \(\Omega(N,V,E)\).

This just makes it more clear that \(\Omega\) is directly related to the thermodynamic properties of the systems.

This is not the place to discuss the microcanonical ensemble in detail. It is not often used in theoretical work, though it does play a major role in molecular dynamics.

\(^5\)We will see other examples of this later. It is due to the large number effect in which for large \(n\), \(\ln \ n\) is negligible compared to \(n\). To see this compare \(10^{23}\) to \(\ln 10^{23} \approx 56\).

\(^6\)For folks like me who grew up thinking that the degeneracy of a given energy level was just an annoying complication of no particular significance, this revelation came as a bit of a shock!
Suffice it to say that the microcanonical ensemble, because of its requirement of fixed energy, ends up being mostly applied to counting problems where then entities being counted all have the same energy.

Boltzmann was able to present a more-or-less convincing argument that the relationship between the microcanonical ensemble and thermodynamics was embodied in the fairly famous equation:

\[ S(N, V, E) = k \ln \Omega(N, V, E), \tag{5.6.2} \]

where \( S \) is the entropy and \( k \) is a constant, known appropriately today as Boltzmann’s constant.\(^7\)

A quick digression. We will study several more ensembles in succeeding chapters. In each of these it will be shown that a general formula for the entropy is given by:

\[ S = -k \sum_{i=1}^{\Omega} P_i \ln P_i, \tag{5.6.3} \]

where the sum is taken over all allowed states and \( P_j \) is the probability that a randomly chosen system in the ensemble will be found in state \( j \). This equation is even correct in the microcanonical ensemble. Here \( P_i = 1/\Omega \). Using this in Equation (5.6.2) gives:

\[ S = -k \sum_{i=1}^{\Omega} P_i \ln P_i = -k \sum_{i=1}^{\Omega} \frac{1}{\Omega} \ln \left( \frac{1}{\Omega} \right) = \frac{k}{\Omega} \Omega \ln \Omega = k \ln \Omega \tag{5.6.4} \]

which shows that this formula applies to the microcanonical ensemble as well.

When a thermodynamic quantity is expressed in terms of its natural values, it not only points in the direction of equilibrium, but it has another property as well. If one has an explicit formula for the property in terms of its natural variables, then all other thermodynamic properties can be obtained from that formula.

In classical thermodynamics we know that if we have a formula for the entropy in terms of \( N, V, \) and \( E \), we can calculate all other thermodynamic functions from it. We can see how to do that from the total differential:

\[ dS = \left( \frac{\partial S}{\partial E} \right) dE + \left( \frac{\partial S}{\partial V} \right) dV + \left( \frac{\partial S}{\partial N} \right) dN, \tag{5.6.5} \]

which, if we evaluate the differentials, gives us:

\[ dS = \frac{1}{T} dE + \frac{p}{T} dV - \frac{\mu}{T} dN, \tag{5.6.6} \]

showing that the temperature, pressure, and chemical potential are also known.

Given Equation (5.6.1) on page 5-7, we can write Equation (5.6.6) as

\[ d \ln \Omega = \frac{1}{kT} dE + \frac{p}{kT} dV - \frac{\mu}{kT} dN. \tag{5.6.7} \]

\(^7\)This equation is carved on Boltzmann’s tombstone, possibly the only equation in human history to be so remembered.
We can complete our formal presentation of the microcanonical ensemble with formal formulas for computing the temperature, pressure, and chemical potential. From Equation (5.6.7) on page 5-8 we see that:

\[
\frac{1}{kT} = \left( \frac{\partial \ln \Omega}{\partial E} \right)_{V,N} \tag{5.6.8}
\]

\[
\frac{p}{kT} = \left( \frac{\partial \ln \Omega}{\partial V} \right)_{E,N} \tag{5.6.9}
\]

\[
-\frac{\mu}{kT} = \left( \frac{\partial \ln \Omega}{\partial N} \right)_{E,V} \tag{5.6.10}
\]

To get actual, useful formulas for \( T \), \( p \), and \( \mu \) one needs a model of whatever system we are interested in. Why a model? Because any real system is way too complicated to represent mathematically in full detail. The physicist Steven Wolfram has said that to compute the next state of the universe requires a computer as large and as complex as the universe. No smaller computer can do that.\(^8\) It is easy to see that the same thing is true for a system as simple as a box of real gas. Every molecule affects the entire system.

To overcome this problem, we never attempt to deal with every single detail of the behavior of a system. We can’t. Instead we make a model of the system.

We pick what we believe are the important phenomena in a system and ignore the rest. The phenomena we pick constitute the model.

For example, the model for an ideal gas is not too complex. Indeed, we will look at that in a later chapter.

In general, however, models for the microcanonical ensemble are difficult to generate, primarily because the restriction to a constant energy is very limiting. Thus it is not much used for modern theoretical work. It does have application to a set of problems that can be called counting problems in which the energy is usually fixed at 0.

### 5.7 The Most Simple Spin System

Here is a very simple model of a very simple spin system. Let us assume that we have \( M \) particles that have two possible spins. We denote these as \( \text{up} \) and \( \text{down} \). At any point we have \( N \) \( \text{up} \) spins and \( M - N \) \( \text{down} \) spins. There is no energy of any kind. The spins are all independent and there is no external field they can interact with.

In this very simple model, \( M \) plays the role of volume \( V \) and \( N \) the role of number \( N \). The energy \( E \) is zero. We want to find \( \Omega(0, M, N) \) for this system.

\(^8\)This is easy to see, since if some parts of the universe were not needed to compute the next state of the entire universe, then those parts can have no interaction whatsoever with the rest of the universe. In that case those parts are not there at all!
This is simple. There are $M!$ ways of arranging the $N$ up-spins and $M - N$ down-spins. But the order of the up-spins doesn’t matter, so we’ve over counted by the $N!$ arrangements of the up-spins. Similarly the arrangement of the down-spins also doesn’t matter, so we’ve over counted by $(M - N)!$ too. The result is then:

$$\Omega(0, M, N) = \binom{M}{N} = \frac{M!}{N!(M - N)!}.$$  \hfill (5.7.1)

We can significantly simplify this by applying Stirling’s approximation:

$$\ln \Omega = M \ln M - M - N \ln N + N - (M - N) \ln(M - N) + (M - N),$$

$$= M \ln M - N \ln N - (M - N) \ln(M - N).$$ \hfill (5.7.2)

Then:

$$\ln \Omega = M \ln M - N \ln N - (M - N) \ln(M - N)$$

$$+ (M - N) \ln M - (M - N) \ln M,$$ \hfill (5.7.3)

and

$$\ln \Omega = -N \ln \theta - (M - N) \ln(1 - \theta),$$ \hfill (5.7.4)

Now substituting $\theta = N/M$ we have:

$$\ln \Omega = -N \ln \theta - (M - N) \ln(1 - \theta),$$ \hfill (5.7.5)

which, after a bit more manipulation becomes:

$$S(0, M, N)/k = \ln \Omega(0, M, N) = -M \{\theta \ln \theta + (1 - \theta) \ln(1 - \theta)\},$$ \hfill (5.7.6)

which, since $\theta$ is a mole fraction of up spins in a two-component system, Equation (5.7.6) gives exactly the thermodynamic entropy of mixing of two components non-interacting components, up spins and down spins.

![Figure 5.1: Spin entropy per spin as a function of fraction upspins](image)

Computation of the chemical potential $\mu/kT$ and the “pressure” $p/kT$ is left as an exercise for the reader.

\footnote{I said it was a simple model of a simple system!}
5.8 Appendix: The Gamma Function and Stirling’s Approximation

Stirling’s approximation is a useful equation that allows writing factorials in terms of continuous functions. The approximation is:

\[ \ln n! = n \ln n - n. \quad (5.8.1) \]

These are the first few terms of an asymptotic expansion for \( \ln n! \). An asymptotic expansion is a series expansion that does not converge to the actual value of the function but also has the property that the ratio of its value and the correct value goes to one as the argument of the function increases in size.

The factorial \( n! \) is defined only for the non-negative integers. The factorial of zero is defined as 1. They are a special case of the gamma function \( \Gamma(z) \), where \( z \) is a complex variable whose real part is always greater than 0. The gamma function is given by:

\[ \Gamma(z) = \int_{0}^{\infty} t^{z-1} e^{-t} dt \quad \text{real part of } z > 0. \quad (5.8.2) \]

Figure 5.2: The Gamma Function

It is simple to show that the gamma function obeys the relationship:

\[ \Gamma(z + 1) = z\Gamma(z). \quad (5.8.3) \]

To do this we need only integrate

\[ \Gamma(z + 1) = \int_{0}^{\infty} t^{z} e^{-t} dt, \quad (5.8.4) \]

by parts. Using the formula \( \int u dv = uv - \int v du \) we can choose \( t^z = u \) and \( e^{-t} dt = dv \). This gives us \( v = -e^{-t} \) and \( dv = -t^{z-1} e^{-t} dt \). So

\[ \Gamma(z + 1) = -t^z e^{-t} \bigg|_{0}^{\infty} + z \int_{0}^{\infty} t^{z-1} e^{-t} dt. \quad (5.8.5) \]
The first term is zero at both limits and the second gives us $2\Gamma(z)$.

We can use Equation (5.8.3) on page 5-11 to show that the gamma function acts like the factorial for integer $z$ by noting that, for example:

$$\Gamma(4) = 3\Gamma(3) = 3 \times 2\Gamma(2) = 3 \times 2 \times 1\Gamma(1), \quad (5.8.6)$$

and knowing that $\Gamma(1) = 1$ we have $\Gamma(4) = 3!$. Showing that $\Gamma(1) = 1$ is easy enough. We need only set $z = 1$ in Equation (5.8.2) on page 5-11 and doing the resulting simple integral. By induction one can then show that

$$\Gamma(z + 1) = z! \quad \text{integer } z > -1 \quad (5.8.7)$$

It is useful to know that

$$\Gamma(1/2) = \sqrt{\pi} \quad (5.8.8)$$

which can be derived by noting that

$$\Gamma(1/2) = \int_0^\infty t^{-1/2}e^{-t}dt = 2\int_0^\infty e^{-u^2}du = \sqrt{\pi} \quad (5.8.9)$$

where we’ve used the substitution $t = u^2$. The last integral is a standard one equal to $\sqrt{\pi}/2$, which leads immediately to Equation (5.8.8).

Indeed, by the use of Equation (5.8.3) on page 5-11 the useful general relation:

$$\Gamma(n + 1/2) = \frac{1 \times 3 \times 5 \cdots (2n - 1)}{2^n} \sqrt{\pi} \quad n = 1, 2, \ldots \quad (5.8.10)$$

is easily discovered.

The gamma function can be differentiated. The result is called the psi or digamma function:

$$\psi(z) = \frac{d\ln \Gamma(z)}{dz} = \frac{\Gamma'(z)}{\Gamma(z)} \quad (5.8.11)$$

where $\Gamma'(z)$ stands for $d\Gamma(z)/dz$.

It can be shown that

$$\psi(1) = \gamma \quad (5.8.12)$$

where $\gamma$ is Euler’s constant 0.5772156649... , an irrational number as famous among mathematicians as $\pi$ and $e$.

For integer $n$:

$$\psi(n) = -\gamma + \sum_{k=1}^{n-1} \frac{1}{k} \quad (5.8.13)$$

which gives us the derivative of a factorial.\(^\text{11}\)

\(^{10}\)Don’t you love it when the Author says it can be shown...

\(^{11}\)Equation (5.8.13) is easy enough to evaluate for $n = 7$, but it can be suspected that you’d not care to evaluate it, term by term, for $n = 6 \times 10^{21}$.\(^\text{21}\)
To make dealing with factorials easier, the British mathematician Stirling derived the asymptotic series expansion the leading terms of which are given in Equation (5.8.1) on page 5-11. It is fairly easy to derive.\footnote{The derivation given here is mainly from D.A. McQuarrie, Mathematical Methods for Scientists and Engineers, University Science Books, 2003, pages 119ff.}

We start with Equation (5.8.4) on page 5-11 and note that the integrand is very sharply peaked. This is because $t^z$ rises rapidly while $e^{-t}$ falls rapidly. Indeed the integrand is zero at both 0 and $\infty$. We now write $t^z$ as $e^{z \ln t}$ and so:

$$
\Gamma(z+1) = \int_0^\infty e^{z \ln t - t} dt
$$

A little numerical exploration shows\footnote{You can try to show this analytically, but it isn’t easy.} that this integrand is peaked at $z = t$, so we expand the exponent $z \ln t - t$ around $z = t$ in a Taylor series. For this we will need the derivatives:

$$
\begin{align*}
 f(t) &= z \ln t - t \\
 f'(t) &= \frac{z}{t} - 1 \\
 f''(t) &= -\frac{z}{t^2} \\
 f'''(t) &= \frac{2z}{t^3}
\end{align*}
$$

so the expansion is:

$$
z \ln t - t \approx z \ln z - z \frac{(t - z)^2}{2z} + \cdots
$$

where terms in $(t - z)^3$ and higher have been neglected.

We now have, for Equation (5.8.14)

$$
\Gamma(z+1) \approx \int_0^\infty e^{z \ln z - z - (t - z)^2/2z} dt
$$

$$
\approx e^{z \ln z - z} \int_0^\infty e^{(t - z)^2/2z} dt
$$

If we now let $(t - z)^2/2z = u^2$, we get, after some manipulation

$$
\Gamma(z+1) \approx (2z)^{1/2} z^z e^{-z} \int_{-z/2}^\infty e^{-u^2} du
$$

Because $z$ is assumed to be large and because the integrand falls of rapidly around its peak at $u = 0$, we can extend the lower limit to $-\infty$ without introducing too much more error. We do this because we can then do the resulting integral\footnote{Necessity is often the cause of much inspiration.} which is:

$$
\int_{-\infty}^\infty e^{-u^2} du = 2 \int_0^\infty e^{-u^2} du = \frac{\pi^{1/2}}{2}
$$
We then finally get our asymptotic approximation:

\[
\Gamma(z + 1) = z! \approx (2\pi z)^{1/2} z^z e^{-z}
\] (5.8.21)

The error in Equation (5.8.21) is too great to allow its use directly.\(^{15}\)

However its logarithm is very useful:

\[
\ln z! \approx z \ln z - z + \frac{1}{2} \ln(2\pi z)
\] (5.8.22)

where normally only the first two terms are used.\(^{16}\)

The more complete Stirling’s Approximation is:

\[
\Gamma(z + 1) = z! \approx (2\pi z)^{1/2} z^z e^{-z} \left(1 + \frac{1}{12z} + \frac{1}{288z^2} + \cdots\right)
\] (5.8.23)

\(^{15}\)Check it yourself. Ten factorial is 3628800, yet Equation (5.8.21) gives 3598695.6, an error of about 30,000...

\(^{16}\)Again, think of \(z\) as \(10^{23}\). In that case the last term is about 100 which is quite negligible.
Chapter 6

The Canonical Ensemble

We now turn our attention to the canonical ensemble. This was first described by Gibbs\(^1\) and turns out, in a practical way, to be much more useful than the microcanonical ensemble.

The reason for this is simple. The microcanonical ensemble assumes that the energy is fixed. That implies that the system can do no mechanical work (and thus it must have rigid walls) and can neither take up or give up heat (and thus its walls must also be adiabatic.) Such a system is completely isolated from its surrounding. This is not a common situation for real systems.\(^2\)

To partially remove the isolation, we take a single component single phase system with walls that are rigid and impermeable (ensuring constant volume and number of particles) but which are also heat conducting. This allows the energy of the system to vary. What we fix instead\(^3\) is the temperature. We do this by putting the system into contact with a constant temperature bath at a temperature \(T\), which then becomes the system temperature.

The energy of this system is not constant, indeed it can and does fluctuate. What is constant is the temperature of the system, held that way by the external heat bath.\(^4\)


\(^2\)Indeed, the most common systems have fixed number, temperature and pressure. We will meet an appropriate ensemble for that situation later.

\(^3\)We need to fix three independent variables for a single component single phase system.

\(^4\)Extensive variables such as volume, energy, and number of particles can be set without reference to anything on the outside of the system. Intensive variables such as the temperature need external control. We will see other examples of this soon enough.
6.1 The Ensemble

We again follow a Gibbsian approach. We take our system defined above and replicate it $A$ times, once again taking $A$ to be a very large number. In doing this we create another ensemble, this one called the **canonical ensemble**. To ensure that these systems all have the same value of $T$ we insert them into the same constant temperature bath at temperature $T$.

![Schematic Canonical Ensemble](image)

Figure 6.1: Schematic Canonical Ensemble. Black boxes represent systems of fixed $N$, $V$, and $T$; the grey walls represent rigid adiabatic walls; the white space represents a heat bath at a temperature $T$.

At some instant of time we take a snapshot of each of these macroscopic systems. Each one will be in one of the possible energy states of the system $E_j$. These are fixed by the volume and number of particles in the system. If we allowed the systems to evolve in time we’d see that a given system may well move from one energy state to another. But the average number in state $E_j$ must remain constant because otherwise the average properties of the ensemble would change at it would not be at equilibrium, which would contradict our assumption of equilibrium.

We will denote the ensemble members in energy state $E_j$ by $a_j$, noting that many different $E_j$’s may have the same value of energy while many others will have totally different energies.

One constraint on our calculations (just as it was for the microcanonical ensemble) is

$$\sum_j a_j = A. \quad (6.1.1)$$

What we want to do is what we did with the microcanonical ensemble. First, we want to assume that all microstates are equally likely, subject two constraints. One is on the number of ensemble members (Equation (6.1.1)) and the other is the constraint that the temperature is constant.

The second thing we want to do is to find what distribution of the systems among the various energy states is the most likely. Again, that means finding which set of $a$’s will most probably occur.

And once we’ve done that we want to find the *probability* that a given ensemble

---

5Considering the huge size of the ensemble, this will have to be a really huge constant temperature bath. Luckily, we only have to imagine it, not build it.

6Remember, there is no time variable in a Gibbsian ensemble.

7That is, the energy levels are quite likely degenerate.
member is in state \( j \). That’s actually the easiest part because if the set of \( a \)'s that has the maximum chance of occurring is

\[
\{ a_1^*, a_2^*, \ldots, a_j^*, \ldots \},
\]

then the probability of finding a random system in state \( j \) is just

\[
P_j = \frac{a_j^*}{\Omega}.
\]

Unlike the microcanonical ensemble, we also have a heat bath to consider. Probably the best way to deal with it is to consider it as a separate system. We can even think of it as system with fixed \( N, V, \) and \( T \) having a large number of possible internal states just like an ensemble member.

Then we can make use of a fantastically clever idea.\(^8\) That idea is to take the entire ensemble plus the heat bath and wrap it in totally rigid impermeable adiabatic walls. (See Figure 6.1 on page 6-2.)

When we do that, what we have is a new single system that is microcanonical!

We can now formulate a constraint that takes the constant temperature bath into account.\(^9\) We do this by recognizing that the total energy of all the systems plus the heat bath must be constant because of the isolating walls around the entire ensemble and bath. Thus

\[
\sum_j a_j E_j + E_{\text{bath}} = \mathcal{E},
\]

where \( \mathcal{E} \) is the total energy of all the systems plus the bath. The \( E_j \)'s are the energies of the systems and \( E_b \) is the energy of the constant temperature bath.

Introducing the heat bath keeps our systems at a constant temperature \( T \) and also gives us an energy constraint that we can use. It also introduces a complication. We not only have \( W_s \), the number of ways of arranging the canonical systems among their energy states, but we also have \( W_b \), which are the number of ways the heat bath can attain the given temperature \( T \).

As a result the total number of ways of arranging system plus bath is:

\[
W_t = W_s W_b,
\]

and it is \( W_t \) that we need to maximize.

The two \( W \)'s are multiplied because they are independent. This is so because \( W_s \) depends \textit{only} on the \( a_j \)'s and \( W_b \) doesn’t depend on them at all.

---

\(^8\)The author wishes that he knew who invented this idea because that person is the unsung hero of statistical thermodynamics.

\(^9\)This is not trivial. Another approach sometimes taken is to omit the heat bath, allowing the other \( \mathcal{A} - 1 \) systems in the ensemble to form a heat bath for whatever system one is looking at. This works. But the problem is that the temperature then gets introduced in a decidedly circular manner since the temperature of the bath becomes the temperature of the system which in turn can be thought of as part of the bath, which sets the temperature of the bath... Here, as we will see, the temperature is determined by the heat bath—as it should be!
We don’t want to bother to know too much about the bath. All it needs to do is maintain a constant temperature. It is enough for us to know that \( W_b \) must depend on \( E_b \).

### 6.2 The Most Probable Occupation Numbers

So what we have is

\[
W_t = W_s \times W_b = \frac{\mathfrak{a}!}{\prod_j a_j} \times W_b(E_b). \tag{6.2.1}
\]

We want to find the set of \( a \)'s that maximize this subject to the conditions:

\[
\sum_j a_j = \mathfrak{a} \quad \text{and} \quad \sum_j a_j E_j + E_b = \mathbf{e}. \tag{6.2.2}
\]

The technique is much like the one for the microcanonical ensemble. First we take logarithms of Equation (6.2.1) and then apply Stirling’s approximation:

\[
\ln W_t = \ln W_b + \ln \mathfrak{a}! - \sum_j (a_j \ln a_j - a_j). \tag{6.2.3}
\]

Then we write the total derivative and set it to zero to find the maximum.\(^{10}\)

\[
d \ln W_t = \left( \frac{\partial \ln W_b}{\partial E_b} \right) dE_b - \sum_j \ln a_j da_j = 0, \tag{6.2.4}
\]

But we have a problem. There are a large but finite number of \( a_j \)'s. Finite because each energy level \( E_j \) has a large but finite degeneracy. If we call the total number of \( a_j \)'s \( \mathfrak{r} \), then Equation (6.2.4) would seem to have \( \mathfrak{r} + 1 \) independent variables. The extra variable is the energy of the bath. However because of the two constraint equations, we actually have only \( \mathfrak{r} - 1 \) independent variables.

But we do have the two constraint equations (6.2.2) and we can use them to eliminate two variables from Equation (6.2.4) leaving us with the required number of independent variables.

The technique is called Lagrange’s Method of Undetermined Multipliers and works here as follows.

We take the two auxiliary equations and differentiate them:

\[
\sum_j da_j = 0 \quad \text{and} \quad \sum_j E_j da_j + dE_b = 0. \tag{6.2.5}
\]

We can remove the two extra variables by multiplying the two constraint equations by \( \alpha \) and \( \beta \), which are so far of unknown value, to give:

\[
\alpha \sum_j da_j = 0 \quad \text{and} \quad \beta \sum_j E_j da_j + \beta dE_b = 0. \tag{6.2.6}
\]

\(^{10}\)We know it will be a maximum because the smallest value \( W_t \) can have is 1.
These are then subtracted\textsuperscript{11} from Equation (6.2.4) on page 6-4 to get:
\[
d\ln W_t = \left[ \left( \frac{\partial \ln W_b}{\partial E_b} \right) - \beta \right] dE_b - \sum_j (\ln a_j + \alpha + \beta E_j) da_j = 0. \tag{6.2.7}
\]

Now we can get rid of the bath conditions by setting $\beta$:
\[
\beta = \left( \frac{\partial \ln W_b}{\partial E_b} \right), \tag{6.2.8}
\]
and we can also get rid of one $a$, say $a_1$ by picking $\alpha$ so that:
\[
\ln a_1 + \alpha + \beta E_1 = 0. \tag{6.2.9}
\]

What is left is
\[
\sum_{j \neq 1} (\ln a_j + \alpha + \beta E_j) da_j = 0. \tag{6.2.10}
\]

But now all of the remaining $a_j$'s are independent of each other. So the only way that Equation (6.2.10) can be zero for all possible choices of the remaining $da_j$'s is if their coefficients in parentheses above are identically zero. Thus we have:
\[
\ln a_j^* + \alpha + \beta E_j = 0, \tag{6.2.11}
\]
for all $j$, including $j = 1$, as can be seen by checking Equation (6.2.9). The asterisk, as usual, indicates that this value for $a_j$ is the one that maximizes $W_{total}$.

Note that these occupation numbers are not equal to a constant, but depend explicitly upon the energies $E_j$ of the system states. This is exactly what we would expect.

We can now simplify things a good bit. For instance:
\[
a_j^* = e^{-\alpha - \beta E_j}, \tag{6.2.12}
\]
which can be summed:
\[
\sum_j a_j^* = \mathcal{A} = e^{-\alpha} \sum_j e^{-\beta E_j}. \tag{6.2.13}
\]

So $\alpha$ can be found from
\[
e^{-\alpha} = \frac{\mathcal{A}}{\sum_j e^{-\beta E_j}}, \tag{6.2.14}
\]
once $\beta$ is known.

So, assuming that we will determine $\beta$, which we will,
\[
a_j^* = \frac{\mathcal{A} e^{-\beta E_j}}{\sum_j e^{-\beta E_j}}. \tag{6.2.15}
\]

\textsuperscript{11}They are subtracted purely for convenience since $\beta$ has a standard definition and we'd like to agree with it...
Two things: first, the fraction $a^*_j/A$ is simply the fraction of the ensemble members in state $j$ and so it is the probability of finding an ensemble member in that state:

$$P_j = \frac{a^*_j}{A}.$$  \hfill(6.1.2)

And using Equation (6.2.15) on page 6-5 for $a^*_j$

$$P_j = \frac{e^{-\beta E_j}}{\sum_j e^{-\beta E_j}}.$$  \hfill(6.2.16)

The sum in the denominator of Equations (6.2.15) on page 6-5 and (6.2.16) is like $\Omega$ in the treatment of the microcanonical ensemble. It turns up so often that it is given a special symbol, $Q$, and a special name, the **canonical partition function**. It is written more formally as:

$$Q(N, V, \beta) = \sum_j e^{-\beta E_j},$$  \hfill(6.2.17)

because it is a function of $N$, $V$, and $\beta$.

### 6.3 The Thermodynamics of the Canonical Ensemble

The canonical partition function $Q(N, V, \beta)$ is directly connected to a thermodynamic quantity, just as $\Omega(N, V, E)$ for the microcanonical ensemble was connected to the entropy.

Indeed, the reader can probably guess which thermodynamic function that is.

Proving it is a bit more difficult. First we will show that we can compute the energy, pressure, and chemical potential from $Q(N, V, \beta)$. We will then form the total derivative of $Q(N, V, \beta)$ and compare that to the appropriate macroscopic thermodynamic function. From that we hope to not only identify $\beta$, but $Q$ as well.

#### 6.3.1 Energy

We calculate the average energy of the systems in the canonical ensemble. The probability that a system is in energy state $E_j$ is $P_j$. Thus:

$$\langle E \rangle = \sum_j P_j E_j.$$  \hfill(6.3.1)

Since $P_j$ is given by Equation (6.2.16), we get:

$$\langle E \rangle = \frac{1}{Q} \sum_j E_j e^{-\beta E_j}.$$  \hfill(6.3.2)

Now consider $(\partial \ln Q / \partial \beta)_{N,V}$:

$$\left(\frac{\partial \ln Q}{\partial \beta}\right)_{N,V} = \frac{\partial}{\partial \beta} \left( \ln \sum_j e^{-\beta E_j} \right) = -\frac{1}{Q} \sum_j E_j e^{-\beta E_j}.$$
so
\[ \langle E \rangle = - \left( \frac{\partial \ln Q(N, V, \beta)}{\partial \beta} \right)_{N, V}. \] (6.3.3)

This isn’t quite the end of the story. Equation (6.3.3) is a formal equation. That is, one cannot use it directly to compute average energies. What it does tell us is that if we can compute \( Q(N, V, \beta) \) for some system, we can then compute the average energy for that system.

Here’s a simple example:

---

**Example 6.1**

The canonical partition function for a simple gaseous system is found to be
\[ Q(N, V, \beta) = \frac{1}{N!} \left( \frac{B}{\beta} \right)^{3N/2} V^N. \]
where \( B \) is a constant. What is the internal energy for such a system?

We first find \( \ln Q \):
\[ \ln Q = - \ln N! + \frac{3N}{2} \ln B - \frac{3N}{2} \ln \beta + N \ln V, \]
and then find
\[ \left( \frac{\partial \ln Q}{\partial \beta} \right) = - \frac{3N}{2\beta}, \]
from which
\[ \langle E \rangle = \frac{3N}{2\beta}. \]

---

### 6.3.2 Pressure

It is clear that the energy levels of a system free to move inside a volume \( V \) depends on that volume. As an example look at the energy levels for a particle in a three-dimensional cubical box:
\[ \epsilon_{nrs} = \frac{(n^2 + r^2 + s^2)h^2}{8mV^{2/3}}, \] (6.3.4)
where \( n, r, \) and \( s \) are quantum numbers, \( m \) is the mass of the particle, and \( V \) is the volume of the box. Of course \( \epsilon \) is the single particle energy.
Imagine that in a canonical ensemble the volumes of all the systems is changed reversibly and adiabatically\(^\text{12}\) by the same small amount. Then, from thermodynamics:

\[
\frac{dE}{dq} = \frac{pdV}{V} + \mu dN, \tag{6.3.5}
\]

where \(q\) is the heat, \(p\) is the pressure, and \(\mu\) is the chemical potential. With impenetrable walls there can be no change in \(N\), so Equation (6.3.5) reduces to

\[
\frac{dE}{dq} = \rho dV, \tag{6.3.6}
\]

and when the change is adiabatic, \(dq\) is zero. Then Equation (6.3.6) can be written:

\[
\frac{dE_j}{p} = \sum_{\beta} \frac{E_j}{N,\beta}, \tag{6.3.7}
\]

where \(p_j\) is the pressure of the system when it is in the \(j\)'th energy state.\(^\text{13}\) Thus we have:

\[
p_j = -\left(\frac{\partial E_j}{\partial V}\right)_{N,T}. \tag{6.3.8}
\]

The average pressure is then:

\[
\langle p \rangle = -\frac{1}{Q(N,V,\beta)} \sum_j \left[ \left(\frac{\partial E_j}{\partial V}\right) e^{-\beta E_j} \right]. \tag{6.3.9}
\]

Note that:

\[
\left(\frac{\partial \ln Q(N,V,\beta)}{\partial V}\right)_{N,\beta} = \frac{1}{Q(N,V,\beta)} \frac{\partial}{\partial V} \sum_j e^{-\beta E_j} = \frac{1}{Q} \sum_j \left(\frac{\partial (-\beta E_j)}{\partial V}\right) e^{-\beta E_j} = -\frac{\beta}{Q(N,V,\beta)} \sum_j \left[ \left(\frac{\partial E_j}{\partial V}\right) e^{-\beta E_j} \right], \tag{6.3.10}
\]

so that

\[
\beta \langle p \rangle = \left(\frac{\partial \ln Q(N,V,\beta)}{\partial V}\right)_{N,\beta}. \tag{6.3.11}
\]

Our example above can be continued.

---

**Example 6.2**

The canonical partition function for a simple gaseous system is found to be

\[
Q(N,V,\beta) = \frac{1}{N!} \left( \frac{B}{\beta} \right)^{3N/2} V^N. \tag{6.3.12}
\]

\(^{12}\)The meaning of this is that the volume change is so slow that the change takes place without any system moving from one energy level to another.

\(^{13}\)There is no reason to suppose that the pressure will remain constant if the energy levels change their numeric values, and in fact it isn’t.
What is the pressure in such a system?

We again find \( \ln Q \):

\[
\ln Q = -\ln N! + \frac{3N}{2} \ln B - \frac{3N}{2} \ln \beta + N \ln V ,
\]

and then find

\[
\beta \langle p \rangle = \left( \frac{\partial \ln Q(N, V, \beta)}{\partial V} \right)_{N, \beta},
\]

from which

\[
\beta \langle p \rangle = \frac{N}{\beta V},
\]

or

\[
\langle p \rangle = \frac{N}{\beta V}.
\]

### 6.3.3 Chemical Potential

If we consider the possibility of changing the number of particles in a system we can again use Equation (6.3.5):

\[
dE = dq - pdV + \mu dN ,
\]

((6.3.5))

We now hold the volume constant but allow \( N \) to change by one particle. When \( N \) changes by one we can expect the internal energy \( E \) to change as well since that particle will have some energy of its own. The change in \( \mu \) will then be:

\[
\mu = \left( \frac{\partial E}{\partial N} \right).
\]

(6.3.12)

We allow all the systems in the ensemble to change the number of particles they contain by \( dN \) so that all the systems remain the same. Since the number of particles in a system has then changed, the number of systems in a given energy level \( E_j \) will change. The change in the chemical potential is given by:

\[
\mu_j = \left( \frac{\partial E_j}{\partial N} \right),
\]

(6.3.13)

where \( dN \) is meant in the sense of a small change in number of particles when there are a huge number present and \( \mu_j \) is the chemical potential associated with the change in number of particles in level \( E_j \).

The average value of the chemical potential is then given by:

\[
\langle \mu \rangle = \frac{1}{Q(N,V,T)} \sum_j \left( \frac{\partial E_j}{\partial N} \right) e^{-\beta E_j} ,
\]

(6.3.14)
Except for a factor of $\beta$ this is exactly what we get if we differentiate $\ln Q$ with respect to $N$, so

$$
\beta\langle \mu \rangle = -\left( \frac{\partial \ln Q(N,V,T)}{\partial N} \right)_{V,\beta},
$$

(6.3.15)

where the details are left to the reader.

We can once more look at our example:

**Example 6.3**

The canonical partition function for a simple gaseous system is found to be

$$
Q(N,V/\beta) = \frac{1}{N!} \left[ \frac{B}{\beta} \right]^{3N/2} V^N.
$$

What is the chemical potential for such a system?

Here it is best to start with $\ln Q$ in the form

$$
\ln Q = -\ln N! + \frac{3N}{2} \ln \left[ \frac{B}{\beta} \right] + N \ln V.
$$

Using Stirling’s approximation we get for the first term on the right $-\ln N! = -N \ln N + N$, resulting in:

$$
\ln Q = -N \ln N + N + \frac{3N}{2} \ln \left[ \frac{B}{\beta} \right] + N \ln V.
$$

Now

$$
\beta\langle \mu \rangle = -\left( \frac{\partial \ln Q(N,V,T)}{\partial N} \right)_{V,\beta},
$$

from which we get,

$$
\beta\langle \mu \rangle = -\ln N + \frac{3}{2} \ln \frac{B}{\beta} + \ln V,
$$

or

$$
\langle \mu \rangle = \frac{1}{\beta} \ln \left[ \left( \frac{B}{\beta} \right)^{3/2} \left( \frac{V}{N} \right) \right].
$$

6.4 The Canonical Partition Function and the Identification of Beta

The canonical partition function $Q(N,V,\beta)$ is an interesting function. Its derivatives with respect to $\beta$, $V$, and $N$ are $E$, $\beta p$, and $\beta \mu$ respectively. In fact the total
derivative of its logarithm is:

\[
d \ln Q(N,V,T) = \left( \frac{\partial \ln Q}{\partial \beta} \right)_{N,V} \ d\beta + \left( \frac{\partial \ln Q}{\partial V} \right)_{N,\beta} \ dV + \left( \frac{\partial \ln Q}{\partial N} \right)_{\beta,V} \ dN,
\]

or, using the results above

\[
d \ln Q(N,V,\beta) = -\langle E \rangle d\beta + \beta \langle p \rangle dV - \beta \langle \mu \rangle dN,
\]

where we expect that \( \beta \) is somehow a function of the temperature.

Is there a thermodynamic function with that total derivative? Of course the answer is yes. But it isn’t a standard function. Those all have dimensions. Equation (6.4.2) is dimensionless.

Dimensionless thermodynamic equations were discussed in Section 1.7. In fact the desired equivalent to Equation (6.4.2) is Equation (1.7.20) on page 1-9:

\[
d \Phi = -E d\beta + \pi dV + \gamma dN,
\]

where \( d\Phi \) is identified in the first of Equations (1.7.27) on page 1-9:

\[
\beta A = d\Phi = -E d\beta + \pi dV + \gamma dN.
\]

Comparing Equations (6.4.2) and (6.4.4) gives us the identifications we need. Thus \( \beta \) becomes:

\[
\beta = \frac{1}{kT},
\]

and \( \pi \) becomes:

\[
\pi = \frac{p}{kT}
\]

and \( \gamma \) is then:

\[
\gamma = -\frac{\mu}{kT}
\]

and finally \( \ln Q(N,V,\beta) \) is then:

\[
\ln Q(N,V,\beta) = -\frac{A}{kT},
\]

or

\[
A(N,V,T) = -kT \ln Q(N,V,T),
\]

where \( A \) is the Helmholtz free energy.

In fact, we’ve already introduced this notation when we discussed dimensionless thermodynamic equations back in Section 1.7 on page 1-7. The notations \( \beta, \pi, \) and \( \gamma \) are somewhat standard and will often be seen.

We have particular satisfaction in identifying \( \beta \) with \( 1/kT \) since \( \beta \) is a property of the heat bath and the heat bath alone. This can be seen by a quick look at Equation (6.2.8) on page 6-5.

Further, though we will not explore it here, imagine that one has two different systems: 1 and 2. These have parameters \( N_1, V_1, E_{1,j} \) and \( N_2, V_2, E_{2,j} \). We make
ensembles out of both, one with $\mathcal{A}_1$ ensemble members, the other with $\mathcal{A}_2$ ensemble members. We place all of these ensemble members into the same heat bath. Equation (6.2.3) on page 6-4 now has an additional set of terms corresponding to systems 2. But there is still only one heat bath term.

The result is obvious. The two ensembles come to equilibrium separately, but at the same temperature. And the only parameter common to both is $\beta = 1/kT$, just as we’d expect.

6.5 The Entropy

The entropy can be computed from $A(N,V,T)$ for any particular system if one has an equation for $A$ explicitly in terms of $N$, $V$, and $T$.

But here are six lines leading to another way of looking at the entropy:

$$P_j = \frac{e^{-\beta E_j}}{Q}.$$  (6.5.1)

$$\ln P_j = -\beta E_j - \ln Q.$$  (6.5.2)

$$- P_j \ln P_j = \beta P_j E_j + P_j \ln Q.$$  (6.5.3)

$$- \sum_j P_j \ln P_j = \beta \langle E \rangle + \ln Q \sum_j P_j.$$  (6.5.4)

$$- \sum_j P_j \ln P_j = \beta \langle E \rangle - \ln Q = \beta E - \beta A = \frac{S}{k},$$  (6.5.5)

and finally

$$S = -k \sum_j P_j \ln P_j = -k \langle \ln P \rangle,$$  (6.5.6)

where the last follows from the definition of a property. Thus the entropy can be thought of as the average value of the logarithm of the occurrence of a particular state. Indeed, we can identify $(-k \ln P_j)$ as the contribution of state $j$ to the total entropy of the system.\(^{14}\)

In words, the entropy of a canonical system depends directly on the probabilities that the system will be found in each of the various states $j$. Of course those probabilities in turn depend on the energies of the states; see Equation (6.5.1).

Since the probabilities are all less than one, the smaller the probability the larger its logarithm is in absolute value. So large entropies depend on the average occupancy

\(^{14}\)Note that if a given $P_j$ is zero that term doesn’t count in the sum since $P_j \ln P_j$ goes to zero as $P_j$ goes to zero and that state does not enter into the sum in Equation (6.5.6).
Chapter 6. The Canonical Ensemble

of state $j$ being very small. One or a relative few highly probable states contribute very little to the entropy.

Our insight is that even opening up a very high energy state to occupation will increase the entropy of the system significantly.

It should also be noted that Equation (6.5.6) on page 6-12 is quite general. We’ve already seen it in Equation (5.6.4) on page 5-8.

6.6 Degeneracy

It is almost always the case that the energy levels available to a system are degenerate. Let’s denote the degeneracy of level $j$ as $\Omega(N, V, E_j)$. This can be as small as 1 or, much more often, a very large number.

In the microcanonical ensemble there is only one allowed energy $E$, so the degeneracy was written $\Omega(N, V, E)$, with no index $j$.

With this notation we can write the canonical partition function $Q(N, V, \beta)$ as

$$Q(N, V, \beta) = \sum_j \Omega(N, V, E_j) e^{-\beta E_j},$$

(6.6.1)

where the sum is now over distinct energy levels instead of counting each quantum state separately.

This form suggests that the canonical partition function $Q$ is some sort of transformation\(^{15}\) of the partition function $\Omega$ for the microcanonical ensemble. This is in fact true and will be explored in a later chapter.

6.7 The Simple Spin System in an External Field

In Section 5.7 on page 5-9 we discussed the most simple spin system possible. Here we are going to add an external magnetic field that interacts with the spins. This adds an energy to the model. Otherwise the model remains the same.

Here’s the model: Let us assume that we have particles that have two possible spins, which we shall denote as up and down. There are $M$ spins and $N$ of them are up and the remainder are down. Again $M$ plays the role of a volume since it gives the extent of the system. We use a symmetric energy scheme in which the energy of the up spins is $\epsilon$ and that of the down spins is $-\epsilon$.

Then for $N$ up spins the total energy $E$ is

$$E = N\epsilon - (M - N)\epsilon = (2N - M)\epsilon.$$  

(6.7.1)

Given $M$ spins with $N$ of them up spins, there are $M!/N!(M-N)!$ ways of arranging

\(^{15}\)It looks like a Laplace Transformation, or would if there was an integral instead of a sum.
them. Thus the canonical partition function for this system is:

\[ Q(\beta, M) = \sum_{N=0}^{M} \frac{M!}{N!(M-N)!} e^{-\beta(2N-M)\epsilon}, \quad (6.7.2) \]

which looks difficult to solve. But it isn’t. If we factor \( \exp(\beta M \epsilon) \) out of the sum we get:

\[ Q(\beta, M) = e^{\beta M \epsilon} \sum_{N=0}^{M} \frac{M!}{N!(M-N)!} e^{-2N\beta \epsilon}, \quad (6.7.3) \]

which the experienced will recognize as the series expansion

\[ (1 + x)^M = \sum_{N=0}^{M} \frac{M!}{N!(M-N)!} x^N, \quad (6.7.4) \]

with \( x = e^{-2\beta \epsilon} \).

The rest is easy.

\[ Q(\beta, M) = e^{\beta M \epsilon} \left[ 1 + e^{-2\beta \epsilon} \right]^M, \quad (6.7.5) \]
\[ Q(\beta, M) = \left[ e^{\beta \epsilon} + e^{-\beta \epsilon} \right]^M, \quad \text{and} \]
\[ Q(\beta, M) = \left[ 2 \cosh(\beta \epsilon) \right]^M. \quad (6.7.6) \]

The last equation comes from the definition of the hyperbolic cosine.\(^{16}\)

The average energy of this system can be found from Equation (6.3.3) on page 6-7. We note that \( \ln Q(\beta, M) \) is given by:

\[ \ln Q(\beta, M) = M \ln[2 \cosh(\beta \epsilon)], \quad (6.7.8) \]

and differentiating with respect to \( \beta \) (and changing the sign) gives:

\[ \langle E \rangle = -M \epsilon \frac{\sinh(\beta \epsilon)}{\cosh(\beta \epsilon)}, \quad (6.7.9) \]
\[ \langle E \rangle = -M \epsilon \tanh(\beta \epsilon). \quad (6.7.10) \]

A graph of the average energy per particle (in units of \( \epsilon \) versus \( T \)) is given in Figure 6.2 on page 6-15.

Note that in the figure the units of energy are \( \langle E \rangle / M \epsilon \) and the horizontal axis is drawn with \( k/\epsilon \) set equal to 1.

This is a somewhat curious result and can use some explanation. At low temperatures all spins are in the \(-\epsilon\) state and hence the average energy per spin is \(-1\) in units of \( \epsilon \). At very high temperatures half the spins are in each of the two states and thus the average energy per spin is 0. However it takes a very high temperature to reach this point, perhaps a temperature of 20 or more in these units.

\(^{16}\)The hyperbolic functions are often unfamiliar but in cases like this they simplify things enormously.
The heat capacity \( C_M \) of such systems is also interesting. To obtain it we must differentiate Equation (6.7.10) on page 6-14 with respect to \( T \). It is best to do this indirectly. Thus we differentiate with respect to \( \beta \) and then differentiate \( \beta \) with respect to \( T \). This last is easy since we find that \( d\beta/dT = -1/kT^2 \). Differentiating a hyperbolic tangent is a bit more difficult but not impossible.\(^\text{17}\) We find that
\[
\frac{d\tanh(ax)}{dx} = a \text{sech}^2(ax).
\]
So:
\[
C_M = \left( \frac{\partial E}{\partial T} \right)_M = \frac{1}{kT^2} \left( \frac{\partial E}{\partial \beta} \right) = \frac{M}{kT^2} \epsilon^2 \text{sech}^2(\epsilon/kT) \tag{6.7.11}
\]

Gaining an understanding of what this function looks like can be helped by looking at Figure 6.3 on page 6-16. Here \( C_M/kM \) is plotted against a reduced temperature obtained by setting \( k/\epsilon \) equal to 1. At low temperatures where all the spins are in the low energy state, the heat capacity is zero because the slope of the energy against temperature curve is zero (see Figure 6.2). Since the higher energy level can never have more particles than the lower one, the two levels must eventually have equal populations and the heat capacity must drop to zero.

\(^{17}\) Especially if you look it up as I did.
Figure 6.3: Reduced Heat Capacity per Particle vs. Reduced Temperature
Chapter 7

Independent Subsystems in the Canonical Ensemble

7.1 Fundamentals

To this point we have focused attention on entire systems. We have assumed that the system energy levels, $E_i$, were or could be known. In fact this is almost never the case since to know the system energy levels we’d have to know the system wave function. And in general we almost never know these since they are usually too difficult to compute.

The primary reason for this is that the particles in a system generally interact. And those interactions (even if we knew them with great accuracy, which we do not) make Schrödinger’s Equation impossible to solve.\footnote{What one has in that case is a system of $3N$ coupled second order partial differential equations to solve simultaneously. Analytic solution is out of the question. Numeric solution, for any value of $N$ we’d regard as reasonable for statistical thermodynamics is just silly as upwards of several moles of memory would be needed, not to mention the the computer power!}

But there is one class of systems that is simple enough so that we can find the system energy levels. This is a model of a system which is made up of \textit{independent subsystems}.

In this model each subsystem in the system is assumed to have no interaction whatsoever with any other subsystem. We will assume for convenience that the independent subsystems are single particles. This works because the motions of polyatomic molecules can, as shown on page 2-8, can be broken into two parts, the motion of the center of mass and motion about the center of mass. We assume here that any motion about the center of mass does not affect any other particle. As a
result we often speak of “particles” rather than “independent subsystems”

Of course all such models of ideal systems involve a certain amount of magic. For instance, if the particles are mobile but are to be contained in a volume $V$, the particles must interact with the walls bounding that volume. And the walls must be made of some sort of particles. But this contradicts the idea that there are no particle-particle interactions.

The standard way around this lies in the fact that particle-wall interactions affect only molecules near the walls. If the system is assumed to be large and not of some strange shape, most molecules are in the bulk and not near the walls. The ratio of molecules near the walls to the total number of molecules at constant density goes to zero as the volume increases.\(^2\) This assumes that the volume is in some sense “regular” and not some pathological shape with walls of huge area. So we assume that the volume is large and the density $N/V$ essentially constant. This is called the thermodynamic limit.

### 7.2 Independent Energies

From quantum mechanics we know that an $N$-particle system in a volume $V$ will have a Hamiltonian operator $H$ that yields the eigenvalue equation:

$$H \Psi_j = E_j \Psi_j,$$

(7.2.1)

where $\Psi_j$ is the system wave function corresponding to the $j$’th $N$ particle system state whose energy is $E_j$.\(^3\)

If the particles in this system are independent, then the system Hamiltonian $H$ breaks down into a sum of single particle Hamiltonians $h$:

$$H = h_1 + h_2 + \cdots + h_N.$$

(7.2.2)

This is so because independent in the quantum mechanical sense means that there are no terms in the Hamiltonian that couple the behavior of particle $i$ to particle $j$ where $i \neq j$. Since the momentum part of the Hamiltonian consists of independent terms, any coupling must come from the potential energy part of the Hamiltonian. Clearly non-interacting particles will be uncoupled. But there still can be an external potential (such as a gravitational field) that gives each particle a potential energy.\(^4\)

If there is no coupling the total energy of the system $E_j$ becomes the sum of the energies of the individual particles:

$$E_j = \epsilon_{j,1} + \epsilon_{j,2} + \ldots$$

(7.2.3)

---

\(^2\)This is trivial to prove in the case of a cubical container of wall length $L$. The volume increases as $L^3$, while the surface area increases as $6L^2$. The ratio of the two goes to zero as $6/L$ as $L$ goes to infinity.

\(^3\)Purists will note that I have slurred over a number of difficulties here since it is very unlikely that any such system will exist in a pure state. In the end that makes no real difference to the argument.

\(^4\)This situation will be examined in a later chapter.
We now assume, for simplicity, that the particles are identical. This means that the single particle Hamiltonians $h$'s for each particle will be identical. Then each particle will have the same set of single particle energy levels.

### 7.3 Single Particle Canonical Partition Functions

Now let $q_a$ be a single particle canonical partition function defined by the following equation:

$$ q_a(V, \beta) = \sum_j e^{-\beta \epsilon_{a,j}} $$

(7.3.1)

where $\epsilon_{a,j}$ denotes the energy of the $a$'th particle when it is in the $j$'th single particle quantum state. Of course $q_a$ is not a function of the number of particles, but it is a function of both the volume $V$ and the temperature $\beta$.

Since each particle has the same set of energy states, all the $q_a$'s sum to exactly the same value and so are the same, independent of the subscript $a$. So we can denote any one of them by the simple symbol $q$.

With this definition it will be true that

$$ Q(N, V, \beta) = \sum_j e^{-\beta E_j} = \frac{1}{N!} q^N. \quad \text{or sometimes just } q^N $$

(7.3.2)

To explain the “or sometimes just” in the equation above takes us ahead of the story. But the short answer is that the $N!$ occurs if the particles are assumed to be indistinguishable in the sense of quantum mechanics. If, for some reason, the particles could be distinguished (perhaps by location, as in a crystal), then there is no $N!$ and we have the “or sometimes just” part of the equation.

The truth of Equation (7.3.2) is best seen by working backwards from it. Ignoring the factorial for now, the right-hand side of Equation (7.3.2) can be written

$$ q^N = \sum_j e^{-\beta \epsilon_{1,j}} \sum_k e^{-\beta \epsilon_{2,k}} \sum_l e^{-\beta \epsilon_{3,l}} \cdots \sum_m e^{-\beta \epsilon_{N,m}} $$

(7.3.3)

where the number tells us which particle is having its energy levels summed and the index $i, j, \text{etc.}$, denotes the actual energy levels.

The first sum in Equation (7.3.3) refers to the first particle, whichever one that is, the second sum to the second particle, etc.

We now multiply Equation (7.3.3) out:

$$ q^N = e^{-\beta \epsilon_{1,1}} e^{-\beta \epsilon_{2,1}} e^{-\beta \epsilon_{3,1}} \cdots + e^{-\beta \epsilon_{1,1}} e^{-\beta \epsilon_{2,2}} e^{-\beta \epsilon_{3,1}} \cdots + e^{-\beta \epsilon_{1,1}} e^{-\beta \epsilon_{2,3}} e^{-\beta \epsilon_{3,2}} \cdots $$

$$ + e^{-\beta \epsilon_{1,2}} e^{-\beta \epsilon_{2,1}} e^{-\beta \epsilon_{3,1}} \cdots + e^{-\beta \epsilon_{1,2}} e^{-\beta \epsilon_{2,2}} e^{-\beta \epsilon_{3,1}} \cdots + e^{-\beta \epsilon_{1,2}} e^{-\beta \epsilon_{2,3}} e^{-\beta \epsilon_{3,2}} \cdots $$

$$ + e^{-\beta \epsilon_{1,3}} e^{-\beta \epsilon_{2,1}} e^{-\beta \epsilon_{3,1}} \cdots + e^{-\beta \epsilon_{1,3}} e^{-\beta \epsilon_{2,2}} e^{-\beta \epsilon_{3,1}} \cdots + e^{-\beta \epsilon_{1,3}} e^{-\beta \epsilon_{2,3}} e^{-\beta \epsilon_{3,2}} \cdots $$

$$ + \cdots. $$

(7.3.4)

5This restriction is easily removed.

6Meaning that they don’t just look alike, they are identical. There is a deeply profound difference between the two.
In Equation (7.3.4) on page 7-3 above, in each energy subscript the first index is the particle number and the second is the energy level. Each line starts with the first particle in a different energy state while the rest of the line contains all the other particles in all possible energy states.

The exponents in each product in each line can be gathered together:

\[ q^N = e^{-\beta_1\epsilon_1 + \epsilon_2 + \epsilon_3 + \ldots} + e^{-\beta_2\epsilon_1 + \epsilon_2 + \epsilon_3 + \ldots} + e^{-\beta_3\epsilon_1 + \epsilon_2 + \epsilon_3 + \ldots} + \ldots \]

(7.3.5)

where again the first subscript is the particle number and the second is the energy level. Perhaps a simpler example would be helpful. Let us assume that we have only three particles, i.e., \( N = 3 \). And let us assume that there are only three energy levels rather than an infinite number. Then, again omitting the \( 1/N! \):

\[ Q(3, V, \beta) = q_1 q_2 q_3 = \sum_{i=1}^{3} e^{-\beta_1 \epsilon_i} \sum_{j=1}^{3} e^{-\beta_2 \epsilon_j} \sum_{l=1}^{3} e^{-\beta_3 \epsilon_l}. \]  

(7.3.6)

What we are trying to do is to show that the product of these sums is equal to the sum of the products.

\[ Q(3, V, \beta) = (e^{-\beta_{11}} + e^{-\beta_{12}} + e^{-\beta_{13}}) \times (e^{-\beta_{21}} + e^{-\beta_{22}} + e^{-\beta_{23}}) \times (e^{-\beta_{31}} + e^{-\beta_{32}} + e^{-\beta_{33}}), \]  

(7.3.7)

and, as a result,

\[ Q(3, V, \beta) = \prod_{j=1}^{3} e^{-\beta_{ij}}, \]  

(7.3.8)
and it can easily be seen that we are summing over every possible system energy.

It should be pointed out that in general, if the series involved are finite, the sum and the product can always be interchanged. Though it isn’t obvious, this is true here. Certainly $N$ is finite, and because $\beta$ is non-zero, $T$ is finite. With $T$ finite, there is a finite upper system energy level that can be occupied.\(^7\)

If the particles are not distinguishable, Equation (7.3.3) on page 7-3 contains a deception.\(^8\) We’ve made sure to point out that the order of the particles was preserved by including the particle index number.

If we were to calculate how many particles are in system energy level $j$, and we will later on, we would know their number, but not which particles they were.

Look for example at the first line of Equation (7.3.5) on page 7-4. The first term has all the particles in energy level 1. The second has one particle in energy level 2 and all the others in energy level 1. And the third term has exactly the same thing! They differ only as to the identity of the particle in energy level 2.

In reality there should only be one such term in Equation (7.3.2) on page 7-3 instead of the $N$ terms that actually occur. Because we have ignored the indistinguishability property of the particles, we have too many terms in Equation (7.3.5) on page 7-4!

Thus this way of making up the system energy levels from the particle energy levels fails. We need a more sophisticated approach.

But there is a quick fix. And that is to realize that we have done the counting by first lining up all the particles in a single line (so that there is a particle 1 and a particle 2) and then produced Equation (7.3.5) on page 7-4 from it. But there are $N!$ ways to produce that line of particles, and each and every one of those will produce Equation (7.3.5) on page 7-4.

The quick fix is to divide Equation (7.3.5) on page 7-4 by $N!$. This works, although why it works is a deeper question that we will examine later.

The result of our combinations and permutations is:

$$Q(N, V, \beta) = \frac{1}{N!} q(V, \beta)^N.$$  \hspace{1cm} (7.3.10)

This is the proper result for indistinguishable non-interacting particles.

The trick of dividing by $N!$ to account for permutations of indistinguishable particles leads to what are often called Boltzmann statistics.

---

\(^7\)The reader should think about this.

\(^8\)This deception, as I have called it, became apparent to Gibbs (though in a slightly different context.) The equations he derived in the ideal gas case for the entropy were not extensive in the number of particles. The cure was to divide by $N!$, as shown below, although Gibbs had no real idea why one did that. Gibbs, of course, used classical mechanics. Had he lived just a bit longer (he died in 1903) he may well have concluded that the classical mechanics he used was wrong and that energy had to be quantized, and particle indistinguishability had to be true.
7.4 Thermodynamics of Independent Subsystems

When dealing with canonical systems made up of independent subsystems there is no need to compute $Q(N,V,\beta)$ before finding $\langle E \rangle$, $\langle p \rangle$, or $\langle \mu \rangle$. These can be computed from $q(V,\beta)$ directly.

The probability that a single independent subsystem will be found in subsystem state $j$ is clearly

$$ P_j = \frac{e^{-\beta \epsilon_j}}{q} $$

where here $P_j$ is the subsystem probability. Thus it is just as easy to find the appropriate formulas for independent subsystems directly from those given in Section 6.3 above. I repeat them here:

$$ \langle E \rangle = -\left( \frac{\partial \ln Q(N,V,\beta)}{\partial \beta} \right)_{N,V} \tag{6.3.3} $$
$$ \langle \beta p \rangle = \left( \frac{\partial \ln Q(N,V,\beta)}{\partial V} \right)_{N,T} \tag{6.3.11} $$
and

$$ \langle \beta \mu \rangle = -\left( \frac{\partial \ln Q(N,V,T)}{\partial N} \right)_{V,\beta} \tag{6.3.15} $$

All we need do is substitute $q^N/N!$ (usually) or $q^N$ (for crystals and similar systems) as appropriate. Using the first form we get simply:

$$ \langle E \rangle = -N \left( \frac{\partial \ln q(V,T)}{\partial \beta} \right)_{V,N} \tag{7.4.2} $$
$$ \langle \beta p \rangle = N \left( \frac{\partial \ln q(V,T)}{\partial V} \right)_{T,N} \tag{7.4.3} $$
and

$$ \langle \beta \mu \rangle = -\ln \left( \frac{q(V,T)}{N} \right) \tag{7.4.4} $$

The entropy is given by:

$$ S = -Nk \sum_j P_j \ln P_j \tag{7.4.5} $$

where the probability $P_j$ is for an independent subsystem and the entropy is for $N$ particles. If $N$ is Avogadro’s number, then the entropy is per mole of such subsystems.

For crystals and similar systems Equation (7.4.4) changes and becomes

$$ \langle \beta \mu \rangle = -\ln q(V,T) \tag{7.4.6} $$

---

9 Don’t confuse this with system probabilities. See the Equation (7.4.5) for the entropy for an example of the difference.

10 The $N!$ matters only in the computation of the chemical potential.
as can be seen by using Equation (6.3.15) on page 6-10 along with $Q(N,V,T) = q(N,V)^N$, since no $N!$ is needed for such systems.

The Helmholtz free energy is

$$-\beta A = \ln Q(N,V,\beta)$$  \hspace{1cm} (7.4.7)

In the case where $Q = q^N/N!$, we have, using Stirling’s Approximation:

$$-\beta A = -N \ln N + N \ln e + N \ln q = N \ln(qe/N)$$  \hspace{1cm} (7.4.8)

so that

$$A = -NkT \ln(qe/N)$$  \hspace{1cm} (7.4.9)
Chapter 8

The Ideal Monatomic Gas

8.1 The Partition Function

We now construct a model for a monatomic gas of \( N \) independent particles contained in a cubical box of volume \( V \).\(^1\) These particles have no internal energies except for the possibility of electronic excitation. An obvious modeling method is to treat the atoms as point particles contained in a cubical box of volume \( V \). We then need to find the energy levels for such a gas. Once we have them we can use them in the canonical partition function to calculate the thermodynamic properties of such a gas.\(^2\)

Since the atoms are assumed to be independent, each atom acts as if the others were not there. The quantum mechanical particle in a box then provides a good model for the translational energy levels in this system. The one added complication is that the atom also has electronic energy levels.

Luckily, the electronic energy levels are completely independent of the kinetic motion of the atom. So the total energy for any atom is the sum of a translational part and an electronic part or:

\[
\epsilon = \epsilon_{tr} + \epsilon_e, \tag{8.1.1}
\]

and it can be seen that the single particle partition function \( q \) can be written as:

\[
q = \sum e^{-\beta \epsilon} = \sum e^{-\beta (\epsilon_{tr} + \epsilon_e)} = \sum e^{-\beta \epsilon_{tr}} e^{-\beta \epsilon_e} = \sum e^{-\beta \epsilon_{tr}} \sum e^{-\beta \epsilon_e} = q_{tr} q_e. \tag{8.1.2}
\]

We will ignore \( q_e \) for now and discuss it only near the end of this chapter in Section 8.5 on page 8-11.

\(^1\)This, of course, restricts the direct applicability to helium, neon, argon, krypton, xenon, and radon. None of these is known for being chemically reactive to any major degree...

\(^2\)Any partition function could be used, the canonical is chosen because it is simple to use here. Indeed, in later chapters we will repeat this calculation using other ensembles.
One of the few simple problems of quantum mechanics is the single particle in a cubical box. We chose a cubical box for convenience. A non-cubical box is a simple adaptation of these calculations and spherical or spheroidal boxes can also be used with a large increase in mathematical complexity and no increase at all in insight into what is going on. The energy levels depend upon three quantum numbers \( n, r, \) and \( s \) according to:

\[
\epsilon_{n,r,s} = \frac{\hbar^2 (n^2 + r^2 + s^2)}{8mL^2},
\]

(8.1.3)

where \( \epsilon \) is the energy, \( \hbar \) is Planck’s constant, \( m \) the mass of the particle, and \( L \) the length of one edge of the cubical box. Since \( V \), the volume of the box is \( L^3 \), we see that the energy levels depend on the volume to the \( 2/3 \) power.

The single independent particle partition function \( q \) is then:

\[
q(V,\beta) = \sum_n \sum_r \sum_s e^{-\beta \hbar^2 (n^2 + r^2 + s^2)/8mL^2},
\]

(8.1.4)

or

\[
q(V,\beta) = \left( \sum_{k=0}^{\infty} e^{-\beta \hbar^2 k^2/8mL^2} \right)^3,
\]

(8.1.5)

where \( k \) is an integer.

In almost all textbooks this sum is done by noting that the energy levels are very densely packed and, using that as an excuse, then converting the sum to an integral. But can one simply change a sum to an integral? The answer is not always! But it works here.

The correct thing to do is to use the Euler-Maclaurin Summation Formula which is discussed at some length in an Appendix to this chapter contained in Section 8.6 on page 8-15. This appendix should be consulted if the reader is not familiar with the Euler-Maclaurin summation formula.

### 8.2 The Evaluation of the Partition Function

The single-particle canonical partition function is:

\[
q(V,\beta) = \left( \sum_{k=0}^{\infty} e^{-\beta \hbar^2 k^2/8mL^2} \right)^3,
\]

(8.1.3)

---

\(^3\)One needs to solve Schrödinger’s equation for the appropriate boundary conditions. A non-cubical box is simple; it is just a composition of three one-dimensional particle-in-a-box results for three different lengths, one in each dimension. Solutions for a spherical box will involve Bessel functions and a spheroidal box, elliptical functions. The results will, however, show no different general behavior.

\(^4\)Even though this is not true as astute students note. As \( k \) increases the spacing between energy levels increases. Thus for large enough \( k \) the spacing will be as large as one wishes.

\(^5\)The author is well aware that almost no textbook ever discusses the Euler-Maclaurin formula. However, he has gotten tired of simply flailing his arms around in class while pretending that the conversion of a sum to an integral is obvious! It isn’t.

\(^6\)Which, in practice, means almost everyone should look at it.
so we need to do the sum

$$\sum_{k=0}^{\infty} e^{-\beta h^2 k^2 / 8mL^2}.$$  \hfill (8.2.1)

To use the Euler-Maclaurin summation formula we need some derivatives. To this end it helps to gather all the constants together in one bunch and call them, say, $a$:

$$a = \frac{\beta h^2}{8mL^2},$$  \hfill (8.2.2)

so that Equation (8.2.1) becomes

$$S = \sum_{k=0}^{\infty} e^{-ak^2},$$  \hfill (8.2.3)

with $S = q(V, \beta)$, and then tabulate some derivatives (see Table 8.1). All the

<table>
<thead>
<tr>
<th>function</th>
<th>at $x = 0$</th>
<th>as $x \to \infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(x) = e^{-ax^2}$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$f'(x) = -2axe^{-ax^2}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$f^{(3)}(x) = [4a^2x(3 - 2ax^2)]e^{-ax^2}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$f^{(5)}(x) = -8a^3x(4a^2x^4 - 20ax^2 + 15)e^{-ax^2}$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 8.1: Derivatives for the Evaluation of the Translational Partition Function

needed derivatives are zero at both limits. The even numbered derivatives are not zero, but those do not occur in the Euler-Maclaurin series. That series is then simply:

$$S = \int_0^{\infty} e^{-ax^2} \, dx + \frac{1}{2}. $$  \hfill (8.2.4)

The integral is easily done:

$$\int_0^{\infty} e^{-ax^2} \, dx = \frac{1}{2} \left( \frac{\pi}{a} \right)^{1/2}. $$  \hfill (8.2.5)

The result is that the sum is best approximated by

$$S = \frac{1}{2} \left( \frac{2\pi mL^2}{\beta h^2} \right)^{1/2} + \frac{1}{2}. $$  \hfill (8.2.6)

Doing an order of magnitude comparison here shows that $S$ is going to be about $10^{10}$ or so not counting the temperature. Thus the $1/2$ can be neglected without worry.

Putting it all together (and not forgetting to cube the result), the partition function $q$ is then:

$$q(V, \beta) = \left( \frac{2\pi m}{\beta h^2} \right)^{3/2} V, $$  \hfill (8.2.7)

where the $V$ comes from the term $L^2$ that occurs inside a square root and is then cubed.
8.3 The Degeneracy Calculation

There is another way to do obtain the single particle partition function $q$. It is a bit more complex but sometimes more useful in other applications. Above, in Section 8.2 on page 8-2 we recognized that a cubical box has three identical sets of energy levels, one for each direction. So we summed over all the energy levels, counting each one separately, and cubed the result.

Instead of doing that we could use the degeneracy of the energy levels to our advantage. Given the degeneracies we can then need to sum only over the different energy levels, each multiplied by the degeneracy.

The translational energy levels can be degenerate in three ways:

1. **Permutation degeneracy** — it is clear that quantum numbers (5, 3, 1) for example will give the same energy as quantum numbers (3, 5, 1) This type of degeneracy only happens when two or more sides of the container are the same length or the lengths are rational multiples of each other.

2. **Accidental degeneracy** — it is also clear that energy levels can have identical values of energy even for differing sets of quantum numbers. For example $\epsilon_2, 2, 5$ is identical in energy to $\epsilon_1, 4, 4$.

3. **Neighborhood degeneracy** — If we were to relax the meaning of identical slightly to allow for almost identical energies, it will turn out that there are many such degeneracies depending on what value we use for almost.\(^7\)

This last point bears close examination.

We can simplify the energy level expression Equation (8.1.3) on page 8-2 by introducing a new variable $R$:

$$R^2 = n^2 + r^2 + s^2,$$

(recall that $n$, $r$, and $s$ are quantum numbers) then:

$$\epsilon = \frac{\hbar^2 R^2}{8mL^2},$$

and conversely

$$R^2 = \frac{8mL^2 \epsilon}{\hbar^2}.$$  \hfill (8.3.3)

If we denote the degeneracy of levels with energy $\epsilon_R$ by $\omega(\epsilon)$, we can write the single particle canonical partition function as:

$$q(\beta, V) = \sum_{\epsilon=0}^{\infty} \omega(\epsilon)e^{-\beta \epsilon},$$ \hfill (8.3.4)

where we are summing over values of $\epsilon$ generated by Equation (8.1.3) on page 8-2. Since $R^2 = n^2 + r^2 + s^2$, $R$ takes on integer values only with $\epsilon(R^2)$ now a function of $R$.

\(^7\)The author is not trying to be facetious. The word *almost* is to be interpreted as a small relative quantity in this sense of $d\epsilon$. 
Often there is no energy level for a given value of $R^2$. In that case $\omega(\epsilon)$ is zero. It is very difficult to calculate $\omega(\epsilon)$ exactly. It is not at all a smooth function of $R^2$. But it can be approximated.

To get a useful approximation we have to realize that since $R^2$ is constructed from $n^2$, $r^2$, and $s^2$, there is a value of $R^2$ for each possible value of the sum of $n^2 + r^2 + s^2$. Now $n$, $r$, and $s$ are restricted to being positive integers. Imagine a three-dimensional graph with axes named $n$, $r$, and $s$. Mark out points on this graph where $n$, $r$, and $s$ have positive integer values. Those points all lie in the octant of the graph where all the axes are positive. Each point in the graph corresponds to a possible set of quantum numbers.

For example, consider the point at $n = 2$, $r = 2$, and $s = 5$. This corresponds to $R = \sqrt{33}$ and the point is at a distance of 5.74456+ from the origin. Another point, this one at $n = 1$, $r = 4$, and $s = 4$ also has $R = \sqrt{33}$ and is also at a distance of 5.74456+ from the origin. Hence these two states have the same energy as does any other point whose squared distance from the origin is 33.

All points that lie on a spherical surface of radius $R = \sqrt{33}$ will have the same energy. These points are all degenerate in energy, and the number of such points is the degeneracy of the energy for which $R^2 = 33$.

Now the value of 33 was chosen more or less at random. What was written is true for any value of $R^2$ you might pick. Of course, for some $R^2$'s the degeneracy might be zero, but for almost all others it will not.

That is in itself amazing. It does not make $\omega(R^2)$ any easier to calculate. Nor is it an insight that understanding that the points at a distance $R$ from the origin lie on a spherical surface at a radius $R$ around the origin. We still have to compute the number of points lying on the surface of that sphere. But here’s the insight: If we knew the volume of a sphere with such a radius we could calculate how many points that sphere would contain in its volume. And once we have the volume, the number of points lying in a surface region between $R^2$ and $R^2 + d(R^2)$ is given for spheres by the derivative of the volume with respect to $R^2$. Or, better, we can express the volume as a function of the energy $\epsilon$ and then look at the spherical shell lying between $\epsilon$ and $\epsilon + d\epsilon$.

Of course this is only approximate since we will include fractional volumes. But the error is not large. And the average value of $R^2$ at normal temperatures is huge, making the error even smaller. Further, the shell has a thickness of $d\epsilon$, so the points included in it are only approximately degenerate. But $d\epsilon$ can be chosen to be as small as we like so the approximation can be as accurate as we like.

Let’s now carry out this program. The volume of a sphere of radius $R$ is just $4\pi R^3/3$. We are only concerned with 1/8’th of this, the part where all axes are

---

Footnotes:

8 If the box is not cubical, the “markings” on the axes of this graph will be differently spaced in each direction but will still define a set of unconnected points.

9 If the box is not a cube, then the surface is not spherical but instead an ellipsoid with three different axes. The analysis remains the same.

10 This doesn’t work for all shapes. Consider a cube, for example. The volume of a cube of side $L$ is $L^3$, the derivative of that is $3L^2$, but the area of the sides is $6L^2$. 
positive. Calling this volume Φ, we then have:

\[ \Phi = \frac{\pi R^3}{6} = \frac{\pi}{6} \left( \frac{8mL^2\epsilon}{\hbar^2} \right)^{3/2}. \]  

(8.3.5)

The degeneracy factor is then:

\[ \omega(\epsilon)d\epsilon = \frac{d\Phi}{d\epsilon} = \frac{\pi}{4} \left( \frac{8mL^2}{\hbar^2} \right)^{3/2} \epsilon^{1/2} d\epsilon. \]

(8.3.6)

It is important to gauge the size of this number. If we take \( \epsilon \) as \( 3kT/2 \), \( T = 300K \), \( m \approx 10^{-25} \text{ kg} \), \( L = 0.01 \text{ meters} \) (10 cm), and \( d\epsilon \) to be about 0.0001 \( \epsilon \) (a 0.01% thickness of the shell), then \( \omega(\epsilon) \) is about \( 10^{26} \).

Considering that we have a much less than a mole of gas under these conditions, it is easy to see that there are far more degenerate levels available to a gas molecule than there are molecules. Thus the average occupation number of a given quantum level is hugely less than 1. This is an important point because as long as it is true, quantum mechanical effects can be ignored.\(^\text{11}\)

With this result for the degeneracy, the single particle canonical partition function is then the sum

\[ q(\beta, V) = \sum_\epsilon \omega(\epsilon)e^{-\beta\epsilon} = \sum_\epsilon \frac{\pi}{4} \left( \frac{8mL^2}{\hbar^2} \right)^{3/2} \epsilon^{1/2} e^{-\beta\epsilon}. \]  

(8.3.7)

It only remains now to do this sum.

Just as before, almost all textbooks do the sum by noting that the energy levels are very densely packed and, using that as a rationale, then converting the sum to an integral which can be readily done. We’ve already done this more correctly by applying the Euler-Maclaurin summation formula and discovering that the integral really is an excellent approximation to the sum. The result is the same one we’ve already obtained:

\[ q(V, \beta) = \left( \frac{2\pi m}{\beta \hbar^2} \right)^{3/2} V. \]  

((8.2.7))

The actual degeneracy bears some examination. It is given by Equation (8.3.6) slightly modified by turning \( L \) into the appropriate power of \( V \):

\[ \omega(1, V, \epsilon) = \frac{\pi}{4} \left( \frac{8m}{\hbar^2} \right)^{3/2} V \epsilon^{1/2}, \]  

(8.3.8)

where the 1 as an argument to \( \omega \) is a reminder that this is the degeneracy for a single particle in a box of volume \( V \) when the particle has an energy \( \epsilon \). We have already discovered that this number is normally huge.\(^\text{12}\) This result will be used several times in later discussions.

\(^\text{11}\)I don’t want to seem mysterious here. Although we will discuss this in more detail later, the point is that all particles are either fermions or bosons. The former are restricted by the Pauli Exclusion Principle to be alone in an energy level. The latter are not. But as long as the average occupation of an energy level is much less than one, the difference between fermions and bosons can be ignored.

\(^\text{12}\)It is not huge when we are close to absolute zero. However, in that case we need to take quantum mechanics more explicitly into account.
8.4 The Thermodynamics of a Monatomic Ideal Gas

With the single particle partition function given:

\[ q(V, \beta) = \left( \frac{2\pi m}{\beta \hbar^2} \right)^{3/2} V, \]  

((8.2.7))

it is possible to evaluate the thermodynamics of a monatomic ideal gas.

For convenience (and because it has physical significance) it is usual to express \( q(\beta, V) \) in one of two different ways.

The first is in terms of a length \( \Lambda \), defined as:

\[ \Lambda = \left( \frac{\beta \hbar^2}{2\pi m} \right)^{1/2} = \left( \frac{\hbar^2}{2\pi m k T} \right)^{1/2}, \]  

((8.4.1))

then \( q \) is given by:

\[ q(\beta, V) = \frac{V}{\Lambda^3}. \]  

((8.4.2))

In this case \( \Lambda \) (which has the dimensions of length) is known as the thermal de Broglie wavelength. The reason is this: quantum mechanically, the de Broglie wavelength is given by:

\[ \lambda = \frac{\hbar}{m \hat{x}} = \left( \frac{\hbar^2}{2mE} \right)^{1/2}. \]  

((8.4.3))

When the temperature is \( T \), the typical energy of a monatomic ideal gas molecule is \( 3kT/2 \). In these terms

\[ \lambda = \left( \frac{\hbar^2}{3mkT} \right)^{1/2}, \]  

((8.4.4))

which (except for a factor of about the square root of two) is the thermal de Broglie wavelength.

Since quantum mechanically, particles whose de Broglie waves overlap will show quantum effects, then as long as the typical interparticle distance is greater than \( \Lambda \), quantum effects can be neglected.

As an example using argon at 300K, the de Broglie wavelength is approximately \( 7 \times 10^{-13} \) meters. The average separation between argon atoms at 1 atm and 300K is about \( 3.4 \times 10^{-9} \) meters, or almost 4700 times greater. Thus we expect quantum effects for argon under these conditions to be negligible.\(^\text{14}\)

<table>
<thead>
<tr>
<th>Gas</th>
<th>T(K)</th>
<th>( \Lambda ) (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>He</td>
<td>4.2</td>
<td>4.25 \times 10^{-10}</td>
</tr>
<tr>
<td>Ne</td>
<td>27.2</td>
<td>7.46 \times 10^{-11}</td>
</tr>
<tr>
<td>Ar</td>
<td>87.4</td>
<td>2.96 \times 10^{-11}</td>
</tr>
</tbody>
</table>

\(^{13}\)The author has always called this the “cute” formula because it looks a bit like arrow heads balanced on each other. Your view may differ.

\(^{14}\)Similar computations quickly reveal the conditions under which a given gas (if it remains a gas) will behave non-classically.
The second way of expressing $q(\beta, V)$ involves defining a temperature $\Theta_t$.

$$\Theta_t = \frac{\hbar^2}{2\pi mk},$$  \hspace{1cm} (8.4.5) 

so that:

$$q(\beta, V) = \left( \frac{T}{\Theta_t} \right)^{3/2} V.$$  \hspace{1cm} (8.4.6)

Here $\Theta_t$ is a characteristic temperature usually known as the translational theta temperature. As long as $T \gg \Theta_t$, quantum effects can be neglected. Of course $\Theta_t$ is trivially related to $\Lambda$ by:

$$\Lambda^2 = \frac{\hbar^2}{2\pi mkT} = \frac{\Theta_t}{T}.$$  \hspace{1cm} (8.4.7)

<table>
<thead>
<tr>
<th>Gas</th>
<th>$\Theta_t$ (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>He</td>
<td>$7.46 \times 10^{-19}$</td>
</tr>
<tr>
<td>Ne</td>
<td>$1.48 \times 10^{-19}$</td>
</tr>
<tr>
<td>Ar</td>
<td>$2.46 \times 10^{-20}$</td>
</tr>
</tbody>
</table>

Table 8.3: $\Theta_t$ for Various Monatomic Gases

Which tells us that we can ignore quantum effects except at very low temperatures.

Now we evaluate $Q(N, V, \beta)$ which is:

$$Q(N, V, \beta) = \frac{1}{N!} q^N = \frac{1}{N!} \left( \frac{2\pi m}{\beta \hbar^2} \right)^{3N/2} V^N$$  \hspace{1cm} (8.4.8)

and for the first time we see a “real” partition function expressed in terms of its natural variables $N$, $V$, and $\beta = 1/kT$.

The thermodynamics depends on $\ln Q$,

$$\ln Q(N, V, \beta) = -N \ln N + N + N \ln q(V, \beta),$$  \hspace{1cm} (8.4.9)

where Stirling’s approximation has been used. Substituting Equation (8.4.1) on page 8-7 into this leads to:

$$\ln Q(N, V, \beta) = -N \ln N + N + N \ln V - 3N \ln \Lambda.$$  \hspace{1cm} (8.4.10)

The expected value of the energy of the system is given by:

$$\langle E \rangle = -\left( \frac{\partial \ln Q}{\partial \beta} \right)_{N,V}.$$  \hspace{1cm} (8.4.11)
Since only $\Lambda$ is a function of $\beta$, this is:

$$\langle E \rangle = \frac{3N}{\Lambda} \left( \frac{\partial \Lambda}{\partial \beta} \right). \quad (8.4.12)$$

Now $\Lambda = \left( \frac{\hbar^2 \beta}{2 \pi m} \right)^{1/2}$, then

$$\left( \frac{\partial \Lambda}{\partial \beta} \right) = \frac{1}{2} \left[ \frac{2\pi m}{\hbar^2 \beta} \right]^{1/2} \left[ \frac{\hbar^2}{2\pi m} \right]. \quad (8.4.13)$$

Putting the pieces together and doing some algebra gives:

$$\langle E \rangle = \frac{3}{2} NkT, \quad (8.4.14)$$

which is a very well-known result. From this the constant volume heat capacity, $C_V$ is then $3Nk/2$, another well-known result.

We readily find the pressure since:

$$\langle \beta p \rangle = \left( \frac{\partial \ln Q}{\partial V} \right)_{N,\beta} = \frac{N}{V}, \quad (8.4.15)$$

so that

$$pV = NkT. \quad (8.4.16)$$

We would be very unhappy if this were not the case.

The chemical potential is interesting: We have:

$$\langle \beta \mu \rangle = - \left( \frac{\partial \ln Q}{\partial N} \right)_{V,\beta}, \quad (8.4.17)$$

and since

$$\ln Q(N,V,\beta) = -N \ln N + N + N \ln V - 3N \ln \Lambda, \quad ((8.4.10))$$

we then have:

$$-\beta \mu = -\ln N - 1 + 1 + \ln V - 3 \ln \Lambda, \quad (8.4.18)$$

$$\mu = kT \ln \left( \frac{2\pi m kT}{\hbar^2} \right)^{3/2} \frac{V}{N}. \quad (8.4.19)$$

Note that the chemical potential is an intensive quantity and has an absolute value. In classical thermodynamics, the chemical potential depends on an arbitrary choice of standard state. This does not contradict thermodynamics. It happens because as is usually the case in statistical thermodynamics, we have implicitly chosen a state of zero energy. We did it here when we picked the zero for the energy levels for the particle in a box.

So we have that $d \ln Q$ is given by:

$$d \ln Q(N,V,\beta) = \left( \frac{\partial \ln Q}{\partial \beta} \right) d\beta + \left( \frac{\partial \ln Q}{\partial V} \right) dV + \left( \frac{\partial \ln Q}{\partial N} \right) dN, \quad (8.4.20)$$
which is:
\[
d\ln Q(N,V,\beta) = -\langle E \rangle d\beta + \langle \beta p \rangle dV - \langle \beta \mu \rangle dN. \tag{8.4.21}
\]
The dimensionless equation corresponding to this is Equation (1.7.20) on page 1-9. Since a function can have only one total differential, this identifies \( \ln Q \) with \( \Phi \) and thus
\[
-\beta A(N,V,\beta) = \ln Q(N,V,\beta), \tag{8.4.22}
\]
and also identifies \( \beta \) as:
\[
\beta = \frac{1}{kT}. \tag{8.4.23}
\]
The main thermodynamic functions are now easy. Since
\[
A = -kT \ln Q(N,V,\beta), \tag{8.4.24}
\]
then
\[
A(N,V,\beta) = -kT(-N \ln N + N \ln V - 3N \ln \Lambda), \tag{8.4.25}
\]
which is in fact an extensive quantity as can be seen with a slight rearrangement:
\[
A(N,V,\beta) = -NkT \ln \left[ \frac{Ve}{N\Lambda^3} \right] = -NkT \ln \left[ \frac{2\pi m V e}{\beta h^2 N} \right], \tag{8.4.26}
\]
where it should be remembered that 1 is identical to \( \ln e \). Since \( A = E - TS \), \( S \) works out to be:
\[
S = Nk \ln \left[ \left( \frac{2\pi mkT}{\hbar^2} \right)^{3/2} e^{5/2} \frac{V}{N} \right], \tag{8.4.27}
\]
which is known as the Sakur-Tetrode equation. While the entropy given there is extensive, it also has the wrong behavior as \( T \to 0 \). Of course this is because the classical approximation fails as \( T \to 0 \). As the temperature goes to zero, the essentially quantum mechanical nature of the system cannot be ignored. This happens when \( T \) begins to approach the theta temperature \( \Theta_t \) given in Equation (8.4.5) on page 8-8.

The story of the Sakur-Tetrode equation is no longer well-known. It begins with two articles for the *Annelen der Physik*, probably the premier physics journal in the world at the time.

...one [article] published in September, 1911, and the other a few months later. One author was Otto Sackur, 31 years old, a rising young physical chemist at the University of Breslau. The other was Hugo Tetrode, 17 years old, the precocious son of the president of the Dutch National Bank.\(^\text{15}\)

It is well worth reading the entire story.\(^\text{16}\)

\(^{15}\)Taken from the August/September 2009 (Volume 18, Number 8) issue of the APS News, a publication of the American Physical Society. The article was written by Richard Williams.

\(^{16}\)It is on-line at http://www.aps.org/publications/apsnews/200908/physicshistory.cfm.
8.5 The Electronic Partition Function

Except for the hydrogen atom, there is no formula for the electronic energy levels of an atom. Thus the evaluation of

\[ q_e = \sum_j e^{-\beta \epsilon_j} \]  

(8.5.1)

requires us to have values for the actual energy levels. Luckily, this information is readily available for low lying levels, which are the important ones. Some of this data is reproduced in Table 8.4 on page 8-13.

From an examination of this data several things are clear. One is that electronic energies are measured from the electronic ground state as zero energy, rather than from the ionized atom as zero energy. The other is that from the magnitude of the listed energies we note that only the two or three lowest levels are important at temperatures below 2000 K.

On the other hand, we have an enormous problem. Electronic partition functions do not converge. For a sum of terms consisting of exponentials to converge, the exponent must go to minus infinity so that the terms go to zero.

Electronic energy levels don’t do this. While the choice of zero point will not change anything, the common convention for monatomic atoms is to take the electron at infinity as zero energy. This is called the ionization limit. The exponential of this is one. So the partition function eventually becomes the sum of an infinite number of finite terms, each approximately one. Which means that it diverges.

We can easily see this by examining Equation (8.5.1).

This is both puzzling and upsetting. A non-convergent partition function really ruins the idea of statistical thermodynamics. That’s upsetting.

And yet we can use an approximate partition function to compute properties that agree very well with experiment. That’s puzzling.

The answer is really simple. We have been misled by our use of quantum mechanics! From quantum mechanics the electronic energy levels of the hydrogen atom are relatively easily found to be:

\[ \epsilon_n = -\frac{m_e e^4}{8 \varepsilon_0 h^2 n^2} = -\frac{e^2}{8\pi \varepsilon_0 a_0} \frac{1}{n^2}, \]  

(8.5.2)

with degeneracy

\[ \omega_n = 2n^2 \]  

(8.5.3)

where \( m_e \) is the mass of an electron, \( e \) the charge on the electron, \( \varepsilon_0 \) the permittivity of free space (also known as the electric constant), \( h \) is Planck’s constant, and \( a_0 \) is the Bohr radius of hydrogen, 5.29177249 × 10^{-11} meters.

Here the ground state, \( \epsilon_1 \) is below the ionization limit \( (n \to \infty) \) which is the standard zero of energy for monatomic atoms. But it is not the standard used in

---

17 Though the author has never seen this problem discussed in any textbook on the subject.
statistical thermodynamics. We can convert this to the common standard where the ground state is zero by adding the ionization energy $−\epsilon^*$ (a positive number) to all values. Then the infinite number of energy levels lie between zero and $|\epsilon^*|$, as we said above.

The electronic energy levels for all other atoms behave in a similar manner.

What’s wrong with this is that the radial wave function in quantum mechanics has infinity as its outer boundary condition. That means that we are solving for a hydrogen atom (or any other atom) alone in an otherwise empty universe. The electron in a real hydrogen atom is always confined in some way to a finite region of space. Here, in particular, we are dealing with a canonical ensemble of systems. Thus any electrons in any atoms are certainly confined to the system’s volume $V$.

What is needed is to solve the quantum mechanical problem of a hydrogen atom not in an empty universe but in a box of volume $V$. Since the problem is spherically symmetric, it is easiest to use a spherical box. The appropriate radial boundary condition is then that the wave function must become zero at a radius $R$ from the origin.

When we do this for a box of macroscopic size we find that the first large number of energy levels are identical to those we find for the unfettered hydrogen atom. But when the radius of the electron orbitals becomes comparable in magnitude to the radius of the box, the energy levels change. And when $r \gg R$, the energy levels become particle in a box energy levels. And particle in a box energy levels go to infinity as the quantum number increases. This is exactly what we want since then the exponential terms in Equation (8.5.1) on page 8-11 then go to zero and the sum then converges.

So the overall behavior of the electronic partition function in real life is that the first term is always one and that the next few terms get to be quite small compared to one – though they never become zero. But now, after a large but finite number of terms, we start getting particle in a box energy levels. These increase in energy so the exponentials start rapidly to get smaller and smaller, going to zero at the limit. And the partition function converges.

And the value of the partition function is almost always simply the value of the sum of the first few terms. The number of terms needed depends on the temperature, the higher $T$, the more terms are needed.

In real life we need only the first few terms of Equation (8.5.1) on page 8-11. Thus $q_e$ can be written as:

$$q_e = \omega_1 + \omega_2 e^{-\beta\epsilon_2} + \omega_3 e^{-\beta\epsilon_3} + \cdots.$$  \hspace{1cm} (8.5.4)

where $\omega_i$ is the degeneracy of the $i$’th electronic energy level.

---

18 This isn’t anybody’s fault. So many of these fields grew up independently of each other that reconciling all their differing standards is impossible now.

19 The hydrogen atom electronic wave functions in a spherical box involve Bessel functions and the particle in a box functions are not ordinary sines and cosines either, but the result is basically the same.

20 Unless the ground state is degenerate.
The table below lists some of the energy levels for selected monatomic gases:

Table 8.4: **Monatomic Gas Electronic States.** The data in this Table is not complete. Lines starting with an asterisk (*) indicate the first of a number of levels with almost exactly the same energy.

<table>
<thead>
<tr>
<th>Atom</th>
<th>Electron config</th>
<th>Term symbol</th>
<th>( \omega )</th>
<th>( \mathbf{E} ) (kJ/mole)</th>
<th>( \mathbf{E} ) (kJ/mole)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>1s(^1)</td>
<td>( ^1S_{1/2} )</td>
<td>2</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td></td>
<td>2s(^1)</td>
<td>( ^2S_{1/2} )</td>
<td>2</td>
<td>82258.9190</td>
<td>984.035</td>
</tr>
<tr>
<td></td>
<td>2p(^1)</td>
<td>( ^2P_{1/2} )</td>
<td>2</td>
<td>82258.9544</td>
<td>984.035</td>
</tr>
<tr>
<td></td>
<td>2p(^3)</td>
<td>( ^2P_{3/2} )</td>
<td>4</td>
<td>82259.2850</td>
<td>984.039</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>( ^3P_{1/2} )</td>
<td>2</td>
<td>97492.2112</td>
<td>1166.266</td>
</tr>
<tr>
<td>He</td>
<td>1s(^2)</td>
<td>( ^1S_0 )</td>
<td>1</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td></td>
<td>1s(^2)2s(^1)</td>
<td>( ^3S_1 )</td>
<td>3</td>
<td>159855.9745</td>
<td>1912.302</td>
</tr>
<tr>
<td></td>
<td>1s(^1)2s(^1)</td>
<td>( ^1S_0 )</td>
<td>1</td>
<td>166277.4403</td>
<td>1989.120</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>( ^3P_2 )</td>
<td>5</td>
<td>169086.7666</td>
<td>2022.727</td>
</tr>
<tr>
<td>Li</td>
<td>1s(^2)2s(^2)</td>
<td>( ^2S_{1/2} )</td>
<td>2</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td></td>
<td>1s(^2)2p(^1)</td>
<td>( ^2P_{1/2} )</td>
<td>2</td>
<td>14903.622</td>
<td>178.287</td>
</tr>
<tr>
<td></td>
<td>1s(^2)2p(^1)</td>
<td>( ^2P_{3/2} )</td>
<td>4</td>
<td>14903.957</td>
<td>178.291</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>( ^3P_2 )</td>
<td>5</td>
<td>27206.066</td>
<td>325.457</td>
</tr>
<tr>
<td>C</td>
<td>2s(^2)2p(^3)</td>
<td>( ^3P_0 )</td>
<td>1</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td></td>
<td>2s(^2)2p(^3)</td>
<td>( ^3P_1 )</td>
<td>3</td>
<td>16.41671</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td>2s(^2)2p(^3)</td>
<td>( ^3P_2 )</td>
<td>5</td>
<td>43.41350</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>( ^1D_2 )</td>
<td>5</td>
<td>10192.666</td>
<td>121.931</td>
</tr>
<tr>
<td>N</td>
<td>2s(^2)2p(^4)</td>
<td>( ^4S_{3/2} )</td>
<td>4</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td></td>
<td>2s(^2)2p(^3)</td>
<td>( ^2D_{5/2} )</td>
<td>6</td>
<td>19224.464</td>
<td>229.976</td>
</tr>
<tr>
<td></td>
<td>2s(^2)2p(^3)</td>
<td>( ^2D_{3/2} )</td>
<td>4</td>
<td>19233.177</td>
<td>230.080</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>( ^2P_{1/2} )</td>
<td>2</td>
<td>28838.920</td>
<td>344.990</td>
</tr>
<tr>
<td>O</td>
<td>2s(^2)2p(^4)</td>
<td>( ^3P_2 )</td>
<td>5</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td></td>
<td>2s(^2)2p(^4)</td>
<td>( ^3P_1 )</td>
<td>3</td>
<td>158.265</td>
<td>1.893</td>
</tr>
<tr>
<td></td>
<td>2s(^2)2p(^4)</td>
<td>( ^3P_0 )</td>
<td>1</td>
<td>226.997</td>
<td>2.715</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>( ^1D_2 )</td>
<td>5</td>
<td>15867.862</td>
<td>189.822</td>
</tr>
<tr>
<td>F</td>
<td>2s(^2)2p(^5)</td>
<td>( ^2P_3/2 )</td>
<td>4</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td></td>
<td>2s(^2)2p(^5)</td>
<td>( ^2P_{1/2} )</td>
<td>2</td>
<td>404.141</td>
<td>4.835</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>( ^2P_{3/2} )</td>
<td>6</td>
<td>102405.714</td>
<td>1225.044</td>
</tr>
<tr>
<td>Ne</td>
<td>2p(^b)</td>
<td>( ^1S_0 )</td>
<td>1</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td></td>
<td>1s(^1)2s(^1)</td>
<td>( ^1S_0 )</td>
<td>3</td>
<td>134459.2871</td>
<td>1608.490</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>( ^1S_2 )</td>
<td>5</td>
<td>134818.6405</td>
<td>1612.789</td>
</tr>
</tbody>
</table>

The energy in wavenumbers in the table above were taken from data supplied by the US National Institute of Science and Technology (NIST) at the internet URL: http://physics.nist.gov/PhysRefData/Handbook/

The energy in Joules per mole was computed from the wavenumber data using 0.0119626556 kJ/cm as the conversion factor. The conversion value was computed from the data in Appendix B.

Some few details were taken from Norman Davidson, *Statistical Mechanics*, McGraw-Hill, 1962, page 110. Note that the values given in the table on page 110 of that book are now obsolete. The entries in the table above are the latest data available.
As an example, let us look at the electronic partition function for hydrogen atoms. The first two terms of the sum are:

\[ q_e = 2 + 2e^{-\frac{984035}{RT}} + \cdots, \quad (8.5.5) \]

where the factors of two come from the degeneracy of the levels (spin) and the energy is taken from Table 8.4 on page 8-13 and is in joules per mole. Note that at 300 K the second term is about $10^{-171}$ in size. On the other hand fluorine gives us:

\[ q_e = 4 + 2e^{-\frac{4835}{RT}} + \cdots = 4 + 2e^{-1.9380625} + \cdots = 4 + 0.2679756 + \cdots \quad (8.5.6) \]

at 300 K. Luckily the next level is so much higher in energy that it simply doesn’t count. So we need the first two terms for fluorine.

How does all this affect the thermodynamic functions for a monatomic ideal gas? The answer is simple. One just computes $q_e$ for the monatomic gas involved. For the ordinary monatomic gases $q_e$ is simply 1. In all other cases one defines $Q(N, V, T)$ this way:

\[ Q(N, V, T) = \frac{1}{N!} (q_t q_e)^N, \quad (8.5.7) \]

and then applies the formulas of Section 8.4 to the result.
8.6 Appendix: The Euler-Maclaurin Summation Formula

8.6.1 The Summation Formula

The goal here is to explain a bit about the Euler-Maclaurin summation formula, discuss some of its quirks, and to give a number of examples of its use. We will not derive it, as that would take us a bit off our path. We will give a peek at where it comes from.\(^1\)

People frequently convert otherwise undoable integrals to sums. The process is called numerical integration. For example, given \( \int_a^b f(x) \, dx \) to be evaluated between \( a \) to \( b \), the simplest possible (and horribly wrong) method is to approximate the area under the integral by a rectangle of length \( b - a \) and height the average of \( f(a) \) and \( f(b) \). This gives the rule:

\[
\int_a^b f(x) \, dx = (b - a) \left[ \frac{f(a)}{2} + \frac{f(b)}{2} \right] + E,
\]  
(8.6.1)

where \( E \) is an error term.

This method can be improved by taking not one rectangle, but a series of \( N \) rectangles each of width \( h = (b - a)/N \). Then repeating Equation (8.6.1) \( N \) times, once for each interval, gives:

\[
\int_a^b f(x) \, dx = h \left[ \frac{f(a)}{2} + f(a + h) + f(a + 2h) + \cdots + f(b - h) + \frac{f(b)}{2} \right] + E,
\]  
(8.6.2)

where there are \( N + 1 \) total function evaluations. There is still an error term \( E \), but it will be smaller as the thinner strips give a better approximation to the integral. This numerical integration rule is known as the trapezoidal rule and it can be rewritten more succinctly as:

\[
\int_a^b f(x) \, dx = \sum_{n=a}^{b} f(n) - h \left[ \frac{f(a)}{2} + f(b) \right] + E,
\]  
(8.6.3)

where we’ve chosen \( N = (b - a) \) so that \( h = 1 \).

But we are not interested in approximating an integral with a sum. Rather, we are interested in the reverse, the approximating of a sum by an integral. Thus we write:

\[
\sum_{n=a}^{b} f(n) = \int_a^b f(x) \, dx + \frac{1}{2} \left[ f(a) + f(b) \right] + E,
\]  
(8.6.4)

where \( E \) is again an error term of some sort.

Right away we see that evaluating a sum by an integral isn’t a simple thing. Even if the error term is negligible, which it usually isn’t, there is still the factor involving \( |f(a) + f(b)|/2 \) to be considered.

\(^1\)At least, one of the many places it comes from.
A decent method of approximating a sum by an integral (or vice versa) was worked out in 1732 by Leonhard Euler to find the sums of slowly convergent series and independently by Colin Maclaurin, at about the same time, who used it to evaluate integrals.\textsuperscript{22}

The result is an expansion of the error term $E$ in a series:

$$E = \sum_{k=2}^{m} \frac{B_k}{k!} \left[ f^{(k-1)}(b) - f^{(k-1)}(a) \right] + R_m,$$

(8.6.5)

where $R_m$ is the value of the remainder of the (possibly) infinite series caused by truncating it after the $m$‘th term. Here $f^{(s)}(c)$ denotes the $s$‘th derivative of the function $f(x)$ evaluated at the point $c$. If $s = 0$, it is the function itself that is to be evaluated, and if $s = -1$ it is the integral of the function that is evaluated. The terms $B_k$ are known as the \textbf{Bernoulli numbers},\textsuperscript{23}\ the first few of which are given later in Table 8.5 on page 8-17. It should be noted that all Bernoulli numbers with odd subscripts greater than 1 are zero and so are not listed in that Table.

Before discussing this result, let’s plug Equation (8.6.5) into Equation (8.6.4) on page 8-15 so we can keep track of where we are:

$$\sum_{n=a}^{b} f(n) = \int_{a}^{b} f(x)dx + \frac{1}{2} \left[ f(a) + f(b) \right] + \sum_{k=2}^{m} \frac{B_k}{k!} \left[ f^{(k-1)}(b) - f^{(k-1)}(a) \right] + R_m.$$

(8.6.6)

In general this is an infinite series that one arbitrarily cuts off at the term involving $B_m$, resulting in a truncation error. But sometimes the function being summed only has a finite number of non-zero derivatives. If all derivatives beyond the $m - 1$’st are zero, then the expansion is exact and there is no truncation error at all.

The Bernoulli numbers are given in many sources and many computer mathematics programs have routines for calculating them. So working with Equation (8.6.6) isn’t difficult in practice. And indeed, the Bernoulli numbers occur often in numerical analysis.

The Bernoulli numbers look as if they are going to stay small as they go on. But this is illusory. Indeed, they have all sorts of values. For example $B_{24}$ is $-86580.25311+$, which isn’t small at all.

Also note that, except for $B_1$, all odd Bernoulli numbers are zero.

\textsuperscript{22}The history of this is somewhat interesting and involves perhaps a bit of national pride. According to Knopp, (Knopp, Konrad, \textit{Theory and Application of Infinite Series}, Dover, 1990, footnote 3 on page 531), this was first hinted at by Euler in a paper published in 1732. Euler gave some examples in 1738, but did not publish a full proof until 1741. The formula was independently discovered by McLaurin and published in 1742. Knopp refers to it as the “Euler Equation.” So does Knuth (Knuth, Donald E, \textit{The Art of Computer Programming: Volume 1, Fundamental Algorithms}, Addison-Wesley, 1966) who is very careful about attributions. Nevertheless, the compound name \textit{Euler-Maclaurin} has stuck in most of the literature in English.

\textsuperscript{23}These were basically discovered by Jakob Bernoulli and published posthumously in 1713. Curiously, these numbers were discovered independently by the Japanese mathematician Seki Kowa (1642-1708) and also published posthumously in 1712.
CHAPTER 8. THE IDEAL MONATOMIC GAS 8-17

\[
\begin{array}{c|c}
B_0 & 1 \\
B_2 & 1 / 6 \\
B_6 & 1 / 42 \\
B_{10} & 5 / 66 \\
B_{14} & 7 / 6 \\
B_{18} & 43867 / 798 \\
\end{array}
\]

\[
\begin{array}{c|c}
B_1 & -1 / 2 \\
B_4 & -1 / 30 \\
B_8 & -1 / 30 \\
B_{12} & -691 / 2730 \\
B_{16} & -3617 / 510 \\
B_{20} & -174611 / 330 \\
\end{array}
\]

Table 8.5: The First Few Bernoulli Numbers

Note that all Bernoulli numbers with odd subscripts greater than one are zero.

In use here the Bernoulli numbers \(B_n\) are always divided by \(n!\), so it is useful to have a table of those numbers which we will call \(C_n\):

\[
\begin{array}{c|c}
C_0 & 1.0000000000 \\
C_1 & -0.5000000000 \\
C_2 & 0.0833333333 \\
C_4 & -0.0013888889 \\
C_6 & 0.0000330688 \\
C_8 & -0.0000008267 \\
C_{10} & 0.000000209 \\
C_{12} & -0.0000000005 \\
\end{array}
\]

Table 8.6: The First Few Euler-Maclaurin Coefficients

The next coefficients are zero to 10 significant figures, but they are not really zero, they are just small. If, as will be seen later, the derivatives are very large, the corresponding terms can’t be neglected at all.

The truncation error term in the Euler-Maclaurin summation formula was not worked out until 1823.\(^{24}\) The error \(R_m\) is given by:

\[
R_m = \frac{B_{m+2}}{(m+2)!} f^{(m)}(\xi), \quad (a \leq \xi \leq b).
\] (8.6.7)

This isn’t too useful in this form, but it can be shown that if \(f^{m+2}(x)\) goes monotonically to zero as \(x\) goes to \(b\), the remainder is less than the first discarded term. This doesn’t sound too useful, but in practice the upper limit is often infinity and the summand\(^{25}\) \(f^{m+2}(x)\) often is of the form of \(e^{-x}\) (or something similar) and does go to zero as \(x\) increases.

The Euler-Maclaurin formula, Equation (8.6.6) on page 8-16 occurs in many forms in the literature. You will find instances of it where the sum goes up to \(b-1\) instead of \(b\). Or instances in which the sum starts at \(a+1\). Or places where the term in \([f(b) + f(a)]\) isn’t there, or is there in somewhat different form. The most common “other” form has a first term involving \(B_1\). That’s been avoided here.\(^{26}\)

\(^{24}\)According to Graham, Knuth, and Patashnik (Graham, Knuth, and Patashnik, Concrete Mathematics, Addison-Wesley, 1988, page 457) and Knopp (Knopp, Konrad, Theory and Application of Infinite Series, Dover, 1990, footnote 3 on page 521) this was done by S. D. Poisson in that year.

\(^{25}\)That’s the thing being summed. It’s a technical term and I tend to use those when the material gets a little complex. Besides, you have to learn those terms anyway, so why not now?

\(^{26}\)Why? Because there seems to be a mild controversy over whether \(B_1\) should be 1/2 or \(-1/2\). While here I list \(B_1 = -1/2\) (all other \(B\)’s with odd subscripts are zero) I thought it best to avoid this altogether.
8.6.2 The Useful Formula

We recognize that in using the Euler-Maclaurin summation formula in practice, only a few terms will be used. It is useful to evaluate those as far as possible and write them out.

Thus, for our purposes, the Euler-Maclaurin summation formula becomes:

\[
\sum_{n=a}^{b} f(n) = \int_{a}^{b} f(x)dx + \frac{1}{2} [f(b) + f(a)] + \frac{1}{12} [f'(b) - f'(a)] - \frac{1}{720} [f^{(3)}(b) - f^{(3)}(a)] + \frac{1}{30240} [f^{(5)}(b) - f^{(5)}(a)] + \cdots, \quad (8.6.8)
\]

and the truncation error is evaluated separately. A number of examples follow.

8.6.3 Examples

Example 8.1

Consider the series

\[
S = \sum_{k=0}^{9} e^{-k}, \quad (8.6.9)
\]

which is easily summed to give the correct answer \( S = 1.581905 \ldots \), to six decimal places. The corresponding integral

\[
\int_{0}^{9} e^{-x}dx = -e^{-9} + 1 = 0.999877 \ldots \quad (8.6.10)
\]

which is nowhere near the correct answer. Even if we include the “trapezoidal term” \( [f(b) + f(a)]/2 \) which here is \( [e^{-9} + 1]/2 = 0.500062 \ldots \), the result is only \( 1.499938 \ldots \), which is still wrong by a fair amount.

So we see why the Euler-Maclaurin formula is needed.

To use it we make a table:

<table>
<thead>
<tr>
<th>function</th>
<th>at ( x = 0 )</th>
<th>at ( x = 9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(x) = e^{-x} )</td>
<td>1</td>
<td>( e^{-9} )</td>
</tr>
<tr>
<td>( f'(x) = -e^{-x} )</td>
<td>(-1)</td>
<td>(-e^{-9})</td>
</tr>
<tr>
<td>( f^{(3)}(x) = -e^{-x} )</td>
<td>(-1)</td>
<td>(-e^{-9})</td>
</tr>
<tr>
<td>( f^{(5)}(x) = -e^{-x} )</td>
<td>(-1)</td>
<td>(-e^{-9})</td>
</tr>
</tbody>
</table>

so that

\[
\sum_{k=0}^{9} e^{-k} = \int_{0}^{9} e^{-x}dx + \frac{1}{2} [e^{-9} + 1] + \frac{1}{12} [-e^{-9} + 1] - \frac{1}{720} [-e^{-9} + 1] + R_n, \quad (8.6.11)
\]
and since \( f^{(5)}(x) = e^{-x} \to 0 \) as \( x \to \infty \), the error is less than

\[
\frac{1 - e^{-11}}{30240} = 3.31 \times 10^{-5}.
\]

The terms in the Euler-Maclaurin series here are as follows, with \( T_0 \) being the integral, \( T_1 \) the trapezoidal contribution and the rest of the \( T_n \)'s the contribution from the Euler-Maclaurin correction:

| \( T_0 \) | 0.999877 | contribution from integral |
| \( T_1 \) | 0.500062 | contribution from trapezoidal correction |
| \( T_2 \) | 0.083323 | Bernoulli contributions |
| \( T_4 \) | -0.001389 |
| \( T_6 \) | 0.000033 |

Adding these up gives a result of 1.581906, with an error of one unit in the sixth decimal place, which is well within the error calculated error.

Before leaving this example, note that if the upper limit on the sum were infinity, the result would be \( 1/(1 - e^{-1}) \), which is 1.58197..., but the integral would give exactly 1.

Here’s another example:

\[
S = \sum_{k=1}^{\infty} \frac{k}{(1 + k\pi)^3}.
\]

Example 8.2

Consider the sum:

\[
S = \sum_{k=1}^{\infty} \frac{k}{(1 + k\pi)^3}.
\]

Don’t even bother trying to sum this analytically. Even computer computation doesn’t help much.\(^{29}\)

\(^{27}\)Carrying one more term would give the exact answer to six decimal digits.

\(^{28}\)One may certainly wonder why so much time is being spent on a curious but practically unknown formula? The answer is that in doing theoretical work, one can go for a very long time without needing the Euler-Maclaurin summation formula. But when you do need it, you need it very badly. So the author feels that it is worth a few moments of your time!

\(^{29}\)Admittedly, these numbers did not take long to compute.
<table>
<thead>
<tr>
<th>Number of Terms</th>
<th>Partial Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0140766028</td>
</tr>
<tr>
<td>10</td>
<td>0.0270192690</td>
</tr>
<tr>
<td>100</td>
<td>0.0296351966</td>
</tr>
<tr>
<td>1,000</td>
<td>0.0299223665</td>
</tr>
<tr>
<td>10,000</td>
<td>0.0299513617</td>
</tr>
<tr>
<td>100,000</td>
<td>0.0299542640</td>
</tr>
<tr>
<td>1,000,000</td>
<td>0.0299545443</td>
</tr>
<tr>
<td>10,000,000</td>
<td>0.0299545833</td>
</tr>
<tr>
<td>100,000,000</td>
<td>0.0299545862</td>
</tr>
<tr>
<td>1,000,000,000</td>
<td>0.0299545865</td>
</tr>
</tbody>
</table>

But from these we can’t even really tell what the accuracy is. If we look at the series, we see that it should eventually converge as \( 1/(n\pi)^2 \), but this is not an alternating series and thus the error is not bounded by the first neglected term.

Doing the Euler-Maclaurin summation requires more work. First we need to compute the derivatives, and that’s admittedly hard.30

Here are some of the needed quantities:

<table>
<thead>
<tr>
<th>function at ( f(k) )</th>
<th>at ( f(1) )</th>
<th>at ( f(\infty) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(k) = k(1 + k\pi)^{-3} )</td>
<td>0.01407660276</td>
<td>0</td>
</tr>
<tr>
<td>( f'(k) = (1 - 2\pi x)(1 + \pi x)^{-4} )</td>
<td>-0.01795669132</td>
<td>0</td>
</tr>
<tr>
<td>( f^{(3)} = [12\pi^2(3 - 2\pi x)(1 + \pi x)^{-6} )</td>
<td>-0.07704990317</td>
<td>0</td>
</tr>
<tr>
<td>( f^{(5)} = [360\pi^4(5 - 2\pi x)(1 + \pi x)^{-8} )</td>
<td>-0.5198192003</td>
<td>0</td>
</tr>
<tr>
<td>( f^{(7)} = [20160\pi^6(7 - 2\pi x)(1 + \pi x)^{10} )</td>
<td>9.356703044</td>
<td>0</td>
</tr>
<tr>
<td>( f^{(9)} = [1814400\pi^8(7 - 2\pi x)(1 + \pi x)^{12} )</td>
<td>1836.468386</td>
<td>0</td>
</tr>
</tbody>
</table>

It can be seen that this is ugly. The derivatives keep growing. That may rule out using the first dropped term as an error bound.

Now using the fact that both the function and all its derivatives are zero as \( x \to \infty \):

\[
S = \sum_{k=1}^{\infty} k = \int_{1}^{\infty} \frac{x}{(1 + \pi x)^3} \, dx + \frac{1}{2} f(1) + \frac{1}{12} f^{(1)}(1) - \frac{1}{720} f^{(3)}(1) + \frac{1}{30240} f^{(5)}(1) + \cdots. \quad (8.6.14)
\]

The integral can be done. It is:

\[
\int_{1}^{\infty} \frac{x}{(1 + \pi x)^3} \, dx = -\frac{1 + 2\pi x}{2\pi^2(1 + \pi x)^2} \bigg|_{1}^{\infty} = 0.0215108145 \ldots, \quad (8.6.15)
\]

and the negative result should not be surprising. When evaluated the value at infinity is zero, so the indicated result is then itself subtracted, making the answer positive.

30I’ll admit to using a computer calculus program to do the differentiations.
the sum can then be evaluated:

\[ S = 0.0215108145 + 0.0070383014 + 0.0014963909 \\
- 0.0001070138 + 0.0000171898 + 0.0000077354 \\
- 0.0000383395 + \ldots = 0.0299250788 + . \quad (8.6.16) \]

But that’s not necessarily the answer, though it is certainly similar, though a bit less than the actual million term sum.

If we look at the partial sums here’s what we get:

<table>
<thead>
<tr>
<th>Term</th>
<th>Partial Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0215108145</td>
</tr>
<tr>
<td>1</td>
<td>0.0285491159</td>
</tr>
<tr>
<td>2</td>
<td>0.0300455069</td>
</tr>
<tr>
<td>4</td>
<td>0.0299384931</td>
</tr>
<tr>
<td>6</td>
<td>0.0299556829</td>
</tr>
<tr>
<td>8</td>
<td>0.0299634183</td>
</tr>
<tr>
<td>10</td>
<td>0.0299250788</td>
</tr>
<tr>
<td>12</td>
<td>0.0300530433</td>
</tr>
</tbody>
</table>

So it can be seen that the partial sums will flop around a bit. We have no real error estimate. About the best that can be said is that the answer seems to be 0.0299 ± 0.0001 or so.

This looks as if it is a useless exercise. But we now know at least a reasonable estimate for the sum of the series, and that’s more than we knew going in.

It looks as if the series summation gives a slightly better result, \( S = 0.0299545865 + \), but that’s really only a lower limit. Don’t forget that an infinite number of terms have been omitted, and they can add up to quite a bit.

So all in all, looking at all the evidence, which is always what one ought to do, we can assert that the sum is at least 0.0299545865, but not, from the Euler-Maclaurin expansion, apt to be too much bigger.

Both of the previous examples have been very simple. In actual practice, the series to be summed depend on more than one parameter. So the answer isn’t a number at all.

Here is a more realistic example.

**Example 8.3**

We shall assume that we have a system of independent particles whose particle energy levels \( \epsilon_n \) are given by:

\[ \epsilon_n = an^3 \quad a > 0, \quad (8.6.17) \]
where \( n \) is a quantum number starting at zero and \( a \) is a constant typically depending on Boltzmann’s constant or Planck’s constant as well as other things. As a result \( a \) is usually quite small as we’d expect for a single particle energy level. For example, as can be seen from Equation (8.1.3) on page 8-2 for the energy levels of a particle in a box, \( a \) is given by:

\[
a = \frac{\beta h^2}{8mL^2},
\]

so that the numerator is of the order of \( 10^{-68} \) and the denominator (via the particle mass) of the order of \( 10^{-466666666} \). Thus overall \( a \approx 10^{-22} \).

Assuming non-degenerate energy levels, the single particle partition function for these energies is:

\[
q = \sum_{n=0}^{\infty} e^{-an^3},
\]

Of course this is a made-up example. There don’t seem to be any systems with this energy level structure.

The problem now is to do this sum. And as we quickly learn, it has no known simple form as a result.

So we resort to the Euler-Maclaurin summation formula for help.

The first thing we need to do is discover if the integral

\[
I = \int_{0}^{\infty} e^{-ax^3} \, dx
\]

can be done.\(^{31}\) It looks like a formidable integral. The best way to do integrals of this sort is to use one or another of the symbolic mathematics packages available on the internet.\(^{32}\) The result turns out to be:

\[
I = \int_{0}^{\infty} e^{-ax^3} \, dx = \frac{1}{a^{1/3}} \Gamma(4/3),
\]

where \( \Gamma(x) \) is the Gamma function of \( x \) discussed in Section 5.8 on page 5-11. The numerical result is:

\[
I = \frac{1}{a^{1/3}} \times 0.892979511569249 + .
\]

The next step is to do the needed derivatives and evaluate them at zero and infinity:

<table>
<thead>
<tr>
<th>function</th>
<th>at ( x = 0 )</th>
<th>as ( x \to \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(x) = e^{-ax^3} )</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( f'(x) = -3ax^2e^{-ax^3} )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( f^{(3)}(x) = [-27a^3x^6 + 54a^2x^3 - 6a]e^{-ax^3} )</td>
<td>(-6a)</td>
<td>0</td>
</tr>
<tr>
<td>( f^{(5)}(x) = [-243a^5x^{10} + 1620a^4x^7 - 2160a^3x^4 + 360a^2x]e^{-ax^3} )</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^{31}\) Why? Because if it can’t be done, we can’t use the Euler-Maclaurin formula, that’s why!

\(^{32}\) I’ve become fairly fond of the Wolfram Alpha site at [http://www.wolframalpha.com/](http://www.wolframalpha.com/) which I used frequently here not only for doing the integral, but for doing the derivatives as well.
The Euler-Maclaurin series is then:
\[
\sum_{n=0}^{\infty} e^{-an^3} = \frac{1}{a^{1/3}} \Gamma(4/3) + \frac{1}{2} + \frac{6a}{720} + \cdots \tag{8.6.23}
\]
which is a usable approximate equation. We can see how well it does in this case using a short computer program to sum the series Equation (8.6.19) on page 8-22 directly.\footnote{This is a dangerous thing to rely upon. There are many series which seem to converge fairly quickly only to have later terms blow up. One should have a bit of theory handy to provide some comfort. In this case, we don’t.}

<table>
<thead>
<tr>
<th>(a)</th>
<th>Terms</th>
<th>Sum</th>
<th>Euler-Maclaurin</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0010</td>
<td>200</td>
<td>9.429786</td>
<td>9.429803</td>
<td>0.000017</td>
</tr>
<tr>
<td>0.0100</td>
<td>200</td>
<td>4.644760</td>
<td>4.644927</td>
<td>0.000167</td>
</tr>
<tr>
<td>0.1000</td>
<td>200</td>
<td>2.423037</td>
<td>2.424699</td>
<td>0.001662</td>
</tr>
<tr>
<td>1.0000</td>
<td>200</td>
<td>1.368215</td>
<td>1.401313</td>
<td>0.033098</td>
</tr>
</tbody>
</table>

So we see that for small \(a\), the Euler-Maclaurin approximation is quite good, but it rapidly becomes less so as \(a\) increases. What do we do then?

The answer is that the series rapidly converges for larger \(a\). For \(a = 0.1\), the \(n = 7\) term is zero to 10 decimal places. And for \(a = 1\), the \(n = 2\) term is zero to 10 decimal places. So\footnote{And we will see examples of this later on.} for (relatively) large \(a\) we just use the series for \(q\) directly.

It should be noted that the seventh derivative is zero at both limits, but the ninth, and every third derivative after that will not be zero at zero. The ninth term contributes \(a^3/180\) to Equation (8.6.23), which for \(a < 0.056\) changes our answer by less than \(1 \times 10^{-6}\).
Chapter 9

Ideal Diatomic Gases

9.1 Introduction

If we look at diatomic molecules, we see a large increase in complexity compared to single atoms. Where single atoms have only translational motion\(^1\) diatomic molecules can rotate and vibrate as well. So the first thing we need to do is concern ourselves with the total number of motions a diatomic molecule can exhibit.

Each atom has three degrees of freedom. Two atoms have three each, or six total. The fact that the atoms are bonded together doesn’t change this. To see this imagine that we have separated out the center of mass of the diatomic so that we have three coordinates, say \(X\), \(Y\), and \(Z\) for the center of mass and three further coordinates, \(r\), \(\theta\), and \(\phi\) to specify the distance between the two atoms and the orientation in space of one atom relative to the other.

We’ve already dealt with the motion of the center of mass. That’s the same for a monatomic molecule as it is for a diatomic, or polyatomic molecule too for that matter. Indeed all that changes is the mass involved. It becomes the mass of the entire molecule.

So we can ignore \(X\), \(Y\), and \(Z\) and deal only with the internal motions of the molecule. One way to do this is using \(r\), \(theta\), and \(phi\). But chemists prefer to break the motions up in another way: vibration and rotation.

Rotation is simple. A molecule in space can rotate at most around three independent space axes, the \(X\), \(Y\), and \(Z\) axes to be specific. Classically all molecules, even monatomic ones rotate around three axes. But quantum mechanics, as often happens, has different rules. It allows a diatomic molecule to have only two axes of rotation. The third, conventionally taken as the bond axis, has no rotation.

\(^1\)Excepting, of course, the electronic contribution to the partition function.
associated with it at all.

Why is this so? Because atoms do not have distinguishing marks. Rotation around the bond axis is undetectable. The wave function for the molecule does not change and so, in a very real sense, nothing is happening.

The other two rotational axes are perpendicular to the bond axis and are at right angles to it and each other. The rotational and bond axis all meet at the center of mass because if they did not, a rotation around one of the rotational axes would otherwise cause the center of mass to move. But that’s not allowed since we have separated out that motion as translational, which it is.

So a diatomic molecule has only two rotations. These rotations account for two angles in the degrees of freedom. There is a third degree of freedom and that must be a vibration along the bond axis that leaves the center of mass unmoved.

So a diatomic molecule has two rotations and one vibration.

It would be simplest to model vibration and rotation individually. That is, to assume that the vibrational and rotational motions don’t affect each other. For large vibrational or rotational energies this is clearly not true. Rapid rotation will stretch the molecular bond, changing the vibrational energies. And vibration will affect rotation since it changes the moment of inertia of the molecule on which the rotations depend.

In spite of this it is useful to assume that these are independent. This allows the energies to simply add. That is, the energy of the molecule $\epsilon$ is then a sum:

$$\epsilon = \epsilon_{\text{trans}} + \epsilon_{\text{vib}} + \epsilon_{\text{rot}} + \epsilon_{\text{elect}},$$

(9.1.1)

and as a result the exponential in the single particle canonical partition function $q$ factors so that we end up with:

$$q = q_{\text{trans}} + q_{\text{vib}} + q_{\text{rot}} + q_{\text{elect}}.$$ 

(9.1.2)

We’ve already dealt with $q_{\text{trans}}$ in Chapter 8 on page 8-1 and we will see that the treatment of $q_{\text{elect}}$ is very similar to the treatment of the electronic partition function in the same chapter. So here we concentrate on vibration and rotation.

### 9.2 Vibration

We want to treat vibration as the ideal vibration treated in quantum mechanics. We want this because we can then know the vibrational energy levels. The ideal quantum mechanical vibrator is parabolic in shape open upward and is symmetric around the minimum point. This model can be solved exactly in quantum mechanics.

---

2As will be discussed later, any linear molecule (acetylene, for example) has only two rotations because again, a rotation around the bond axis produces no change in the wave function.
But the actual bond potential energy in a diatomic molecule is not symmetric around the equilibrium bond length. It generally goes to infinity as the distance between the atoms goes to zero, and goes to zero as the distance between the atoms goes to infinity. In between there is a single minimum dipping below zero bond energy.\(^3\) What is done is to match the actual bond potential energy with a parabola that fits exactly at the minimum energy point. That means that the minimum points are the same and that the curvature at the minimum is the same.

Thus with the distance to the energy minimum denoted by \(r_o\), we have fixed \(r_o\) to be the real \(r_o\), the potential energy derivative to be \(du/dr = 0\) (so it will be an extremum), and \(d^2/dr^2\), the curvature, to match the real curvature. Since a parabola has three constants, these three conditions are sufficient to fix the parabola.

We actually do this by creating a Taylor series expansion of the potential around \(r_o\). If \(u(r)\) is the potential energy:

\[
\begin{align*}
\quad u(r) & = u(r_o) + \left. \frac{du}{dr} \right|_{r_o} (r - r_o) + \frac{1}{2!} \left. \frac{d^2u}{dr^2} \right|_{r_o} (r - r_o)^2 + \cdots. \\
\end{align*}
\]  
(9.2.1)

If we assume that the minimum energy is zero,\(^4\) then the first term on the right in Equation (9.2.1) is zero. The first derivative in the expansion is also zero since the potential has zero slope there. The second derivative is just a constant, the curvature of the potential curve at its minimum. If we truncate the series at this point we have a formula that is quadratic in \((r - r_o)^2\), of the same sort as in the harmonic oscillator:

\[
\begin{align*}
\quad u(r) & = \frac{1}{2!} \left. \frac{d^2u}{dr^2} \right|_{r_o} (r - r_o)^2. \\
\end{align*}
\]  
(9.2.2)

This approximation can always be made, though to what extent it will be true is subject to analysis in each case in which it is applied.

Now that we’ve assumed that the potential is harmonic we know the quantum mechanical result. There are an infinite set of non-degenerate vibrational quantum states with energies:

\[
\epsilon_n = (n + 1/2) \hbar \nu, 
\]  
(9.2.3)

where \(n = 0, 1, 2, \ldots\) and \(\nu\) is the classical fundamental vibrational frequency given by

\[
\nu = \left( \frac{1}{2\pi} \right) \left( \frac{k}{\mu} \right)^{1/2}, 
\]  
(9.2.4)

where \(\mu\), the reduced mass, is:

\[
\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}, 
\]  
(9.2.5)

with \(m_1\) and \(m_2\) the masses of the two atoms; and \(k\), is the classical spring constant.

The lowest energy level is \(\epsilon_0 = \hbar \nu/2\) which is above the zero point of the potential. At times some folks look at this and wonder if it is possible to get the zero point

\(^3\)Otherwise there’d be no bond!

\(^4\)One can always add a constant to the energy later to compensate for this if one wants a different zero point.
energy out of a system of oscillators. In fact, there is no zero-point energy because the quadratic potential is a mathematical fiction. The actual zero point energy is, in fact, $\epsilon_0 = \hbar \nu / 2$. This is often called $D_e$ by spectroscopists but it has no physical meaning. It is usually given as a positive number as if the bottom $u(r_o) = 0$ of the potential had some real meaning. It is simply the (positive) distance from $u = 0$ to the zero’th energy level of the vibrator.

Sometimes one will see the notation $D_o$. That is the “actual” energy of the bottom of the potential well. The relationship between these is clearly:

$$D_e = D_o + \frac{1}{2} \hbar \nu,$$  \hspace{1cm} (9.2.6)

or, put another way, Equation (9.2.2) on page 9-3 could be written as

$$u(r) = -D_e + \frac{1}{2} \frac{d^2 U}{dr^2} \bigg|_{r_o} (r - r_o)^2,$$  \hspace{1cm} (9.2.7)

where the minus sign is included since $D_e$ is invariably given as a positive number and $u(r_o)$ is almost always negative at that point.\(^5\)

The vibrational partition function $q_v$ is:

$$q_v = \sum_{n=0}^\infty e^{-\beta \epsilon_n} = \sum_{n=0}^\infty e^{-\beta(n+1/2)\hbar \nu},$$  \hspace{1cm} (9.2.8)

which can be re-arranged to be

$$q_v = e^{-\beta \hbar \nu / 2} \sum_{n=0}^\infty e^{-n\beta \hbar \nu}.$$  \hspace{1cm} (9.2.9)

Now if we let $x = \exp(-\beta \hbar \nu)$, this becomes

$$q_v = e^{-\beta \hbar \nu / 2} \sum_{n=0}^\infty x^n,$$  \hspace{1cm} (9.2.10)

where the sum, a geometric series, is easily seen to be $1/(1 - x)$ if $|x| < 1$. And $|x| < 1$ since $\beta$, $\hbar$, and $\nu$ are all intrinsically positive numbers.

It then follows that $q_v$ is:

$$q_v = \frac{e^{-\beta \hbar \nu / 2}}{1 - e^{-\beta \hbar \nu}}.$$  \hspace{1cm} (9.2.11)

This can be put into a somewhat more useful form by grouping the constants. If $\Theta_v$ is defined by:

$$\Theta_v = \frac{\hbar \nu}{k},$$  \hspace{1cm} (9.2.12)

\(^5\)Confused yet? You are not alone. These notational problems arise because spectroscopy was a more or less independent field when the notations were developed.
where \( \Theta_v \) of course has Kelvins as units. In these terms we have for Equation (9.2.11) on page 9-4:

\[
q_v = \frac{e^{-\Theta_v/2T}}{1 - e^{-\Theta_v/T}}.
\]  

(9.2.13)

Of course \( \Theta_v \) is another characteristic temperature. (See, for example Equations (8.4.5) on page 8-8 and (8.4.6) on page 8-8. As long as \( T \) is small compared to \( \Theta_v \), quantum effects can be ignored. So it is useful to look at a small table of \( \Theta_v \) for various diatomic molecules such as Table 9.1 on the next page.

<table>
<thead>
<tr>
<th>Species</th>
<th>( \Theta_v (K) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>H(_2)</td>
<td>6215</td>
</tr>
<tr>
<td>HCl</td>
<td>4227</td>
</tr>
<tr>
<td>N(_2)</td>
<td>3374</td>
</tr>
<tr>
<td>CO</td>
<td>3100</td>
</tr>
<tr>
<td>Cl(_2)</td>
<td>810</td>
</tr>
<tr>
<td>I(_2)</td>
<td>310</td>
</tr>
</tbody>
</table>

Table 9.1: \( \Theta_v \) for Selected Diatomic Molecules

The probability of finding a molecule in vibrational state \( n \) is

\[
p_n = \frac{e^{-\beta \epsilon_n}}{q_v} = \frac{e^{-\beta(n+1/2)h\nu}}{q_v},
\]  

(9.2.14)

and so the probability of finding a molecule in the \( n = 0 \) (ground) state is:

\[
p_0 = \frac{e^{-\beta(1/2)h\nu}}{q_{vib}} = \frac{e^{-\Theta_v/2T}}{q_v}.
\]  

(9.2.15)

With \( q_v \) given by Equation (9.2.13), this becomes

\[
p_0 = 1 - e^{-\Theta_v/T},
\]  

(9.2.16)

and so the probability of finding a molecule in \( \text{any} \) vibrational

\[
p_n = \frac{e^{-\beta \epsilon_n}}{q_v} = \frac{e^{-\beta(n+1/2)h\nu}}{q_v},
\]  

(9.2.17)

and so the probability of finding a molecule in the \( n = 0 \) (ground) state is:

\[
p_0 = \frac{e^{-\beta(1/2)h\nu}}{q_{vib}} = \frac{e^{-\Theta_v/2T}}{q_v}.
\]  

(9.2.18)

With \( q_v \) given by Equation (9.2.13), this becomes

\[
p_0 = 1 - e^{-\Theta_v/T},
\]  

(9.2.19)

and so the probability of finding a molecule in \( \text{any} \) vibrational state greater than the ground state is then:

\[
p_{(n>0)} = e^{-\Theta_v/T},
\]  

(9.2.20)
hence the importance of the characteristic temperature $\Theta_v$. For example in hydrogen with $\Theta_v = 6215\text{K}$, only about $1 \times 10^{-7}\%$ of the molecules at 300K are not in the ground state. Even at 1000K only about 0.2% of hydrogen molecules are not in the vibrational ground state.

For iodine the situation is different. With $\Theta_v = 310\text{K}$, at 300K about 36% of the molecules are in elevated vibrational states.

To the extent that the vibrational partition function can be treated separately, a separate vibrational energy can be found from $q_v$:

$$\langle \epsilon_v \rangle = k\Theta_v \left( \frac{1}{2} + \frac{1}{e^{\Theta_v/T} - 1} \right),$$

and $\langle E_v \rangle$ can be found from Equation (9.2.21) by multiplying it by $N$, the number of diatomic molecules in the system.\(^6\)

![Figure 9.1: Diatomic Molecule Vibrational Energy](image)

With the average vibrational energy per molecule known, the constant volume vibrational heat capacity per molecule can be found from $C_V = \partial \langle \epsilon_v \rangle / \partial T$ and is:

$$C_{v,V} = k \left( \frac{\Theta_v}{T} \right)^2 \frac{e^{\Theta_v/T}}{(e^{\Theta_v/T} - 1)^2},$$

which again has to be multiplied by $N$ for the heat capacity of all $N$ particles.

It is left as an exercise for the reader to show that as the temperature increases without limit, $C_{v,V}$ in Equation (9.2.22) approaches $k$ as a limit. The physical explanation for this is that all the energy levels become uniformly populated as the temperature increases and a further increase in energy can no longer move a particle to a higher vibrational energy level.\(^7\) In fact this limit is never reached because (1)

---

\(^6\)A reminder as to why this works: the molecules are independent and do not affect each other in any way.

\(^7\)It is again left to the reader to show that in a canonical ensemble no energy level can have a higher occupation number than one with less energy.
the potential energy isn’t really a parabola and anharmonicity becomes more and more important as \( T \) increases, (2) the bond will break before the temperature becomes too high, and (3) in most cases electrons will move to higher energy levels changing the entire potential energy surface of the diatomic molecule.

At typical laboratory temperatures, most diatomic molecules are in their vibrational ground states (see Table 9.1) and their theoretical behavior in higher states is, well, theoretical.

### 9.3 Rotation

A diatomic molecule has only two rotational degrees of freedom. There are still three axes, two go through the center of mass and are perpendicular to the bond axis. These are ”real” rotational axes. The third rotational axis is the bond axis itself.

However, since the atoms are considered to be spheres (or points, depending on your point of view), there is infinite rotational symmetry for any rotation about the bond axis. Thus any rotation of any magnitude at all is equivalent to no rotation because the rotation produces no measurable change in the diatomic molecule.

Thus we are left with only two rotational degrees of freedom in a diatomic molecule.

It is easy to see that the two rotational degrees of freedom are degenerate. The two rotational axes are arbitrary in the sense that the first one can be picked to point in any direction that is perpendicular to the bond axis. Call that direction \( A \). The other axis (call it direction \( B \)) is then determined since it must be perpendicular to both already chosen axes. But another viewer can chose her first axis to point
in the direction $B$ and then her second will point in the direction $A$. Or the choice of the first could have been in a different direction altogether. Clearly the choice of which axis is $A$ and which is $B$ cannot affect the description of the system, so the two rotations must be totally equivalent.

So for rotation in a diatomic molecule we have two rotational motions which are degenerate.

If the molecule is heteronuclear, that’s the end of the story. If, however, the molecule is homonuclear, the complications continue. Thus, after some general remarks, we will consider the heteronuclear case first and then deal with homonuclear diatomics.

### 9.3.1 Evaluation of the Rotational Partition Function

#### Generalities

There is no potential energy associated with rotation about the center of mass. Thus the rotation is free. If we assume that the rotation is independent of vibration, then the problem of the rotation of a diatomic molecule can be modelled by the rotation of two atoms of masses $m_1$ and $m_2$ separated by a fixed distance and rotating around their common center of mass.

This problem is one of the few exactly soluble problems in quantum mechanics. The results are simple: there are an infinite number of rotational quantum states indexed by a quantum number traditionally called $J$, where $J = 0, 1, 2, \ldots$. The energy in quantum state $J$ is given by:

$$
\epsilon_J = \frac{\hbar^2 J(J+1)}{8\pi^2 I},
$$

where $I$ is the moment of inertia

$$
I = \mu r_o^2.
$$

With $\mu$ being the reduced mass

$$
\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2},
$$

and $r_o$ the equilibrium internuclear distance.\(^8\)

Each of these energy levels is degenerate with a degeneracy factor $g_J$:

$$
g_J = 2J + 1,
$$

with the states with odd $J$ being anti-symmetric\(^9\) and those with even $J$ being symmetric.\(^{10}\)

---

\(^8\)Known to chemists as the bond length.

\(^9\)Anti-symmetric means that the wave function associated with the odd quantum number $J$ will change sign when the molecule is rotated by 180°.

\(^{10}\)Symmetric is the opposite of anti-symmetric; the wave function does not change sign on a 180° rotation.
In almost all cases, rotational energies are relatively small and at any reasonable temperature most diatomic molecules are in a fairly high rotational state. This means that the rotational partition function is normally “classical”. But hydrogen is an exception.\textsuperscript{11}

To jump ahead of the story, there is a characteristic rotational temperature $\Theta_r$ for each molecule. This is analogous to the vibrational characteristic temperature $\Theta_v$. If the actual temperature is higher that $\Theta_r$, the resultant partition function can be taken to be classical. Otherwise a more detailed analysis is needed. For hydrogen $\Theta_r$ is 85.4 K. Hydrogen is still a gas at this temperature. The next highest $\Theta_r$ is for HCl.

Table 9.2 gives some values of the characteristic rotational temperatures of simple diatomic molecules. As we should expect, the lighter the molecule, the higher the characteristic rotational temperature. This is primarily due to the lesser moment of inertia of light molecules.

<table>
<thead>
<tr>
<th>Species</th>
<th>$\Theta_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>H$_2$</td>
<td>85.4</td>
</tr>
<tr>
<td>HCl</td>
<td>15.2</td>
</tr>
<tr>
<td>HI</td>
<td>9.0</td>
</tr>
<tr>
<td>N$_2$</td>
<td>2.86</td>
</tr>
<tr>
<td>O$_2$</td>
<td>2.07</td>
</tr>
<tr>
<td>Cl$_2$</td>
<td>0.346</td>
</tr>
<tr>
<td>I$_2$</td>
<td>0.054</td>
</tr>
</tbody>
</table>

Table 9.2: $\Theta_r$ for Selected Diatomic Molecules

\textbf{Heteronuclear Diatomic Molecules}

The rotational partition function $q_r$, including the degeneracy of each state $J$, is:

$$q_r = \sum_{J=0}^{\infty} (2J + 1) e^{-\beta(J+1)\hbar^2/8\pi^2I}.$$  \hspace{1cm} (9.3.5)

It is best to simplify this by setting $\Theta_r$, the \textit{characteristic rotational temperature} to be

$$\Theta_r = \frac{\hbar^2}{8\pi^2Ik},$$  \hspace{1cm} (9.3.6)

so that Equation (9.3.5) becomes

$$q_r = \sum_{J=0}^{\infty} (2J + 1) e^{-J(J+1)\Theta_r/T},$$  \hspace{1cm} (9.3.7)

which, as we ought now to expect, can not be summed in any convenient way. Thus we once more need the \textit{Euler-Maclaurin summation formula}.\textsuperscript{11} Indeed, the only exception.

\textsuperscript{11}Indeed, the only exception.
CHAPTER 9. IDEAL DIATOMIC GASES

Function | at $J = 0$ | at $J = \infty$
--- | --- | ---
$f(J) = (2J + 1)e^{-J(J+1)s}$ | 1 | 0
$f'(J) = -(4J^2 + 4J + 1)s - 2J e^{-J(J+1)s}$ | 2$-s$ | 0
$f^{(3)}(J)$ | $-s(s^2 - 12s + 12)$ | 0
$f^{(5)}(J)$ | $-s^2(s^3 - 30s^2 + 180s - 120)$ | 0
$f^{(7)}(J)$ | $-s^3(s^4 - 56s^3 + 840s^2 - 3360s + 1680)$ | 0

Table 9.3: Derivatives for the Evaluation of the Rotational Partition. Note: $s$ has been written for $\Theta_r/T$ and the actual equations omitted for the third, fifth, and seventh derivatives.

Generating these derivatives is tedious.\textsuperscript{12} The integral can be done by substituting $x$ for $J(J + 1)$ whence we then obtain:

$$
\int_0^\infty e^{-\Theta_r x/T}\,dx = \frac{T}{\Theta_r} = \frac{8\pi^2 I}{\beta h^2}. \quad (9.3.8)
$$

With $s = \Theta_r/T$ we then have for the rotational partition function:

$$
q_r = \frac{1}{s} + \frac{1}{2} + \left[ \frac{1}{6} + \frac{s}{12} \right] + \left[ -\frac{s^3}{720} + \frac{s^2}{60} - \frac{s}{60} \right] + \left[ \frac{s^5}{30240} - \frac{s^4}{1008} + \frac{s^3}{168} - \frac{s^2}{252} \right] + \left[ \frac{s^7}{1209600} + \frac{s^6}{21600} - \frac{s^5}{1440} + \frac{s^4}{360} - \frac{s^3}{720} \right] + \cdots \quad (9.3.9)
$$

The first term above comes from the integral and each of the others (in square brackets) from the corresponding term of the Euler-Maclaurin expansion. Note that we can see which powers of $s$ are involved in each term in square brackets. The next term, not shown in Equation (9.3.9), would involve the ninth derivative (the eighth evaluates to zero) and would contain $s$ to the ninth power and conclude with $s$ to the fourth power. Thus Equation (9.3.9), as shown, contains all the terms involving $s$ to the third power. If we truncate there and gather like powers of $s$ together we get:

$$
q_r = \frac{1}{s} + \frac{1}{2} - \frac{1}{6} + \frac{s}{12} - \frac{s}{60} + \frac{s^2}{60} - \frac{s^2}{252} - \frac{s^3}{720} + \frac{s^3}{168} - \frac{s^3}{720} + \cdots \quad (9.3.10)
$$

which reduces to:

$$
q_r = \frac{1}{s} \left[ 1 + \frac{s}{3} + \frac{s^2}{15} + \frac{4s^3}{315} + \frac{s^4}{315} + \cdots \right], \quad (9.3.11)
$$

\textsuperscript{12}I cheated. I used a computer symbolic math program called \textit{Derive}, which, among other things, makes far fewer differentiation errors than I do.
or, in terms of \( \Theta_r/T \):

\[
q_{\text{rot}} = \frac{T}{\Theta_r} \left[ 1 + \frac{1}{3} \frac{\Theta_r}{T} + \frac{1}{15} \left( \frac{\Theta_r}{T} \right)^2 + \frac{4}{315} \left( \frac{\Theta_r}{T} \right)^3 + \frac{1}{315} \left( \frac{\Theta_r}{T} \right)^4 + \cdots \right],
\]

(9.3.12)

so that as the temperature increases the terms in the square brackets rapidly become small. Thus the high temperature approximation for the vibrational partition function for a heteronuclear diatomic molecule is simply

\[
q_r = \frac{T}{\Theta_r}.
\]

(9.3.13)

Equation (9.3.12) can be seen to be the high-temperature or “classical” result \( T/\Theta_r \) multiplied by a correction series. This form is very accurate if the temperature is reasonably higher than \( \Theta_r \). In this range convergence is aided both by the powers of \( \Theta_r/T \), which will be less than 1, and by the coefficients in front of each power, which can be shown to converge even if \( \Theta_r/T \) is one.

To see this consider hydrogen chloride, a gas condensing to a liquid at 187.9 K. For this gas \( \Theta_r \) is 15.2. So \( s = 0.0809 \). Then

\[
q_r = 12.362[1 + 0.0270 + 4.4 \times 10^{-4}] = 12.697
\]

so one never needs more than the first correction term (and often not even that) for a heteronuclear diatomic molecule. Indeed, HI, the next highest \( \Theta_r \) has \( s = 0.03 \) and the classical result, \( q_r = 33.33 \) has an error of one in the fourth significant figure at 300 K.

But for completeness, and to show how things are handled, below the temperature where \( \Theta_r/T \) becomes 1, the convergence depends totally on the ratio of the coefficients to the powers of \( \Theta_r/T \).

For that case we can develop another series, one specifically aimed at low temperatures by simply taking Equation (9.3.7) on page 9-9 and writing out the first few terms:

\[
\sum_{J=0}^{\infty} (2J+1)e^{-J(J+1)\Theta_r/T} = 1 + 3e^{-2\Theta_r/T} + 5e^{-6\Theta_r/T} + 7e^{-12\Theta_r/T} + \cdots.
\]

(9.3.14)

As \( T \to 0 \), rotation becomes “frozen” out, which is what we’d expect. That is, there is not enough thermal energy to kick systems into higher rotational levels.

The regions of utility of our two formulas, (9.3.12) and (9.3.14) overlap and in the region where the temperature is near the characteristic temperature, both should be used to check each other.

**Homonuclear Diatomic Molecules**

This case presents difficulties. For quantum mechanical reasons only half the possible rotational quantum numbers actually occur.\(^{13}\)

\(^{13}\)The short explanation for this is that the overall wave function of the molecule is either symmetric or antisymmetric with respect to a 180° rotation of the molecule. The rotational wave functions are symmetric for even \( J \) and antisymmetric for odd \( J \). Only one set can occur in a given homonuclear molecule in a given electronic state.
CHAPTER 9. IDEAL DIATOMIC GASES 9-12

There are two cases here: even rotational quantum number only and odd rotational quantum numbers only. We take the case of even $J$ first. In hydrogen, where these effects can be measured, such hydrogen molecules are called *para-hydrogen* or $p$-hydrogen for short. Which case applies depends on the nuclear spins of the hydrogen atoms.

To treat this we modify the equation for the rotational partition function by substituting $2k$ for $J$, thus ensuring that only even numbers can occur as we increment $k$.

Doing this and also once again writing $s$ for $\Theta_r/T$, we get:

$$q_{r,p} = \sum_{k=0}^{\infty} (4k + 1)e^{-2k(2k+1)s}.$$  \hspace{1cm} (9.3.15)

We still need to use the Euler-Maclaurin summation formula. The details will not be written out here as they closely follow the heteronuclear case. The integral (which is the high-temperature limit) is:

$$\int_{0}^{\infty} (4x + 1)e^{-2x(2x+1)s}ds = \frac{1}{2s},$$  \hspace{1cm} (9.3.16)

or exactly half of the high-temperature limit for the heteronuclear case.

The rotational partition function expression analogous to Equation (9.3.9) on page 9-10 is

$$q_{r,p} = \frac{1}{s} + \left[ \frac{1}{2} \right] + \left[ \frac{s}{6} - \frac{1}{3} \right] + \left[ -\frac{s^3}{90} + \frac{2s^2}{15} - \frac{2s}{15} \right] + \left[ \frac{s^5}{945} - \frac{2s^4}{63} + \frac{4s^3}{21} - \frac{8s^2}{63} \right] + \left[ -\frac{s^7}{9450} + \frac{4s^6}{675} - \frac{4s^5}{45} + \frac{16s^4}{45} - \frac{8s^3}{45} \right] + \cdots,$$  \hspace{1cm} (9.3.17)

which reduces to:

$$q_{r,p} = \frac{1}{2s} \left[ 1 + \frac{s}{3} + \frac{s^2}{15} + \frac{4s^3}{315} + \frac{s^4}{315} + \cdots \right],$$  \hspace{1cm} (9.3.18)

which can be seen to be exactly $1/2$ of Equation (9.3.11) on page 9-10. Thus, unsurprisingly

$$q_{r,p} = \frac{1}{2} q_{\text{rot}}.$$  \hspace{1cm} (9.3.19)

A low temperature series can be developed as before. We have, by expansion of Equation (9.3.15):

$$q_{r,p} = \sum_{k=0}^{\infty} (4k + 1)e^{-2k(2k+1)s} = 1 + 5e^{-6\Theta_r/T} + 9e^{-20\Theta_r/T} + \cdots.$$  \hspace{1cm} (9.3.20)
The other case we have to examine is the one where only odd values of $J$ occur. In hydrogen this is called ortho-hydrogen or $o$-hydrogen for short.

We can repeat the analysis above or we can simply write down the answer by inspection. Since the two homonuclear partition functions must add up to the heteronuclear case, we have for the high-temperature partition function:

$$q_{\text{rot},o} = \frac{1}{2} q_{\text{rot}},$$

(9.3.21)

with the series expansion

$$q_{r,o} = \frac{1}{2s} \left[ 1 + \frac{s}{3} + \frac{s^2}{15} + \frac{4s^3}{315} + \frac{s^4}{315} + \cdots \right].$$

(9.3.22)

The only difference comes in the low-temperature expansion which is:

$$\sum_{k=0}^{\infty} (4k+3)e^{-(2k+1)(2k+2)\Theta_r/T} = 3e^{-2\Theta_r/T} + 7e^{-12\Theta_r/T} + 11e^{-30\Theta_r/T} + \cdots.$$  \hspace{1cm} (9.3.23)

And so all is well from the point of view of the theoretician.

But there is one glaring problem. In the real world\textsuperscript{14} the zero of molecular energy in diatomics is taken as that of the lowest energy state the system can reach. And for $o$-hydrogen, that isn’t the same zero as that for $p$-hydrogen. The former is $\exp(-2\Theta_r/T)$ above the latter. (Look at Equation (9.3.23)!!).

So when doing low temperature calculations involving Equation (9.3.23), a factor of $\exp(-2\Theta_r/T)$ is removed from the partition function, thus making it

$$q_{r,o}^* = 3 + 7e^{-10\Theta_r/T} + 11e^{-28\Theta_r/T} + \cdots,$$

(9.3.24)

where the * on the partition function indicates that it is an “energy corrected” partition function.

\textbf{9.4 The Thermodynamics of Rotation}

Since rotation is almost always classical, the partition function is simple:

$$q_r = \frac{T}{\sigma \Theta_r} = \frac{8\pi^2 I}{\sigma \beta h^2},$$

(9.4.1)

where $\sigma = 1$ for heteronuclear diatomics and $\sigma = 2$ for homonuclear diatomics.

In these terms the thermodynamic functions are simple. The Helmholtz free energy contribution from rotation, $A_r$ is:

$$A_r = -NkT \ln \frac{T}{\sigma \Theta_r},$$

(9.4.2)

\textsuperscript{14}The one in which experimentalists live.
and the energy is
\[ \langle E_r \rangle = NkT \]. \hspace{1cm} (9.4.3)

The heat capacity due to rotation in a diatomic is:
\[ C_{V,r} = Nk \]. \hspace{1cm} (9.4.4)

Note that neither the energy nor the heat capacity depend on \( \sigma \). The entropy works out to be
\[ S_r = Nk \ln \frac{T_e}{\sigma \Theta_r} \]. \hspace{1cm} (9.4.5)
9.5 Appendix: Homonuclear Rotation

9.5.1 Singlets and Triplets

From quantum mechanics we have the following rules:

1. All nuclei have spin quantum numbers of magnitude $J\hbar$ and degeneracy $2J+1$.

2. Allowed values for spin:
   
   (a) The allowed values of $J$ for bosons are the integers $J = 0, 1, \cdots$.
   
   (b) The allowed values of $J$ for fermions are the half-integers $J = 1/2, 3/2, \cdots$.

3. Interchange of particles:
   
   (a) In any molecule interchange of symmetrically placed bosons leaves the wave function $\Psi$ unchanged, i.e. $\Psi \rightarrow \Psi$.
   
   (b) In any molecule interchange of symmetrically placed fermions results in the wave function changing sign, i.e. $\Psi \rightarrow -\Psi$.

If we apply this to nuclei we then have the following rule:

The overall wave function for a homonuclear diatomic molecule whose nuclei are fermions, i.e. $H_2$, will be antisymmetric with respect to a rotation of $180^\circ$.

The overall wave function for a homonuclear diatomic molecule whose nuclei are bosons, i.e. $D_2$, will be symmetric with respect to a rotation of $180^\circ$.

Note that $D_2$ stands for deuterium.

Now let’s think about a homonuclear diatomic molecule where the nuclei are fermions. If we interchange two electrons in $H_2$ the result must change the sign of the overall wave function. So:

$$\Psi \rightarrow -\Psi$$  \hspace{1cm} (9.5.1)

But the translational, rotational, vibrational, and nuclear spin wave functions are unchanged by this. So the overall electron wave function must change sign:

$$\psi_e = -\psi_e$$  \hspace{1cm} (9.5.2)

Now the wave function for the electron is itself composed of two parts, $\psi_{\text{orbit}}$ and $\psi_{\text{spin}}$:

$$\psi_e = \psi_{\text{orbit}}\psi_{\text{spin}}$$  \hspace{1cm} (9.5.3)

Thus since the interchange must be antisymmetric since electrons are fermions:

- if $\psi_{\text{orbit}}$ is anti then $\psi_{\text{spin}}$ must be sym
- if $\psi_{\text{orbit}}$ is sym then $\psi_{\text{spin}}$ must be anti
For H$_2$, the ground state $\psi_{\text{orbit}}$ is symmetric and so $\psi_{\text{spin}}$ must be antisymmetric. The two electrons then have a net spin of zero and the spin degeneracy $\omega_{\text{spin}} = 2J + 1 = 1$. This state is called a singlet.

On the other hand, the first excited state has $\psi_{\text{orbit}}$ antisymmetric, so $\psi_{\text{spin}}$ must be symmetric, the electron spins are parallel, and the net spin is 1. Hence the spin degeneracy $\omega_{\text{spin}} = 2J + 1 = 3$. This state is called a triplet.

9.5.2 Rotational Symmetry

Let us consider a molecule made up of two identical atoms. The rules for the construction of allowed wave functions for this molecule are as given in 9.5.1. Thus the overall wave function for fermions must be antisymmetric while that for bosons must be symmetric.

In the first approximation, the wave function of such a system is the product of translational, vibrational, electronic, rotational, and nuclear wave functions. The exact wave function may have a slightly different form than this product, but no perturbation which may have been neglected in this factoring can change the symmetry character of the overall wave function.$^{15}$

The coordinates of the center of mass of the molecule are unchanged by an interchange of two identical nuclei, so the translational wave function must also be completely unchanged (symmetric) with this interchange.

The vibrational coordinate is also unaffected by the interchange of the nuclei. The vibrational wave function is also then symmetric with regard to the exchange of nuclei.

The electronic wave function is already antisymmetric with regard to the interchange of electrons. It may be either symmetric or antisymmetric with regard to the interchange of nuclei.

For most diatomics, the ground state electronic wave function is symmetric with regard to nuclear interchange.

Thus each of these three functions is unaffected by the exchange of nuclei so their product must also be unaffected by this permutation.

The symmetry character of the wave functions for the entire molecule will be that of the product of the rotational and nuclear spin wave functions.

What happens to the rotational wave function if nuclei are exchanged?

It turns out that this is a fairly easy question to answer as the rotational wave function contains sines and cosines of these angles. Simple examination shows that rotational wave functions having quantum number $J$ even are symmetric with

$^{15}$Most of the material here is simply presented as fact and no proof is offered. Details can be found in any reasonable quantum mechanics text.
respect to nuclear interchange, while rotational wave functions having quantum number \( J \) odd are \textit{antisymmetric} with regard to nuclear interchange.

The next task is to examine the symmetry character of the nuclear spin wave functions.

If the magnitude of the spin (in units of \( \hbar \)) is \( I \), then there are \( 2I + 1 \) possible spin wave functions. For two particles there are \((2I + 1)^2\) possible spin functions.

Of the total it can be shown that \((I + 1)(2I + 1)\) are symmetric and \(I(2I + 1)\) are antisymmetric.

If one considers the simple case if \( I = 0 \), then only one spin function exists for each nuclei and consequently only one for the molecule. This single function is symmetric. Most diatomic molecule ground states are singlets and have \( I = 0 \).

<table>
<thead>
<tr>
<th>If the nuclei are \textit{bosons}, the total wave function for the molecule must be symmetric, and with a symmetric electronic wave function, \textit{only symmetric rotational wave functions will be allowed}. Thus only even values of ( J ) will be allowed.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>If the nuclei are \textit{fermions}, the total wave function for the molecule must be antisymmetric, and with an antisymmetric wave function \textit{only antisymmetric rotational wave functions will be allowed}. Thus only odd values of ( J ) will be allowed.</th>
</tr>
</thead>
</table>

With antisymmetric electronic wave functions, as happens in the case of electronic triplets,\(^{16}\) the situation would be reversed and only odd \( J \)'s would appear with bosons and even \( J \)'s with fermions.

In any case only half of the possible rotational energy levels will exist.

Thus, for determining rotational energies, the nuclear spin wave functions and the rotational wave functions are inextricably linked.

\(^{16}\)Which are often the first excited states of diatomic molecules
Chapter 10

Ideal Polyatomic Gases

10.1 Introduction

An atom has three degrees of freedom. That’s a shorthand way of saying that its motion in the three mutually perpendicular directions are independent of each other. A degree of freedom is an allowed movement, including rotation, that a particle can make. Degrees of freedom persist even if there are restraints on the motion as long as the dimensionality of space is not changed. An atom on a two-dimensional surface has only two degrees of freedom.

Thus two atoms then have a total of six degrees of freedom, three for each atom. A molecule containing \( n \) atoms has \( 3n \) degrees of freedom.

The presence of chemical bonds between atoms does not change the number of degrees of freedom.

If we make use of a notion from mechanics we can gain an important insight. We (mentally) change coordinates from those that describe the position of each particle in the diatomic molecule to a set of three coordinates describing the position of the center of mass of the molecule and a set of three describing the orientation of the molecule about its center of mass.

It was shown in section 2.6 on page 2-8 that this can always be done, no matter how many atoms are present in the molecule or how they are bonded together. The same holds true in quantum mechanics; the Hamiltonian will always separate into the motion of the center of mass and the motion about the center of mass. It won’t be derived here. But if you think back to how the quantum mechanical Hamiltonian is derived from the classical Hamiltonian by the substitution of operators for variables, it is clear the center of mass coordinates will separate out because there is no
potential energy associated with the center of mass.\(^1\)

In general we have three coordinates specifying the position of the center of mass and \(3n - 3\) specifying the positions of the \(n\) atoms \textit{about} the center of mass.

The motion of the center of mass is independent of any motions about the center of mass. This is easy to see since the relative positions contain no reference to the center of mass at all, and vice versa.

As a result we can always separate out the motion of the center of mass of any polyatomic molecule. This motion is the \textit{translational} motion of the molecule as a whole and we’ve already seen how to compute the translational energy and thermodynamics of the center of mass. It is identical to that for a single atom treated in Chapter 8.

Thus a polyatomic molecule has \(3n - 3\) \textit{internal} degrees of freedom. Of these, only three or fewer are \textit{rotational} degrees of freedom. The remainder are all \textit{vibrational} degrees of freedom.

A molecule can have no more than three independent rotational axes through the center of mass. In general any arbitrary rotation can be decomposed into rotations about these three axes. These axes are chosen to be the \textit{principle axes} of the molecule. Those are axes about which the moment of inertia is a minimum. They can always be found, though we will not have to do that here.

Linear molecules have fewer rotational axes. For example a diatomic linear molecule (like \(\text{H}_2\)) or a polyatomic linear molecule (like acetylene, \(\text{HCCH}\)) have only two. The principle axes of a linear molecule are of necessity one along the bond axis of the molecule and the other two are any two axes through the center of mass at right angles to each other and at right angles to the bond axis.

The rotational axis along the bond axis is, in quantum mechanics, \textit{not} a physical rotational axis. Any rotation around that axis results in no change whatsoever to the molecule, its wave function, or its properties. This is because of the fact that atoms can have no markings and the simple rotation of an atom can’t be distinguished from no rotation at all, even in principle.\(^2\)

Thus linear molecules have only two rotational degrees of freedom, all others have three. And accounting for the three translational degrees of freedom there are then \(3n - 6\) or \(3n - 5\) vibrational degrees of freedom, depending on the shape of the molecule.

So polyatomic molecules differ from atoms in that atoms have only translational degrees of freedom\(^3\) while a polyatomic molecule can have all of:

1. translational energy
2. rotational energy

\(^1\)If there is a potential energy associated with the center of mass, things \textit{might} be very different.
\(^2\)In other words, this is a quantum mechanical effect. Classical mechanics always implicitly assumes that any particle can be marked.
\(^3\)Electronic energies are not considered degrees of freedom in this sense.
3. vibrational energy

If we assume that these energies are additive, that is that the particle energy $\epsilon$ is made by adding the energies of each of these separate contributions:

$$\epsilon = \epsilon_{\text{translation}} + \epsilon_{\text{rotation}} + \epsilon_{\text{vibration}}, \quad (10.1.1)$$

then we can treat each of these forms of energy separately and add the results.

Assuming that the polyatomic molecule does not interact with any other molecules in the system, then the intermolecular potential energy is identically zero at all distances. That being assumed, we can then factor the canonical partition function $Q$ into independent molecular parts $q$ such that once again:

$$Q(N,V,T) = \frac{1}{N!} q(V,T)^N, \quad (10.1.2)$$

with now

$$q(V,T) = \sum_j e^{-\beta \epsilon_j} = \sum e^{-\beta (\epsilon_t + \epsilon_r + \epsilon_v)} = \sum e^{-\beta \epsilon_t} e^{-\beta \epsilon_r} e^{-\beta \epsilon_v} = \sum e^{-\beta \epsilon_t} \sum e^{-\beta \epsilon_r} \sum e^{-\beta \epsilon_v} = q_t q_r q_v, \quad (10.1.3)$$

where only $q_t$ is a function of both the volume and the temperature. The other $q$'s are functions of temperature alone.\(^5\)

This separation has a further effect on the calculation of the thermodynamics of a polyatomic molecule. Given Equation (10.1.2) we then get:

$$\ln Q(N,V,T) = -\ln N! + N \ln q = -\ln N! + N \ln q_t + N \ln q_r + N \ln q_v. \quad (10.1.4)$$

As a result the total energy of a polyatomic molecule also is, not unexpectedly, a sum such that:

$$\langle \epsilon \rangle = \langle \epsilon_t \rangle + \langle \epsilon_r \rangle + \langle \epsilon_v \rangle, \quad (10.1.5)$$

so that it makes sense to talk about the rotational energy or the vibrational energy of a polyatomic molecule. But this is an idealization since in reality these energies, especially at higher temperatures, are not really independent.

\(^4\)This does not have to be true. In fact, as already noted in the discussion of diatomic molecules, vibrational and rotational energies interact since rapid rotation stretches the bond and increases the bond length and higher vibrational levels change the moment of inertia of the molecule because the vibrations are only approximately harmonic. However, the translational motion is rigorously always separate.

\(^5\)Unless the volume of the container is so small that it interferes with vibration or rotation or both. In that case we are not in a classical situation and quantum statistics must be used.
By the way, it is customary to include the $1/N!$ factor with the translational partition function. Thus we write $q_t/N!$ for it.

We’ve already computed $q_t$ in Chapter 8 and so don’t have to repeat that. Of course the mass $m$ in Chapter 8 becomes the total mass of the molecule.

It remains now to find $q_r$ and $q_v$.

## 10.2 Vibration in Polyatomic Molecules

Vibrations in polyatomic molecules are, in principle, not in any way different than vibrations in diatomic molecules. The only practical difference lies in the fact that the vibrations generally are not along the chemical bonds. Usually a given vibrational mode will involve a number or even all the chemical bonds.

This is so because vibrations in a single bond most always will move the center of gravity. We can’t allow such vibrations because we have removed the motion of the center of gravity from these calculations.

If a molecule has $c$ vibrational degrees of freedom, then it is always possible to find $c$ distinct sets of independent motions involving vibrations that do not cause the center of mass to move. These motions are called normal vibrational modes or normal modes for short.

Discovering the vibrational modes is a tedious exercise in mathematics. Suffice it to say that once these are determined and the fundamental frequencies $\nu$ known, the rest is easy.\(^6\)

If we assume that we have the fundamental frequencies $\nu_i$ for the $s$ vibrational modes, then we have, from Equation (9.2.12) on page 9-4 for each mode:

$$\Theta_{v_i} = \frac{\hbar \nu_i}{k}, \quad (10.2.1)$$

and since the vibrational modes are independent, then

$$q_v = \prod_{i=1}^{s} \frac{e^{\Theta_{v_i}/2T}}{1 - e^{-\Theta_{v_i}/T}}. \quad (10.2.2)$$

The vibrational energy per molecule is

$$\langle \epsilon_v \rangle = k \sum_{i=1}^{s} \left( \frac{\Theta_{v_i}}{2} + \frac{\Theta_{v_i}}{e^{\Theta_{v_i}/T} - 1} \right), \quad (10.2.3)$$

and the vibrational heat capacity per molecule is:

$$C_{v,V} = k \sum_{i=1}^{s} \left[ \frac{\Theta_{v_i}}{T} \right]^{2} \frac{e^{\Theta_{v_i}/T}}{(e^{\Theta_{v_i}/T} - 1)^2}. \quad (10.2.4)$$

\(^6\)Actually, determining the fundamental frequencies $\nu$ is a rather large problem, difficult in molecules with tens or atoms and essentially impossible in molecules like DNA because of the huge number of vibrational modes!
10.3 Rotation in Polyatomic Molecules

10.3.1 Principle Axes and the Moment of Inertia Tensor

In general a molecule has three axes of rotation. These are always chosen so that they meet at the center of mass of the molecule. This ensures that rotation will not move the center of mass.\(^7\) In rotation moment of inertia plays a role analogous to the role played by mass in translation. The smaller the moment of inertia about a given axis, the easier it is to accelerate the rotation about that axis.

The moment of inertia about an axis for a molecule with \(n\) atoms is given by:

\[
I_{xx} = \sum_{i=1}^{n} m_i (y_i^2 + z_i^2),
\]

where \(I_{xx}\) is the moment of inertia about the \(x\)-axis, \(m_i\) is the mass of the \(i\)'th atom, and \(y_i\) and \(z_i\) are the respective distances of the \(i\)'th atom from the \(y\) and \(z\) axes.

Similarly we have \(I_{yy}\), involving distances \(x_i\) and \(z_i\), and \(I_{zz}\) involving distances \(x_i\) and \(y_i\).

However, there are other components to the inertia. These are the products of inertia \(I_{xy}\), \(I_{yz}\), and \(I_{zx}\) given by:

\[
I_{xy} = -\sum_{i=1}^{n} m_i x_i y_i,
\]

and two similar equations for \(I_{yz}\) and \(I_{zx}\). Here again \(x_i\) and \(y_i\) are the distances to the \(x\) and \(y\) axes, respectively.

The result is the moment of inertia tensor:

\[
\mathbf{I} = \begin{bmatrix}
I_{xx} & I_{xy} & I_{xz} \\
I_{yx} & I_{yy} & I_{yz} \\
I_{zx} & I_{zy} & I_{zz}
\end{bmatrix},
\]

which is symmetric since, as can be seen from Equation (10.3.3), \(I_{xy} = I_{yx}\).

If the coordinates are rotated with respect to the molecule, the value of the elements change. Since the matrix is symmetric and real, a position of the coordinates can always be found that makes all the off-diagonal elements zero. The axes are then called the principal axes of the molecule.

There are three such choices, corresponding to a different (right-handed) permutation of the coordinate names. The standard choice is that

\[
I_{xx} \leq I_{yy} \leq I_{zz}.
\]

\(^7\)Otherwise rotation would move the center of mass and so rotation would not be independent of translation.
10.3.2 The Rotational Partition Functions

In general, as we’ve noted before, linear polyatomic molecules have two rotational degrees of freedom, non-linear polyatomic molecules have three.

In the case of non-linear polyatomic molecules, there are three moments of inertia. The standard notation has them arranged in the following order:

\[ I_A \leq I_B \leq I_C. \]  \hfill (10.3.5)

These are related to the spectroscopic symbols \( \tilde{A}, \tilde{B}, \) and \( \tilde{C} \) by:

\[ \tilde{A} = \frac{h}{8\pi^2 c I_A}, \quad \tilde{B} = \frac{h}{8\pi^2 c I_B}, \quad \tilde{C} = \frac{h}{8\pi^2 c I_C}. \]  \hfill (10.3.6)

where \( c \) is the speed of light.

Thus relationship (10.3.5) can be written:

\[ \tilde{A} \geq \tilde{B} \geq \tilde{C}. \]  \hfill (10.3.7)

Keeping this in mind, we have the standard definitions given in Table 10.1.

| \( I_A = I_B = I_C \) | symmetric top | \( \tilde{A} = \tilde{B} = \tilde{C} \) |
| \( I_A = I_B < I_C \) | oblate spheroidal top | \( \tilde{A} = \tilde{B} > \tilde{C} \) |
| \( I_A < I_B = I_C \) | prolate spheroidal top | \( \tilde{A} > \tilde{B} = \tilde{C} \) |
| \( I_A < I_B < I_C \) | asymmetric top | \( \tilde{A} > \tilde{B} > \tilde{C} \) |

Table 10.1: Names for Rigid Rotators

A prolate spheroid is one that is (American) football shaped. And oblate spheroid is one that is a somewhat squashed basketball.

The energy state structure and degeneracies for the first three types can be found fairly easily from quantum mechanics.\(^9\)

Symmetric Top: \( \epsilon_J = \tilde{B} J(J + 1), \quad g_J = (2J + 1)^2, \quad J = 0, 1, 2, \ldots. \)  \hfill (10.3.8)

Equation (10.3.8) also holds for linear molecules but with degeneracy \( g_J = 2J + 1. \)

Oblate: \( \epsilon_{J,K} = \tilde{B} J(J + 1) + (\tilde{C} - \tilde{B}) K^2, \quad g_J = 2J + 1, \quad J = 0, 1, 2, \ldots, \quad K = 0, \pm 1, \pm 2, \ldots, \pm J. \)  \hfill (10.3.9)
CHAPTER 10. IDEAL POLYATOMIC GASES

10.3 The Canonical Rotational Partition Function

and

\[ \epsilon_{J,K} = \hat{B} J (J + 1) + (\hat{A} - \hat{B}) K^2, \quad g_J = 2J + 1, \]

\[ J = 0, 1, 2, \ldots, \quad K = 0, \pm 1, \pm 2, \ldots, \pm J. \]  \tag{10.3.10}

Note that these have two associated quantum numbers, \( J \) and \( K \), both of which affect the actual energy.

The case of the asymmetric top is different. There is no simple formula for the energy levels, and energy level calculations have to be done for each specific molecule.

The canonical rotational partition functions for the four varieties of rotator presented here (the linear molecule has already been discussed) are simple in the high temperature limit. The first three can easily, and the asymmetric top with some difficulty, can be shown to give:

\[ q_r(T) = \frac{\pi^{1/2}}{\sigma} \left( \frac{T}{\Theta_{r,A}} \right)^{1/2} \left( \frac{T}{\Theta_{r,B}} \right)^{1/2} \left( \frac{T}{\Theta_{r,C}} \right)^{1/2}, \]  \tag{10.3.11}

where the \( \Theta_r \) are the rotational characteristic temperatures defined by

\[ \Theta_{r,A} = \frac{\hbar^2}{8\pi^2 k I_A}, \]  \tag{10.3.12}

and where \( A \) could be \( A, B, \) or \( C \) with a corresponding change in the moment of inertia around that axis, \( I_A, I_B, \) or \( I_C. \) The equation contains a \( \sigma \), which is the usual symmetry number.

In using Equation (10.3.11) one must remember to make the corresponding \( \Theta_r \)'s equal if the corresponding moments of inertia \( I \) are the same. Thus the symmetric top, with all moments of inertia the same, becomes

\[ q_r(T) = \frac{\pi^{1/2}}{\sigma} \left( \frac{T}{\Theta_{r,A}} \right)^{3/2}. \]  \tag{10.3.13}

10.4 The Electronic Partition Function in Polyatomic Molecules

The electronic partition function for polyatomic molecules is, unsurprisingly:

\[ q_e = [\omega_1 + \omega_2 e^{-\beta \epsilon_2} + \omega_3 e^{-\beta \epsilon_3} + \cdots] e^{\beta D_e}, \]  \tag{10.4.1}

where \( \omega_j \) is the degeneracy of the \( j \)’th electronic state and \( \epsilon_j \) is the corresponding electronic energy. The quantity \( D_e \) is the (negative) binding energy of the molecule.\(^{11}\)


\(^{11}\)Remember: as is true for diatomic molecules, the zero of energy for polyatomic molecules is taken to be the energy of the separated atoms. This energy is negative in magnitude.
As is well-known, it is not possible to solve the correct quantum mechanical equations for any polyatomic molecule. Of course electronic energy levels for such molecules can be approximately computed in various ways. But it is far more simple (and accurate) to use energies obtained by spectroscopy.

10.5 The Thermodynamics of Polyatomic Molecules

The following section is probably redundant, but it is useful to have all of these formulas together in one spot.

The partition function for a polyatomic molecule is not complex, but it is quite long.\(^{12}\) It is best to consider that \(q\) for an ideal polyatomic molecule in the rigid-rotor harmonic oscillator approximation as a product of partition functions for each separate type of motion. Thus

\[
q = q_t q_v q_r q_e
\]  

(10.5.1)

where the terms are, from left to right, the translational, vibrational, rotational, and electronic molecular partition functions.

In these terms we can write:

\[
Q = \frac{q_t^N}{N!} q_v^N q_r^N q_e^N.
\]  

(10.5.2)

10.5.1 Translation

We have for translation:\(^{13}\)

\[
q_t = \left(\frac{2\pi M kT}{\hbar^2}\right)^{2/3} V,
\]  

(10.5.3)

and, using obvious notations:

\[
A_t(N,V,T) = -NkT \ln \frac{q_t}{N},
\]  

(10.5.4)

Then, using Equation (10.5.3)

\[
E_t = \frac{3}{2} NkT \quad C_{V,t} = \frac{3}{2} Nk,
\]  

(10.5.5)

\[
p = \frac{NkT}{V}.
\]  

(10.5.6)

\(^{12}\) Indeed, the author used to love arranging his lectures so as to leave the completely written out partition function on the classroom blackboard for the edification of the class next using the room.

\(^{13}\) This is the classical result. Of course the volume would have to be very small or the temperature microscopic to have to use the appropriate quantum expression.
Note that the pressure does not have a subscript because $V$ occurs only in the translational partition function. Hence neither vibration nor rotation (or electronic excitation) affects the pressure of an ideal gas.

Also:

$$\mu_t = -kT \ln \frac{q_t}{N}, \quad (10.5.7)$$

and

$$S_t = Nk \ln \left[ \left( \frac{2\pi MkT}{h^2} \right)^{3/2} \frac{V_t^{5/2}}{N} \right]. \quad (10.5.8)$$

**10.5.2 Vibration**

The vibrational partition function $q_v$ is given by:

$$q_v = \prod_{j=1}^{r} \frac{e^{-\Theta_{v,j}/2T}}{(1 - e^{-\Theta_{v,j}/T})}, \quad (10.5.9)$$

where $r$ is either $3n - 5$ or $3n - 6$ depending on the molecule’s being $n$ atoms in a line or not, respectively.

Then:

$$A = -NkT \ln q_v, \quad (10.5.10)$$

$$E_v = Nk \sum_{j=1}^{r} \left( \frac{\Theta_{v,j}}{2} + \frac{\Theta_{v,j}e^{-\Theta_{v,j}/T}}{(1 - e^{-\Theta_{v,j}/T})} \right), \quad (10.5.11)$$

and

$$C_{V,v} = NK \sum_{j=1}^{r} \left[ \frac{(\Theta_{v,j}/T)^2}{e^{-\Theta_{v,j}/T} - 1} + \ln(1 - e^{-\Theta_{v,j}/T}) \right], \quad (10.5.12)$$

where $\Theta_{v,j} = h\nu_j/k$.

There is no pressure due to vibration as vibration does not depend on volume. The vibrational contribution to the chemical potential is:

$$\mu_v = -kT \ln \frac{q_v}{N}, \quad (10.5.13)$$

and the entropy due to vibration is:

$$S = Nk \sum_{j=1}^{r} \left[ \frac{\Theta_{v,j}/T}{e^{\Theta_{v,j}/T} - 1} - \ln(1 - e^{-\Theta_{v,j}/T}) \right], \quad (10.5.14)$$

---

14 Derivation of this formula is left to the reader.

15 And so is the derivation of this one.
10.5.3 Rotation

Since rotation is almost always classical, the partition function is simple:

\[ q_r = \frac{\pi^{1/2}}{\sigma} \left( \frac{T^3}{\Theta_A \Theta_B \Theta_C} \right)^{1/2}, \]  
(10.5.15)

where \( \sigma \) is the symmetry number of the molecule and the \( \Theta \)'s are the rotational temperatures. Two or even all three \( \Theta \)'s may be the same in a given molecule. In case of a linear molecule, one \( \Theta \) will be missing.

In these terms the thermodynamic functions are simple. The Helmholtz free energy contribution from rotation, \( A_r \) is:

\[ E_r = \frac{3}{2} N k T. \]  
(10.5.16)

This is \( E_r = N k T \) in the case of a linear molecule due to the loss of one rotational degree of freedom. The heat capacity due to rotation is:

\[ C_{V,r} = \frac{3}{2} N k, \]  
(10.5.17)

or, for a linear molecule, \( C_{V,r} = N k. \)

The chemical potential for rotation is, as usual:

\[ \mu_v = -kT \ln \frac{q_r}{N}. \]  
(10.5.18)

The entropy contribution from rotation is:

\[ S_r = N k \ln \frac{\pi^{1/2} e^{3/2}}{\sigma} \left( \frac{T^3}{\Theta_A \Theta_B \Theta_C} \right)^{1/2} \]  
(10.5.19)
Chapter 11

The Grand Canonical Ensemble

11.1 Introduction

We have looked at systems with independent variables $N$, $V$, and $E$ (the micro-canonical ensemble), and systems with independent variables $N$, $V$, and $T$ (the canonical ensemble). These are by far not the only possible ensembles.

Indeed, there are six common (but not equally useful) ensembles for single component, single phase systems alone. If more components are added, the number of possible ensembles increases dramatically.

This chapter discusses an ensemble that allows the number of particles in a system to vary. This is often quite useful, as shall be seen.

11.2 The Equations for the Ensemble

If we pick $\mu$, $V$, and $T$ as the independent variables, the resulting ensemble is called the grand canonical ensemble.

Clearly this is an open system since the number of particles is allowed to vary. But it is a feasible ensemble and systems with such independent variables can be constructed experimentally. The walls need to be fixed and rigid as well as diathermal. And the walls also have to be permeable to the particles that make up the system. Perforated steel sheeting would suit admirably, as would a number of other choices.

To control the temperature and the chemical potential we need both a constant
temperature bath and a bath at a constant chemical potential. These can even be combined. It isn't always easy to control chemical potential, but it can be done in several ways. The most obvious is to place our system into a huge bath, much larger than the system, containing the same particles that are in the system but at the required chemical potential. Setting that chemical potential in the bath can almost always be done. Electrochemical means is but one technique.\footnote{What can and cannot be controlled depends both on the chemical species in the system and the state of current technology. We will take the view here that it can be done for any system of interest and leave the details to be examined at the time particular systems are treated.}

We now (mentally) replicate this system a huge number of times, and once again reduce the problem to that of the microcanonical ensemble by wrapping the ensemble and its assorted baths in an adiabatic impermeable wrapper.

Our problem then is to once again find the set of occupation numbers that maximizes $W_{\text{total}}$ the number of ways of arranging $\mathbf{A}$ ensembles among the possible system quantum states subject to both the constraint of constant temperature and constant chemical potential.

The occupation numbers are a bit more complex here than previously. Since the number of particles is variable, we have a whole series of possible energy levels, one for each possible number of particles in the system, including zero particles in the system!

For instance if we have 14 particles, then we might use $a_{14,3}$ to denote the number of ensemble members with 14 particles which is in the third quantum state for such systems. And $a_{254,7}$ would be the number of ensemble members with 254 particles in the seventh quantum state for such systems.

The set of occupation numbers can be written:

$$
a_{0,0}
a_{1,1}, a_{1,2}, \ldots, a_{1,j}, \ldots
a_{2,1}, a_{2,2}, \ldots, a_{2,j}, \ldots
\ldots
a_{k,1}, a_{k,2}, \ldots, a_{k,j}, \ldots
\ldots
$$

where the term $a_{0,0}$ is not a mistake. It is perfectly possible for a system to contain no particles at all. Of course then its energy is zero.

What we have to maximize is:

$$W_{\text{total}} = W_{\text{tbath}} \times W_{\mu \text{ bath}} \times W_{\text{sys}}. \quad (11.2.1)$$

Now $W_{\text{tbath}}$, the ways of internally arranging the constant temperature bath, depends on the energy of the bath $E_{\text{tbath}}$ And $W_{\mu \text{ bath}}$, the ways of internally arranging the constant chemical potential bath, depends on the number of particles $N_{\mu \text{ bath}}$ in the bath. And $W_{\text{sys}}$, the number of ways of internally arranging the members of the system, depends on the usual multinomial coefficient. These three terms are
otherwise independent of each other. So Equation (11.2.1) on page 11-2 becomes

\[ W_{\text{total}} = W_{\text{tbath}}(E_{\text{tbath}}) \times W_{\mu \text{ bath}}(N_\mu \text{ bath}) \times \frac{A!}{a_{0,0}!a_{1,1}! \cdots a_{N,j}!}, \]  

(11.2.2)

and we have to maximize with respect not only to the \( a \)’s but to the energy of the temperature bath and to the number of particles in the chemical potential bath. Thus there are \( A + 2 \) variables.

But not all of these are independent. There are three constraints:

\[ \sum_{N=0}^{\infty} \sum_{j=1}^{\infty} a_{N,j} = A, \]  

(11.2.3)

\[ \sum_{N=0}^{\infty} \sum_{j=1}^{\infty} a_{N,j} E_{N,j} + E_{\text{tbath}} = C, \]  

(11.2.4)

\[ \sum_{N=0}^{\infty} \sum_{j=1}^{\infty} a_{N,j} N_j + N_\mu \text{ bath} = \mathfrak{R}, \]  

(11.2.5)

where \( A, C \) and \( \mathfrak{R} \) are constants.

We denote the two baths separately above for convenience. As will be seen each “separate” bath contributes its own special properties to the systems. In reality the baths might be combined into one bath, but that need not bother us here.

### 11.3 Solution of the Equations

The solution takes place essentially by the same steps that were used for the canonical ensemble, except now there are three constraints.

The first step is to convert Equation (11.2.2) to logarithmic form. This gives:

\[ \ln W_{\text{total}} = \ln W_{\text{tbath}}(E_{\text{tbath}}) + \ln W_{\mu \text{ bath}}(N_\mu \text{ bath}) \]

\[ + \ln A! - \sum_{N} \sum_{j} (a_{N,j} \ln a_{N,j} - a_{N,j}), \]  

(11.3.1)

where Stirling’s approximation has been used on the last term.

To maximize this we take the total derivative of \( W_{\text{total}} \) and set it to zero:

\[ d \ln W_{\text{total}} = \left( \frac{\partial \ln W_{\text{tbath}}(E_{\text{tbath}})}{\partial E_{\text{tbath}}} \right) dE_{\text{tbath}} \]

\[ + \left( \frac{\partial \ln W_{\mu \text{ bath}}(N_\mu \text{ bath})}{\partial N_\mu \text{ bath}} \right) dN_\mu \text{ bath} - \sum_{N} \sum_{j} a_{N,j} d a_{N,j} = 0. \]  

(11.3.2)
We now take the total derivative of each of Equations (11.2.3) to (11.2.5) on page 11-3 and multiply them by $\alpha$, $\beta$, and $\gamma$, respectively.

\[
\alpha \sum_{N=0}^{\infty} \sum_{j=1}^{\infty} da_{N,j} = 0, \quad (11.3.3)
\]

\[
\beta \left[ \sum_{N=0}^{\infty} \sum_{j=1}^{\infty} E_{N,j} da_{N,j} + dE_{\text{bath}} \right] = 0, \quad (11.3.4)
\]

\[
\gamma \left[ \sum_{N=0}^{\infty} \sum_{j=1}^{\infty} N da_{N,j} + dN_{\mu \text{ bath}} \right] = 0. \quad (11.3.5)
\]

These are then subtracted\footnote{Again, this is done for convenience. To add instead just change the signs of $\alpha$, $\beta$, and $\gamma$.} from Equation (11.3.2) on page 11-3 to give:

\[
d \ln W_{\text{total}} = \left[ \left( \frac{\partial \ln W_{\text{bath}}(E_{\text{bath}})}{\partial E_{\text{bath}}} \right) - \beta \right] dE_{\text{bath}}
\]

\[
+ \left[ \left( \frac{\partial \ln W_{\mu \text{ bath}}(N_{\mu \text{ bath}})}{\partial N_{\mu \text{ bath}}} \right) - \gamma \right] dN_{\mu \text{ bath}}
\]

\[- \sum_{N} \sum_{j} \left[ \ln a_{N,j} + \alpha + \beta E_{N,j} + \gamma N \right] da_{N,j}. \quad (11.3.6)
\]

We can make the $\text{tbath}$ term disappear by setting

\[
\beta = \left( \frac{\partial \ln W_{\text{bath}}(E_{\text{bath}})}{\partial E_{\text{bath}}} \right), \quad (11.3.7)
\]

and the $\mu$ bath term do the same by setting

\[
\gamma = \left( \frac{\partial \ln W_{\mu \text{ bath}}(N_{\mu \text{ bath}})}{\partial N_{\mu \text{ bath}}} \right), \quad (11.3.8)
\]

and also one of the $a$’s vanish by taking, for example,

\[
\ln a_{3,1} + \alpha + \beta E_{3,1} + 3\gamma = 0. \quad (11.3.9)
\]

Equation (11.3.6) then reduces to:

\[
\sum_{N} \sum_{j} \left[ \ln a_{N,j} + \alpha + \beta E_{N,j} + \gamma N \right] da_{N,j} = 0, \quad (11.3.10)
\]

for all $N$ and $j$ except for $(3,1)$. But since Equation (11.3.9) also fits Equation (11.3.10) we can include $(3,1)$ here without error.

Now we make our usual argument in which we claim that since all of the $da$’s are independent, the only way Equation (11.3.10) can be satisfied for all values of the $da_{N,j}$ will be if the quantity in the square brackets is identically zero. This give us:

\[
\ln a_{N,j} + \alpha + \beta E_{N,j} + \gamma N = 0, \quad (11.3.11)
\]
for all $N$ and $j$ so that
\begin{equation}
  a_{N,j} = e^{-\alpha} e^{-\beta E_{N,j}} e^{-\gamma N}.
\end{equation}

As before, $\alpha$ can be found from the first constraint equation, Equation (11.3.3) on page 11-4:
\begin{equation}
  \sum_N \sum_j a_{N,j} = e^{-\alpha} \sum_N \sum_j e^{-\beta E_{N,j}} e^{-\gamma N} = \mathcal{A},
\end{equation}

so
\begin{equation}
  e^{-\alpha} = \frac{\mathcal{A}}{\sum_N \sum_j e^{-\beta E_{N,j}} e^{-\gamma N}}.
\end{equation}

It is convenient to define
\begin{equation}
  \Xi(\gamma, V, \beta) = \sum_N \sum_j e^{-\beta E_{N,j}} e^{-\gamma N}.
\end{equation}

This quantity is known as the **grand canonical partition function**. The symbol for it, $\Xi$, is pronounced ksee. It plays the same role in the grand canonical ensemble that $Q(N, V, \beta)$ does for the canonical ensemble.

With this definition Equation (11.3.14) becomes
\begin{equation}
  e^{-\alpha} = \frac{\mathcal{A}}{\Xi(\gamma, V, \beta)}.
\end{equation}

The probability if finding a system in the ensemble with $N$ particles and energy level $j$ is then:
\begin{equation}
  P_{N,j} = \frac{a_{N,j}}{\mathcal{A}} = \frac{e^{-\beta E_{N,j}} e^{-\gamma N}}{\Xi(\gamma, V, \beta)}.
\end{equation}

### 11.4 Thermodynamics of the Grand Canonical Ensemble

Given Equation (11.3.17) it is simple to work out the thermodynamics of the system and, at the same time, identify $\beta$ and $\gamma$. The method follows that for the canonical ensemble.

The average value of the energy $\langle E \rangle$ is
\begin{equation}
  \langle E \rangle = \sum_N \sum_j P_{N,j} E_{N,j} = \frac{1}{\Xi(\mu, V, \beta)} \sum_N \sum_j E_{N,j} e^{-\beta E_{N,j}} e^{-\gamma N}
\end{equation}

which can be seen to be the same as:
\begin{equation}
  \langle E \rangle = - \left( \frac{\partial \ln \Xi}{\partial \beta} \right)_{V, \gamma}.
\end{equation}

We can immediately identify this energy with the internal energy of the system.
In the same manner it can be shown that
\[ \beta \langle p \rangle = \left( \frac{\partial \ln \Xi}{\partial V} \right)_{\beta,\gamma}, \] (11.4.3)
where we identify \( \langle p \rangle \) with the thermodynamic pressure \( p \), and
\[ \langle N \rangle = -\left( \frac{\partial \ln \Xi}{\partial \gamma} \right)_{\beta,V}. \] (11.4.4)
where we identify \( \langle N \rangle \) with the thermodynamic number of particles in the system.

We now have what we need to write down the total derivative of the grand canonical partition function \( \Xi(\beta, V, \gamma) \):
\[ d \ln \Xi(\mu, V, T) = \left( \frac{\partial \ln \Xi}{\partial \beta} \right)_{V,\gamma} d\beta + \left( \frac{\partial \ln \Xi}{\partial V} \right)_{\beta,\gamma} dV + \left( \frac{\partial \ln \Xi}{\partial \gamma} \right)_{\beta,V} d\gamma. \] (11.4.5)
or, in terms of Equations ((11.4.2) – (11.4.4)):
\[ d \ln \Xi(\mu, V, T) = -Ed\beta + \beta pdV - Nd\gamma. \] (11.4.6)
The equivalent thermodynamic equation is Equation (1.7.24) on page 1-9 which is:
\[ \frac{pV}{kT} = -E \left( \frac{1}{kT} \right) d\beta + \beta pdV - N \left( \frac{\mu}{kT} \right) d\gamma, \] (11.4.7)
from which we can make the identifications:
\[ \beta = \frac{1}{kT} \quad \text{and} \quad \gamma = -\frac{\mu}{kT}, \] (11.4.8)
and, most importantly:
\[ pV = kT \ln \Xi(T, V, \mu). \] (11.4.9)
Last, we find a formula for the entropy. This is most easily done from the thermodynamic relationship:
\[ E = TS - pV + N\mu, \] (11.4.10)
which can be rewritten as:
\[ S = \frac{E}{T} - \frac{N\mu}{T} - \frac{pV}{T}, \] (11.4.11)
and so \( S \) can be found from
\[ S = -\frac{1}{T} \left( \frac{\partial \ln \Xi}{\partial \beta} \right) + \frac{\mu}{T} \left( \frac{\partial \ln \Xi}{\partial \gamma} \right) + k \ln \Xi, \] (11.4.12)
although it is almost always easier to compute \( E, N, \) and \( p \) and then compute \( S \) from Equation (11.4.11).
11.5 Entropy and Probability

Equation (11.3.17) on page 11-5 is:

\[ P_{N,j} = \frac{a_{N,j}}{A} = \frac{e^{-\beta E_{N,j} - \gamma N}}{\Xi(\gamma, V, \beta)}. \]  

(11.3.17)

We can manipulate this to get:

\[ \ln P_{j,N} = -\beta E_{j,N} - \gamma N - \ln \Xi, \]  

(11.5.1)

\[-P_{j,N} \ln P_{j,N} = \beta E_{j,N} P_{j,N} + \gamma N P_{j,N} - P_{j,N} \ln \Xi, \]

\[-\sum_{j,N} P_{j,N} \ln P_{j,N} = \beta \sum_{j,N} E_{j,N} P_{j,N} + \gamma \sum_{j,N} N P_{j,N} - \ln \Xi \sum_{j,N} P_{j,N}, \]

\[-\sum_{j,N} P_{j,N} \ln P_{j,N} = \beta \langle E \rangle + \gamma \langle N \rangle - \ln \Xi. \]

Given Equation (11.4.7) on page 11-6 and our other identifications we have:

\[-\sum_{j,N} P_{j,N} \ln P_{j,N} = E_{kT} + \mu N kT + A kT \]  

(11.5.2)

We had previously gotten Equation (1.7.8) on page 1-8:

\[ S = \frac{1}{k} E + \frac{p}{kT} V - \frac{\mu}{kT} N \]  

((1.7.8))

so that we have, once again\(^3\)

\[ S = -k \sum_{j,N} P_{j,N} \ln P_{j,N} \]  

(11.5.3)

11.6 The Relationship to the Canonical Ensemble

There are some interesting relationships among our ensembles.

The definition of the grand canonical partition function \( \Xi \) is:

\[ \Xi(\gamma, V, \beta) = \sum_N \sum_j e^{-\beta E_{N,j} - \gamma N} \]  

((11.3.15))

If we do the sums separately we can write:

\[ \Xi(T, V, \mu) = \sum_N e^{\mu N/kT} \sum_j e^{-E_{j,N}/kT} \]  

(11.6.1)

which simplifies to be:

\[ \Xi(T, V, \mu) = \sum_N Q(N, V, T) e^{\mu N/kT} \]  

(11.6.2)

\(^3\)Indeed, it is always true for all ensembles
Since $Q$ itself can be written in terms of $\Omega$, we have

$$\Xi(T, V, \mu) = \sum_N \left( \sum_j \Omega(N, V, E_j)e^{-E_j/kT} \right)e^{\mu N/kT} \quad (11.6.3)$$

This sort of summation of one quantity (for example $\Omega$ to get another (for example $Q$) is called a Laplace Transformation.

### 11.7 A Direct Consequence of Having Independent Subsystems

We have seen earlier (Chapter 7) that having independent subsystems leads to expressing the canonical partition function in terms of single subsystem partition functions.

Using this in the grand canonical partition function leads directly to an interesting and very general result.

The grand canonical partition function can be written as:

$$\Xi(\beta, V, \gamma) = \sum_N Q(N, V, \beta)e^{-\gamma N}, \quad (11.7.1)$$

where $\gamma = -\mu/kT$. Now for convenience let

$$\lambda = e^{-\gamma} = e^{\mu/kT}, \quad (11.7.2)$$

where $\lambda$ is known as the absolute activity because

$$\ln \lambda = \mu/kT \quad (11.7.3)$$

and so

$$\mu = kT \ln \lambda. \quad (11.7.4)$$

Since a chemical potential is normally written in terms of the thermodynamic activity $a$ as:

$$\mu = \mu^o + kT \ln a, \quad (11.7.5)$$

(where the activity is given per molecule instead of per mole), it is clear that $\lambda$ is an activity per particle in a system in which $\mu^o$ is identically zero in the standard state. Thus the name absolute activity.

With this definition of $\lambda$ in mind (and, by the way we shall use it often in the rest of this work), we have:

$$\Xi(\beta, V, \gamma) = \sum_N Q(\beta, V, N)\lambda^N$$

$$= \sum_N \frac{1}{N!} q(V, \beta)^N \lambda^N$$

$$= \sum_N \frac{1}{N!} (q\lambda)^N$$

$$= e^{q\lambda}, \quad (11.7.6)$$
where the last line comes from comparison of the line just above it to the series expansion of \(\exp x\).

So we have evaluated the grand partition function on the assumption that (1) the particles are independent and (2) that Boltzmann statistics apply.

But we are not done.

\[
\ln \Xi(\beta, V, \gamma) = q\lambda = qe^{-\gamma}. \tag{11.7.7}
\]

The expected number of particles in a member of a grand canonical ensemble is:

\[
\langle N \rangle = -\left(\frac{\partial \ln \Xi}{\partial \gamma}\right), \tag{11.7.8}
\]

so we get in this case:

\[
\langle N \rangle = qe^{-\gamma}. \tag{11.7.9}
\]

So, comparing to Equation (11.7.7) we have

\[
\langle N \rangle = \ln \Xi(\beta, V, \gamma). \tag{11.7.10}
\]

But ln \(\Xi = \beta pV\), so we arrive at

\[
\langle N \rangle = \beta pV, \tag{11.7.11}
\]

which is:

\[
pV = \langle N \rangle kT, \tag{11.7.12}
\]

the ideal gas equation!

Now this isn’t just a cute result. It is in fact profound. What we’ve shown⁴ is that if a system is made up of independent particles and if the system obeys Boltzmann statistics, the particles obey the ideal gas law. Nothing is said about the state or condition of the particles, their energy levels, or whatever. We come away with the conclusion that it is independence that makes a system ideal and nothing else. This is a good thing to remember.

---

⁴After what must be the longest derivation of \(pV = NkT\) in history.
Chapter 12

The Equivalence of Ensembles

Now that we have considered three different partition functions it is useful to stop and take a more general look.

If we have a system at equilibrium, then all of its thermodynamic variables have fixed values. We know this experimentally. It should not matter what partition function we use to compute those values, we must always get the same answers.

So thermodynamically if we take a one-component single phase system and consider it to have fixed independent variables \( N, V, \) and \( T, \) it also has fixed dependent variables \( \mu, p \) and \( E. \) The values we calculate statistically for these three should agree with the measured values.

We could also consider that same system as having a fixed \( \mu, V \) and \( T, \) in which case the values of \( N, p, \) and \( E \) we calculate should agree with the experimental values.

As a practical matter then, it should not matter which ensemble we use for computations. And indeed this is true. One gets the same results no matter what ensemble is used. In practice we to use the ensemble that makes the computations the easiest.

However, there is another implication here. That is that all ensembles should be equivalent in some mathematical sense. Yet they certainly look different.

In this chapter we will investigate the mathematical relationship among ensembles, old and new.
12.1 Expansion of the Grand Canonical Partition Function

We’ve already seen that the grand canonical partition function can be written as:

$$\Xi(\gamma, V, \beta) = \sum_N Q(N, V, \beta) e^{-\gamma N}, \quad (12.1.1)$$

so that the grand canonical partition function can be considered as being built up of canonical partition functions for systems of every possible particle number.

To make this explicit $\Xi$ can be written:

$$\Xi(\gamma, V, \beta) = Q(0, V, \beta) + Q(1, V, \beta)e^{-\gamma} + Q(2, V, \beta)e^{-2\gamma} + ..., \quad (12.1.2)$$

which points up the close relationship between the two ensembles. But in fact we know that there is a single most probable value of $N$, so only one of the $Q$’s is really important. The others just don’t matter! How can this be? To understand this we have to make several detours.

12.2 Generalized Laplace Transform

A type of transformation that is often very useful in statistical thermodynamics is the generalized Laplace transform.

An “ordinary” Laplace transform is an integral transform in which one function is transformed into another (with different independent variables) by integration. Thus if $f(x)$ is a function, then

$$F(s) = \int_0^\infty f(x)e^{-sx}dx, \quad (12.2.1)$$

is its Laplace transform. Note that the old independent $x$ variable has been swapped for the new independent variable $s$. The function $F(s)$ can be transformed back into $f(x)$ by means of the inverse Laplace transform

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} F(s) e^{sx}ds, \quad (12.2.2)$$

where the integration is carried out in the right-hand half of the complex plane.¹

As a trivial example, if $f(x) = 1$, then its Laplace transform $F(s)$ is:

$$F(s) = \int_0^\infty e^{-sx}dx = \frac{1}{s}. \quad (12.2.3)$$

A generalized Laplace transform is a Laplace transform in which one integrates over $x$ if $x$ is a continuous variable and sums over $x$ if it is a discrete variable.

¹Students reading this should not worry overmuch about it. I include the inverse transformation only for the sake of being complete. We are not going into the general theory of Laplace transforms.
The case in which \( x \) is discrete is of great interest to us. Consider \( f(n) = n \) where \( n \) is an integer. Then the generalized transform of \( f(n) \) is

\[
F(s) = \sum_{n=0}^{\infty} ne^{-sn}.
\] (12.2.4)

This sum can be found in closed form.

The first step in solving Equation (12.2.4) is to recognize that \( e^{-s} \) is just a constant as far as the summation is concerned. Let us call that constant \( a \). Then Equation (12.2.4) becomes:

\[
F(a) = \sum_{n=0}^{\infty} na^n.
\] (12.2.5)

The second step is to realize that the related series

\[
G(a) = \sum_{n=0}^{\infty} a^n,
\] (12.2.6)

is a geometric series and converges to a known sum as long as \( a \) satisfies \(-1 \leq a < 1\). Here \( a \) can’t be less than 0 as long as \( s \) is positive. And since \( s \) is arbitrary, all we need do is ensure that we never use a negative \( s \).

The sum in Equation (12.2.6) is:

\[
G(a) = \sum_{n=0}^{\infty} a^n = \frac{1}{1-a} \quad -1 < a < 1.
\] (12.2.7)

The two series, Equations (12.2.5) and (12.2.7) are related by differentiation. If we differentiate Equation (12.2.7) with respect to \( a \) we get

\[
\frac{dG(a)}{da} = \sum_{n=0}^{\infty} na^{n-1},
\] (12.2.8)

which, when multiplied by \( a \) gives

\[
a \frac{dG(a)}{da} = \sum_{n=0}^{\infty} na^n = F(a).
\] (12.2.9)

Since \( G(a) = 1/(1-a) \), then it is simple to determine that \( F(a) \) is

\[
F(a) = \frac{a}{(1-a)^2},
\] (12.2.10)

and so

\[
F(s) = \frac{e^{-s}}{(1-e^{-s})} = \frac{1}{(e^s - 1)}.
\] (12.2.11)

An example of a generalized Laplace transform as part of statistical thermodynamics is:

\[
\Xi(\gamma, V, \beta) = \sum_{N} Q(N, V, \beta)e^{-\gamma N}
\] (12.2.12)

where \(-\gamma\) plays the role of \( s \) and \( N \) plays the role of \( x \).
12.3 Transformations Among Ensembles

We start with an example. We will transform the canonical partition function $Q$ to the grand canonical partition function $\Xi$.

An appropriate starting point is:

$$\beta A(N,V,\beta) = -\ln Q(N,V,\beta), \quad (12.3.1)$$

or, what is the same thing:

$$e^{-\beta A(N,V,\beta)} = Q(N,V,\beta). \quad (12.3.2)$$

Let’s now do a generalized Laplace transform on this, replacing the variable $N$ with the variable $\mu$. To do this we multiply both sides of Equation (12.3.2) by $e^{-\gamma N}$ and sum over all $N$:

$$\sum_N e^{-\beta A(N,V,\beta)} e^{-\gamma N} = \sum_N Q(N,V,\beta) e^{-\gamma N}. \quad (12.3.3)$$

The right-hand side is the grand canonical partition function $\Xi(\gamma,V,\beta)$.

$$\sum_N e^{-\beta A(N,V,\beta) - \gamma N} = \Xi(\gamma,V,\beta). \quad (12.3.4)$$

On the left-hand side we recognize that the exponent is $\beta p V$ (which is a function of $N$) so the result of the Laplace transform is:

$$\sum_N e^{\beta p V} = \Xi(\gamma,V,\beta). \quad (12.3.5)$$

If the left-hand side were simply $e^{\beta p V}$ without the summation we’d have the result we expected, for then we’d have

$$\beta p V = \ln \Xi.$$  

But we have a summation sign. That’s not quite what we expected. How can:

$$\sum_N e^{\beta p V} = e^{\beta p V},$$

since when written down it even looks silly.

12.3.1 The Maximum Term Method

It turns out that the summation sign that was a problem in the previous section can be dropped under special circumstances. Or, put another way, sometimes a single

---

2This may not be obvious. In that case refer to Section 1.7 on page 1-7 for details. Also, $\beta p V$ is a function of $N$ because $p$ is a function of $N$ since $p$ depends on $\partial E_j/\partial V$ and $E_j$ is a function of $N$. See Subsection 6.3.2 on page 6-7 for details.
term in a summation is as large as the entire sum! Seeing this requires another 
digression.\footnote{Yes, the author knows that these frequent digressions interrupt the
flow of the text. But he knows that the digressions involve new material. And he
chooses not to simply lump these mathematical details together in separate chapter that
nobody would read (at least without falling asleep). What he’s doing is called
\textit{motivation}...}

We are dealing with very large numbers here. And very large numbers sometimes
have interesting properties.

Let’s first see how large $e^{\beta pV}$ really is. Take a typical system with a pressure $p$ of
101325 pascals,\footnote{One atmosphere to you chemists.} a volume $V$ of 0.020 cubic meters,\footnote{That’s 20 liters.} and a temperature $T$ of 300 K. The Boltzmann factor $k$ is $1.38 \times 10^{-23}$ in these units. Thus $\beta pV$ is about $4.9 \times 10^{23}$.

What we are dealing with is \textit{not} $4.9 \times 10^{23}$, it is

$$e^{4.9 \times 10^{23}}$$

which is a number that is a power of ten with about $2 \times 10^{23}$ digits in its exponent! And that’s the value of just one term in the sum in Equation (12.3.4) on page 12-4.

Now we need to digress a bit further and consider the following theorem:

Let

$$S = \sum_{N=1}^{M} T_N$$

(12.3.6)

where we assume that the terms $T_N > 0$ for all $N$. Note that our sum of $e^{\beta pV}$ fits this condition.

All the terms are positive, Hence $S$ is surely greater than or equal to the largest
term $T_{\text{max}}$ in the sum. And it is surely smaller than or equal to $M$ times $T_{\text{max}}$. So

$$T_{\text{max}} \leq S \leq MT_{\text{max}}$$

(12.3.7)

We now take logs to get

$$\ln T_{\text{max}} \leq \ln S \leq \ln T_{\text{max}} + \ln M$$

(12.3.8)

Now assume that $T_{\text{max}}$ is roughly as large as $e^M$ (or even larger), then

$$T_{\text{max}} \approx e^M$$

(12.3.9)

and since clearly for large $M$

$$M \gg \ln M$$

(12.3.10)

so

$$\ln T_{\text{max}} \approx M \gg \ln M$$

(12.3.11)

and the factor of $\ln M$ on the right-hand side of Equation (12.3.8) is then negligible
compared to $\ln T_{\text{max}}$. Compare, for example $10^{23}$ and $\ln 10^{23} = 52.96$. Surely 52.96 is
enormously smaller than $10^{23}$. 
Equation (12.3.8) on page 12-5 now becomes the remarkable

$$\ln T_{\text{max}} \leq \ln S \leq \ln T_{\text{max}} \quad (12.3.12)$$

so that we inescapably have the conclusion that to a fantastically good approxima-

$$\ln S = \ln T_{\text{max}} \quad (12.3.13)$$

and so

$$S = T_{\text{max}} \quad (12.3.14)$$

This is a rather astounding result. It does not hold in general. It only holds for
series of all positive terms where the the largest term is huge. But that is exactly
the case in Equation (12.3.5) on page 12-4! Thus we can simply replace the sum on
the left-hand side of that equation with its largest term to get:

$$e^{\beta p^* V} = \Xi(\gamma, V, \beta) \quad (12.3.15)$$

where $p^*$ is the value of $p$ that causes $e^{\beta p V}$ to be a maximum. So now we have:

$$\beta p^* V = \ln \Xi(\gamma, V, \beta) \quad (12.3.16)$$

But what value does $p^*$ have? First, on a microscopic level $p^*$ must be defined here
by

$$\beta p^* = \left( \frac{\partial \ln Q}{\partial V} \right)_{N,\beta} \quad (12.3.17)$$

Now that pressure would have the numerical value appropriate for a system with
the given values of $N$, $V$, and $\beta$. That is, it would be the \textit{equilibrium} value of $p$.
Which is exactly what we want.

So we come out of this with a simple rule: When the terms of a sum (such as in
a partition function) are all positive and at least some of the terms are huge in
numerical value, we can replace the sum with the largest term if we also replace the
dependent variable by its equilibrium value as given by the independent variables.

### 12.3.2 A New Ensemble

Just as an example let us find the partition function of an ensemble whose indepen-
dent variables are $E$, $V$, and $\mu$. The partition function for this ensemble is denoted
$\Gamma(E, V, \mu)$ as in Equation (1.7.22) on page 1-9.

We can start with the microcanonical ensemble $\Omega(E, V, N)$:

$$\beta T S(E, V, N) = \ln \Omega(E, V, N) \; , \quad (12.3.18)$$

and then write

$$e^{\beta TS} = \Omega(E, V, N) \; . \quad (12.3.19)$$

Now we multiply both sides by $e^{-\gamma N}$:

$$e^{\beta TS - \gamma N} = \Omega(E, V, N)e^{-\gamma N} \; , \quad (12.3.20)$$
and sum over all \( N \)
\[
\sum_N e^{\beta TS - \gamma N} = \sum_N \Omega(E, V, N) e^{-\gamma N}.
\tag{12.3.21}
\]

The right-hand side will be our partition function \( \Gamma \)
\[
\sum_N e^{\beta TS - \gamma N} = \Gamma(E, V, \mu),
\tag{12.3.22}
\]
where the \( N \)-dependence on the right-hand side has been summed away and of course \( \gamma \) is a direct function of \( \mu \).

Now we invoke the laws of large numbers and claim that the sum on the left is equal to its largest term. This gives:
\[
e^{\beta TS - \gamma N} = \Gamma(E, V, \mu).
\tag{12.3.23}
\]
which can be seen to be exactly Equation (1.7.12) on page 1-9.Unfortunately, this ensemble has no name.

### 12.3.3 A Cautionary Example: The Isobaric-Isothermal Ensemble

As we’ve seen, in general it is not necessary to go through the entire Lagrangian multiplier derivation in order to find new ensembles. It is quicker simply to apply the transforms we’ve discussed above.

But this does not always work smoothly. To illustrate this let’s find a partition function whose independent variables are \( N, p, \) and \( T \). Actually, we need the variables \( N, \beta p, \) and \( \beta \), because we need dimensionless variables.\(^6\)

However, There is a difficulty. As was first pointed out by W. Byers Brown,\(^7\) while temperature and chemical potential are true ensemble variables having no existence for individual molecules, pressure is in fact a mechanical variable and can be defined quite well both classically and quantum mechanically for single systems.\(^8\)

All this is in a way quite besides the point, though it is a warning flag. The problem in the Isobaric-Isothermal ensemble lies in the volume. Volume is a continuous variable. This leads to some serious problems with Gibbsian ensembles\(^9\) even if we artificially force the volume to be “quantized”.

Nevertheless, it is instructive to try to carry out the procedure developed in this chapter and to try to generate the appropriate partition function for this ensemble.

---

\(^6\)The idea of using ensembles having constant pressure rather than constant volume was first advanced by E. A. Guggenheim, *J. Chem. Phys.* 7, 103 (1939).


\(^8\)Note that \( p_j = (\partial E_j/\partial V) \), so that quantum mechanically even a single particle in a box of volume \( V \) can have a pressure.

\(^9\)The problem is more tractable when using the Boltzmann approach as will be seen later when we discuss classical statistical thermodynamics.
We can start our example with any partition function we know, but it is best to start with one that is close to what we want. Here we start with the canonical partition function:

\[-\beta A(N, V, \beta) = \ln Q(N, V, \beta) \quad \text{or} \quad e^{-\beta A} = Q(N, V, \beta) \tag{12.3.24}\]

and now transform it to our desired variables.

To replace \( V \) with \( \beta p \), we multiply both sides by \( e^{-\beta p V} \) and apply a generalized Laplace transform. Since \( V \) is a continuous variable, we will integrate instead of sum:

\[\int e^{-\beta A - \beta p V} dV = \int Q(N, V, \beta) e^{-\beta p V} dV \tag{12.3.25}\]

But look at what has happened! Equation (12.3.25) now has the units of volume! Partition functions need to be dimensionless if for no other reason than we know that we will be taking logarithms of them.

We replace the integral by its largest term:\(^{10}\)

\[e^{-(\beta A + \beta p V)} = \Delta = \int Q(N, V, \beta) e^{-\beta p V} dV \tag{12.3.26}\]

where, as usual, we call this partition function \( \Delta \) because that’s what it is called in the literature.

And now we are in terrible trouble. Multiplying by \( dV \) and integrating has left the right-hand-side of Equation (12.3.26) with units. And partition functions are dimensionless.\(^{11}\)

Hill\(^{12}\) has argued that this can be fixed by multiplying \textit{not} by \( dV \) but by \( d(V/V_o) \), where \( V_o \) is a suitably small reference volume. In fact he suggests using

\[V_o = \Lambda^3 \quad \text{where} \quad \Lambda = \frac{\hbar}{(2\pi mkT)^{1/2}} \tag{12.3.27}\]

If we do this then Equation (12.3.26) becomes instead:

\[e^{-(\beta A + \beta p V)} = \Delta = \frac{1}{V_o} \int Q(N, V, \beta) e^{-\beta p V} dV \tag{12.3.28}\]

Now taking logs we get:

\[-(\beta A + \beta p V) = \ln \Delta = \ln \int Q(N, V, \beta) e^{-\beta p V} dV - \ln V_o \tag{12.3.29}\]

where we assume that we’ve left the units of \( V_o \) inside the integral (making it dimensionless) while now having \( V_o \) represent only a numerical value.

\(^{10}\)Wait a minute! How can we do that? What we just proved was for sums, not integrals! That’s true. But if we consider an integral as a sum of thin strips of integrand, then our result applies for such integrals as well.

\(^{11}\)Note that taking the largest term in the integral on the right-hand-side of Equation (12.3.26) has removed the problem there.

We now dispose of $V_o$ by noting that it is of the order of $N$ times smaller than the integral.

This actually works. We’ll postpone a discussion of why for a few moments.

It remains to see what thermodynamic function Equation (12.3.29) on page 12-8 really is (if any).

We have:

$$-\beta(A + pV) = -\beta(E - TS + pV)$$
$$= -\beta(TS - pV + \mu N - TS + pV)$$
$$= -\beta \mu N$$
$$= -\beta G$$

where $G$ is the Gibbs free energy. The result is:

$$-\beta G = \ln \Delta = \ln \int Q(N, V, \beta) e^{-\beta pV} dV$$

Note that in Equation (12.3.31) we are treating $dV$ as dimensionless.

Why does this work? Because what we’ve done is essentially make the volume discrete and then to sum over the discrete volumes rather than integrate. We’ve introduced a basic “quantum” of volume, $\Lambda^3$, and measured volumes in terms of it.

This is not a satisfactory solution and the reader should not think that it is.

What is interesting is that the author can find no proper treatment of the Isobaric-Isothermal system in any detail\(^{13}\) in the available literature today. This is, among other things, rather unconscionable. It is his personal opinion that the proper way to deal with any ensemble involving integrating over the volume is to assume that volume is quantized. That is, that there exists a smallest volume element $v_o$ and that all other volumes are simply multiples of this. Then instead of integrating over the volume, we sum and have no problems at all\(^{14}\).

### 12.4 Summary: The Relationship Among Ensembles

One way to think about this is to realize that we can start with any ensemble whose thermodynamic connection we know and transform it to any other set of thermodynamic variables. We can find the partition function for a two-component system with variables $p$, $T$, $N_1$ and $\mu_2$, where $\mu_2$ is the chemical potential associated with component 2.

\(^{13}\) Many authors mention it and give (without derivation) Equation (12.3.31) as the definition of $\Delta$ without realizing that Equation (12.3.31) is incomplete as it stands.

\(^{14}\) This is, in effect, what happens when we quantize energy and sum over energies instead of integrating. If we did integrate, we’d have a $dE$ in our integrals and the resulting partition function would have the units of energy.
Or we can show that the partition function for the ensemble with the independent variables \( p, T, \) and \( \mu \) is zero.

So we are done with undetermined multipliers and the like. If we need an ensemble, we’ll just transform one of our regular ones to fit the job.

There is a good thermodynamic reason as to why this works. If one has a (say) single component single phase system one wants to describe, three variables will do it. Which three are chosen does not matter.\(^{15}\) One person can choose \( N, V, \) and \( T. \) Another can chose \( N, p, \) and \( T. \) If the pressure experimentally found by the first person is the pressure used by the second, the two systems will be identical in all thermodynamic respects.

And so mathematically. A system itself has no particular preference for any set of independent variables. Those are a choice imposed by the human studying the system on the system. Two different humans can choose two different sets of independent variables. And we’d be shocked if the results they calculated for other variables (say a heat capacity, for instance) were different.

\(^{15}\)Unless all three are intensive in which case the volume of the system is not defined and we are in some trouble. Indeed, the partition function for such a system is zero.
Chapter 13

The Ideal Crystal

13.1 Introduction

The ideal crystal is an example of a system that is not made up of independent subsystems but which can be converted to independent subsystems if we take a different view of the system.

The mechanics of doing this are a bit difficult and so we will approach this topic in a somewhat unconventional way. ¹

The first thing to say is that the ideal crystal is much more of a hypothetical system than an ideal gas. But we shall ignore this and pretend that real crystals are ideal enough for this model to work. ² We will assume that we have a complete crystal with no imperfections and no missing atoms or molecules. And we will assume that all the atoms or molecules are identical. We will also assume that any motion of the atoms or molecules is insufficient to cause them to leave their positions within the crystal and “wander around” inside of it. And we will assume that the forces between particles in the crystal are harmonic. Last, we shall ignore any internal structure of any of the constituents.

The last assumption isn’t necessary, but it does allow us to focus on the main problem, that of vibrations in a crystal. If one wants also to consider the internal structure, one simply puts it in with the assumption that the internal structure (internal rotations and vibrations) do not influence the vibrations of the particles making up the crystal itself.

At any finite temperature (and, considering zero point energy, even at absolute zero) all the particles of the system are vibrating around their equilibrium position

¹But one, I hope, that will provide motivation for the difficulties that will come.
²Which seems to be true, at least in a rough way.
in the crystal. And we have assumed that the magnitude of the motion is such that
the particle never is very far from its equilibrium position.

The forces felt by any one of these particles depends not only on the position of the
particle itself, but on the position of its nearest-neighbor particles and, perhaps,
even on particles further away. It is not possible to consider any one of the basic
atoms or molecules of the crystal as independent of any of the others.

Thus the crystal has to be considered as a whole. And if we do this and treat the
system classically, we will end up with simultaneous equations of motion involving
all the particles of the system. This huge number of equations will have to be solved
in order to know the properties of the system.

This is an example of a strongly interacting system. And in general such systems
cannot be treated exactly.

However, in this case we can simplify things greatly by focusing not on the particles,
but on their vibrations.

It seems clear that the vibrations are concerted. That is, the motion of one particle
is communicated to its neighbors and cause them to move in reaction. And that
motion is communicated to their neighbors and so on until the entire crystal is
involved.

If the forces acting on the particles is harmonic as we have assumed, then it is always
possible to mathematically decompose the motions of the particles into a number
of independent vibrations existing within the crystal called normal vibrational
modes. Each particle is affected by all the normal mode vibrations, but the normal
mode vibrations are independent of each other.

What that means is that if the amplitude of one normal mode is changed (or perhaps
the mode is eliminated in some way), the remainder of the normal vibrational modes
are not affected in any way.

Thus we change our viewpoint and focus on the normal modes as independent
“subsystems” and pay no attention to the actual particles making up the crystal
and without which no vibration could take place.

This change in viewpoint leads to some very simple and very useful models of
crystals.

13.2 The Models

So we shall focus on the vibrations themselves and not on the particles making
up the crystal. The first question that then comes up is: how many normal mode
vibrations are there?

As we’ve discussed before (see Chapter 10), classical mechanics tells us that there
are three degrees of freedom for every particle. That means that each particle is
free to move in three independent directions. Thus a diatomic molecule has a total
of six degrees of freedom. Three are manifested as motions of the center of mass in the $x$-, $y$-, and $z$-directions. Two of the remaining three are rotations about the center of mass and the last is a vibration about the center of mass.

If we have $N$ particles in our crystal, we then expect to have $3N$ degrees of freedom. Three of these are again the motion of the center of mass of the crystal and three more are rotations of the crystal about the center of mass. The remaining $3N - 6$ are vibrations.

Since $N$ is often of the order of Avogadro’s number, there is no real difference between $3N$ and $3N - 6$ and we shall often simply speak of the $3N$ normal vibrational modes.\footnote{As shall be seen later on when we discuss how we get from particles to vibrations, it will be seen that $3N$ is actually correct!}

To allow for the possibility that not all of these normal vibrational modes have the same frequency, we shall define $g(\nu)$ to be the degeneracy factor for the vibrations. Thus $g(\nu)d\nu$ will be the number of normal mode vibrations that occur between frequency $\nu$ and frequency $\nu + d\nu$.

There are some restrictions on $g(\nu)$, the most important of which is that there must be finite number of them, $3N$, in particular. So we can write:

$$\int_0^\infty g(\nu)d\nu = 3N \quad (13.2.1)$$

We must also understand that $g(\nu)$ is not really a continuous function. It is in fact highly discontinuous since there are only a finite number of vibrations between 0 and infinity. So we might better have written Equation (13.2.1) as a sum, but as long as we understand what that equation means, there should be no confusion.

The rest is (almost) easy. We already know the single “particle” partition function for a vibration:

$$q(V, T) = \frac{e^{\Theta/2T}}{1 - e^{-\Theta/T}} \quad (13.2.2)$$

where $\Theta$ is given by:

$$\Theta = \frac{h\nu}{k} \quad (13.2.3)$$

with $h$ Planck’s constant, $k$ Boltzmann’s constant and $\nu$ the fundamental frequency of the harmonic vibrator.

The canonical partition function is

$$Q(N, V, T) = e^{-NU_o/2kT} \prod_\nu q_\nu(V, T) \quad (13.2.4)$$

where the term $e^{-NU_o/2kT}$ accounts for the binding energy of the crystal. The binding energy of any particular particle\footnote{We must not forget that underneath the vibrations we are still dealing with particles.} is due to its interactions with all the other particles. Counting these up leads to each particle being counted twice, hence the factor of two in the denominator of Equation (13.2.4).
From Equation (13.2.4) on page 13-3 we can get all of the thermodynamic properties. In fact, for historical reasons, we are most interested in the heat capacity.

A long while back Dulong and Petit noticed that the heat capacities of elements made up of single atoms had constant pressure molar heat capacities of almost exactly $3R$, where $R$ is the gas constant. This observation is now known as the Rule of Dulong and Petit.

However, when it became possible to make low temperature measurements, it became clear that this law was disobeyed in that region. Heat capacities fell off toward zero, becoming proportional to $T^3$ at very low temperatures.

Classical statistical mechanics was unable to account for this. In fact the problem is similar to that of black body radiation. And indeed, the same solution works for both. If one assumes that the vibrations are quantized, the fall off to zero in the heat capacity becomes simple to demonstrate.

There is one catch. We can’t do anything unless we know $g(\nu)$. It is in the value of $g(\nu)$ that most of the simple theories vary.

### 13.3 The Einstein Theory

Einstein, with characteristic ability to cut to the heart of a problem, assumed that all of the vibrational frequencies were identical. That is, he took $g(\nu)$ to be a constant. He knew that wasn’t really correct, but he knew that often major results depend on the basic facts of a model and not on the details. The basic fact he wanted to emphasize was quantization of the vibrational energy.

With that, things are fairly simple. We have only one $\nu$ and Equation (13.2.4) on page 13-3 is then:

$$Q(N, V, T) = e^{-NU_0/2kT} q(V, T)^{3N}$$

where the factor $3N$ is because that’s the number of independent vibrations we have. With $q(V, T)$ given by Equation (13.2.2) on page 13-3 we get

$$Q(N, V, T) = e^{-NU_0/2kT} \left[ \frac{e^{-\Theta/2T}}{1 - e^{-\Theta/T}} \right]^{3N}$$

And now the thermodynamics of the Einstein model are very simple.

$$\langle A \rangle = -kT \ln Q = \frac{NU_0}{2} - 3NkT \ln \left[ \frac{e^{-\Theta/2T}}{1 - e^{-\Theta/T}} \right]$$

$$\langle E \rangle = -\left( \frac{\partial \ln Q}{\partial \beta} \right) = \frac{NU_0}{2} + 3Nk\nu + 3NkT \left[ \frac{\Theta/T}{e^{\Theta/T} - 1} \right]$$

5Which leaves out elements such as sulfur but includes most metals.

6And was demonstrated by A. Einstein, Ann. Physik, 22: 180 (1907)

7Actually we have $3N - 6$ if we take three translational and three rotational degrees of freedom for the crystal into account. However, with $N$ of the order of Avogadro’s number, six is quite negligible.
Only the last term in the expression for the energy depends on temperature, so the heat capacity is then:

\[ C_V = \left( \frac{\partial E}{\partial T} \right) = 3Nk \left( \frac{\Theta}{T} \right)^2 \frac{e^{\Theta/T}}{(e^{\Theta/T} - 1)^2} \]  

(13.3.5)

Equation (13.3.5) does contain the law of Dulong and Petit. One has only to take the limit as \( T \to \infty \). In that limit \( \Theta/T \) gets very small so that the exponentials can be expanded in a power series:

\[ e^{\Theta/T} = 1 + \frac{\Theta}{T} + \cdots \]

So the numerator in Equation (13.3.5) can be replaced by 1, but since one is subtracted from the exponential in the denominator, the denominator must be replaced by \( \Theta/T \). The result is then immediate:

\[ C_V \to 3Nk \quad \text{as} \quad T \to \infty \]  

(13.3.6)

This is a Good Thing; had not the result agreed with the known experimental results we’d have heard no more of the theory.

Encouraged, let us look at the low temperature behavior of the heat capacity. Here \( \Theta/T \) becomes very large, the one in the denominator can be neglected as small and we end up with

\[ C_V \to 3Nk \left( \frac{\Theta}{T} \right)^2 e^{-\Theta/T} \quad \text{as} \quad T \to 0 \]  

(13.3.7)

This clearly goes to zero as the temperature goes to zero. So we have done the primary job, the overall behavior of the heat capacity is found and we can be happy.

There is another test. Does the entropy obey the Third Law? It should go to zero as the temperature goes to zero.

Figure 13.1: Graph of \( C_V \) vs \( T/\Theta \) for an Einstein Crystal
The entropy is easy to find:

\[ S = \frac{E}{T} - \frac{A}{T} = 3Nk \left[ \frac{\Theta / T}{e^{\Theta / T} - 1} - \ln(1 - e^{-\Theta / T}) \right] \] (13.3.8)

which indeed\(^8\) does go to zero as \(T\) goes to zero.

Much happiness reigns. And Einstein became even more famous than he already was.\(^9\)

But wait. There is a dark cloud.

Experimentally the heat capacity goes to zero as \(T^3\), and not as given in Equation (13.3.8), which has an exponential fall-off to zero.

Of course, this is not surprising. The Einstein theory is crude and certainly not all the vibrational modes have the same frequency. And so we look for a better theory.\(^10\)

### 13.4 The Debye Theory

#### 13.4.1 The Frequency Distribution

We find a better theory in the Debye model.\(^11\) Debye approached the problem from the standpoint of the continuum mechanics of solids, a well-studied field when he was a student. He knew that for a continuous medium \(g(\nu)\) was proportional to \(\nu^2\) for low frequencies. And those are precisely the frequencies that govern the low temperature behavior because they contain the least energy.

And he realized that the classical crystal becomes “continuous” in the limit of very long wavelength vibrations.

A short wavelength vibration, say one with a wavelength of three or four interparticle distances, can’t really be treated as a wave in a continuous medium. At some point in time it might have one particle at a node, another at the wave maximum, a third at a node again, and the fourth at the wave minimum with nothing in between.

Yet a long wavelength vibration might span \(10^{19}\) particles, which would be a very different story. Particles would be almost continuously placed along the wave.

Of course the long wavelength waves have the lowest energy, and those would be the important ones at low temperatures.

With these insights Debye essentially set \(g(\nu)\) equal to \(A\nu^2\) where \(A\) is a constant.

---

8. Though it is left as an exercise for the reader.

9. The special theory of relativity was published in 1905. His heat capacity paper was published two years later.

10. Einstein was well aware of this deficiency. He wrote in the early days of quantum theory (1907 in this case) and was interested in showing that quantization would solve the heat capacity problem. To do this he used the crudest model that contained the basic features needed.

We can determine \( A \) from Equation (13.2.1):

\[
\int_0^{\nu_{\text{max}}} g(\nu) d\nu = A \int_0^{\nu_{\text{max}}} \nu^2 d\nu = 3N
\]  

(13.4.1)

where \( \nu_{\text{max}} \) is used instead of infinity since with a finite number of normal modes there must be an upper limit to the vibrational frequencies allowed.

Equation (13.4.1) gives us

\[
\frac{A}{3} \nu_{\text{max}}^3 = 3N
\]

from which

\[
A = \frac{9N}{\nu_{\text{max}}^3}
\]  

(13.4.2)

so that we end up with

\[
g(\nu) d\nu = \begin{cases} 
9N(\nu^2/\nu_{\text{max}}^3) d\nu & 0 \leq \nu \leq \nu_{\text{max}} \\
0 & \text{otherwise}
\end{cases}
\]

(13.4.3)

### 13.4.2 The Debye Partition Function

Since the particles supporting the vibrations are in principle identifiable, we know that the partition function \( Q(N,V,T) \) is given by:

\[
Q(N,V,T) = q_1 q_2 q_3 \cdots = \prod_{i=1}^{3N} q_i
\]

and, since \( q_i \) is associated with the \( i \)'th vibration, we also know that:

\[
q_i = \frac{e^{-h\nu_i/2kT}}{1 - e^{-h\nu_i/kT}}
\]

where \( \nu_i \) is the frequency of the \( \nu^{'th} \) vibration. Then

\[
Q(N,V,T) = e^{-\beta NU_0/2} \prod_i \frac{e^{-h\nu_i/2kT}}{1 - e^{-h\nu_i/kT}}
\]  

(13.4.4)

Taking logarithms:

\[
\ln Q = -\frac{\beta NU_0}{2} - \int_0^\infty \left[ \frac{h\nu}{2kT} + \ln(1 - e^{-h\nu/kT}) \right] g(\nu) d\nu
\]

(13.4.5)

where we need \( g(\nu) \) because we are now assuming that the frequencies are continuous.

The energy is then found from:

\[
\left( \frac{\partial \ln Q}{\partial \beta} \right) = -\langle E \rangle = -\frac{NU_0}{2} - \frac{9N}{\nu_{\text{max}}^3} \int_0^{\nu_{\text{max}}} \left( \frac{h\nu}{2} + \frac{h\nu}{e^{\beta h\nu} - 1} \right) \nu^2 d\nu
\]

(13.4.6)
\[ \langle E \rangle = \frac{NU_0}{2} + \frac{9NkT}{\nu_{\text{max}}^3} \int_0^{\nu_{\text{max}}} \left( \frac{\nu}{2kT} + \frac{\nu/kT}{e^{\nu/kT} - 1} \right) \nu^2 d\nu \]  

(13.4.7)

Letting \( x = \nu/kT \) and \( u = \nu_{\text{max}}/kT \)

\[ \langle E \rangle = \frac{NU_0}{2} + \frac{9Nk}{u^3} \int_0^u \left( \frac{x}{2} + \frac{x}{e^x - 1} \right) x^2 dx \]  

(13.4.8)

The first integral in Equation (13.4.8) is easy but the second is not. Because of this the energy is usually written as:

\[ \langle E \rangle = \frac{NU_0}{2} + \frac{9Nh\nu_{\text{max}}}{8} + 3NkT \cdot D(u) \]  

(13.4.9)

where \( D(u) \) is called the Debye function.\(^{12}\) The Debye function is defined as:

\[ D(u) = \frac{3}{u^3} \int_0^u \frac{x^3}{e^x - 1} dx \]  

(13.4.10)

and is tabulated in many places.\(^{13}\)

Some properties of the Debye function are easy to establish.\(^{14}\) Although we will not show it, it is not hard to prove that

\[ D(u) \to \frac{3}{u^3} \int_0^u \frac{x^3}{e^x - 1} \to \frac{\pi^4}{5u^3} \text{ as } T \to 0 \text{ and } u \to \infty \]  

(13.4.11)

and

\[ D(u) = \frac{3}{u^3} \int_0^u \frac{x^3}{(1 + x + \ldots) - 1} dx = 1 \text{ as } T \to \infty \]  

(13.4.12)

Thus as \( T \to 0 \)

\[ \langle E \rangle \to \frac{NU_0}{2} + \frac{9Nh\nu_{\text{max}}}{8} + \frac{3N\pi^4h\nu_{\text{max}}}{5} \left( \frac{kT}{h\nu_{\text{max}}} \right)^4 \text{ as } T \to 0 \]  

(13.4.13)

from which it is easy to see that the heat capacity is going to go as \( T^3 \), as expected. At high temperatures we have:

\[ \langle E \rangle \to \frac{NU_0}{2} + 3NkT \text{ as } T \to \infty \]  

(13.4.14)

which will give us, as it should, the law of Dulong and Petit once again.

The heat capacity can be computed more completely from the Debye function. We need to differentiate Equation (13.4.7) with respect to \( T \). Since the first two terms of that equation are not functions of \( T \) we are left with:

\[ C_V = \left( \frac{\partial E}{\partial T} \right)_{V,N} = \frac{\partial}{\partial T} \left[ 3NkT \cdot D(u) \right] = 3NkD(u) + 3NkT \left( \frac{\partial D(u)}{\partial T} \right) \]  

(13.4.15)

\(^{12}\)It isn’t too shabby to have a function named after you...

\(^{13}\)One needs to be careful with such tables. There are several things called “Debye” functions, all rather simply related to each other and to Equation (13.4.10). Check the definitions of the function in any table that you might use.

\(^{14}\)It needs to be said that the Debye function is as much a function as the sine or cosine or exponential, for that matter. Those have to be tabulated too.
Using the method explained in Appendix 13.7 on page 13-20 we get:

\[
\left( \frac{\partial D(u)}{\partial T} \right) = \frac{3}{T} \left[ D(u) - \frac{u}{e^u - 1} \right]
\]  \hspace{1cm} (13.4.16)

so that the result is:

\[
C_V = 3Nk \left[ 4D(u) - \frac{3u}{e^u - 1} \right]
\]  \hspace{1cm} (13.4.17)

Following convention we define a Debye Theta Temperature $\Theta_D$ by

\[
u = \frac{h\nu_{\text{max}}}{kT} = \frac{\Theta_D}{T} \quad \text{so that} \quad \Theta_D = \frac{h\nu_{\text{max}}}{k}
\]  \hspace{1cm} (13.4.18)

![The Debye Function](image)

Figure 13.2: Graph of the Debye Function vs $T/\Theta_D$

In these terms we have

\[
C_V \rightarrow \frac{12Nk\pi^4}{5} \left( \frac{T}{\Theta_D} \right)^3 \quad \text{as} \quad T \rightarrow 0
\]  \hspace{1cm} (13.4.19)

This reproduces the experimental results quite well, so well that in fact it is used to compute Third Law entropies. It is not easy to get heat capacities below about 10K, so all one needs to do is to assume that $C_V$ goes as $CT^3$ between 0 and 10K. The constant $C$ can be found from the measured heat capacity at 10K.

One final note, the entropy is given by:

\[
S = \frac{9Nk}{u^3} \int_{0}^{u} \left[ \frac{x}{e^x - 1} - \ln(1 - e^{-x}) \right] x^2 dx
\]  \hspace{1cm} (13.4.20)

At low temperatures $u$ gets very large and nothing is lost by setting it to infinity. In that limit $S$ becomes

\[
S \rightarrow \frac{4\pi^4Nk}{5} \left( \frac{T}{\Theta_D} \right)^3
\]  \hspace{1cm} (13.4.21)

which can be seen to go to zero with temperature, as it should.
13.5 The One-Dimensional Crystal

Imagine a collection of $N$ atoms arranged such that each has two nearest neighbors, except possibly for atoms 1 and $N$. This is a one-dimensional crystal. A line through all the atoms doesn’t have to be straight, what is important is the interactions.

In fact here we will assume that the atoms are arranged on a circle so that the nearest neighbors of atom 1 are atom 2 and atom $N$.

And of course the particles don’t have to be atoms. They can be molecules just as easily. However, we don’t want to have to worry about the details of the internal motions of molecules or electronic excitations of either molecules or atoms, for that matter. If one really needs to incorporate internal motions, that’s easy enough to do. The math gets no harder, but the equations get significantly longer.

Let $x_i$ denote the position of the $i$th atom. And let the potential energy of interaction be $u(x_{i+1} - x_i)$. Then the total potential energy is:

$$U = \sum_{i=1}^{N} u(x_{i+1} - x_i) \quad (13.5.1)$$

where it is understood that the subscripts ‘wrap’ at atom $N$ so that atom $N + 1$ is atom 1.

For simplicity right now let $x_{i+1} - x_i = r$ so that $r$ is the distance between neighboring atoms, then the potential energy between pairs of atoms is $u(r)$.

Now for a key assumption. We assume that $u(r)$ has a single minimum at $r = a$. We also assume that $u(r)$ rises rapidly enough from that minimum so that the atoms can never move past each other.

From the first assumption we have that:

$$\frac{du(r)}{dr} = 0 \quad \text{at} \quad r = a \quad (13.5.2)$$

where $a$ is the equilibrium distance between any pair of atoms.

For simplicity we’ll let the first particle be at $x_1 = 0$. Thus $x_i = (i - 1)a$. These particles are moving so at any time their actual positions are:

$$x_i = (i - 1)a + \alpha_i \quad (13.5.3)$$

where $\alpha_i$ is the deviation of particle $i$ from its equilibrium position. With this we have:

$$x_{i+1} - x_i = ia + \alpha_{i+1} - (i - 1)a - \alpha_i = \alpha_{i+1} - \alpha_i + a \quad (13.5.4)$$

and

$$u(x_{i+1} - x_i) = u(\alpha_{i+1} - \alpha_i + a) \quad (13.5.5)$$

so that

$$U = \sum_{i=1}^{N} u(\alpha_{i+1} - \alpha_i + a) \quad (13.5.6)$$
To simplify things let
\[ y_i = \alpha_{i+1} - \alpha_i \quad (13.5.7) \]

We now expand \( U \) in Equation (13.5.6) on page 13-10 in a power series around \( y_i = 0 \):

\[ U = U_0 + \sum_{i=1}^{N} \left( \frac{\partial U}{\partial y_i} \right)_0 y_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{\partial^2 U}{\partial y_i \partial y_j} \right)_0 y_i y_j + \ldots \quad (13.5.8) \]

The constant term is easy to evaluate since all the \( y \)'s are zero there. So
\[ U_0 = \sum_{i=1}^{N} u(a) = Nu(a) \quad (13.5.9) \]

The first derivative is zero by virtue of Equation (13.5.2) on page 13-10.

We evaluate the second derivative by noting that:
\[ \frac{\partial^2 U}{\partial y_i \partial y_j} = \frac{\partial^2}{\partial y_i \partial y_j} \sum_{i=1}^{N} U(y_i + a) \quad (13.5.10) \]

This is zero if and only if \( i = j \) so only the pure second degree derivatives survive. Thus to terms in the second order\(^{15} \) we have:
\[ U = Nu(a) + \frac{k}{2} \sum_{i=1}^{N} y_i^2 \quad (13.5.11) \]

where \( k \) is the value of the second derivative at equilibrium.

Reverting now to our \( \alpha \) notation, we have:
\[ U = Nu(a) + \frac{k}{2} \sum_{i=1}^{N} (\alpha_{i+1} - \alpha_i)^2 \quad (13.5.12) \]

To see how the atoms move, we find the force \( F_i \) on the \( i \)'th atom:
\[ F_i = -\left( \frac{\partial U}{\partial \alpha_i} \right) = m \frac{d^2 \alpha_i}{dt^2} \quad (13.5.13) \]

Differentiation of Equation (13.5.12) gives us several terms in \( \alpha \):
\[ m \frac{d^2 \alpha_i}{dt^2} = -k(\alpha_{i+1} - 2\alpha_i + \alpha_{i-1}) \quad (13.5.14) \]

which is true for all atoms. If we had not put atom \( N \) next to atom 1, we’d have had to take the end atoms into account separately. This is no problem, but it just adds complication without adding any insight.

\(^{15}\)What we’ve done is to take an arbitrary potential with a single minimum and derive the harmonic approximation to it.
We assume that the solution to Equation (13.5.14) on page 13-11 is:

\[
\alpha_i = A \sin \left[ \frac{2\pi x_i}{\lambda} + \theta \right] \sin(2\pi \nu t) = A \sin \left[ \frac{2\pi a(i - 1)}{\lambda} + \theta \right] \sin(2\pi \nu t) \quad (13.5.15)
\]

where \( A \) is the amplitude of the vibration, \( \theta \) is the phase of the vibration, \( \lambda \) is the wavelength of the wave, and \( \nu \) is its frequency. The first term on the right-hand side of Equation (13.5.15) is the vibrational amplitude at point \( x_i = (i - 1)a \) in the lattice and the second term is the time variation of that amplitude.

We now differentiate Equation (13.5.15) twice with respect to time to get:

\[
\frac{d^2 \alpha_i}{dt^2} = -4\pi^2 \nu^2 A \sin \left[ \frac{2\pi a(i - 1)}{\lambda} + \theta \right] \sin(2\pi \nu t) \quad (13.5.16)
\]

Inserting Equation (13.5.16) into the left-hand side of Equation (13.5.14) on page 13-11 and using Equation (13.5.15) with suitable values of \( i \) in the right-hand side of Equation (13.5.14) on page 13-11 we get, after some manipulation

\[
-4\pi^2 \nu^2 m \alpha_i = 2k \left[ \cos \left( \frac{2\pi a}{\lambda} \right) - 1 \right] \alpha_i = -4k \alpha_i \sin^2 \left( \frac{\pi a}{\lambda} \right) \quad (13.5.17)
\]

from which we see that the condition under which Equation (13.5.15) is a solution to Equation (13.5.14) on page 13-11 is

\[
\nu^2 = \frac{1}{\pi^2 m} \sin^2 \left( \frac{\pi a}{\lambda} \right) \quad (13.5.18)
\]

Now the sinusoidal solution in Equation (13.5.15) has meaning only at the actual atoms themselves. Or, put more exactly, only at the positions \( x_i \). If we had other waves with wavelength \( \lambda^* \) satisfying:

\[
\frac{a}{\lambda^*} = 1 + \frac{a}{\lambda} \quad \text{or} \quad \frac{a}{\lambda^*} = 1 - \frac{a}{\lambda} \quad (13.5.19)
\]

these would result in exactly the same solutions as before, except that in the second case in Equation (13.5.19) the displacement would have the opposite sign.

Thus we can restrict our wavelengths to satisfy \( a/\lambda = 1/2 \) or:

\[
\lambda \geq 2a \quad (13.5.20)
\]

We see that the waves in our crystal can be infinitely long\(^{16}\) but they cannot be shorter than \( 2a \). Put another way it does not matter if the wave makes one oscillation between adjacent atoms or ten oscillations. The effect on the atoms is the same and hence the waves are the same.

We can use this restriction in Equation (13.5.18) to get:

\[
\nu^2 = \frac{1}{\pi^2 m} \sin^2 \left( \frac{\pi}{2} \right)
\]

\(^{16}\) Assuming an infinitely long one-dimensional crystal.
In the long wavelength limit, $\lambda$ becomes large and hence $\sin \pi a/\lambda$ becomes very small. In that limit $\sin \pi a/\lambda = \pi a/\lambda$ and Equation (13.5.18) on page 13-12 becomes:

$$\nu^2 = \frac{1}{\pi^2} \frac{k}{m} \left( \frac{\pi a}{\lambda} \right)$$

and

$$\nu \lambda = \left( \frac{k}{m} \right)^{1/2} a = c$$

where here $c$ is the speed of the waves.

What we’ve shown just above is that the speed of the waves\(^{17}\) is constant at the long wavelength end of the spectrum. This is the region of low temperature excitations of the crystal. At shorter wavelengths $\nu \lambda$ is not constant, indeed their relationship is contained in Equation (13.5.18) on page 13-12. If one expands the sine function (after taking the square root of each side) in that equation as a function of its arguments, one can see that the product $\nu \lambda$ will depend on $\lambda$ and the smaller $\lambda$ the more rapidly $c$ will change.\(^{18}\)

We can derive the degeneracy function $g(\nu)$ for the one-dimensional crystal from Equation (13.5.18) on page 13-12, or rather its square root:

$$\nu = \frac{1}{\pi} \left( \frac{k}{m} \right)^{1/2} \sin \frac{\pi a}{\lambda}$$

by first noting that the wavelengths in our crystal are $aN/1$ units long ($a$ is the interatom spacing), $aN/2$ units long, $aN/3$ units long, etc., or in general $\lambda = aN/n$ units long. This rearranges to $a/\lambda = n/N$ and so we can write Equation (13.5.24) as:

$$\nu = \frac{1}{\pi} \left( \frac{k}{m} \right)^{1/2} \sin \frac{\pi n}{N}$$

If we treat $n$ as a continuous variable we have that:

$$d\nu = \frac{1}{N} \left( \frac{k}{m} \right)^{1/2} \cos \frac{\pi n}{N} dn$$

Noting that $\sin \theta = \sin(\pi - \theta)$ there are two identical values of $\nu$, one for a given value of $n$ and the other for $n^* = N - n$. So we can (as we have) limit $n$ to be half of the allowed range and multiply by two to compensate.

---

\(^{17}\)These are, of course, sound waves not electromagnetic ones.

\(^{18}\)This phenomenon, called dispersion can be seen (or rather heard) during thunderstorms when sharp high pitched sound of a thunderclap reaches your ears faster than the low rumbles that come later.
We now identify $dn$ with $g(\nu)$, the number of frequencies in the range from $\nu$ to $\nu + d\nu$. We then get:

$$g(\nu)d\nu = \frac{2N(m/k)^{1/2}}{\cos(n\pi/N)}d\nu$$  \hspace{1cm} (13.5.27)

where it should be remembered that $n \leq N/2$. The notation here can be simplified a good deal. First we let

$$\nu_0 = \left(\frac{k}{m\pi^2}\right)^{1/2}$$  \hspace{1cm} (13.5.28)

where $\nu_0$ is the maximum allowed frequency. Then we rewrite Equation (13.5.25) on page 13-13 as

$$\nu = \nu_0 \sin \frac{n\pi}{N}$$  \hspace{1cm} (13.5.29)

then

$$\cos \frac{n\pi}{N} = \frac{1}{\nu_0}(\nu_0^2 - \nu^2)^{1/2}$$

and the final form for the degeneracy function is:

$$g(\nu)d\nu = \frac{2N}{\pi(\nu_0^2 - \nu^2)^{1/2}}d\nu$$  \hspace{1cm} (13.5.30)

As the frequency increases, the denominator gets smaller and smaller until at the limit there is an infinite discontinuity. But that doesn’t matter. This result could have been deduced from classical mechanics. What matters is the low frequency long wavelength limit where $\nu$ becomes negligible compared to $\nu_0$. When that approximation can be made the right-hand side of Equation (13.5.30) becomes a constant. The Debye assumption for $g(\nu)$ under those conditions is also a constant.

A very similar thing happens in three dimensions. The Debye assumption, Equation (13.4.3) on page 13-7, is correct at low frequencies, which are the important ones at low temperatures, but not correct at high frequencies. But that doesn’t matter because at temperatures where high frequencies become important, the behavior is classical and the details of $g(\nu)$ are not so important.
13.6 Appendix: Behavior of the Debye Function

The variation of the Debye function with temperature is not obvious from its definition, except as $T \to 0$ and $T \to \infty$, it is useful to examine it more carefully.

As a reminder the Debye function is:

$$D(u) = \frac{3}{u^3} \int_0^u \frac{x^3 dx}{e^x - 1}, \quad u = \frac{\Theta_D}{T}$$

((13.4.10))

In general this can’t be integrated unless $u = 0$. So some sort of approximation for the value of $D(u)$ must be found.

13.6.1 High Temperature ($u$ small)

At high temperatures $u$ is small and since $x$ lies between 0 and $u$, then $x$ must be small too. We can get a useful approximation by noting that the exponential in the denominator of Equation (13.4.10) on page 13-8 can then be expanded in a power series. This gives us:

$$\frac{x^3}{e^x - 1} = \frac{x^3}{x + x^2/2! + x^3/3! + \cdots} = \frac{x^2}{1 + x/2! + x^2/3! + \cdots}$$

((13.6.1))

where the denominator was divided by $x$ to get it into “standard” form. Now if we could find the series that is the reciprocal of the series in the denominator of Equation (13.6.1) perhaps we could then integrate the result.

We can in fact do this, but it requires a small discussion of the reciprocals of power series first.

If we start with a power series with coefficients $a_i$ and we assume that the reciprocal is another power series with coefficients $b_i$ we can write:

$$\frac{1}{1 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots} = 1 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots$$

((13.6.2))

Now the $a_n$ on the left are assumed known and the $b_n$ are to be determined. All we need to do is multiply it out to get:

$$1 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots + a_1 x + a_1 b_1 x^2 + a_2 b_2 x^3 + \cdots + a_2 x^2 + a_2 b_1 x^3 + \cdots = 1$$

((13.6.3))

Collecting powers of $x$:

$$1 + (a_1 + b_1) x + (a_2 + a_1 b_1 + b_2) x^2 + (a_3 + a_2 b_1 + a_1 b_2 + b_3) x^3 + \cdots = 1$$

((13.6.4))

and the only way this equality can hold is if each of the coefficients of $x$ vanishes. So we set them equal to zero.

The first, $a_1 + b_1 = 0$ is easy to solve for $b_1$. It must be $-a_1$. The second coefficient is $a_2 + a_1 b_1 + b_2 = 0$, which involves $a_1$ and $a_2$ (which we assume we know) and $b_1$
and $b_2$, which we don’t. So how can we solve this? We are lucky! We do know $b_1$ so we can set it to $-a_1$ and solve for $b_2$.

If we look at the third coefficient we see that our luck continues. That involves three different $b$’s, but we know two of them. In fact, if we write out higher coefficients we see that if we work out the $b$’s, starting from $b_1$, we will always have just enough information to find the next $b$.

I’ve done this already, and the results for the first six $b$’s are:

\[
\begin{align*}
b_1 &= -a_1 \\
b_2 &= a_1^2 - a_2 \\
b_3 &= -a_1^3 + 2a_1a_2 - a_3 \\
b_4 &= a_1^4 - 3a_1^2a_2 + 2a_1a_3 + a_2^2 - a_4 \\
b_5 &= -a_1^5 + 4a_1^3a_2 - 3a_1^2a_3 + 2a_1a_4 - 3a_1a_2^2 + 2a_2a_3 - a_5 \\
b_6 &= a_1^6 - 5a_1^4a_2 + 4a_1^3a_3 + a_1^2(6a_2^2 - 3a_4) + a_1(2a_5 - 6a_2a_3) - a_3^2 + 2a_2a_4 + a_3^2 - a_6
\end{align*}
\]

So we have solved the problem of finding the reciprocal of a power series.\(^{19}\)

Our job now is to apply this to the denominator in Equation (13.6.1) on page 13-15. In that equation we have $a_1 = 1/2!$, $a_2 = 1/3!$, and in general $a_n = (n + 1)!$. So we have for the $b$’s:

\[
\begin{align*}
b_1 &= -\frac{1}{2} \\
b_2 &= \frac{1}{12} \\
b_3 &= 0 \\
b_4 &= -\frac{1}{720} \\
b_5 &= 0 \\
b_6 &= \frac{1}{30240}
\end{align*}
\]

Those of you with good memories will remember that we’ve seen these numbers before. They occur in Equation (8.6.8) on page 8-18 and are clearly related to the Bernoulli numbers given in Table 8.5 on page 8-17. Indeed $b_n$ is given by:

\[
b_n = \frac{B_n}{n!}
\]

where $B_n$ is the $n$’th Bernoulli number.

\(^{19}\)Be a bit careful about using this. The result above is correct from $x = 0$ up to (but not including) the first zero of the polynomial in $a$. 
We now have:

\[ \frac{x^2}{1 + x/2! + x^2/3! + \ldots} = \left[ 1 - \frac{x}{2} + \frac{1}{12} x^2 - \frac{1}{720} x^4 + \frac{1}{30240} x^6 - \frac{1}{1209600} x^8 \cdots \right] = x^2 - \frac{1}{2} x^3 + \frac{1}{12} x^4 - \frac{1}{720} x^6 + \frac{1}{30240} x^8 - \frac{1}{1209600} x^{10} + \cdots \] (13.6.8)

where, of course, all the odd terms (except for the first) are missing because \( B_n = 0 \) for odd \( n \neq 1 \).

So the Debye function at high temperatures can be written as:

\[
D(u) = 3 \int_0^u \frac{x^3 dx}{e^x - 1} = 3 \int_0^u \left[ x^3 - \frac{1}{2} x^3 + \frac{1}{12} x^4 - \frac{1}{720} x^6 \right. \\
+ \left. \frac{1}{30240} x^8 - \frac{1}{1209600} x^{10} + \frac{1}{47900160} x^{12} + \cdots + \frac{B_n}{n!} x^{n+2} + \cdots \right] \, dx \tag{13.6.9}
\]

This can now be integrated term by term to give:

\[
D(u) = 3 \int_0^u \frac{x^3 dx}{e^x - 1} = 3 \int_0^u \left[ x^3 - \frac{1}{2} x^3 + \frac{1}{12} x^4 - \frac{1}{720} x^6 \right. \\
+ \left. \frac{1}{30240} x^8 - \frac{1}{1209600} x^{10} + \frac{1}{47900160} x^{12} + \cdots + \frac{B_n}{(n+2)n!} x^{n+2} + \cdots \right] \, dx \tag{13.6.10}
\]

and so the final result is:

\[
D(u) = 1 - \frac{3u^3}{8} + \frac{u^2}{20} - \frac{u^4}{1680} + \frac{u^6}{97720} - \frac{u^8}{4435200} + \frac{u^{10}}{207567360} + \cdots + \frac{3B_n u^n}{(n+3)n!} + \cdots \tag{13.6.11}
\]

which, when compared to the table in Abramowitz and Stegun,\(^{20}\) agrees to six decimal places for \(0 < u \leq 0.7\).

### 13.6.2 Low Temperature (\( u \) large)

At low temperatures \( u \) is large and the series derived in the previous subsection becomes less and less accurate as \( u \) increases beyond one. So for low temperatures (large \( u \)) we clearly need another approach.

We write the Debye function as:

\[
D(u) = \frac{3}{u^3} \int_0^\infty \frac{x^3 dx}{e^x - 1} - \frac{3}{u^3} \int_u^\infty \frac{x^3 dx}{e^x - 1} \tag{13.6.12}
\]

The first integral can be done and has the value \( \pi^4/15 \). In the second term we can multiply top and bottom of the integrand by \( \exp(-x) \) to get

\[
\frac{x^3 dx}{e^x - 1} = \frac{x^3 e^{-x} dx}{1 - e^{-x}} \tag{13.6.13}
\]

\(^{20}\)Abramowitz and Stegun, eds, \textbf{Handbook of Mathematical Functions}, page 998.
The trick is now to recognize that the denominator is a geometric series

\[ 1 + y + y^2 + y^3 + \cdots = \frac{1}{1 - y} \quad y < 1 \]

and since \( y \) is our \( \exp(-x) \) it is always less than one, so we can write the second integral in Equation (13.6.12) on page 13-17 as:

\[ \frac{x^3e^{-x}}{1 - e^{-x}} = x^3e^{-x} \left[1 + e^{-x} + e^{-2x} + \cdots\right] \quad (13.6.14)\]

which is always valid since \( x \) is always positive.

So \( D(u) \) is then:

\[ D(u) = \frac{\pi^4}{5u^3} - \frac{3}{u^3} \int_u^\infty x^3e^{-x} \left[1 + e^{-x} + e^{-2x} + \cdots\right] \, dx \quad (13.6.15)\]

This can be re-arranged to

\[ D(u) = \frac{\pi^4}{5u^3} - \frac{3}{u^3} \sum_{n=1}^\infty \int_u^\infty x^3e^{-nx} \, dx \quad (13.6.16)\]

The integral is

\[ \int_u^\infty x^3e^{-nx} \, dx = -\left[\frac{x^3}{n} + \frac{3x^2}{n^2} + \frac{6x}{n^3} + \frac{6}{n^4}\right] \left[\frac{e^{-nx}}{u}\right]_u^\infty \]

\[ = \left(\frac{u^3}{n} + \frac{3u^2}{n^2} + \frac{6u}{n^3} + \frac{6}{n^4}\right) e^{-nu} \quad (13.6.17)\]

When all the pieces are put together the Debye function is then:

\[ D(u) = \frac{\pi^4}{5u^3} - \sum_{n=1}^\infty \left[\frac{3}{n} + \frac{9}{un^2} + \frac{18}{u^2n^3} + \frac{18}{u^3n^4}\right] e^{-nu} \quad (13.6.18)\]

This series in principle is exact. Keeping only one or two correction terms suffices for values of \( u > 5 \). For smaller values of \( u \) more and more correction terms are needed. Of course if one programs a computer to do the computation, then using 20 terms in Equation (13.6.17) gives full agreement with the Abramowitz and Stegun tables.\(^\text{21}\) In that case the use of Equation (13.6.11) on page 13-17 for \( 0 \leq u \leq 0.7 \) coupled with the use of Equation (13.6.18) for \( 0.7 \leq u \) suffices to give six digit accuracy for all positive values of \( u \).\(^\text{22}\)

\(^{21}\)Abramowitz and Stegun, eds, **Handbook of Mathematical Functions**, page 998.

\(^{22}\)I have followed the derivation of Mayer and Mayer, *Statistical Mechanics*, First Edition, John Wiley and Sons, 1940. The only real difference is that Equation (13.6.18) is more general, Mayer and Mayer having stopped with \( n = 2 \).
Figure 13.3: The Debye Function $D(u)$

<table>
<thead>
<tr>
<th>$T/\Theta_D$</th>
<th>$D(\Theta_D)/T$</th>
<th>$C_V/Nk$</th>
<th>$T/\Theta_D$</th>
<th>$D(\Theta_D)/T$</th>
<th>$C_V/Nk$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.0</td>
<td>0.0</td>
<td>0.50</td>
<td>0.441128</td>
<td>0.825408</td>
</tr>
<tr>
<td>0.05</td>
<td>0.002435</td>
<td>0.009741</td>
<td>0.55</td>
<td>0.477339</td>
<td>0.852407</td>
</tr>
<tr>
<td>0.10</td>
<td>0.019296</td>
<td>0.075821</td>
<td>0.60</td>
<td>0.509520</td>
<td>0.873796</td>
</tr>
<tr>
<td>0.15</td>
<td>0.059619</td>
<td>0.212992</td>
<td>0.65</td>
<td>0.538225</td>
<td>0.890976</td>
</tr>
<tr>
<td>0.20</td>
<td>0.117597</td>
<td>0.368635</td>
<td>0.70</td>
<td>0.563937</td>
<td>0.904954</td>
</tr>
<tr>
<td>0.25</td>
<td>0.181737</td>
<td>0.503059</td>
<td>0.75</td>
<td>0.587067</td>
<td>0.916460</td>
</tr>
<tr>
<td>0.30</td>
<td>0.244410</td>
<td>0.607703</td>
<td>0.80</td>
<td>0.607963</td>
<td>0.926033</td>
</tr>
<tr>
<td>0.35</td>
<td>0.302212</td>
<td>0.686573</td>
<td>0.85</td>
<td>0.626916</td>
<td>0.934076</td>
</tr>
<tr>
<td>0.40</td>
<td>0.354136</td>
<td>0.745853</td>
<td>0.90</td>
<td>0.644175</td>
<td>0.940893</td>
</tr>
<tr>
<td>0.45</td>
<td>0.400271</td>
<td>0.790823</td>
<td>0.95</td>
<td>0.659949</td>
<td>0.946718</td>
</tr>
<tr>
<td>1.00</td>
<td>0.674416</td>
<td>0.951732</td>
<td>6.00</td>
<td>0.938888</td>
<td>0.998612</td>
</tr>
<tr>
<td>1.50</td>
<td>0.772106</td>
<td>0.978125</td>
<td>6.50</td>
<td>0.943491</td>
<td>0.998818</td>
</tr>
<tr>
<td>2.00</td>
<td>0.824963</td>
<td>0.987611</td>
<td>7.00</td>
<td>0.947449</td>
<td>0.998980</td>
</tr>
<tr>
<td>2.50</td>
<td>0.857985</td>
<td>0.992045</td>
<td>7.50</td>
<td>0.950889</td>
<td>0.999112</td>
</tr>
<tr>
<td>3.00</td>
<td>0.880548</td>
<td>0.994466</td>
<td>8.00</td>
<td>0.953906</td>
<td>0.999219</td>
</tr>
<tr>
<td>3.50</td>
<td>0.896935</td>
<td>0.995930</td>
<td>8.50</td>
<td>0.956574</td>
<td>0.999308</td>
</tr>
<tr>
<td>4.00</td>
<td>0.909373</td>
<td>0.996882</td>
<td>9.00</td>
<td>0.958951</td>
<td>0.999383</td>
</tr>
<tr>
<td>4.50</td>
<td>0.919134</td>
<td>0.997535</td>
<td>9.50</td>
<td>0.961080</td>
<td>0.999446</td>
</tr>
<tr>
<td>5.00</td>
<td>0.926999</td>
<td>0.998003</td>
<td>10.00</td>
<td>0.963000</td>
<td>0.999500</td>
</tr>
</tbody>
</table>

Table 13.1: The Debye Function and Heat Capacity for Reduced Temperatures
13.7 Appendix: Differentiating Functions Defined by Integrals

It sometimes happens that one needs to find the derivative of a function defined by an integral. For example, consider the integral
\[ I(r) = \int_0^r x \, dx , \]  
(13.7.1)
and let us assume that one wants to know \( dI/dr \). Here it is easy. One first does the integral:
\[ I(r) = r^2/2 , \]  
(13.7.2)
and then differentiates with respect to \( r \) to get:
\[ \frac{dI(r)}{dr} = r . \]  
(13.7.3)

But what if the integral cannot be done?

The general case of this is an integral that is a function of \( r \) that could be given by:
\[ I(r) = \int_{a(r)}^{b(r)} f(r,x) \, dx , \]  
(13.7.4)
where \( f(r,x) \) is some function that cannot be readily integrated and \( a \) and \( b \) as well as \( f(r,x) \) depend on \( r \).

The solution is based on Leibnitz’ rule and is given by\(^{23}\)
\[ \frac{d}{dr} \int_{a(r)}^{b(r)} f(r,x) \, dx = f(r, b(r)) b'(r) - f(r, a(r)) a'(r) + \int_{a(r)}^{b(r)} \frac{\partial f(r,x)}{\partial r} \, dx , \]  
(13.7.5)
where \( a'(r) \) and \( b'(r) \) are the first derivatives of the limits \( a \) and \( b \) with respect to \( r \).

Here are some examples:

---

**Example 13.1**

Consider the integral
\[ I = \frac{d}{dr} \int_1^{2r^2} \frac{r}{x} \, dx , \]  
(13.7.6)

\(^{23}\)I'm not going to prove this. The proof is not at all hard but would take us too far afield. My source for this is W. Kaplan’s *Advanced Calculus*, Addison-Wesley, 1952, a book published over 50 years ago but still in print as the 5th edition published in 2002. The page reference is 220 in the original edition. The proof can likely be found in any reasonable calculus book.
The integral can be done easily enough and the result then differentiated. Doing so we get:

\[
I = \frac{d}{dr} \int_1^{2r^2} \frac{r}{x} dx = \frac{d}{dr} r \ln x \bigg|_1^{2r^2} = \frac{d}{dr} (r \ln 2r^2) = 2 + \ln 2 + 2 \ln 2.
\] (13.7.7)

Or we could use Equation (13.7.5):

\[
I = \frac{d}{dr} \int_1^{2r^2} \frac{r}{x} dx = \frac{r}{2r^2} \ln x \bigg|_1^{2r^2} = 2 + \ln x \bigg|_1^{2r^2} = 2 + \ln 2 + 2 \ln 2.
\] (13.7.8)

which gives the same result.

Here’s another example using the Debye Function:

**Example 13.2**

Consider

\[
I = \frac{3}{u^2} \int_0^u \frac{x^3 dx}{e^x - 1},
\] (13.7.9)

where \( u = h/T \), \( x = \lambda_{\max} / kT \), and \( I \) is the Debye function. This example is a bit more complex than the example above first because the \( T \) in \( x \) is not the \( T \) that is being differentiated. It is a dummy variable of integration and could be changed to another letter entirely. And second it is more complex because of the terms out in front of the integral.

All these complications do is require us to be more careful. I rewrite Equation (13.7.9) to show the explicit dependence on \( T \) using \( u = b/T \) with \( b = h/T \):

\[
I = \frac{3T^3}{b^3} \int_0^b \frac{x^3 dx}{e^x - 1}.
\]  

(13.7.10)

Now

\[
\frac{dI}{dT} = \frac{d}{dT} \left[ \frac{3T^3}{b^3} \int_0^b \frac{x^3 dx}{e^x - 1} \right] = \frac{3}{T} \frac{3}{u^3} \int_0^b \frac{x^3 dx}{e^x - 1} + \frac{3}{u^3} \frac{d}{dT} \int_0^b \frac{x^3 dx}{e^x - 1},
\]  

(13.7.11)

which is:

\[
\frac{dI}{dT} = \frac{3}{T} D(u) + \frac{3}{u^3} \frac{d}{dT} \int_0^b \frac{x^3 dx}{e^x - 1}
\]  

(13.7.12)
Now we can concentrate on differentiating the integral itself. We have:

\[
\frac{d}{dT} \int_0^{b/T} \frac{x^3}{e^x - 1} \, dx = \frac{u^3}{e^u - 1} \left( \frac{du}{dT} \right) + \int_0^{b/T} \frac{d}{dT} \frac{x^3}{e^x - 1},
\]

(13.7.13)

where since the integral is not a function of \( T \) the last term is zero. The derivative \( du/dT \) is simply \(-u/T\) so the result is

\[
\frac{d}{dT} D(u) = \frac{3}{T} \left[ D(u) - \frac{u}{e^u - 1} \right].
\]

(13.7.14)
Chapter 14

Simple Quantum Statistics

14.1 Quantum Statistics

In quantum mechanics it is a fact that if two identical particles are interchanged in position, the wave function for that system either changes sign, i.e. $\psi(x, y) = -\psi(x, y)$, or it does not change sign, i.e. $\psi(x, y) = \psi(x, y)$.

In either case, the square of the wave function is unchanged.

Both cases are known. If the sign changes, the wave function is antisymmetric and the particles are said to obey Fermi-Dirac statistics and are called fermions.

If the sign does not change, the wave function is symmetric and the particles are said to obey Bose-Einstein statistics and are called bosons.

Fermions have spins that are half integers such as $1/2$, $3/2$, $5/2$, . . . . Bosons have spins that are whole integers such as $0$, $1$, $2$, . . . .

Fermions are common in chemistry. Electrons, protons, and neutrons are fermions. Bosons are less common in chemistry, though many nuclei are bosons. The most common fundamental boson is the photon.

The property of fermions and bosons most important in statistical thermodynamics is that fermions have the property that no more than one fermion may be in a given quantum state at a given time,\footnote{This is the source of the Pauli exclusion principle, among other phenomenon.} while bosons have the property that any number of them may be in the same quantum state at the same time.

The problem of identicality has another ramification in statistical mechanics. Macroscopically, it takes a turn of $360^\circ$ to return a circle to its original configuration. To see this simply put a mark on an otherwise plane circle. There is no way to rotate the
mark back to its starting position other than to rotate it through 360°.

Things are different in quantum mechanics. Think about isotopically pure benzene. Rotation of a benzene molecule around an axis perpendicular the molecular plane through 60° results in returning it to its starting position. There is no way to mark an isotopically pure benzene molecule. Thus a “benzene circle” contains only 60°.

Up to this point we have been treating systems as essentially classical, correcting at times for the indistinguishability problem by dividing by \( N! \). To illustrate the fundamental differences among fermions, bosons, and classical particles obeying Boltzmann statistics, consider a system having only two particles and five possible states:

<table>
<thead>
<tr>
<th>state</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>±</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>±</td>
<td>±</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>+</td>
</tr>
</tbody>
</table>

where + indicates a legal boson state and a ± a legal state for both fermions and bosons. All \( \frac{25}{2} = 12.5 \) classical states are legal, where we’ve divided by 2 to “correct” for indenticality. Why? Because all the boxes are valid classical states.

But quantum mechanically the state \( \{1,2\} \) is identical to the state \( \{2,1\} \). So the empty boxes are really not there at all!

Further, the state \( \{2,2\} \) is not a legal fermion state, though it is a legal boson state. So all in all there are just 15 Bose-Einstein states and 10 Fermi-Dirac states.

Our problem is to construct a partition function for the quantum cases. Using the canonical ensemble is very hard because we have two major constraints. If \( n_k \) is the number of particles in quantum state \( k \), then

\[
\sum_k n_k = N
\]

where \( N \) is the (fixed) number of particles in the system, and

\[
\sum_k n_k \epsilon_k = E_j
\]

where \( E_j \) is the \( j \)’th system quantum state and \( \epsilon_k \) is the \( k \)'th particle energy level in the system. What this last equation means is that given a set of particle energy levels \( \epsilon_k \), then the numbers \( n_0, n_1, n_2, \ldots \) are not only the numbers of particles in each particle energy level, but that set of \( n \)'s must add up to \( N \).

Then we have:

\[
Q(\beta, V, N) = \sum e^{-\beta \sum_k n_k \epsilon_k}
\]

\(^2\)I reject out of hand the term used in some texts, Boltzons. It does not roll readily off the English-adjusted tongue...
where it is assumed that the values of the $n_k$ satisfy both Equations (14.1.1) on page 14-2 and (14.1.2) on page 14-2.

This is a tall order. Clearly most randomly chosen sets of $n_k$ won’t do this.

The way out of this is to use the grand canonical ensemble. That ensemble sums over all $N$ and in doing so will include all possible sets of numbers $\{n_k\}$ somewhere because each set adds up to some value of $N$.

The grand canonical partition function can be written as:

$$\Xi(\lambda, V, \beta) = \sum_{N=0}^{\infty} Q(N, V, \beta) \lambda^N$$

(14.1.4)

as in Equation (11.7.6) on page 11-8.

Inserting Equation (14.1.3) on page 14-2 into this gives:

$$\Xi(\lambda, V, \beta) = \sum_{N=0}^{\infty} \sum_j e^{-\beta \sum k n_k \epsilon_k} \lambda^N$$

(14.1.5)

where the second sum is done in such a way as to satisfy Equations (14.1.1) on page 14-2 and (14.1.2) on page 14-2.

To quote Huang,\(^3\)

It is to be noted that the double summation just given is equivalent to summing each $n_k$ independently. To prove this we must show that every term in one case appears once and only once in the other, and vice versa. This is easily done mentally.

Of course, it isn’t “easily done”, but we will defer further discussion of this point until later to avoid interrupting the main argument.

Assuming then that we have proven the equality to our satisfaction, then:

$$\Xi(\lambda, V, \beta) = \sum_{n_0} \sum_{n_1} \sum_{n_2} \cdots \left[ (\lambda e^{-\beta \epsilon_0})^{n_0} (\lambda e^{-\beta \epsilon_1})^{n_1} \cdots \right]$$

(14.1.6)

$$= \left[ \sum_{n_0} (\lambda e^{-\beta \epsilon_0})^{n_0} \right] \left[ \sum_{n_1} (\lambda e^{-\beta \epsilon_1})^{n_1} \right] \cdots$$

(14.1.7)

$$= \prod_k \left[ \sum_n (\lambda e^{-\beta \epsilon_k})^n \right].$$

(14.1.8)

If we define $\xi$ to be:

$$\xi = \sum_n e^{-\beta \epsilon_n} \lambda^n,$$

(14.1.9)

where $x_i$ is a single level grand partition function, then

$$\Xi(\lambda, V, \beta) = \begin{cases} \xi_0 \xi_1 \xi_2 \cdots \xi_N & \text{if the levels are different,} \\ \xi^N & \text{if the levels are identical.} \end{cases}$$

(14.1.10)

Example 14.1

Here’s an example to clarify the “easily done” quote from Huang. Let us work out a simple situation in which there are only three quantum levels, each of which can have 0, 1, or 2 particles in it. So \( k \), which indexes the energy levels of this system, is zero, one, or two and \( n \), which is the number of particles in a given level ranges from 0 to 3. So Equation (14.1.4) on page 14-3 is

\[
\Xi(\lambda, V, \beta) = \sum_{N=0}^{6} Q(N, V, \beta) \lambda^N ,
\]  

(14.1.11)

where 6 is the maximum number of particles in the system. Of course the sum is constrained by the fact that we can have no more than two particles in a given level. In fact it is useful to write out some of the \( Q \)'s: Of course \( Q(0) = 1 \) (that’s the system with no particles in it). Then

\[
Q(1) \lambda = [e^{-\beta \epsilon_1} + e^{-\beta \epsilon_2} + e^{-\beta \epsilon_3}] \lambda,
\]  

(14.1.12)

where this sum isn’t infinite since there are only three levels and there is only one particle in this system. Next:

\[
Q(2) \lambda^2 = \left[ e^{-2\beta \epsilon_1} + e^{-\beta(\epsilon_1+\epsilon_2)} + e^{-2\beta \epsilon_2} + e^{-\beta(\epsilon_1+\epsilon_3)} + e^{-\beta(\epsilon_2+\epsilon_3)} + e^{-2\beta \epsilon_3} \right] ,
\]  

(14.1.13)

where those six terms give all the possibilities for two particles in three levels.

We can simplify the notation a bit by letting

\[
x_k = e^{-\beta \epsilon_k} \lambda .
\]  

(14.1.14)

It is now simple to write the \( Q \)'s:

\[
Q(0) = 1, \quad (14.1.15)
\]
\[
Q(1) \lambda = x_1 + x_2 + x_3, \quad (14.1.16)
\]
\[
Q(2) \lambda^2 = x_1^2 + 2x_1x_2 + x_2^2 + x_3x_2 + x_2x_3 + x_3^2, \quad (14.1.17)
\]
\[
Q(3) \lambda^3 = x_1^3x_2 + x_1^2x_2x_3 + x_1x_2^2x_3 + x_1x_2x_3^2 + x_1^2x_2x_3 + x_1x_2x_3^2 + x_3^3x_2 + x_3^2x_3x_2 + x_3^2x_2x_3, \quad (14.1.18)
\]
\[
Q(4) \lambda^4 = x_1^4x_2 + 4x_1^3x_2x_3 + 4x_1^2x_2^2x_3 + 6x_1^2x_2^2x_3 + 4x_1x_2^2x_3^2 + 4x_1x_2x_3^2 + x_1^2x_2x_3^2 + 8x_2^2x_3^2, \quad (14.1.19)
\]
\[
Q(5) \lambda^5 = x_1^5x_2x_3 + 5x_1^4x_2^2x_3 + 10x_1^3x_2^2x_3^2 + 10x_1^2x_2^3x_3 + 5x_1x_2^4x_3 + x_1x_2^3x_3^2, \quad (14.1.20)
\]
\[
Q(6) \lambda^6 = x_1^6x_2x_3, \quad (14.1.21)
\]

To see what is going on, compare Equation (14.1.17) to Equation (14.1.13). That should help.

It should also be pointed out that the equations above take the restrictions on the number of particles per level into account. So using them Equation (14.1.11) is perfectly correct.
What we want to show now is that Equation (14.1.8) on page 14-3 results from what we have obtained so far. Written using \(x\)'s instead of exponentials and \(\lambda\)'s, we have:

\[
\Xi(\lambda, V, \beta) = 1 + x_1 + x_2 + x_3 + x_1^2 + x_1 x_2 + x_2^2 + x_1 x_3 + x_2 x_3 + x_3^2 \\
+ x_1^2 x_2 + x_1 x_2^2 + x_2^2 x_3 + x_1 x_2 x_3 + x_2^2 x_1 + x_3^2 x_2 \\
+ x_1^2 x_2^2 + x_1^2 x_2 x_3 + x_1 x_2^2 x_3 + x_1^2 x_3 + x_1 x_2 x_3^2 + x_2^2 x_3 \\
+ x_1^2 x_2^2 x_3 + x_1^2 x_2 x_3^2 + x_1 x_2^2 x_3^2 + x_1^2 x_2^2 x_3^2, \quad (14.1.22)
\]

which, believe it or not, factors into

\[
\Xi(\lambda, V, \beta) = (1 + x_1 + x_1^2)(1 + x_2 + x_2^2)(1 + x_3 + x_3^2). \quad (14.1.23)
\]

You may have to multiply that out in order to see that this is true, but it is true nevertheless.

And equation (14.1.23) is, in fact, our desired result. It can be more abstractly written as:

\[
\Xi(\lambda, V, \beta) = \prod_{k=1}^{3} \left[ \sum_{n=0}^{2} x_k^n \right] = \prod_{k=1}^{3} \left[ \sum_{n=0}^{2} (\lambda e^{-\beta \epsilon_k})^n \right], \quad (14.1.24)
\]

which is precisely what we want.

Back now to the problem at hand. So we have:

\[
\Xi(\beta, V, \gamma) = \prod_{k} \left[ \sum_{n} (\lambda e^{-\beta \epsilon_k})^n \right]. \quad (14.1.25)
\]

This contains a sum over \(n\), running from 0 to the maximum number of particles that can be in state \(k\). In the case of fermions each state can hold at most 1 particle, so Equation (14.1.25) becomes simply

\[
\Xi_{FD}(\beta, V, \lambda) = \prod_{k} \left[ 1 + \lambda e^{-\beta \epsilon_k} \right], \quad (14.1.26)
\]

But for bosons any number of particles in a single state. So \(n\) is infinity. We then get:

\[
\Xi_{BE}(\beta, V, \lambda) = \prod_{k} \sum_{n=0}^{\infty} [\lambda e^{-\beta \epsilon_k}]^{n_k}. \quad (14.1.27)
\]

This last equation can be simplified. If \(x_k\) is substituted for \(\lambda e^{-\beta \epsilon_k}\), we get:

\[
\Xi_{BE}(\beta, V, \lambda) = \prod_{k} \sum_{n=0}^{\infty} x_k^{n_k}, \quad (14.1.28)
\]

and we see that the sum is a geometric series whose sum is:

\[
\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad -1 \leq x < 1,
\]
as long as \( x_k = \lambda e^{-\beta \epsilon_k} \) is less than 1. This is clearly the case for \( \exp(-\beta \epsilon_k) \) since both \( \beta \) and \( \epsilon_k \) are positive. However, what about \( \lambda \)?

The largest exponential term is the ground state \((k = 0)\) so \( \lambda \) must be less than \( \exp(\beta \epsilon_0) \). Since \( \lambda = \exp(\beta \mu) \), this is equivalent to saying that \( \mu \) must be less than the ground state energy \( \epsilon_0 \). If, as is often the case, \( \epsilon_0 = 0 \), then \( \mu < 0 \) or \( \lambda < 1 \).

With this restriction on \( \lambda \), Equation (14.1.28) on page 14-5 reduces to:

\[
\Xi_{BE} = \prod_k \left[ 1 - \lambda e^{-\beta \epsilon_k} \right]^{-1} \tag{14.1.29}
\]

These two equations can be written together as:

\[
\Xi(\beta, V, \gamma) = \prod_k \left[ 1 \pm \lambda e^{-\beta \epsilon_k} \right]^{\pm1} \tag{14.1.30}
\]

We will use this equation with the convention that the Fermi-Dirac sign is on the top and the Bose-Einstein sign is on the bottom.

### 14.2 Simple Results

Some simple results stem from Equation (14.1.30). If we take logs we get

\[
\ln \Xi = (\pm) \sum_k \ln \left[ 1 \pm \lambda e^{-\beta \epsilon_k} \right] . \tag{14.2.1}
\]

Now since:

\[
\langle N \rangle = \left( \frac{\partial \ln \Xi}{\partial \gamma} \right) = -\lambda \left( \frac{\partial \ln \Xi}{\partial \lambda} \right) , \tag{14.2.2}
\]

it follows that

\[
\langle N \rangle = \sum_k \left[ \frac{\lambda e^{-\beta \epsilon_k}}{1 \pm \lambda e^{-\beta \epsilon_k}} \right] . \tag{14.2.3}
\]

Note the form of this. The average number of particles in the system is a sum over the quantum states \( k \) of the system. Each term in the sum pertains to a different quantum \( k \). Now each quantum state also has an average population. In fact it must be true that:

\[
\langle N \rangle = \sum_k \langle n_k \rangle , \tag{14.2.4}
\]

where \( \langle n_k \rangle \) is the expected number (average) of particles in quantum state \( k \). Then from Equation (14.2.3) we can infer that \( \langle n_k \rangle \) is given by:

\[
\langle n_k \rangle = \frac{\lambda e^{-\beta \epsilon_k}}{1 \pm \lambda e^{-\beta \epsilon_k}} . \tag{14.2.5}
\]

It can be seen that for Fermi-Dirac statistics (top sign) the denominator is always larger than the numerator and the average number of particles in a state is one or less. For Bose-Einstein statistics (bottom sign), the numerator is less than one and so the average number of particles in a state can be greater than one. This is exactly what we’d expect.
Another similar result is a simple way to calculate the average energy $\langle E \rangle$:

$$
\langle E \rangle = \sum_k n_k \epsilon_k = \sum_k \left[ \frac{\epsilon_k \lambda e^{-\beta \epsilon_k}}{1 \pm \lambda e^{-\beta \epsilon_k}} \right].
$$

(14.2.6)

The simplest result is saved for last. Since $\beta PV = \ln \Xi$,

$$
\beta \langle p \rangle V = \pm \sum_k \ln \left[ 1 \pm \lambda e^{-\beta \epsilon_k} \right].
$$

(14.2.7)

This is not the Ideal Gas law!

### 14.3 The Ideal Gas Limit

The equations derived above up to this point are exact for systems made up of independent subsystems. Comparison to the corresponding expressions for classical statistics (the Boltzmann expressions) shows that these are different. Thus under conditions that favor the display of quantum effects (low temperatures or high densities) it must be expected that the classical equations will not be obeyed.

But in the other limit, high temperatures and low densities, we can expect that the classical results will be obtained. In particular, it is to be hoped that both the Fermi-Dirac and the Bose-Einstein equations, different though they may look, will reduce to Boltzmann statistics. And we hope that Equation (14.2.7) will turn into the ideal gas law.

In fact, it does. Any situation that results in the average occupation of a quantum energy level being small will do it. Consider Equation (14.2.5) on page 14-6:

$$
\langle n_k \rangle = \frac{\lambda e^{-\beta \epsilon_k}}{1 \pm \lambda e^{-\beta \epsilon_k}}.
$$

(14.2.5)

The only way $\langle n_k \rangle$ will be much less than 1 for any energy level $\epsilon_k$ will be for $\lambda$ to be much less than 1. This works for both Fermi-Dirac and Bose-Einstein statistics. And when it is true that $\lambda$ is much less than 1, the denominator in both cases becomes essentially 1. Then Equation (14.2.5) on page 14-6 becomes

$$
\langle n_k \rangle = \lambda e^{-\beta \epsilon_k},
$$

(14.3.1)

which has to be less than 1 too since $\lambda$ is small. Note that the right hand side of Equation (14.3.1) will be small for $k$ large since our energies are numbered in ascending order. So if $\epsilon_k$ is large, $\langle n_k \rangle$ is small in any case. It is only the situations in which $k = 1$ or some other very small integer that we have to worry about $\lambda$.

If Equation (14.3.1) is now summed over all energy levels we get:

$$
\langle N \rangle = \sum_k \langle n_k \rangle = \lambda \sum_k e^{-\beta \epsilon_k} = \lambda q,
$$

(14.3.2)

where $q$ is the classical single particle partition function. Thus this equation is identical to Equation (11.7.9) on page 11-9, the classical result.
We can solve Equation (14.3.2) on page 14-7 for $\lambda$:

$$\lambda = \frac{\langle N \rangle}{q}.$$  \hfill (14.3.3)

Using this value of $\lambda$ in Equation (14.3.1) on page 14-7 then yields:

$$\frac{\langle n_k \rangle}{\langle N \rangle} = \frac{e^{-\beta \epsilon_k}}{q},$$  \hfill (14.3.4)

the right-hand side of which will be recognized as the Boltzmann expression for the probability of finding a given particle in quantum energy level $k$. This is, of course, a classical result.

Once Equation (14.3.4) is obtained, all other results obtained at small $\lambda$ must also be classical. Just to confirm this, when $\lambda$ is small, Equation (14.2.7) on page 14-7 can be expanded in a power series. The series needed is:

$$\ln(1 \pm x) = \pm x + x^2 \pm x^3 + x^4 \ldots,$$  \hfill (14.3.5)

which is valid for $-1 \leq x < 1$. The result of the expansion, carrying only the first term in $\lambda$ is:

$$\beta \langle p \rangle V = \pm \sum_k \ln \left[ 1 \pm \lambda e^{-\beta \epsilon_k} \right] \approx \lambda \sum_k e^{-\beta \epsilon_k} = \lambda q,$$  \hfill (14.3.6)

which, by Equation (14.3.2) on page 14-7 is equal to $\langle N \rangle$. So the result is:

$$\beta \langle p \rangle V = \langle N \rangle,$$  \hfill (14.3.7)

which is clearly the ideal gas law.

So the quantum statistical expressions, by some colossal coincidence, both give rise to the same high temperature low density result, the classical statistical thermodynamics of Boltzmann and Gibbs, at least for systems composed of independent subsystems.

Is there really magic in this? No, at least if it is looked at in the right way. It is an experimental fact that systems of independent subsystems, under classical conditions, do obey the same laws. There were not two or more different sets of ideal laws discovered, there was only one. Thus, had one put the question to Gibbs or Boltzmann, they would have replied that, if low temperature high density laws were different from the laws they knew, they would nevertheless have to come down to the classical laws in the classical region. Well, there are, and they do.
Chapter 15

Ideal Quantum Gases

15.1 Introduction

We took a brief look at ideal quantum gases in Section 14.2 on page 14-6 of Chapter 14. Those results were, with the Fermi-Dirac case as the upper sign and the Bose-Einstein case as the lower sign:

\[
\ln \Xi = (\pm) \sum_k \ln \left[ 1 \pm \lambda e^{-\beta \epsilon_k} \right] \quad (15.1.1)
\]

\[
\langle N \rangle = \sum_k \left[ \frac{\lambda e^{-\beta \epsilon_k}}{1 \pm \lambda e^{-\beta \epsilon_k}} \right] \quad (15.1.2)
\]

\[
\langle n_k \rangle = \frac{\lambda e^{-\beta \epsilon_k}}{1 \pm \lambda e^{-\beta \epsilon_k}} \quad (15.1.3)
\]

\[
\langle E \rangle = \sum_k n_k \epsilon_k = \sum_k \left[ \frac{\epsilon_k \lambda e^{-\beta \epsilon_k}}{1 \pm \lambda e^{-\beta \epsilon_k}} \right] \quad (15.1.4)
\]

and

\[
\beta \langle p \rangle V = \pm \sum_k \ln \left[ 1 \pm \lambda e^{-\beta \epsilon_k} \right] \quad (15.1.5)
\]

Here \( \lambda = \exp (\mu/kT) \).

What we’d like to do is find expressions for the various thermodynamic quantities that are free of \( \lambda \). For example we’d like an expression for \( p \) in terms of \( V \) and \( N \).

The general scheme is to manipulate the equations above into some form and then eliminate \( \lambda \) by solving one of the above equations for it and then using that solution to remove it from another.
15.2 Weakly Degenerate Ideal Fermi-Dirac Gas

The needed formulas from Section 15.1 are:

\[
\langle N \rangle = \sum_k \left[ \frac{\lambda e^{-\beta \epsilon_k}}{1 + \lambda e^{-\beta \epsilon_k}} \right] \quad (15.2.1)
\]

and

\[
\beta \langle p \rangle V = \sum_k \ln \left[ 1 + \lambda e^{-\beta \epsilon_k} \right] \quad (15.2.2)
\]

with

\[
\lambda = e^{\mu/kT} \quad (15.2.3)
\]

where \( \lambda \) is the absolute activity.

None of the sums in the first two equations can be done exactly. We can, however, convert them to integrals over the allowed energy levels by introducing the degeneracy of states \( \omega(\epsilon) \) discussed in Section 8.3 of Chapter 8. The result obtained there was:

\[
\omega(\epsilon)d\epsilon = \frac{\pi}{4} \left( \frac{8mL^2}{\hbar^2} \right)^{3/2} \epsilon^{1/2}d\epsilon \quad ((8.3.6))
\]

Using this, Equations (15.2.1) and (15.2.2) can be recast to:

\[
N = 2\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^\infty \lambda \epsilon^{1/2}e^{-\beta \epsilon} \left[ 1 - \lambda e^{-\beta \epsilon} + \lambda^2 e^{-2\beta \epsilon} - \lambda^3 e^{-3\beta \epsilon} + \ldots \right] d\epsilon \quad (15.2.4)
\]

and

\[
\beta pV = 2\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^\infty \epsilon^{1/2} \ln(1 + \lambda e^{-\beta \epsilon})d\epsilon \quad (15.2.5)
\]

where the \( \langle \rangle \) brackets have been removed for convenience.

Now these equations can’t be integrated either.\footnote{1}{Then why did I bother to change from a sum to an integral? That’s a good question isn’t it?}

What can be done with these equations is that they can be expanded in a power series and then integrated term by term.\footnote{2}{And that’s a good answer, don’t you think?} There are a number of assumptions involved in this and we shall have to be a bit careful in doing this.

First we handle the expansion of Equation (15.2.4): We have

\[
\frac{1}{1+y} = 1 - y + y^2 - y^3 + \ldots \quad |y| < 1 \quad (15.2.6)
\]

so that with the substitution

\[
y = \lambda e^{-\beta \epsilon} \quad (15.2.7)
\]

the \( N \) equation becomes

\[
N = 2\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^\infty \epsilon^{1/2}e^{-\beta \epsilon} \left[ 1 - \lambda e^{-\beta \epsilon} + \lambda^2 e^{-2\beta \epsilon} - \lambda^3 e^{-3\beta \epsilon} + \ldots \right] d\epsilon \quad (15.2.8)
\]
An examination of Equations (15.2.1) on page 15-2 and (15.2.2) on page 15-2 shows that \( \lambda \) can take on any value in \( 0 \leq \lambda \leq \infty \) since \( \epsilon \) is always positive (or, in the case of the ground state, zero). Thus for any fixed \( \lambda \), \( e^{-\beta\epsilon} \) will eventually dominate and the sums will converge.

However, the expansion in Equation (15.2.6) on page 15-2 holds only for \( |y| < 1 \), unlike that for an exponential which is valid for any value of its argument. So Equation (15.2.7) on page 15-2 is limited to small values of \( \lambda \). In particular, as we can see from the substitution (15.2.7) on page 15-2,

\[
|\lambda| < 1 \quad \text{(weakly degenerate Fermi gas)} \tag{15.2.9}
\]

since then the maximum value of \( \lambda e^{-\beta\epsilon} \) is then 1.

There is another consideration. We are taking an integral, Equation (15.2.4) on page 15-2, expanding it in a series, Equation (15.2.8) on page 15-2 and then we will integrate it term by term. This does not always work unless the series is absolutely convergent. In this case, being an alternating series, it is absolutely convergent. So we are in good shape.

With the constants segregated out, Equation (15.2.8) on page 15-2 becomes

\[
N = 2\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \sum_{n=1}^{\infty} (-1)^{n+1} I_n, \quad n = 1, 2, \ldots \infty \tag{15.2.10}
\]

where

\[
I_n = \left( \frac{1}{n\beta} \right)^{3/2} \lambda^n \int_0^{\infty} x^{1/2} e^{-x} dx \tag{15.2.11}
\]

Here we’ve substituted \( x \) for \( n\beta\epsilon \). The integral is just a gamma function. See Section 5.8 starting on page 5-11. The result is:

\[
I_n = \left( \frac{1}{\beta} \right)^{3/2} \left( \frac{\lambda^n}{n^{3/2}} \right) \Gamma(3/2) \tag{15.2.12}
\]

Making use of the fact that \( \Gamma(3/2) = \pi^{1/2}/2 \) and that the thermal wavelength \( \Lambda \) defined by Equation (8.4.1) on page 8-7 is \( (h^2\beta/2\pi m)^{1/2} \), we get our final and surprisingly simple-looking result:

\[
N = \frac{V}{\Lambda^3} \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\Lambda^n}{n^{3/2}} \tag{15.2.13}
\]

Now the expansion of Equation (15.2.5) on page 15-2. This turns out to be surprisingly similar to the \( N \) expansion we’ve just done in spite of the fact that the functions look so different.

Here we need to expand the logarithm. The relevant series is:

\[
\ln(1 + y) = y - \frac{1}{2} y^2 + \frac{1}{3} y^3 - \ldots \quad |y| < 1 \tag{15.2.14}
\]

\(^3\)Consult any advanced calculus text for proof. The series being integrated must be uniformly convergent. See, for example, D. V. Widder, *Advanced Calculus*, 1947, Chapter IX.

\(^4\)So why did I bring it up? Education. I brought it up to educate!
We expand the logarithm in Equation (15.2.5) on page 15-2 to yield:
\[
\ln(1 + \lambda e^{-\beta \epsilon}) = \lambda e^{-\beta \epsilon} - \frac{1}{2} \lambda^2 e^{-2\beta \epsilon} + \frac{1}{3} \lambda^3 e^{-3\beta \epsilon} - \ldots
\]  
which then gives us
\[
\beta p V = 2\pi \left(\frac{2m}{\hbar^2}\right)^{3/2} V \int_0^\infty e^{1/2} \left[\lambda e^{-\beta \epsilon} - \frac{1}{2} \lambda^2 e^{-2\beta \epsilon} + \frac{1}{3} \lambda^3 e^{-3\beta \epsilon} + \ldots\right] d\epsilon
\]  
we rewrite this:
\[
\beta p V = 2\pi \left(\frac{2m}{\hbar^2}\right)^{3/2} V \sum_{n=1}^\infty (-1)^{n+1} J_n \quad n = 1, 2, \ldots, \infty
\]  
where
\[
J_n = \frac{1}{\beta^{3/2} n^{5/2}} \int_0^\infty x^{1/2} e^{-x} dx
\]  
where \(x = n/\beta \epsilon\). The integral is once again a gamma function and
\[
J_n = \frac{1}{\beta^{3/2} n^{5/2}} \Gamma(3/2)
\]  
Plugging this into Equation (15.2.17) along with the value for \(\Gamma(3/2)\) gives the final result:
\[
\beta p V = \frac{V}{\Lambda^3} \sum_{n=1}^\infty (-1)^{n+1} \frac{\lambda^n}{n^{5/2}}
\]
The plan now is to take Equation (15.2.13) on page 15-3 and revert it. That is, we will solve it for \(\lambda\) as a power series in \(N\). Reversion of series is discussed in some detail in Appendix 15.7 on page 15-21.
Then that will be used to eliminate \(\lambda\) from Equation (15.2.20).
The actual reversion operation is tedious. The result is:
\[
\lambda = a_1 \frac{\Lambda^3}{V} N + a_2 \left(\frac{\Lambda^3}{V}\right)^2 N^2 + \cdots + a_n \left(\frac{\Lambda^3}{V}\right)^n N^n + \cdots
\]
and the first seven coefficients \(a_1, \ldots, a_7\) are given below to 10 significant figures:

<table>
<thead>
<tr>
<th>(a_n)</th>
<th>(\times 10^m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_1)</td>
<td>1.</td>
</tr>
<tr>
<td>(a_2)</td>
<td>3.535533906</td>
</tr>
<tr>
<td>(a_3)</td>
<td>5.754991027</td>
</tr>
<tr>
<td>(a_4)</td>
<td>5.763960401</td>
</tr>
<tr>
<td>(a_5)</td>
<td>4.019494152</td>
</tr>
<tr>
<td>(a_6)</td>
<td>2.098189887</td>
</tr>
<tr>
<td>(a_7)</td>
<td>8.063131085</td>
</tr>
</tbody>
</table>

\(^5\)the computer program Derive was used to generate these results.
It can be seen that the coefficients get rapidly smaller, falling off eventually by a factor of about ten for each coefficient.

The coefficients above were obtained by evaluating the exact but highly messy expressions:

\[ a_1 = 1 \] (15.2.22)
\[ a_2 = \frac{1}{2^{3/2}} \] (15.2.23)
\[ a_3 = \frac{1}{4} - \frac{1}{3^{3/2}} \] (15.2.24)
\[ a_4 = \frac{1}{8} + \frac{5\sqrt{2}}{32} - \frac{5\sqrt{6}}{36} \] (15.2.25)
\[ a_5 = -\frac{\sqrt{5}}{25} - \frac{7\sqrt{3}}{24} + \frac{3\sqrt{2}}{16} + \frac{95}{288} \] (15.2.26)
\[ a_6 = -\frac{7\sqrt{10}}{100} - \frac{19\sqrt{6}}{72} - \frac{7\sqrt{3}}{72} + \frac{1463\sqrt{2}}{3456} + \frac{7}{16} \] (15.2.27)
\[ a_7 = -\frac{\sqrt{7}}{49} + \frac{8\sqrt{15}}{225} - \frac{9\sqrt{5}}{50} - \frac{\sqrt{6}}{4} - \frac{1325\sqrt{3}}{2592} + \frac{15\sqrt{2}}{32} + \frac{443}{384} \] (15.2.28)

We now plug Equation (15.2.21) on page 15-4 into Equation (15.2.20) on page 15-4. The result is:

\[ \beta pV = b_1 N + b_2 \left( \frac{A^3}{V} \right) N^2 + \ldots + b_n \left( \frac{A^3}{V} \right)^{n-1} N^n + \ldots \] (15.2.29)

The first seven coefficients are:

\[ b_1 = 1 \] (15.2.30)
\[ b_2 = \sqrt{2} \] (15.2.31)
\[ b_3 = \frac{1}{8} - \frac{2\sqrt{3}}{27} \] (15.2.32)
\[ b_4 = -\frac{\sqrt{6}}{12} + \frac{5\sqrt{2}}{64} + \frac{3}{32} \] (15.2.33)
\[ b_5 = -\frac{4\sqrt{5}}{125} - \frac{\sqrt{3}}{6} + \frac{\sqrt{2}}{8} + \frac{317}{1728} \] (15.2.34)
\[ b_6 = -\frac{\sqrt{10}}{20} - \frac{5\sqrt{6}}{36} - \frac{5\sqrt{3}}{72} + \frac{1687\sqrt{2}}{6912} + \frac{35}{128} \] (15.2.35)
\[ b_7 = -\frac{6\sqrt{7}}{343} + \frac{2\sqrt{15}}{75} - \frac{3\sqrt{5}}{25} - \frac{\sqrt{6}}{6} - \frac{1019\sqrt{3}}{3888} + \frac{9\sqrt{2}}{32} + \frac{173}{256} \] (15.2.36)

\[ ^{6}\text{Again, this is difficult but basically trivial algebra.} \]
It should be again noted that these results are exact. The coefficients are, rounded to 10 significant figures:

\[
\begin{array}{l}
  b_1 = 1.0 \\
  b_2 = 1.767766953 \times 10^{-1} \\
  b_3 = -3.30059820 \times 10^{-3} \\
  b_4 = 1.112893285 \times 10^{-4} \\
  b_5 = -3.540504095 \times 10^{-6} \\
  b_6 = 8.386347040 \times 10^{-8} \\
  b_7 = -3.662061883 \times 10^{-10}
\end{array}
\]

The numbers in the displays above should be accurate to within one digit in the last place given. They were computed to 15 digits and rounded from there. However, extreme care is needed in computing these numbers as they are the result of huge cancellations between positive and negative terms, most much larger than the result. For example, \( b_3 \) is \( 1/8 - 2\sqrt{3}/27 \). The first term is 0.125000, the second \(-0.128300\) (both to six decimal places. The result is 0.003300, to six places and it can be seen that the result has lost two significant figures. The situation is even worse in the computation of \( b_7 \) where the individual terms are of the order of 0.1 and the result of the order of \( 10^{-10} \). Clearly nine significant digits have been lost in that case.

Thus the numbers in the display above should be used with great caution. And the reason for inclusion of the exact results is now clear.\(^7\)

If we introduce the number density \( \rho = N/V \) Equation (15.2.29) on page 15-5 can be written as a virial equation of the form:

\[
\frac{p}{kT} = \rho + B_2(T)\rho^2 + B_3(T)\rho^3 + \cdots
\]

which here is:

\[
\frac{p}{kT} = \rho + \frac{\Lambda^3}{2\sqrt{2}}\rho^2 + \left(1 - \frac{2}{3\sqrt{2}}\right)\Lambda^6\rho^3 + \cdots
\]

from which it can be seen that the second virial coefficient of a Fermi-Dirac gas is always positive. The remainder of the series converges (for appropriately small values of \( \lambda \)) and is an alternating series. Thus the error is less than the first neglected term, which is less than the terms included. Thus the pressure in a weakly degenerate Fermi-Dirac gas is always higher than in the corresponding classical ideal gas.

This can be understood from the Pauli Exclusion principle. There are, of course, many more quantum states than particles. But still, the fact that no two particles can be in the same state forces some to be in states higher than they otherwise would – increasing the energy and hence the pressure.

All thermodynamic quantities can be obtained by differentiation of the grand canonical partition function \( \Xi(\beta, V, \mu) \), which is:

\[
\beta pV = \ln \Xi(\beta, V, \mu)
\]

\(^7\)At some point in the future I will reevaluate the numerical values of the \( b \)'s to ensure greater precision.
We already know $\beta pV$ from Equation (15.2.20) on page 15-4. Equation (15.2.29) on page 15-5 cannot be used to determine thermodynamic properties since the variables there are $\beta, V,$ and $N$, which are not the ‘natural’ variables for the grand partition function. So we must use Equation (15.2.20) on page 15-4 instead.

This is reasonable since in determining the other thermodynamic functions it is assumed that we know $\beta, V,$ and $\mu = (1/\beta) \ln \lambda$.

We can, however, obtain the energy $E$ from Equation (15.2.4) on page 15-2. The derivation is much like what we have done above. The energy is given by:

$$E = \sum_k \frac{\epsilon_k \lambda e^{-\beta \epsilon_k}}{1 + \lambda e^{-\beta \epsilon_k}}$$  \hspace{1cm} (15.2.40)

Using the degeneracy $\omega(\epsilon)$ from Equation (8.3.6) on page 8-6 and expanding the denominator as in Equation (15.2.8) on page 15-2 we get:

$$E = 2\pi \left(\frac{2m}{\hbar^2}\right)^{3/2} V \int_0^\infty \lambda e^{3/2/\beta} \left[1 - \lambda e^{-\beta \epsilon} + \lambda^2 e^{-2\beta \epsilon} - \lambda^3 e^{-3\beta \epsilon} + \ldots\right] d\epsilon$$  \hspace{1cm} (15.2.41)

Note the presence of the factor $\epsilon^{3/2}$ instead of $\epsilon^{1/2}$ as in Equation (15.2.8) on page 15-2.

As before we write this as:

$$E = 2\pi \left(\frac{2m}{\hbar^2}\right)^{3/2} V \sum_{n=0}^\infty (-1)^{n+1} I'_n$$  \hspace{1cm} (15.2.42)

where

$$I'_n = \lambda^n \int_0^\infty e^{3/2/\beta} e^{-n\beta \epsilon} d\epsilon$$  \hspace{1cm} (15.2.43)

Making the substitution $x = n\beta \epsilon$, we get

$$I'_n = \left(\frac{1}{n\beta}\right)^{5/3} \lambda^n \int_0^\infty x^{3/2} e^{-x} dx = \left(\frac{1}{n\beta}\right)^{5/3} \lambda^n \Gamma(5/2)$$  \hspace{1cm} (15.2.44)

Since $\Gamma(5/2) = 3\sqrt{\pi}/4$, when we plug this into Equation (15.2.42) we get:

$$E = 2\pi \left(\frac{2m}{\hbar^2}\right)^{3/2} \frac{3\sqrt{\pi}}{4} V \sum_{n=1}^\infty (-1)^{n+1} \frac{\lambda^n}{n^{5/2}}$$  \hspace{1cm} (15.2.45)

Putting the bits together and introducing $\Lambda$ once again, we get

$$E = \frac{3}{2\beta} \frac{V}{\Lambda^3} \sum_{n=1}^\infty (-1)^{n+1} \frac{\lambda^n}{n^{5/2}}$$  \hspace{1cm} (15.2.46)

We could now go eliminate $\lambda$ using Equation (15.2.21) on page 15-4, but we don’t have to! If we compare Equation (15.2.46) to Equation (15.2.20) on page 15-4 we see immediately that:

$$E = \frac{3}{2} pV$$  \hspace{1cm} (15.2.47)

---

8 See Equation (8.4.1) on page 8-7 for the definition of $\Lambda$. 

so that once $p$ is known, so also $E$ is known. Further, remember that $p$ is always greater than the Boltzmann ideal gas $p$, so the energy of the weakly degenerate Fermi-Dirac gas is also always larger than that of the classical ideal gas.

## 15.3 Strongly Degenerate Ideal Fermi-Dirac Gas

### 15.3.1 Absolute Zero

The strongly degenerate Fermi-Dirac gas isn’t ideal at all. Here strong degeneracy means conditions under which quantum phenomenon are important, which in turn means low temperatures (for suitable values of ‘low’.)

We begin with Equation (15.1.3) on page 15-1 for Fermi particles:

$$\langle n_k \rangle = \frac{\lambda e^{-\beta \epsilon_k}}{1 + \lambda e^{-\beta \epsilon_k}} \quad ((15.1.3))$$

and write $\lambda$ as $e^{\beta \mu}$, where $\mu$ is the chemical potential.\textsuperscript{9} Then Equation (15.1.3) on page 15-1 becomes

$$\langle n_k \rangle = \frac{1}{1 + e^{\beta (\epsilon_k - \mu)}} \quad (15.3.1)$$

Since there can be only a single Fermi-Dirac particle in a given energy state,\textsuperscript{10} $\langle n_k \rangle$ is, in effect, the probability that there is a particle in state $k$.

At high temperatures this goes over to the weakly degenerate Fermi-Dirac gas discussed in Section 15.2. But as $T \to 0$ something strange happens. First, note that $\mu$ itself is a function of temperature. We will denote $\mu$ at absolute zero as $\mu_0$.

Second, the sign of the exponent on the exponential in the denominator at absolute zero depends on whether $\epsilon_k > \mu_0$ or not. If it is, then the exponent is a positive number that goes to infinity as $T \to 0$. Otherwise it goes to zero as $T \to 0$.

As a result, at absolute zero $\langle n_k \rangle$ obeys

$$\langle n_k \rangle = \begin{cases} 1 & \text{if } \epsilon_k < \mu_0 \\ 0 & \text{otherwise} \end{cases} \quad (15.3.2)$$

and we see that all levels up to $\mu_0$ in energy are occupied and all levels above $\mu_0$ are empty. We are used to exactly this behavior for electrons in atoms and molecules as well as electrons in metals, even at normal temperatures. This implies that, for example, 300 K, is “cold” to an electron gas.

Using the degeneracy $\omega(\epsilon)$ from Equation (8.3.6) on page 8-6

$$\omega(\epsilon) d\epsilon = \frac{\pi}{4} \left( \frac{8mL^2}{\hbar^2} \right)^{3/2} \epsilon^{1/2} d\epsilon \quad ((8.3.6))$$

\textsuperscript{9}This is, of course, the definition of $\lambda$.

\textsuperscript{10}Actually, there can be \textit{two} particles in an energy state if we consider spin, which we must. This will be taken care of later.
allows us to write, at absolute zero:

\[ N = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^{\mu_0} e^{1/2} d\epsilon \]  
(15.3.3)

where a factor of two has been added to take spin into account. The integral is trivial and the result is:

\[ N = \frac{8\pi}{3} \left( \frac{2m}{\hbar^2} \right)^{3/2} V \mu_0^{3/2} \]  
(15.3.4)

which can be rearranged to give:

\[ \mu_0 = \frac{\hbar^2}{2m} \left( \frac{3}{8\pi} \right)^{2/3} \left( \frac{N}{V} \right)^{2/3} \]  
(15.3.5)

where \((N/V)\) is the number density of the material.

In the study of metals where to a first approximation the electrons are treated as uncharged, \(\mu_0\) is called the Fermi energy and \(\mu_0/k\) the Fermi temperature. The Fermi temperature denotes the temperature below which the system is effectively at absolute zero. For most metals the Fermi temperature \(T_F\) is a few thousand degrees.

At absolute zero the energy can be written as:

\[ E_0 = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^{\mu_0} \epsilon^{3/2} d\epsilon = \frac{3}{5} N \mu_0 \]  
(15.3.6)

This was written as \(E_0\) to emphasize that this is a result strictly true only at absolute zero. It represents a zero point energy just like the energy \(h\nu/2\) for a quantum oscillator. It is the energy an electron gas would retain even if cooled to absolute zero.

The pressure of a Fermi-Dirac gas at absolute zero can be expected to be non-zero because there is a zero point energy and hence a zero point pressure.

We had previously obtained Equation (15.2.5) on page 15-2 reproduced here with the additional spin factor of two:

\[ \beta\langle p \rangle V = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^{\mu_0} \epsilon^{1/2} \ln(1 + \lambda e^{-\beta\epsilon}) d\epsilon \]  
(15.3.7)

which is

\[ \beta p V = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^{\mu_0} \epsilon^{1/2} \ln(1 + e^{\beta(\mu_0 - \epsilon)}) d\epsilon \]  
(15.3.8)

In the range of the integral \((\mu_0 - \epsilon)\) is positive since \(\mu_0 > \epsilon\). And since \(\beta\) is large, the exponential is much greater than 1. Thus the 1 can be dropped and the logarithm taken to give:

\[ \beta p V = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^{\mu_0} \epsilon^{1/2}(\mu_0 - \epsilon) d\epsilon \]  
(15.3.9)

\(^{11}\)Because there is a positive charge on a metal atom for each negative charge on the electron. Thus while the electrons move in a periodic rather than zero potential, the movement is still free enough to be considered “ideal”.

---

11 Because there is a positive charge on a metal atom for each negative charge on the electron. Thus while the electrons move in a periodic rather than zero potential, the movement is still free enough to be considered “ideal”.
The integral is simple. The result is:

\[ p_0 = \frac{2}{5} \left( \frac{N}{V} \right)^{1/3} \frac{\hbar^2}{2m} \left( \frac{3}{8\pi} \right)^{2/3} = \frac{2}{5} \frac{N\mu_0}{V} \]  

(15.3.10)

which is usually of the order of about a million atmospheres...

15.3.2 The Merely Cold Ideal Fermi-Dirac Gas

The development here follows that of McQuarrie. The reader should also compare Huang.\(^\text{12}\) We will take the zero temperature results and expand them in a series involving a small parameter.

First we reproduce the equations for \(N\) and \(E\):

\[ N = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^{\mu_0} \epsilon^{1/2} d\epsilon \]  

(15.3.3)

and

\[ E_0 = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^{\mu_0} \epsilon^{3/2} d\epsilon \]  

(15.3.6)

We obtained these by assuming that the temperature was absolute zero. That worked because the probability of occupation of a state, \(n_k\) was basically a step function, one that is one from 0 to some point, say \(\mu_0\) and then zero above that. Call that function \(f(\epsilon)\) where

\[ f(\epsilon) = \langle n_k \rangle = \frac{1}{1 + e^{\beta(\epsilon - \mu)}} \]  

(15.3.1)

which is still basically Equation (15.3.1) on page 15-8.

With this in mind we can rewrite Equations (15.3.3) on page 15-9 and (15.3.6) on page 15-9 as:

\[ N = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^{\infty} f(\epsilon)\epsilon^{1/2} d\epsilon \]  

(15.3.11)

and

\[ E = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^{\infty} f(\epsilon)\epsilon^{3/2} d\epsilon \]  

(15.3.12)

Both of these are of the form:

\[ I = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^{\infty} f(\epsilon)h(\epsilon) d\epsilon \]  

(15.3.13)

where \(f(\epsilon)\) is given by Equation (15.3.1) on page 15-8 and \(h(\epsilon)\) is either \(\epsilon^{1/2}\) for \(I = N\) or \(\epsilon^{3/2}\) for \(I = E\).

So far this is nothing unusual, but this is where the trick comes in. We integrate Equation (15.3.13) on page 15-10 by parts to get:

\[
I = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \left[ f(\epsilon) H(\epsilon) \right]_0^\infty - \int_0^\infty f'(\epsilon) H(\epsilon) d\epsilon
\]  

(15.3.14)

where \( H(\epsilon) \) is the integral of \( h(\epsilon) \):

\[
H(\epsilon) = \int_0^\epsilon h(\epsilon) d\epsilon
\]  

(15.3.15)

and \( f'(\epsilon) \) is the derivative of \( f(\epsilon) \). The first term in square brackets in Equation (15.3.14) is zero because \( H(\epsilon) \) is zero when \( \epsilon \) is zero and \( f(\epsilon) \) is zero when \( \epsilon \) is infinity.

The result is then:

\[
I = -4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^\infty f'(\epsilon) H(\epsilon) d\epsilon
\]  

(15.3.16)

What we’ve gained from this can be seen by looking at a graph of \( f(\epsilon) \). At absolute zero it is flat and equal to 1 up to \( \epsilon = \mu_0 \) where it suddenly falls to 0 and stays there. Its derivative is then zero from 0 to \( \mu_0 \) and zero from \( \mu_0 \) to infinity. It is not zero at \( \mu_0 \). There it is a point at infinity.

But when we are not at absolute zero some of the fermions can move to higher energy levels. They can’t move too high because we still assume that the temperature is very low. But the step function behavior of \( f(\epsilon) \) becomes modified. The sharp upper corner at \( \epsilon = \mu_0 \) becomes somewhat rounded and the foot of the graph at that same point also becomes rounded. The derivative is no longer non-zero at a single point, but now looks something like a Gaussian, a sharply peaked function with the peak at \( \epsilon = \mu \). Note that this is \( \mu \) and not \( \mu_0 \) because \( \mu \) is also a function of temperature and will have a slightly different value once \( T \) is no longer zero.

As a result the integrand in Equation (15.3.16) can be seen to be zero everywhere except in the small region around \( \epsilon = \mu \). So we can expand the integrand in an infinite series around that point, which is now what we are going to do.

We expand \( H(\epsilon) \) in a Taylor series around \( \epsilon = \mu \). This gives us:

\[
H(\epsilon) = H(\mu) + \left( \frac{\partial H}{\partial \epsilon} \right) (\epsilon - \mu) + \frac{1}{2} \left( \frac{\partial^2 H}{\partial \epsilon^2} \right) (\epsilon - \mu)^2 + \cdots
\]  

(15.3.17)

Note that the derivatives are evaluated at \( \epsilon = \mu \) and are no longer functions of \( \epsilon \).

The integral in Equation (15.3.16) is now:

\[
I = 4\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \left[ H(\mu) L_0 + \left( \frac{\partial H}{\partial \epsilon} \right) L_1 + \frac{1}{2} \left( \frac{\partial^2 H}{\partial \epsilon^2} \right) L_2 + \cdots \right]
\]  

(15.3.18)

---

13If you consider my taking simple formulas and making them more complex 'usual'.

14This is called a delta function, \( \delta(x) \), which is zero everywhere except at \( x = 0 \) where it is infinite. It also has the property that \( \int_{-\infty}^{\infty} \delta(x) dx = 1 \), a property we will use later.
where
\[ L_n = -\int_0^\infty (\epsilon - \mu)^n f'(\epsilon) d\epsilon \] (15.3.19)

The first term, \( L_0 \) is 1 since \( f'(\epsilon) \) is, as noted above, a delta function. For the others we can set the lower limit in Equation (15.3.19) to \(-\infty\) since the contribution of the part from \(-\infty\) to 0 is negligible. Thus:
\[ L_n = -\int_{-\infty}^\infty (\epsilon - \mu)^n f'(\epsilon) d\epsilon \quad n = 1, 2, \ldots \] (15.3.20)

If we now let \( x = \beta(\epsilon - \mu) \) we have
\[ L_n = \frac{1}{\beta^n} \int_{-\infty}^\infty \frac{x^n e^x}{(1 + e^x)^2} dx \] (15.3.21)

Note that except for the factor of \( x^n \), the integrand is symmetric around \( x = 0 \). Thus for odd \( n \) \( L_n = 0 \) because the contribution from \( x \) negative equals the contribution from \( x \) positive. So \( L_n \) exists only for even \( n \).

These integrals can be represented in simple form. For instance
\[ \int_{-\infty}^\infty \frac{x^2 e^x}{(1 + e^x)^2} = \frac{\pi^2}{3} \] (15.3.22)

Here’s a listing of the first few values of \( L_n \):
\[ L_0 = 1 \] (15.3.23)
\[ L_2 = \frac{\pi^2}{3\beta^2} \] (15.3.24)
\[ L_4 = \frac{7\pi^4}{15\beta^4} \] (15.3.25)
\[ L_6 = \frac{31\pi^6}{21\beta^6} \] (15.3.26)
\[ L_8 = \frac{127\pi^8}{15\beta^8} \] (15.3.27)
\[ L_{10} = \frac{2555\pi^{10}}{33310\beta^{10}} \] (15.3.28)

Now \( H(\epsilon) \) is defined by Equation (15.3.15) on page 15-11. That integral is:
\[ H(\epsilon) = \frac{2}{a + 2} \epsilon^{1+a/2} \] (15.3.29)

where \( a \) is either 1 if we are dealing with equations for \( N \) or 3 if we are dealing with equations for \( E \). The general equation for \( H^{(n)} \), the \( n \)’th derivative of \( H \) is:
\[ H^{(n)}(\epsilon) = \left( \frac{a}{2} \right) \left( \frac{a}{2} - 1 \right) \left( \frac{a}{2} - 2 \right) \cdots \left( \frac{a}{2} - n + 2 \right) \epsilon^{(a/2)-n+1} \quad n \geq 2 \] (15.3.30)
With all these bits and pieces in hand we substitute our results into Equation (15.3.18) on page 15-11, not forgetting that we are evaluating the derivatives $H^{(n)}(\epsilon)$ at $\epsilon = \mu$ to obtain:

$$N = \frac{8\pi}{3} \left( \frac{2m}{\hbar^2} \right)^{3/2} V \mu^{3/2} \left[ 1 + \frac{\pi^2}{8} (\beta \mu)^{-2} + \cdots \right]$$

(15.3.31)

A similar equation applies to $E$.

Using Equation (15.3.5) on page 15-9 for $\mu_0$ allows us to write:

$$\mu_0 = \mu \left[ 1 + \frac{\pi^2}{8} (\beta \mu)^{-2} + \cdots \right]^{2/3} = \mu \left[ 1 + \frac{\pi^2}{12} (\beta \mu)^{-2} + \cdots \right]$$

(15.3.32)

where the last expression above comes from raising the middle expression to the $2/3$ power. (See Appendix 15.7 for details.)

This is really an expression giving $\mu_0/\mu$ as a function of $(\beta \mu)^{-2}$. If we revert this series we get $\mu/\mu_0$ as a power series in $\eta = 1/\beta \mu$:

$$\frac{\mu}{\mu_0} = 1 - \frac{\pi^2}{12} \eta^2 + \cdots$$

(15.3.33)

Since for many metals $\eta$ is of the order of 0.01 we see that $\mu$ changes very slowly with temperature. As a result little error is made by using $\mu_0$ instead of $\mu$ throughout the range of temperatures at which a metal is solid.

The energy can be found in the same way, except that the parameter $a$ is now $3/2$. The result is:

$$E = E_0 \left[ 1 + \frac{5\pi^2}{12} \eta^2 + \cdots \right]$$

(15.3.34)

The heat capacity of the conduction electrons in a metal is then:

$$C_V = \frac{\pi^2 NkT}{2(\mu_0/k)} = \frac{\pi^2}{2} Nk \left( \frac{T}{T_F} \right)$$

(15.3.35)

where $T_F$ is the Fermi Temperature. Evaluation of this shows that the heat capacity is about $10^{-4}T$ Joules/K for metals.

### 15.4 The Weakly Degenerate Bose-Einstein Gas

The equations involved are once again given in Section 15.1 on page 15-1. Specialized to the case at hand they are:

$$\beta \langle p \rangle V = \ln \Xi = -\sum_k \ln \left[ 1 - \lambda e^{-\beta \epsilon_k} \right]$$

(15.4.1)

$$\langle N \rangle = \sum_k \left[ \frac{\lambda e^{-\beta \epsilon_k}}{1 - \lambda e^{-\beta \epsilon_k}} \right]$$

(15.4.2)

and for the average occupation of a particular state $k$:

$$\langle n_k \rangle = \frac{\lambda e^{-\beta \epsilon_k}}{1 - \lambda e^{-\beta \epsilon_k}}$$

(15.4.3)
The only difference between these equations and those for the Fermi-Dirac weak degeneracy case in Section 15.2 is the occurrence of a minus sign instead of a plus sign in certain places. That will make no difference here, but it will when we discuss the strongly degenerate case.

The reason is that with a minus sign each of the equations above contain the term $1 - \lambda e^{-\beta \epsilon}$ which can become zero under certain circumstances. One does not like taking the logarithm of zero or having a zero in the denominator of an expression.\(^\text{15}\)

Beyond this the development follows exactly the same path as in Section 15.2 except that the series which there all have positive terms, here have terms with alternating signs.

As a result, only the outline of the development will be given.

The degeneracy of states with energy $\epsilon$ is:\(^\text{16}\)

$$\omega(\epsilon) d\epsilon = \frac{\pi}{4} \left( \frac{8 m L^2}{\hbar^2} \right)^{3/2} \epsilon^{1/2} d\epsilon$$

and the equations above can be converted to integrals:

$$N = 2\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^\infty \frac{\lambda \epsilon^{1/2} e^{-\beta \epsilon}}{1 - \lambda e^{-\beta \epsilon}} d\epsilon$$

and

$$\beta p V = -2\pi \left( \frac{2m}{\hbar^2} \right)^{3/2} V \int_0^\infty \epsilon^{1/2} \ln(1 - \lambda e^{-\beta \epsilon}) d\epsilon$$

where the $\langle \rangle$ brackets have been removed for convenience.

These equations are then expanded in a power series, except now we have $1/(1-y) = 1 + y + y^2 + \cdots$ and $\ln(1-y) = y + (1/2)y^2 + (1/3)y^3 + \cdots$ with $y = \lambda \exp(-\beta \epsilon)$. After integration and some substitutions we have:

$$N = \frac{V}{\Lambda^3} \sum_{n=1}^\infty \frac{\lambda^n}{n^{3/2}}$$

$$\beta p V = \frac{V}{\Lambda^3} \sum_{n=1}^\infty \frac{\lambda^n}{n^{5/2}}$$

The series Equation (15.4.7) is now reverted to give:

$$\lambda = a_1 \frac{\Lambda^3}{V} N - a_2 \left( \frac{\Lambda^3}{V} \right)^2 N^2 + \cdots + (-1)^{n+1} a_n \left( \frac{\Lambda^3}{V} \right)^n N^n + \cdots$$

where the $a_n$ have the values given for Equation (15.2.21) on page 15-4.

Equation (15.4.9) is then plugged into Equation (15.4.8) giving:

$$\beta p V = b_1 N - b_2 \left( \frac{\Lambda^3}{V} \right)^2 N^2 + \cdots + (-1)^{n+1} b_n \left( \frac{\Lambda^3}{V} \right)^{n-1} N^n + \cdots$$

\(^\text{15}\)Or perhaps one does, if you like to see things blow up in ugly ways.

\(^\text{16}\)I will not repeat ‘as before’ each time. Just assume that the phrase is there.
where the coefficients $b_n$ have the values given for Equation (15.2.29) on page 15-5.

If we now introduce the number density $\rho = N/V$ and write Equation (15.4.10) on page 15-14 as a virial equation:

$$\frac{p}{kT} = \rho + B_2(T)\rho^2 + B_3(T)\rho^3 + \cdots \quad (15.4.11)$$

we then have:

$$\frac{p}{kT} = \rho - \frac{\Lambda^3}{2^{3/2}}\rho^2 + \left(\frac{1}{8} - \frac{2}{3^{5/2}}\right)\Lambda^6\rho^3 + \cdots \quad (15.4.12)$$

from which it can be seen that an ideal Bose-Einstein gas, below the region in which it acts like a ‘regular’ ideal gas, has a pressure that is less than the ideal gas. This is the opposite to the Fermi-Dirac case.

It is interesting that if you average the Fermi-Dirac case, Equation (15.2.38) on page 15-6 and the Bose-Einstein case, Equation (15.4.12) what you get is:

$$\frac{p}{kT} = \rho + \left(\frac{1}{8} - \frac{2}{3^{5/2}}\right)\Lambda^6\rho^3 + \cdots \quad (15.4.13)$$

where all the even numbered virial coefficients cancel, but the odd ones do not.

The other thermodynamic properties, as in the Fermi-Dirac case, follow in a similar way.

### 15.5 The Strongly Degenerate Bose-Einstein Gas

It is here that the problem with the first term in Equations (15.4.1) on page 15-13 – (15.4.3) on page 15-13 occur. If we assume that the ground state energy $\epsilon_0$ is zero, then $\exp(-\beta\epsilon_0)$ is 1 and $\lambda \exp(-\beta\epsilon_0) = \lambda$. Thus we have, for the first term in the sums, either $\ln(1 - \lambda)$ in Equation (15.4.1) on page 15-13 or $\lambda/(1 - \lambda)$ in Equation (15.4.2) on page 15-13. And it is clear that a problem can occur as $\lambda \to 1$, since the first term will no longer be small.

This is best seen in Equation (15.4.3) on page 15-13 where if $\lambda \to 1$, $\langle n_0 \rangle$ goes to infinity! In the weakly degenerate cases (and the Boltzmann case) we expect $\langle n_k \rangle$ to be less than 1 always, this tells us that something strange is going on.

We deal with this by splitting off the first term in Equations (15.4.7) on page 15-14 and (15.4.8) on page 15-14 and rewriting the series slightly:

$$\rho = \frac{N}{V} = \frac{1}{\Lambda^3} \sum_{n=1}^{\infty} \frac{\lambda^n}{n^{3/2}} + \frac{\lambda}{V(1 - \lambda)} \quad (15.5.1)$$

$$\frac{p}{kT} = \frac{1}{\Lambda^3} \sum_{n=1}^{\infty} \frac{\lambda^n}{n^{3/2}} - \frac{1}{V} \ln(1 - \lambda) \quad (15.5.2)$$

The average number of particles in the ground state is

$$\langle n_0 \rangle = \frac{\lambda}{1 - \lambda} \quad (15.5.3)$$
so to have sensible results we must have $0 \leq \lambda < 1$.

Since we already know that we can have more than one particle in an energy level if the particles are bosons, we can estimate the worst case by assuming that all the particles are in the ground state $\langle n_0 \rangle$. Then $\langle n_0 \rangle = N$ and

$$N = \frac{\lambda}{1 - \lambda} \quad (15.5.4)$$

and so

$$\lambda = \frac{N}{N + 1} = \frac{1}{1 + 1/N} \approx 1 - \frac{1}{N} \quad (15.5.5)$$

Thus, with $N$ of the order of $10^{23}$, the maximum value for $\lambda$ is something like $1 - 10^{-23}$.

The sum in Equations (15.5.1) on page 15-15 and (15.5.2) on page 15-15 are of some interest. They are examples of a polylogarithm. There is probably more than you want to know about them in Section 15.8 on page 15-25. We will refer to them using a sort of obvious notation:

$$g_s(\lambda) = \sum_{n=1}^{\infty} \frac{\lambda^n}{n^s} \quad (15.5.6)$$

where here $s = 3/2$ or 5/2.

As the definition of $g_{3/2}(\lambda)$ shows, for very small values of $\lambda$ $g_{3/2}(\lambda)$ is linear in $\lambda$ with a slope of 1. As $\lambda$ increases, the function begins to increase in value faster than linear and requires that more and more terms be kept to estimate it to any desired accuracy.

When $\lambda = 1$ the function is given by:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \quad (15.5.7)$$

which is called the zeta function or, sometimes, the Riemannian zeta function. This is also discussed in some detail (the function is important and comes up in various places) in Section 15.8 on page 15-25.

What is important to us here is that at $\lambda = 1$ we have $\zeta(3/2) = 2.6123753507 \cdots$, so that this is the value taken by $g_{3/2}(1)$.

In principle, to find the equation of state, we first solve Equation (15.5.1) on page 15-15 for $\lambda$ as a function of $\rho$ and then substitute this into Equation (15.5.2) on page 15-15 to eliminate $\lambda$.

Sadly, this can’t be done analytically and must be approached graphically, numerically, or by more subtle arguments.

We choose the latter approach. We rewrite Equation (15.5.1) on page 15-15 slightly:

$$\rho \Lambda^3 = g_{3/2}(\lambda) + \left( \frac{\Lambda^3}{V} \right) \frac{\lambda}{1 - \lambda} \quad (15.5.8)$$
and consider the magnitude of the terms involved. For helium-4 at 4.2 K, \( \Lambda = 4.3 \times 10^{-10} \) m. Then \( \Lambda^3 = 7.7 \times 10^{-29} \) m\(^3\). The volume will be of the order, at worst, of a milliliter or so. Thus \( V \approx 10^{-6} \) m\(^3\). So \( \Lambda/V \) is then approximately \( 7 \times 10^{-23} \) or smaller.

Knowing that the maximum value of \( g_{3/2}(\lambda) \) is about 2.6, then the first term on the right in Equation (15.5.8) on page 15-16 is about 2.6 while the second is \( 7 \times 10^{-23} \lambda/(1 - \lambda) \).

So most of the time the second term is totally ignorable. When that is true, we have the weakly degenerate case considered in Section 15.4 on page 15-13 above.

We need consider the second term when \( \lambda/(1 - \lambda) \) gets to be roughly \( 10^{20} \) or so.

As far as \( g_{3/2}(\lambda) \) is concerned, this value of \( \lambda \) is the same as \( g_{3/2}(1) \), the difference being in the 20th decimal place.

So we have the interesting situation that, as far as \( g_{3/2}(\lambda) \) is concerned, \( \lambda = 1 \) when anything interesting starts to happen in the ground state.

We now rewrite Equation (15.5.1) on page 15-15 as:

\[
N = \left( \frac{V}{\Lambda^3} \right) g_{3/2}(1) + n_0
\]

where \( n_0 \), the number of bosons in the ground state, has been written for \( \lambda/(1 - \lambda) \).

This can be rearranged to be:

\[
\frac{n_0}{N} = 1 - \frac{g_{3/2}(1)}{\rho \Lambda^3}
\]

We extract the \( T \) dependence of \( \Lambda \):

\[
\Lambda^3 = \left( \frac{\hbar^2}{2\pi mkT} \right)^3 / 2 = \frac{A}{T^{3/2}}
\]

which defines the constant \( A \). Then Equation (15.5.10) can be written:

\[
\frac{n_0}{N} = 1 - \left( \frac{g_{3/2}(1)}{\rho A} \right) T^{3/2}
\]

If we now define \( T_0 \) by:

\[
T^{3/2} = \frac{\rho A}{g_{3/2}(1)} = \left( \frac{\hbar^2}{2\pi mk} \right)^{3/2} \frac{\rho}{g_{3/2}(1)}
\]

or, more simply, \( T_0 \) is the temperature at which \( \rho \Lambda^3 = g_{3/2}(1) \).

The results for the Bose-Einstein ideal gas at fixed density \( \rho \) can then be written:

\[
\frac{n_0}{N} = \begin{cases} 
0 & T > T_0 \\
1 - (T/T_0)^{3/2} & T < T_0
\end{cases}
\]
On the other hand, if the temperature is fixed and $\rho$ is varied, it can be shown by an almost identical argument starting with Equation (15.5.10) on page 15-17 that

$$\frac{n_0}{N} = \begin{cases} 0 & \rho < \rho_0 \\ 1 - \rho_0/\rho & \rho > \rho_0 \end{cases} \quad (15.5.15)$$

We now turn to a slightly rewritten Equation (15.5.2) on page 15-15:

$$\frac{pV}{kT} = \left( \frac{V}{\Lambda^3} \right) g_{5/2}(\lambda) - \ln(1 - \lambda) \quad (15.5.16)$$

Once again we look at the relative sizes of the terms. The left-hand side, $pV/kT$, is, except for $k$, of the order of 1 or, more exactly $nR$, where $R$ is the molar gas constant and $n$ the number of moles. This is very roughly of the order of 1. The presence of Boltzmann’s constant $k$ makes the left-hand side of the order of $10^{23}$.

As discussed above, $\Lambda^3/V$ is of the order of $7 \times 10^{-23}$, so $V/\Lambda^3$ is of the order of $10^{22}$. This is quite compatible with the left-hand side.

What of $\ln(1 - \lambda)$. As discussed above, we do not get into the highly degenerate range until $(1 - \lambda)$ is of the order of $10^{-20}$. However, we are now taking a logarithm. So with that value for $1 - \lambda$, the logarithm is only about 46. This is totally negligible compared to $10^{22}$.

Thus in the case of Equation (15.5.16), the last term can be ignored!

Does this mean that $pV/kT$ is the same as for the weakly degenerate case? No, not at all.

It is true that the weakly degenerate case applies to the point where $\lambda$ is essentially 1. This happens when $T = T_0$. But then $g_{5/2}(\lambda)$ becomes $g_{5/2}(1)$ and is no longer a function of $\lambda$ at all.

So we have:

$$\frac{p}{kT} = \begin{cases} g_{5/2}(\lambda)/\Lambda^3 & T > T_0 \\ g_{5/2}(1)/\Lambda^3 & T < T_0 \end{cases} \quad (15.5.17)$$

The first thing to note is that the volume dependence seems to have vanished. It hasn’t really. It can be seen in Equation (15.2.21) on page 15-4 on page 15-4. So it is contained in $\lambda$ and exists in the case $T > T_0$.

But in the degenerate case, there is no longer any real $\lambda$ dependence. Thus, while the pressure still depends on the temperature (through $\Lambda$), it no longer depends on the volume $V$. Put another way, below $\rho_0$, the pressure remains constant as the volume continues to shrink.\footnote{Indeed, as the volume decreases to zero at constant $N$, all the molecules go into the ground state and take up no volume at all! This amazing result is an artifact of our model where we assumed that the molecules themselves have no volumes.}
15.6 The Photon Gas

Photons are interesting. Among their interesting properties are the following:

1. Photons have one unit of spin and are hence bosons.
2. Photons have a spin degeneracy of two, not three. That is because the vibration represented by a photon have only two vibrational modes, not three. The translational mode does not exist.\(^{18}\)
3. Photons do not interact with each other.\(^{19}\)
4. Photons, unlike “normal” particles, have zero rest mass.
5. The number of photons present in a system is not conserved.
6. Photons of one frequency can be re-emitted at another frequency.

We imagine a gas of photons inside a rigid container with adiabatic walls which are perfect mirrors on the inside. Since photons do not interact with each other, a perfect black body of negligible volume is assumed to be inside the container.\(^{20}\) Thus the total energy contained inside the system is constant.

We imagine an ensemble of such systems, each of fixed \(E\) and \(V\). We cannot specify \(N\) except as an average.

A consideration of the number of standing waves inside a cubical box of length \(L\) shows that the number having energies between \(\epsilon\) and \(\epsilon + d\epsilon\) is:

\[
\omega(\epsilon)d\epsilon = \frac{V\epsilon^2d\epsilon}{\pi^2\epsilon^3c^3} \tag{15.6.1}
\]

The total energy of the system is:

\[
E = \sum_k n_k \epsilon_k \tag{15.6.2}
\]

The partition function \(Q(V,T)\) is then:

\[
Q(V,T) = \sum_{\{n_k\}} e^{-\beta E(\{n_k\})} = \sum_{\{n_k\}} e^{-\beta \sum_k n_k \epsilon_k} \tag{15.6.3}
\]

where \(\{n_k\}\) stands for the set of \(n_k\)’s making up the total energy. There are, of course, many sets that satisfy that constraint. We’ve seen this before in situations where the number of particles was fixed. In addition to the constraint Equation (15.6.2) we had the constraint that

\[
N = \sum_{\{n_k\}} n_k
\]

\(^{18}\)This is a relativistic effect. One can think of a photon as moving so fast that no translational mode can propagate and only the two lateral modes exist.

\(^{19}\)This is true under “normal” conditions. Under conditions of high photon density and with a material object present, non-linear behavior can be found.

\(^{20}\)Alternatively, one can consider the walls to be perfect black bodies.
had to be satisfied. There is no such constraint here. So Equation (15.6.3) on page 15-19 becomes simply:

$$Q(V,T) = \prod_k \left( \sum_{n=0}^{\infty} e^{-\beta n \epsilon_n} \right) \prod_k \frac{1}{1-e^{-\beta \epsilon_k}} \quad (15.6.4)$$

and then

$$\ln Q = -\sum_k \ln(1 - e^{-\beta \epsilon_k}) = -\sum \ln(1 - e^{-\beta \epsilon}) \quad (15.6.5)$$

We can’t evaluate the sum directly, but we can do the equivalent integral with the aid of Equation (15.6.1) on page 15-19. We get:

$$\ln Q = -\frac{V}{\pi^2 c^4 \hbar^4} \int_0^\infty \epsilon^2 \ln(1 - e^{-\beta \epsilon}) d\epsilon \quad (15.6.6)$$

which, like its brother, the ideal Bose-Einstein gas, turns out to involve a zeta function:

$$\ln Q = -\frac{V}{\pi^2 c^4 \hbar^4} \frac{2}{\beta} \sum_n \frac{1}{n^4} = \frac{2V}{\pi^2 (c\hbar)^3} \zeta(4) \quad (15.6.7)$$

And \(\zeta(4) = \frac{\pi^4}{90}\).

We are now basically done. The energy, which is fixed, is:

$$E = \frac{\pi^2 V (kT)^4}{15(\hbar c)^3} \quad (15.6.8)$$

which verifies the well-known energy dependence on the fourth power of the temperature.

The heat capacity of the photon gas is then simply:

$$C_V = \frac{4\pi^2 V k^4 T^3}{15(\hbar c)^3} \quad (15.6.9)$$

The pressure of the photon gas is:

$$p = kT \left( \frac{\partial \ln Q}{\partial V} \right)_T = \frac{2(kT)^4}{\pi^2 (c\hbar)^3} \zeta(4) = \frac{\pi^2 (kT)^4}{45(\hbar c)^3} \quad (15.6.10)$$

which, curiously, is independent of the volume. The pressure is very small, but easily measurable.

The entropy can be computed from the usual formulas, The result is:

$$S = \frac{4\pi^2 V k (kT)^3}{45(\hbar c)^3} \quad (15.6.11)$$

which goes to zero as \(T \to 0\) as it should for a good quantum gas.

Curiously, if we calculate \(G = \langle N \rangle \mu\) from

$$\langle N \rangle \mu = E - TS + pV$$

we find that \(\langle N \rangle \mu = 0\), and, since \(N\) certainly isn’t zero, we then have that:

$$\mu = 0 \quad (15.6.12)$$

which is probably the simplest equation in this entire work!
15.7 Appendix: Operations with Series

Infinite series occur in statistical thermodynamics with some frequency. At times one needs to manipulate such series in various ways. Here are some collected formulas with hints on how they were derived.

15.7.1 Powers of Series

Let us take
\[ y = 1 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots \]  
(15.7.1)
as our 'standard' series. Any power series can be put into this form by simply dividing through by the constant term and renaming the variables.

The square of this series is obtained by simply writing it as
\[ y^2 = (1 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots)(1 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots) \]
and then multiplying to get:
\[ y^2 = 1 + 2a_1 x + [a_1^2 + 2a_2]x^2 + [2a_1 a_2 + 2a_3]x^3 + \ldots \]
so that if we write the result as:
\[ y^2 = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \ldots \]  
(15.7.2)
we have the following results:

<table>
<thead>
<tr>
<th>Coefficient Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_0 ) 1</td>
</tr>
<tr>
<td>( c_1 ) 2a_1</td>
</tr>
<tr>
<td>( c_2 ) ( a_1^2 + 2a_2 )</td>
</tr>
<tr>
<td>( c_3 ) 2a_1 a_2 + 2a_3</td>
</tr>
<tr>
<td>( c_4 ) 2a_1 a_3 + a_2^2 + 2a_4</td>
</tr>
</tbody>
</table>

If we cube Equation (15.7.1) we get:
\[ y^3 = 1 + 3a_1 x + (3a_1^2 + 3a_2)x^2 + (a_1^3 + 6a_1 a_2 + 3a_3)x^3 + (a_1^2 a_2 + 2a_1 a_3 + a_2^2 + a_4)x^4 + \ldots \]
or in the form:
\[ y^3 = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \ldots \]  
(15.7.3)
we have:

\(^{21}\)The material in this section was taken from Abramowitz and Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, 1964, Section 3.6. Later versions of this exceptional valuable book are available from Dover Publications.
In general the \( n \)’th power of Equation (15.7.1) on page 15-21 is

\[
y^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \ldots
\]  

(15.7.4)

and

\[
\begin{array}{|c|c|}
\hline
\text{Coefficient} & \text{Value} \\
\hline
c_0 & 1 \\
c_1 & 3a_1 \\
c_2 & 3a_1^2 + 3a_2 \\
c_3 & a_1^3 + 6a_1a_2 + 3a_3 \\
c_4 & a_1^3a_2 + 2a_1a_3 + a_2^2 + a_4 \\
\hline
\end{array}
\]

Also of use is the product of two infinite series

\[
c_0 + c_1 x + c_2 x^2 + c_2 x^3 + \ldots = (1 + a_1x + a_2x^2 + a_3x^3 + \ldots) \times (1 + b_1x + b_2x^2 + b_3x^3 + \ldots)
\]  

(15.7.5)

produces:

\[
\begin{array}{|c|c|}
\hline
\text{Coefficient} & \text{Value} \\
\hline
c_0 & 1 \\
c_1 & na_1 \\
c_2 & (1/2)n(n-1)a_1 + na_2 \\
c_3 & n(n-1)a_1a_2 + n(n-1)(n-2)a_1^3 + na_3 \\
c_4 & n(n-1)(n-2)(n-3)a_1^3/24 + n(n-1)(n-2)a_1^2a_2/2 \\
& + n(n-1)a_2^2/2 + n(n-1)a_1a_3 + na_4 \\
\hline
\end{array}
\]

which is produced by simply multiplying out the two series in Equation (15.7.5)

The quotient of two series

\[
c_0 + c_1 x + c_2 x^2 + c_2 x^3 + \ldots = \frac{(1 + a_1x + a_2x^2 + a_3x^3 + \ldots)}{(1 + b_1x + b_2x^2 + b_3x^3 + \ldots)}
\]  

(15.7.6)

is obtained by multiplying through by the denominator series of Equation (15.7.6) and then using the product result above. This gives:

\[
\begin{array}{|c|c|}
\hline
\text{Coefficient} & \text{Value} \\
\hline
c_0 & 1 \\
c_1 & b_1 + a_1 \\
c_2 & b_2 + a_1b_1 + a_2 \\
c_3 & b_3 + a_1b_2 + a_2b_1 + a_3 \\
c_4 & b_4 + a_1b_3 + a_2b_2 + a_3b_1 + a_4 \\
\hline
\end{array}
\]
15.7.2 Reversion of Series

At times one needs to revert a series. That is, given the series:
\[ y = ax + bx^2 + cx^3 + dx^4 + ex^5 + fx^6 + gx^7 + \ldots \]  
(15.7.7)

one wants to find how \( x \) depends on \( y \). That is, given Equation (15.7.7), one wants to find
\[ x = Ay + By^2 + Cy^3 + Dy^4 + Ey^5 + Gy^6 + Gy^7 + \ldots \]  
(15.7.8)

where \( A, B, \) etc., are known functions of \( a, b, \) etc. Note the absence of a constant term in both Equations (15.7.7) and (15.7.8).

Surprisingly, this can be done in many cases. Each of the terms \( A, B, \) etc. turn out to be functions of only a finite number of \( a, b, \) etc.

Working this out is a bit complex if more than two or perhaps three terms are wanted. The method is simple enough. One simply takes Equation (15.7.7) and plugs it in for \( y \) wherever \( y \) occurs in Equation (15.7.8). Thus a version of this process would look like:
\[ x = A(ax + bx^2 + cx^3 + \ldots) + B(ax + bx^2 + cx^3 + \ldots)^2 + C(ax + bx^2 + cx^3 + \ldots)^3 + \ldots \]

One then multiplies this out and collects like powers of \( x \).

The result is an identity. The left hand side \( x \) equals some coefficient times \( x \), some other coefficient times \( x^2 \), and so on. The coefficient of \( x \) must be 1 since the right-hand side is 1. The coefficients of all other terms must vanish.

It can easily be seen that the first coefficient is \( aA \), hence \( A = 1/a \). The second coefficient (that of \( x^2 \)) is \((Ab + Ba^2)\), which must be zero, so \( B = -b/a^3 \). To solve this we must already know \( A \), but we do know it from the first step.

At each step along the way it turns out that the new coefficient that we are looking for is given in terms of coefficients already known. The results are lengthy:
\[
A = 1/a \quad (15.7.9)
\]
\[
B = -b/a^3 \quad (15.7.10)
\]
\[
C = (2b^2 - ac)/a^5 \quad (15.7.11)
\]
\[
D = (5abc - a^2d - 5b^3)/a^7 \quad (15.7.12)
\]
\[
E = (6a^2bd + 3a^2c^2 + 14b^4 - a^3e - 21abc^2)/a^9 \quad (15.7.13)
\]
\[
F = (7a^3be + 7a^3cd + 84ab^3c - a^4f - 28a^2bc^3 - 42b^5 - 28a^3b^2d)/a^{11} \quad (15.7.14)
\]
\[
G = (8a^4bf + 8a^4ce + 4a^4d^2 + 120a^2b^3d + 180a^2b^2c^2 + 132b^6 - a^5g - 36a^3b^2e - 72a^3bcd - 12a^3c^3 - 330ab^4c)/a^{13} \quad (15.7.15)
\]
As one can see, these are fairly horrible relationships, getting worse as one goes up in order. For serious work the reader is strongly recommended to gain access software that will do this automatically. The two best are Mathematica and Maple, with Derive a close third for most work.\footnote{The first two of these are inordinately expensive. If one can purchase them through a school program, so much the better. Derive is much cheaper, but not cheap and will do the job.}

For the reader who wishes to check a program to revert series, the series:

\[ y = x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + 6x^6 + 7x^7 + \ldots \]

reverts to:

\[ x = y - 2y^2 + 5y^3 - 14y^4 + 42y^5 - 132y^6 + 429y^7 + \ldots \]
15.8 Appendix: The Zeta Function and Generalizations

15.8.1 The Zeta Function

In 1737, Leonhard Euler wrote a paper in which he discussed several interesting series, one being the sum of the reciprocals of the positive integers, another the sum of the squares of such integers.

However, it was not until Riemann’s paper in 1859 that series of the sort studied by Euler were shown to be part of class of functions we today call the Riemann Zeta function.

Riemann’s zeta function is one of those amazing mathematical entities which are exceptionally useful, beautiful, and slightly mysterious all at once.

The zeta function can be defined in a number of ways, all equivalent, though that may not be obvious at first glance. These include the form we are most interested in, the summation formula:

\[ \zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots + \frac{1}{k^s}, \quad (15.8.1) \]

the integral formula:

\[ \zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{u^{s-1}}{e^u - 1} du, \quad (15.8.2) \]

and the prime product formula

\[ \zeta(s) = \prod_{k=1}^\infty \frac{1}{1 - \frac{1}{p_k^s}}, \quad \text{where } p_k \text{ is the } k\text{’th prime, starting with 2} \quad (15.8.3) \]

Basic properties of the zeta function are perhaps most easily seen from Equation (15.8.1). Clearly it does not converge for \( s = 1 \), since Equation (15.8.1) then becomes the harmonic series:

\[ \zeta(1) = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} + \cdots \quad (15.8.4) \]

which diverges, but does converge for any value of \( s > 1 \). In fact, the zeta function exists for complex values of \( s \) as well, as long as the real part of \( s \) is greater than 1.

Further, examination of the values of \( \zeta(s) \) show that it rapidly approaches 1 from above, being already about 1.08 at \( s = 4 \) and becoming much closer as \( s \) increases.

Values of the zeta function can be computed by various methods. The best method so far developed is by Peter Borwein:

\[ \zeta(s) = -\frac{1}{d_n^s(1 - 2^{1-s})} \sum_{k=0}^{n-1} \frac{(-1)^k (d_k - d_n)}{(k + 1)^s} + \gamma_n(s) \quad (15.8.5) \]

\(^{23}\)A fact proven in almost all calculus books.

\[ d_k = n \sum_{i=0}^{k} \frac{(n+i-1)!4^i}{(n-i)!(2i)!} \] \hspace{1cm} (15.8.5a)

\[ \gamma_n(s) \leq \frac{3}{(3+\sqrt{8})^n} \frac{1}{1-2^{1-s}} \] \hspace{1cm} (15.8.5b)

This is for \( s \) real, which is the case we need. Here \( \gamma_n(s) \) is the error estimate, which implies that for \( p \) decimal digits in the result \( n \) ought to be taken as roughly \( 1.3n \). With \( n \) in hand and \( s \) already known, Equation ((15.8.5)a) is then used to compute \( d_k \) for \( k = 0 \) up to \( k = n \). The resulting values are then used in Equation (15.8.5) on page 15-25 to produce values of \( \zeta(s) \).

Zeta is also tabulated in Abramowitz and Stegun, which, if you have access to this valuable reference, might be the easiest way to obtain various values.

For comparison, some values of \( \zeta \) are given to 10 decimals below.

\[
\begin{align*}
\zeta(1) &= \infty \\
\zeta(2) &= \frac{\pi^2}{6} \\
\zeta(3) &= 1.2020569032 \\
\zeta(4) &= \frac{\pi^4}{90} \\
\zeta(5) &= 1.0369277511 \\
\zeta(6) &= \frac{\pi^6}{945} \\
\zeta(7) &= 1.0083492774 \\
\zeta(8) &= \frac{\pi^8}{9450} \\
\zeta(9) &= 1.0008383928 \\
\zeta(10) &= \frac{\pi^{10}}{93555}
\end{align*}
\]

The even values are “exact” because in these cases the value of \( \zeta(s) \) given can be obtained by doing the appropriate integrals. The other values are approximate values obtained by computations using Equations like (15.8.5) on page 15-25. It is not known if these are are transcendental numbers, except in the case of \( \zeta(3) \), also known as Apéry’s constant, which has been proven to be transcendental.

It may be of use to see how Equations (15.8.1) on page 15-25—(15.8.3) on page 15-25 can be derived from each other. The integrand in Equation (15.8.2) on page 15-25 can be written as:

\[
\frac{u^{s-1}}{e^u - 1} = \frac{e^{-u}u^{s-1}}{1 - e^{-u}} = e^{-u}u^{s-1} \sum_{k=0}^{\infty} e^{-ku} = \sum_{k=1}^{\infty} u^{s-1}e^{-ku}
\] \hspace{1cm} (15.8.6)

where the series expansion for \( e^{-u} \) has been used in the denominator. Using this in Equation (15.8.2) on page 15-25:

\[
\zeta(s) = \frac{1}{\Gamma(s)} \int_{0}^{\infty} \frac{u^{s-1}}{e^u - 1} \, du = \frac{1}{\Gamma(s)} \sum_{k=1}^{\infty} \int_{0}^{\infty} u^{s-1}e^{-ku} \, du
\] \hspace{1cm} (15.8.7)

Letting \( x = ku \) and doing some small manipulation leads to:

\[
\zeta(s) = \frac{1}{\Gamma(s)} \sum_{k=1}^{\infty} \frac{1}{k^s} \int_{0}^{\infty} x^{s-1}e^{-x} \, dx
\] \hspace{1cm} (15.8.8)

\( ^{25} \)Abramowitz and Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, National Bureau of Standards, 1964, Chapter 23.

\( ^{26} \)Given here to ten decimal places.
The integral is the Gamma Function of $s$ and what is left is Equation (15.8.1) on page 15-25. So the secret lay in the series expansion of $e^{-s}$, which, in retrospect, seems quite natural.

The seeming mysterious equivalence of Equation (15.8.1) on page 15-25 and Equation (15.8.3) on page 15-25 turns out to be just as simple. The $p_k$ in Equation (15.8.3) on page 15-25 are the successive prime numbers, the first 10 of which are:

$$2, 3, 5, 7, 9, 11, 13, 17, 19, 23$$

The basic theorem of arithmetic says that every whole number $r$ can be uniquely factored into primes to various powers. Thus, for instance, 12 is $2^2 \cdot 3^1$, meaning that it is the product of two 2’s and one 3.

Equation (15.8.3) on page 15-25 is the product of terms like:

$$\frac{1}{1 - p_k^{-s}}$$

where $p_k$ is the $k$’th prime. The secret, once again, is to expand this as a power series:

$$\frac{1}{1 - p^{-s}} = 1 + \frac{1}{(p^1)^s} + \frac{1}{(p^2)^s} + \frac{1}{(p^3)^s} + \cdots$$  \hspace{1cm} (15.8.9)

and to do this for every prime $p_k$ in Equation (15.8.3) on page 15-25. Thus we have for $k = 2$, $p_k = 2$:

$$\frac{1}{1 - 2^{-s}} = 1 + \frac{1}{(2^1)^s} + \frac{1}{(2^2)^s} + \frac{1}{(2^3)^s} + \cdots$$

So the first few products are:

$$\frac{1}{1 - 2^{-s}} \frac{1}{1 - 3^{-s}} \frac{1}{1 - 5^{-s}} \cdots = 
\left[ 1 + \frac{1}{(2^1)^s} + \frac{1}{(2^2)^s} + \frac{1}{(2^3)^s} + \cdots \right] \times
\left[ 1 + \frac{1}{(3^1)^s} + \frac{1}{(3^2)^s} + \frac{1}{(3^3)^s} + \cdots \right] \times
\left[ 1 + \frac{1}{(5^1)^s} + \frac{1}{(5^2)^s} + \frac{1}{(5^3)^s} + \cdots \right] \times \cdots$$

Now when these are multiplied out, the first term is 1. The second term is $1/2^s$ times a bunch of 1’s. The next is $1/3^s$, again times a bunch of 1’s. The fourth is $1/(2^2)^s$, the fifth $1/5^s$, and so on. What we have then is:

$$\zeta(s) = \prod_{k=1}^{\infty} \frac{1}{1 - p_k^{-s}} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots$$  \hspace{1cm} (15.8.10)

which is what we set out to show.

Truly, the zeta function is interesting!
15.8.2 The Dirichlet Eta Function

The Dirichlet eta function is just like the zeta function except that the terms alternate in signs:

\[ \eta(s) = 1 - \frac{1}{2^s} + \frac{1}{3^s} + \cdots = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^s} \]  

(15.8.11)

The eta function does exist at \( s = 1 \); the alternating series converges rather nicely there. Examination of tabulated values for \( \eta(s) \) show that it rapidly approaches 1 from below, being always less than 1 for all \( s \).

Some values of \( \eta(s) \) are given below:

\[
\begin{align*}
\eta(0) &= \frac{1}{2} \\
\eta(1) &= \ln 2 \\
\eta(2) &= \frac{\pi^2}{12} \\
\eta(3) &= (3/4)\zeta(3) \\
\eta(4) &= 7\pi^4/720 \\
\eta(5) &= (15/16)\zeta(5) \\
\eta(6) &= 31\pi^6/30240 \\
\eta(7) &= (63/64)\zeta(7) \\
\eta(8) &= 127\pi^8/1209600 \\
\eta(9) &= (255/256)\zeta(9) \\
\eta(10) &= 73\pi^{10}/6842880 \\
\eta(11) &= (1023/1024)\zeta(11)
\end{align*}
\]

There is a close connection between \( \eta(s) \) and \( \zeta(s) \). This can be seen by subtracting \( \eta(s) \) from \( \zeta(s) \):

\[
\zeta(s) - \eta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s} - \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^s} = \sum_{k=1}^{\infty} \frac{[1 - (-1)^{k+1}]}{k^s} = \sum_{k=2, 4, \ldots}^{\infty} \frac{2}{k^s} 
\]

(15.8.12)

from which we easily find that

\[
\eta(s) = (1 - 2^{1-s})\zeta(s) 
\]

(15.8.14)

which in turn explains some of the entries in the list of values for \( \eta(s) \) above.

15.8.3 The Polylogarithm

We are interested in a function very similar to the zeta function, Equation (15.8.1) on page 15-25, except this one has an additional parameter. It could be denoted as \( \zeta(s, \lambda) \) but for various reasons the notation \( g_s(\lambda) \) will be used here. The definition is:

\[ g_s(\lambda) = \lambda + \frac{\lambda^2}{2^s} + \frac{\lambda^3}{3^s} + \cdots = \sum_{k=1}^{\infty} \frac{\lambda^k}{k^s} \]  

(15.8.15)
In the literature this function is known as the \textit{polylogarithm}. It is a slight generalization of the zeta function and, indeed, it might be called the (slightly) generalized zeta function. But there are many generalizations of the zeta function so that it is perhaps better to call it by its most known name.\footnote{The polylogarithm is also sometimes called the \textit{de Jonquières function}}. This function is usually denoted by $\text{Li}_s(\lambda)$. We will not use that name here.\footnote{Primarily because we will also have a second closely related function generalized from the eta function which we will call $f_s(\lambda)$.}

It can be trivially seen that for $\lambda = 1$, the polylogarithm becomes the zeta function.

Equation (15.8.15) on page 15-28 converges for all $s > 0, 0 \leq \lambda \leq 1$, except that if $\lambda = 1$, $s$ must be larger than 1.

For any $\lambda > 1$, we have
\[ \lambda^k > k^s \quad (15.8.16) \]
for sufficiently large $k$. This can be seen by taking logarithms:
\[ k \ln \lambda > s \ln k \]
since for fixed $\lambda$ and $s$, $k$ increases faster than $\ln k$. Thus eventually the later terms of Equation (15.8.15) on page 15-28 get larger and larger and the series diverges.

Some values of $g_s(\lambda)$ can be seen by inspection. We have, for example:
\[ g_s(0) = 0 \quad g_\infty(\lambda) = \lambda \quad (15.8.17) \]
Other values can be deduced from Equation (15.8.1) on page 15-25:
\[ g_0(\lambda) = \sum_{k=0}^{\infty} \lambda^k = \frac{\lambda}{1-\lambda} \quad (15.8.18) \]
and
\[ g_1(\lambda) = -\ln(1-z) \quad (15.8.19) \]

The function $g_s(\lambda)$ can be differentiated with respect to $\lambda$:
\[ \frac{d}{d\lambda} g_s(\lambda) = \sum_{k=1}^{\infty} \frac{\lambda^k}{k^s} = \sum_{k=1}^{\infty} \frac{\lambda k^{k-1}}{k^s} = \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{k^{s-1}} = \frac{1}{\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k^{s-1}} = \frac{1}{\lambda} g_{s-1}(\lambda) \quad (15.8.20) \]
from which we have
\[ \lambda \frac{d g_s(\lambda)}{d\lambda} = g_{s-1}(\lambda) \quad (15.8.21) \]

A similar function, $f_s(\lambda)$ can be defined:
\[ f_s(\lambda) = \lambda - \frac{\lambda^2}{2^s} + \frac{\lambda^3}{3^s} + \cdots = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\lambda^k}{k^s} \quad (15.8.22) \]
which is $g_s(\lambda)$ with alternating signs.
Again there are some special values:

\[ f_s(0) = 0 \quad f_\infty(\lambda) = \lambda \quad (15.8.23) \]

and

\[ f_0(\lambda) = \sum_{k=0}^{\infty} = \frac{\lambda}{1 + \lambda} \quad (15.8.24) \]
Chapter 16

Classical Statistical Mechanics

16.1 Liouville’s Equation

We’ve already discussed classical mechanics and introduced both Lagrangian and Hamiltonian formulations of mechanics. And we’ve introduced phase space. We need now to return to those ideas in order to understand the origins of statistical thermodynamics.

Consider a classical system containing \( N \) particles. We’ll let \( q \) stand for the \( 3N \) generalized coordinates and \( p \) for the \( 3N \) generalized momenta.

At any given instant the exact state of this system is given by a single point in a \( 6N \)-dimensional phase space.

The motion of the particles in this system are governed by the \( 6N \) first-order differential equations:

\[
\dot{q}_j = \frac{\partial H}{\partial p_j}, \quad \dot{p}_j = -\frac{\partial H}{\partial q_j} \tag{16.1.1}
\]

so that, in principle if \( 6N \) initial conditions

\[
q = q_o, \quad p = p_o \tag{16.1.2}
\]

are known, the motion of each particle in the system is then known for all time. In phase space the point representing the state of the system moves in time as dictated by Hamilton’s equations, tracing out a trajectory that can never intersect itself. This last is obvious since if a given point is the point of intersection, there can be, by Equations (16.1.1), only one next point. Since a crossing implies that there are two next points, the one originally taken and the one to be taken now, crossings must be impossible.
By the same token, two different trajectories also cannot intersect, since, with the systems identical, such an intersection also implies two next points.

In addition to not crossing itself, the trajectory of the system may or may not be closed. If it is closed, then the points lying on the trajectory will be revisited over and over again. The logic of this also depends on Equation (16.1.1) on page 16-1.

If the trajectory is not closed, as is usually the case, then the system never repeats its state, but moves forever without ever exactly repeating itself. It can be shown that under these conditions the system will return to the neighborhood of a previous point on the trajectory infinitely often. How long this takes depends on the size of the neighborhood.¹

Now let’s consider an ensemble consisting of \( A \) macroscopically identical copies of our original system. These systems are independent and do not interact in any way. Each of course has identical macroscopic values of (say) \( N, V, \) and \( E \). And each will have drastically different sets of microscopic variables. This is equivalent to saying that to each macrostate of a system there correspond a large number of microstates. To make this plausible² note that two macroscopically identical systems of gas molecules having identical Maxwell-Boltzmann velocity distributions will have the macroscopic same properties. Thus we can change the properties of individual molecules how we will as long as we keep the velocity distribution unchanged.

We will represent each of these systems as a point in our \( 6N \)-dimensional phase space. The ensemble will then be a cloud of points in phase space.

A very useful quantity is \( \rho(p, q, t) \), the density of ensemble members in a region of phase space of size \( d\tau = dq_1 \ldots dq_{3N} dp_1 \ldots dp_{3N} \). This quantity will in general vary with position and momentum and perhaps change in time.

Often \( \rho \) is normalized so that it its integral over the whole phase space is the total number of ensemble members \( A \). Thus:

\[
\int_\tau \rho(q, p, t) \, dq \, dp = A \quad (16.1.3)
\]

where \( \tau \) indicates that the integration is to be taken over the entire phase space. Here \( \rho \) is a number density of ensemble members in phase space.

Sometimes \( \rho \) is also normalized to 1. In that case it is the fraction of ensemble members in a given volume element, or, what is the same thing, the probability that a randomly chosen system is in a given volume element. This is really a probability and to avoid confusion with Equation (16.1.3) we will normalize \( P(q, p, t) \) by:

\[
\int_\tau P(q, p, t) \, dq \, dp = 1 \quad (16.1.4)
\]

This is often useful since \( \rho \) depends on \( A \), which is a purely arbitrary (but large) number. The two are simply related: \( \rho = AP \).

¹This idea is embodied in the Poincaré Recurrence Theorem.
²We will demonstrate it later when we talk about quantum systems because the demonstration is trivial there.
We’re interested in the phase space density of states because with it we can calculate the value of any mechanical property.\(^3\) If \(R\) is a mechanical property, then its expected value is:

\[
\langle R \rangle = \frac{\int_{p,q} R(p,q) \rho(q,p,t) \, dq \, dp}{\int_{p,q} \rho(q,p,t) \, dq \, dp}
\]

where clearly here we’ve used \(\rho\) as the density of ensemble members.

Let us now ask how \(\rho\) changes in time as the various trajectories develop in phase space.\(^4\) First, we consider the rate of change of \(\rho\) with time at any fixed point \(p,q\) in phase space. This isn’t really a point, but a very small multidimensional hypercube with one corner at \(p,q\) and a diagonally opposite corner at \(p+d\,p, q+d\,q\). The volume of this hypercube is \(d\tau = dq_1 \cdots dq_N \, dp_1 \cdots dp_N\).

The number of ensemble members inside this volume at any time \(t\) is then given by:

\[
\delta N = \rho dq_1 \cdots dq_N \, dp_1 \cdots dp_N = \rho d\tau
\]

This number will change in time because the number of ensemble members entering the volume element in unit time through one face will not necessarily be the same as the number leaving though the other faces in that time element.

To be specific let us consider the two faces perpendicular to the \(q_1\) axis, one located at \(q_1\), the other at \(q_1 + dq_1\).

Now in time \(dt\), all systems outside the face at \(q_1\) and moving toward it with speed \(\dot{q}_1 = dq_1/dt\) will cross the boundary and enter the volume element. The number of such systems is:

\[
\rho\dot{q}_1 dq_2 \cdots dq_N \, dp_1 \cdots dp_N
\]

The same argument can be applied to the systems leaving the volume element through the face at \(q_1 + dq_1\). At that face \(\rho\) is slightly different, being at a slightly different location. We can develop that difference in a power series using an obvious notation:

\[
\rho(q_1 + dq_1) = \rho(q_1) + \left( \frac{\partial \rho}{\partial q_1} \right) dq_1 + \frac{1}{2!} \left( \frac{\partial^2 \rho}{\partial q_1^2} \right) (dq_1)^2 + \cdots
\]

Because the change from \(q_1\) to \(q_1 + dq_1\) was an infinitesimal one anyway, we can safely neglect terms in \((dq_1)^2\) and higher. Similarly \(\dot{q}_1\) is also slightly different and can be written (in the same obvious notation) as:

\[
\dot{q}_1(q_1 + dq_1) = \dot{q}_1(q_1) + \left( \frac{\partial \dot{q}_1}{\partial q_1} \right) dq_1 + \frac{1}{2!} \left( \frac{\partial^2 \dot{q}_1}{\partial q_1^2} \right) (dq_1)^2 + \cdots
\]

Then the equivalent of Equation (16.1.7) for the face at \(q_1 + dq_1\) is:

\[
\left[ \rho + \left( \frac{\partial \rho}{\partial q_1} \right) dq_1 \right] \left[ \dot{q}_1 + \left( \frac{\partial \dot{q}_1}{\partial q_1} \right) dq_1 \right] dq_2 \cdots dq_N \, dp_1 \cdots dp_N
\]

\(^3\)A mechanical property is one that can be computed from a knowledge of the coordinates and momenta of a system.

\(^4\)Much of the following is taken from Richard C. Tolman, _The Principles of Statistical Mechanics_, Dover Publications, 1979, being a reprint of the 1938 edition published by Oxford.
Let us multiply Equation (16.1.8) on page 16-3 out, dropping terms involving products of differentials. We then get:

\[
\dot{\rho} \frac{\partial q_1}{\partial q_1} + \dot{q}_1 \frac{\partial \rho}{\partial q_1} dq_1 \ldots dq_{3N} dp_1 \ldots dp_{3N} \tag{16.1.11}
\]

The change in number of ensemble members in the small volume due to ingress and egress through these two sides is found by subtracting Equation (16.1.11) from Equation (16.1.7) on page 16-3 to get:

\[
- \left[ \rho \left( \frac{\partial \dot{q}_1}{\partial q_1} \right) + \dot{q}_1 \left( \frac{\partial \rho}{\partial q_1} \right) \right] dq_1 \ldots dq_{3N} dp_1 \ldots dp_{3N} \tag{16.1.12}
\]

where \(dq_1\) has been factored out of the square brackets.

Exactly the same arguments can be applied to the walls at \(p_1\) and \(p_1 + dp_1\). This will, in the end, give:

\[
- \left\{ \rho \left[ \left( \frac{\partial \dot{q}_1}{\partial q_1} \right) + \left( \frac{\partial \dot{p}_1}{\partial p_1} \right) \right] + \left[ \left( \frac{\partial \rho}{\partial q_1} \right) \dot{q}_1 + \left( \frac{\partial \rho}{\partial p_1} \right) \dot{p}_1 \right] \right\} dq dp \tag{16.1.13}
\]

If we sum over both all coordinate and position coordinates we get the change \(dN/dt\) in number of ensemble members in the little volume element per unit time:

\[
\frac{dN}{dt} = - \sum_{i=1}^{3N} \rho \left[ \left( \frac{\partial q_i}{\partial q_i} \right) + \left( \frac{\partial p_i}{\partial p_i} \right) \right] \left[ \left( \frac{\partial \rho}{\partial q_i} \right) \dot{q}_i + \left( \frac{\partial \rho}{\partial p_i} \right) \dot{p}_i \right] d\mathbf{q} d\mathbf{p} \tag{16.1.14}
\]

This simplifies a lot since we know that

\[
\dot{q}_i = \left( \frac{\partial H}{\partial p_i} \right) \quad \text{and} \quad \dot{p}_i = - \left( \frac{\partial H}{\partial q_i} \right) \tag{3.1.11, 3.1.12}
\]

then, since the order of differentiation doesn’t matter:

\[
\left( \frac{\partial \dot{q}_i}{\partial q_i} \right) = \frac{\partial^2 H}{\partial p_i \partial q_i} = \frac{\partial^2 H}{\partial q_i \partial p_i} = - \left( \frac{\partial \dot{p}_i}{\partial p_i} \right) \tag{16.1.15}
\]

and so

\[
\left[ \left( \frac{\partial \dot{q}_i}{\partial q_i} \right) + \left( \frac{\partial \dot{p}_i}{\partial p_i} \right) \right] = 0 \tag{16.1.16}
\]

With this simplification Equation (16.1.14) becomes:

\[
\frac{dN}{dt} = - \sum_{i=1}^{3N} \left[ \left( \frac{\partial \rho}{\partial q_i} \right) \dot{q}_i + \left( \frac{\partial \rho}{\partial p_i} \right) \dot{p}_i \right] d\mathbf{q} d\mathbf{p} \tag{16.1.17}
\]

Now we divide by the phase space volume element \(d\mathbf{q} d\mathbf{p}\). Doing so will convert \(dN\), the number of ensemble members in the volume element to \(dN/d\mathbf{q} d\mathbf{p}\), the number
per unit volume. This is, of course \( \rho \). Since there is still a division by \( dt \), what we have is:

\[
\left( \frac{\partial \rho}{\partial t} \right)_{\bar{p}, \bar{q}} = -3N \sum_{i=1}^{3N} \left[ \left( \frac{\partial \rho}{\partial q_i} \right) \dot{q}_i + \left( \frac{\partial \rho}{\partial p_i} \right) \dot{p}_i \right]
\]  
(16.1.18)

The meaning of the partial derivative on the left in Equation (16.1.18) is that it is the change in density in the volume element at a specific point \( \bar{p} \bar{q} \) fixed in phase space.

Equation (16.1.18) is known as Liouville’s Theorem and applies to all incompressible fluids. It is of particular importance in the foundations of statistical mechanics.

If we now substitute the definitions of \( \dot{q}_i \) and \( \dot{p}_i \) in terms of the Hamiltonian into Liouville’s Theorem and move the sum to the left-hand side we get:

\[
\left( \frac{\partial \rho}{\partial t} \right)_{\bar{p}, \bar{q}} + 3N \sum_{i=1}^{3N} \left[ \left( \frac{\partial H}{\partial p_i} \right) \left( \frac{\partial \rho}{\partial q_i} \right) - \left( \frac{\partial H}{\partial q_i} \right) \left( \frac{\partial \rho}{\partial p_i} \right) \right] = 0
\]  
(16.1.19)

In classical mechanics the second term in Equation (16.1.19) is called a Poisson bracket and is written

\[
\sum_{3N} \left[ \left( \frac{\partial H}{\partial p_i} \right) \left( \frac{\partial \rho}{\partial q_i} \right) - \left( \frac{\partial H}{\partial q_i} \right) \left( \frac{\partial \rho}{\partial p_i} \right) \right] = [\rho, H]
\]  
(16.1.20)

The notation looks like the commutator in quantum mechanics. In fact it is the classical analog of the quantum mechanical commutator and plays a similar role. Using this notation we can write Liouville’s Equation in its more customary form:

\[
\left( \frac{\partial \rho}{\partial t} \right)_{\bar{p}, \bar{q}} + [\rho, H] = 0
\]  
(16.1.21)

The Liouville equation contains all of Hamiltonian mechanics in it. Indeed, it is not too much to say that it is the fundamental equation of classical statistical mechanics. And it is especially useful in time-dependent situations.

### 16.1.1 Incompressible flow in Phase Space

The Liouville equation also allows some interesting and useful deductions. The most important of these is the following:

If we form the total derivative of \( \rho(\bar{p}, \bar{q}, t) \) we get:

\[
\left( \frac{\partial \rho}{\partial t} \right) + 3N \sum_{i=1}^{3N} \left[ \left( \frac{\partial \rho}{\partial q_i} \right) \dot{q}_i + \left( \frac{\partial \rho}{\partial p_i} \right) \dot{p}_i \right] = 0
\]  
(16.1.22)

---

5Liouville, *Journ. de Math.* 3, 349 (1838)

6It would be more correct to call the quantum mechanical commutator the analog of the Poisson bracket, mainly because the Poisson bracket was developed first and the quantum mechanical entity a direct analog of it.
we see that it must equal zero as the right hand side is identically zero.

The quantity \( \frac{d\rho}{dt} \) is the rate at which the density of ensemble points in our volume element changes as the volume element moves in time. The fact that it is zero means that the density in that volume element does not change.

This is another way of saying that the “fluid” composed of phase space points is \textit{incompressible}.

\subsection*{16.1.2 Conservation of Extension in Phase}

Let us return to Equation (16.1.6) on page 16-3

\[ \delta N = \rho dq_1 \ldots dq_{3N} dp_1 \ldots dp_{3N} = \rho d\tau \quad ((16.1.6)) \]

which is the number of ensemble members inside the volume element \( d\tau \).

Let’s turn this around and for a moment regard a volume element \( d\tau \) to be defined by the phase points inside of it.\footnote{For the vividly minded among you, imagine that these points painted red.} As time goes by these points move in slightly different directions. Thus the volume element changes shape with time. However its density remains \( \rho \). Since trajectories cannot suddenly start at any arbitrary time, or disappear either, no new system points can suddenly appear inside or vanish from this volume element. And since Liouville’s Theorem applies to this volume element even if its shape changes, no new trajectories enter or leave this volume element either.

Thus, taking the time derivative of Equation (16.1.6) on page 16-3 we must have:

\[ \frac{d(\delta N)}{dt} = \frac{d\rho}{dt} \delta \tau + \rho \frac{d(\delta \tau)}{dt} = 0 \quad (16.1.23) \]

Of course the term \( \frac{d\rho}{dt} \) is zero as shown by Equation (16.1.22) on page 16-5. Thus \( \frac{d(\delta \tau)}{dt} \) must be zero as well and we have:

\[ \frac{d(\delta \tau)}{dt} = \frac{d}{dt} \int \cdots \int dq_1 \ldots dp_{3N} = 0 \quad (16.1.24) \]

This result means that the size of our volume element does not change in time. So while our original region can become distorted, indeed quite contorted, its volume does not change.\footnote{Again, for the vividly minded, our little region defined by the red dots changes shape but not volume.} Gibbs called this the \textit{conservation of extension in phase}.

\section*{16.2 The Virial Theorem}

The \textit{Virial Theorem} is not, strictly speaking, part of Hamiltonian mechanics. Nevertheless this is a reasonable place to discuss it, even though we won’t use it for...
a while.\footnote{After all, we are not developing a text involving all knowledge. We pick and choose what to discuss primarily on the basis of what will be needed later – and fit the chosen material in where we can. This looked like a good spot...}

Newton’s Second Law is:

\[ F = \dot{p} = m\ddot{a} \quad (16.2.1) \]

We\footnote{I’m following H. Goldstein, \textit{Classical Mechanics}, Addison-Wesley, 1953 in this.} consider the quantity

\[ G = \sum_i p_i \cdot r_i \quad (16.2.2) \]

where the subscript \( i \) denotes a particle in a system of many particles and the dot indicates the vector dot product. The total derivative of \( G \) with respect to time is:\footnote{The reader may be wondering where we are going with all this. For now just follow along. The road is neither long nor tedious.}

\[ \frac{dG}{dt} = \sum_i \dot{\mathbf{r}}_i \cdot \mathbf{p}_i + \sum_i \dot{\mathbf{p}}_i \cdot \mathbf{r}_i \quad (16.2.3) \]

The first term in Equation (16.2.3) is:

\[ \sum_i \dot{\mathbf{r}}_i \cdot \mathbf{p}_i = \sum_i m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i = \sum_i m_i s_i^2 = 2K \quad (16.2.4) \]

where \( s \) is the scalar speed of the particle and \( K \) is the kinetic energy. The second term can be rearranged using Equation (16.2.1):

\[ \sum_i \dot{\mathbf{p}}_i \cdot \mathbf{r}_i = \sum_i \mathbf{F}_i \cdot \mathbf{r}_i \quad (16.2.5) \]

so that the expression for \( dG/dt \) (Equation (16.2.3)) can be written as:

\[ \frac{d}{dt} \sum_i \mathbf{p}_i \cdot \mathbf{r}_i = 2K + \sum_i \mathbf{F}_i \cdot \mathbf{r}_i \quad (16.2.6) \]

It is instructive to look at the long time average of this. We get that by integrating over a time interval from 0 to \( t \) and then dividing by \( t \):

\[ \frac{1}{t} \int_0^t \left( \frac{dG}{dt} \right) dt = \left\langle \frac{dG}{dt} \right\rangle = \left\langle 2K \right\rangle + \left\langle \sum_i \mathbf{F}_i \cdot \mathbf{r}_i \right\rangle \quad (16.2.7) \]

or

\[ \left\langle 2K \right\rangle + \left\langle \sum_i \mathbf{F}_i \cdot \mathbf{r}_i \right\rangle = \frac{1}{t} \left[ G(t) - G(0) \right] \quad (16.2.8) \]

where the right hand side comes from the fact that \( G \) is an exact differential (see Equation (16.2.2)) and hence dependent only on the initial and final values.

Now Equation (16.2.8) is of no use without a bit of insight.\footnote{Isn’t that so often the case in life.} If there is an upper bound to \( G \), then \( [G(t) - G(0)] \) is less than some quantity \( B \) and the right hand side of Equation (16.2.8) will be less than or equal to \( B/t \). Since we can take \( t \) to be as large as we like, \( B/t \) can be made as small as we like. For the limit as \( t \to \infty \), the right hand side then becomes zero. In that case we have:

\[ \left\langle K \right\rangle = -\frac{1}{2} \left\langle \sum_i \mathbf{F}_i \cdot \mathbf{r}_i \right\rangle \quad (16.2.9) \]
This is known as the Virial Theorem and the right-hand side is the virial of Clausius. But is $G$ in fact bounded? Since $G$ involves both $r$ and $p$ we can consider these separately.

For any system confined to a box of finite dimensions, $r$ is clearly limited to the maximum dimension of of the box.

For any system with a finite energy $E$, $p$ is clearly limited since even if one particle has all the energy in the system, it’s momentum $p = (2mE)^{1/2}$, which is bounded.

What if we deal with systems with no fixed energy? In such systems the probability of finding a particle with energy $E$ falls off as $\exp(-E/kT)$. This causes the long-time integral to converge to a finite value. So $G$ is in fact bounded and Equation (16.2.9) on page 16-7 is correct.

\[\text{Equation (16.2.9)}\]

\[\text{I know that it would be best if I’d prove that right now, but we’ve not developed the machinery.}\]
Chapter 17

The Classical Microcanonical Ensemble

17.1 Introduction

Here we rederive the microcanonical ensemble, but now using classical mechanics. We ignore quantum effects. And, of course, we get the same high-temperature results we obtained in Chapter 5 on page 5-1, as we must.

There are several reasons for doing this. One is historical, to see how these things were done. The other is because many molecular dynamics calculations are done using classical mechanics.

17.2 The Specification of a System

The Gibbs Phase Rule for a single phase single component system tells us that such a system has three degrees of freedom. To specify the extent of the system, at least one variable must be extensive.

The specification is usually done by placing constraints on the walls of the system. For instance we can give a system a specified volume by having it surrounded by fixed rigid walls. And we can specify a number of particles in a system by placing that number of particles inside walls that are impenetrable to particles.

Fixing the energy is a bit more complex. This requires that the walls be both

---

1 Though we do simplify things a bit here and there.

2 Or you’d have a system of undefined extent, whose thermodynamic properties would be equally difficult to define.
adiabatic and rigid. That is, that they do not allow heat to pass into or out of the system and they do not allow the system to do any work.

It isn’t hard to see how to control volume, number, and energy all at the same time. All these are extensive variables.

Control of intensive variables is more complex. We can control temperature by using walls that conduct heat if we also place our system into a constant temperature bath. That does not interfere with controlling volume or number of particles, but a system cannot have both its energy and its temperature controlled in this way. Similarly pressure can be controlled by using movable walls and placing the system into a constant pressure bath. But clearly we cannot control pressure and volume simultaneously.

And we can control the chemical potential of a given component by using a wall made of a semi-permeable membrane that allows passage of that component. And we also must place the system in a bath containing the specified component at a concentration such that it has the desired chemical potential.

In fact we can in principle allow several or all components to have a fixed chemical potential by suitable choice of membrane and bath. But it is clear that we cannot also control the number of particles of that component.

In other words thermodynamic variables come in conjugate pairs: \( p \) and \( V \), \( N \) and \( \mu \), and \( E \) and \( T \). Only one of each pair of variables can be specified. The other will then be what it is in the system.

There are times when certain types of systems are not possible. For example chemists call a system with fixed energy, volume, and number a closed system. But such a system may allow electromagnetic fields to penetrate. Thus a battery driven radio transmitter inside such a system could still lose energy to the outside. Such a system is not closed in spite of having what might seem as appropriate walls.

While we can shield against electrical and magnetic fields, we cannot shield against gravitational ones. Thus if gravitational effects are taken into account, there can be no closed systems.

### 17.3 Problems

The classical approach to statistical mechanics suffers from one major problem. It attempts to apply Newton’s Laws, which are essentially macroscopic laws, to microscopic situations.

In many cases this works out well. But in others, it fails completely.

---

3Think of a system contained in a balloon in the atmosphere. The pressure is constrained to remain atmospheric.

4As a practical matter not all combinations of materials may be simultaneously controllable, mainly due to limitation on our technology rather than any theoretical reason for the limitation.
These failures were discovered by early researchers. But because quantum mechanics was still in the future, they did not really grasp the implications of these failures.

The first of these failures had to do with specification of position and momentum. Classically it was known that the probability of a particle being exactly at a given point \( q \) is zero. This is easy to see. For example consider a line segment running from \( q = 0 \) to \( q = 1 \) that contains one particle. If the probability of finding a particle exactly at \( q \) were finite, then the probability that the particle was somewhere in the given interval would be that constant times the number of points in the interval, which is infinite. This is not a satisfactory result.

The classical “fix” for this was to assign a finite probability not to a point but to a range of points, usually given as running from \( q \) to \( q + dq \). In doing this \( dq \) is thought of as being a very small but finite range. The same thing applies to the position \( q \).

Classically though the chance of finding that particle with a position uncertainty \( dq \) and a momentum uncertainty \( dp \) could be made as small as desired as long as it remained non-zero.

So while in reality we can’t talk about a system being at a point in phase space, we can talk about it being inside a volume element of arbitrarily small dimensions.

On the other hand we know that systems on a microscopic level are really governed by quantum mechanics. And so we also know that \( dq \) \( dp \) is in fact given by the Heisenberg Uncertainty Principle:

\[
dq \, dp \geq \frac{\hbar}{2\pi}
\]  

(17.3.1)

and so clearly the joint uncertainty \( dq \, dp \) can not be made as small as desired. However, in classical mechanics in principle the volume elements in phase space could be made as small as desired.

Today, because of quantum mechanics, we know we cannot talk about a system being at a point in phase space at all. The system at best is a blob in phase space. And the blob has dimensions of roughly \( \hbar^3 N \), where \( N \) is the number of particles in the three-dimensional system.

This is not just a difference in approach. It turns up, as we shall soon see, in the division of phase space into microstates. A microstate is a small hypervolume element in phase space supposedly large enough to contain one or more trajectories (Boltzmann) or system points (Gibbs). Unfortunately, the physical results obtained from a classical statistical mechanical calculation often depend on the size of this hypervolume element – and there is no way, short of experiment, to determine the size of it.

Experimental measurements showed that nature required a hypervolume for a microstate that was approximately \( \hbar^3 N \), where \( N \) was the number of particles in the system and \( \hbar \) turns out to be Planck’s Constant. That the hypervolume was exactly \( \hbar^3 N \) was first postulated by O. Sakur in 1911 and verified by comparison of theory
to experiment a year later by Tetrode\textsuperscript{5} using data on gaseous mercury.\textsuperscript{6}

Knowing the Uncertainty Principle, as we do, this is not a surprise to us.

To avoid the sort of complication and handwaving endemic to early presentations of classical statistical mechanics, we will choose the dimensions of our microstates to be $h^{3N}$ right from the start.\textsuperscript{7}

A second problem is of the same nature. When dealing with systems composed of identical molecules, it was classically assumed that these molecules were distinguishable. That is, in principle one could model each molecule as a small shape and, with a very very small pen, write a number on that shape. Thus the molecules could be distinguished. For instance, given three molecules on a line, it was known that there were six different ways of arranging them:

$$(1, 2, 3) \quad (1, 3, 2) \quad (3, 1, 2) \quad (2, 1, 3) \quad (2, 3, 1) \quad (3, 2, 1)$$

But in fact you can’t write numbers on molecules. So the six different configurations\textsuperscript{8} are actually only one! In classical statistical mechanics one has to correct for this manually by inserting an $N!$ in the appropriate place when dealing with systems of indistinguishable molecules.

The third problem is in a way more benign. It is due to the quantization of energy levels. Classical statistical mechanics works well at high temperatures but fails badly as the temperature drops. For example we know that $C_V$, the constant volume heat capacity should go to zero as the temperature goes to zero. The classical $C_V$ does not go to zero.

The cure for this is to use classical formulas properly. That is, at sufficiently high temperatures. Beyond that there is nothing that can be done.\textsuperscript{9}

### 17.4 The Microcanonical Ensemble

Let us specify a system by fixing its energy, volume, and number of particles. We now duplicate that system so as to create an ensemble of $\mathcal{A}$ such systems. This ensemble is known as the **microcanonical ensemble**.

How are the system points of this ensemble distributed in phase space?

\textsuperscript{5}Tetrode has three syllables.

\textsuperscript{6}The information on Sakur and Tetrode here was taken from R.K. Pathria, *Statistical Mechanics*, Pergamon, 1972, page 43.

\textsuperscript{7}Most modern texts, if they mention classical statistical mechanics at all, simply assume that the volume of a microstate is $h^{3N}$ without any discussion. Early writers had a great deal of difficulty with this volume.

\textsuperscript{8}Actually $3!$. The number of configurations of $n$ particles is, of course, $n!$ as the reader doubtless knows.

\textsuperscript{9}It should be noted that the famous example of this is *black body radiation*. The appropriate description was quantized as was shown by Planck in 1900. But it took until Einstein’s treatment of the photoelectric effect in 1905 (for which he received the Nobel Prize in 1921) for its implications to even start to affect classical statistical mechanics.
First, we have to recognize that in a mathematical sense we cannot specify the energy with infinite precision. The chance of our producing a system with, for example, an energy of exactly 105 1/3 kilojoules is zero.\(^\text{10}\) The best we can do is to produce a total energy \(H\) such that:

\[
E \leq H(p, q) \leq E + dE
\]

(17.4.1)

where \(H\) is the Hamiltonian (the total energy) and \(dE\) is a small (infinitesimal) amount of energy.

This means that all the systems points of our ensemble are located in phase space on a thin surface shell of thickness \(dE\) and constant energy \(E\). The probability \(P\) of finding a system in a given microstate is then:\(^\text{11}\)

\[
P(p, q) = \begin{cases} 
\text{constant} & \text{if } E \leq H(p, q) \leq E + dE \\
0 & \text{otherwise}
\end{cases}
\]

(17.4.2)

This needs to be justified. It is clear from the definition of the microcanonical ensemble that the energy \(E\) is fixed. Thus the conditions on the right of Equation (17.4.2) are correct. But why do we make the probability of finding a system in a given microstate a constant?

The answer is, once again, the Principle of Democratic Ignorance, which was discussed back in Chapter 5 on page 5-1 in Section 5.3 on page 5-3.

If \(R(p, q)\) is a mechanical property of these systems, and if the ensemble is large enough to sample all parts of the energy surface in phase space, then the value of \(R\) that we would expect to observe, \(\langle R \rangle\) is given by:

\[
\langle R \rangle = \int \cdots \int R(p, q)P(p, q) dq dp
\]

(17.4.3)

The expected value of \(R\), \(\langle R \rangle\) is not the only possible “average” value of \(R\) that we might measure. But if the distribution of values of \(R\) is sharply peaked around \(\langle R \rangle\), all the different measures of the average such as the most probable value,\(^\text{12}\) the mode,\(^\text{13}\) etc., will have the same numerical value. One measure of the sharpness of the peak is the relative standard deviation:

\[
\frac{\langle R^2 \rangle - \langle R \rangle^2}{\langle R \rangle^2} << 1
\]

(17.4.4)

As long as Equation (17.4.4) holds, the actual measure used for the average does

\(^{10}\)The same argument applies to the volume as well. However the number of particles is a different matter as even before quantum mechanics was discovered it was recognized that matter is quantized into small units called atoms. Thus while we can’t produce exactly any decimal number of particles, we can produce a definite integer number of particles.

\(^{11}\)In this and the following material I am using the approach of Kerson Huang, Statistical Mechanics, Second Edition, Wiley, 1987

\(^{12}\)The value of \(R\) that occurs most often in the ensemble

\(^{13}\)The value of \(R\) that is the middle value of all the observed values of \(R\).
not matter; they will all be the same. When Equation (17.4.4) on page 17-5 does not hold, things need to be investigated more closely.\(^{14}\)

The fundamental quantity that gives us the connection between the microscopic world and macroscopic thermodynamics in the microcanonical ensemble is the number of microstates in the thin energy shell where \(P\) is non-zero.

We will denote this number by \(\Omega(N, V, E)\), a dimensionless quantity. The connection to thermodynamics is simply this:

\[
S(N, V, E) = k \ln \Omega(N, V, E) \tag{17.4.5}
\]

where \(S\) is the macroscopic entropy and \(k\) is a constant.\(^{15}\)

Why would we think that the logarithm of \(\Omega\) would have anything at all to do with the entropy?

Right now, Equation (17.4.5) is an assumption.\(^{16}\) The primary justification for it now is that it works. One area in which it works is that it allows \(S\) to be an extensive property of a system. If, for example, a system is made up of two parts whose entropies are \(S_1\) and \(S_2\), then the entropy of the entire system is \(S = S_1 + S_2\).

To demonstrate this, we will consider a system divided into two independent subsystems.

The microcanonical ensemble corresponding to the first subsystem will be assumed to have \(N_1\) particles, a volume \(V_1\) and an energy \(E_1\) lying between \(E_1\) and \(E_1 + dE_1\). The ensemble for the second will have \(N_2\) particles, a volume \(V_2\) and an energy \(E_2\) lying between \(E_2\) and \(E_2 + dE_1\).

Then:

\[
S_1(N_1, E_1, V_1) = k \ln \Omega(E_1) \quad \text{and} \quad S_2(N_2, E_2, V_2) = k \ln \Omega(E_2) \tag{17.4.6}
\]

Now the composite system made up of the two subsystems will have an \(\Omega\) given by:

\[
\Omega(E_1 + E_2) = \Omega(E_1)\Omega(E_2) \tag{17.4.7}
\]

The \(\Omega\)'s occur as a product because the subsystems are independent. For instance if your right hand is an independent subsystem that can be in one of three microstates and your left hand is an independent subsystem that can be in one of four microstates, then the two together can be in a total of \(3 \times 4 = 12\) different microstates.

Given this then

\[
S = S_1 + S_2 = k \ln \Omega(E_1) + k \ln \Omega(E_2) = k \ln[\Omega(E_1)\Omega(E_2)] \tag{17.4.8}
\]

\(^{14}\)Examples of situations where Equation (17.4.4) on page 17-5 does not hold includes at phase transitions and in systems either with a high boundary to volume ratio or a small number of particles (or both.)

\(^{15}\)It will turn out that \(k\) is Boltzmann’s Constant, as we shall shortly see.

\(^{16}\)It does not have to be an assumption. It will be shown to be true. But we can’t do that right now.
which was to be proved.

To compute $S$ we must compute $\Omega$. The computation depends on the sort of system for which we do the computation. Here we will do this for the case of an ideal gas.

We make the gas ideal by assuming that there is no potential energy in the Hamiltonian for the system. Thus the Hamiltonian for this system in cartesian coordinates is:

$$H(p, q) = \sum_i \frac{p_i^2}{2m}$$  \hspace{1cm} (17.4.9)

where $m$ is the mass of the ideal gas molecules.

The Hamiltonian is the total energy of the system and is a constant in the microcanonical ensemble.

That means that all the system points (and indeed all the trajectories) lie on a $6N - 1$ dimensional surface in the $6N$ dimensional phase space for these systems. This surface has the specified energy $E$ and a “thickness” of $dE$.

This surface has a hyperarea which we will denote by $A(E)$.

One way to compute $\Omega$ is to compute the hyperarea $A(E)$ and then divide it by the hyperarea of a single microstate, $\delta v$.

$$\Omega = \frac{A(E)}{\delta v}$$  \hspace{1cm} (17.4.10)

Note that there is no way in classical mechanics to unambiguously decide on a value for $\delta v$. But as discussed in Section 17.3 on page 17-2, we will use the value $\delta v = \hbar^{3N}$ as the hyperarea of a microstate since we know from quantum mechanics that this is correct.

What we need now is the hyperarea of the $(6N - 1)$-dimensional shell. That’s given by:

$$A(E) = \int \cdots \int_{E \leq H(p, q) \leq E + dE} dq \, dp$$  \hspace{1cm} (17.4.11)

where $H(p, q)$ is the Hamiltonian for our systems. The limits on the integration reflect the parts of phase space where $P(q, p)$ is non-zero.

The fact that the sums of the squares of the momenta (divided by $m$) must add up to a constant suggests that a change to spherical coordinates might simplify this problem. Then we could take the sums of the squares of the momenta as the square of the radius of a hypersphere. The hyperarea is simply related to the hypervolume of the sphere in the same way that the area of a sphere is related to the volume of the sphere.

So we find the hyperarea by first calculating the volume of a hypersphere of volume
\( \mathcal{W} \) of radius \( E \) given by:\textsuperscript{17}

\[
\mathcal{W}(E) = \int \ldots \int_{E \leq H} dq \, dp
\]

(17.4.12)

Because the systems are ideal gases, there is no potential energy and the Hamiltonian is independent of the coordinates \( q \). Then the integral Equation (17.4.12) becomes the product of two integrals, one over the positions \( q \) and the other over the momenta \( p \).

The integral over the positions is trivial. We’ve assumed that the systems have a fixed volume, so the integration over \( dx \, dy \, dz \) for any particle must give the volume \( V \). And since there are \( N \) such particles, the result is \( V^N \).

We now have simply:

\[
\mathcal{W}(E) = V^N \int \ldots \int_{E \leq H} dp
\]

(17.4.13)

Now we do the integrations over the momenta. Here we really want to go into spherical coordinates, so we first change variables, letting \( y_i = p_i/(2m)^{1/2} \). Then \( dp_i = (2m)^{3N/2} dy_i \) for all \( i \). The momentum integral is then

\[
I = (2m)^{3N/2} \int \ldots \int_{E \leq H} dy
\]

(17.4.14)

Now from the definition of \( y \) we can easily see that

\[
\sum_{i=1}^{3N} y_i^2 = E
\]

(17.4.15)

so what we really have in Equation (17.4.14) is the contents of hypersphere of radius \( R \) where

\[
R^2 = E = \sum_{i=1}^{3N} y_i^2
\]

(17.4.16)

The volume of such an \( n \)-dimensional hypersphere is (see Section 17.7 on page 17-14):

\[
V_n = \frac{\pi^{n/2}}{(n/2)!} R^n
\]

(17.4.17)

so that in our case of \( 3N \)-dimensions

\[
I = \frac{(2m\pi)^{3N/2}}{(3N/2)!} R^{3N}
\]

(17.4.18)

and thus:

\[
\mathcal{W}(R) = V^N \frac{(2m\pi)^{3N/2}}{(3N/2)!} R^{3N}
\]

(17.4.19)

\textsuperscript{17}Note that this works for spheres, but not for all multidimensional figures such as hypercubes. See Appendix 17.7 on page 17-14 for a discussion.
Converting back to energy, since \( R = E^{1/2} \) (see Equation (17.4.17) on page 17-8) yields:

\[
\mathcal{M}(E) = V^N \frac{(2m\pi E)^{3N/2}}{(3N/2)!} \tag{17.4.20}
\]

We’ve now done the hard part. To get the hyperarea of the shell we need only differentiate Equation (17.4.20) with respect to \( E \):

\[
\mathfrak{A}(E) = \frac{3NV^N}{2} \frac{(2m\pi E)^{3N/2-1}}{(3N/2)!} \tag{17.4.21}
\]

Of course we don’t want \( \mathfrak{A}(E) \), we want \( \Omega \) which is \( \mathfrak{A}/\hbar^{3N} \):

\[
\Omega = V^N \frac{3N}{2} \frac{(2m\pi E/h)^{3N/2-1}}{(3N/2)!} = V^N \frac{(2m\pi E/h^2)^{3N/2-1}}{(3N/2-1)!} \tag{17.4.22}
\]

Since \( S = k \ln \Omega \), we get:

\[
S(N,V,E)/k = N \ln V + \ln(2\pi m) + \left( \frac{3N}{2} - 1 \right) \ln \left( \frac{2m\pi E}{\hbar^2} \right) - \ln(3N/2 - 1)! \tag{17.4.23}
\]

which can be simplified greatly. First, we recall that \( N \) is a macroscopic number of particles, something of the order of \( 10^{23} \). Compared to that 1 is so small as to be ignorable. So we will ignore it. Further \( \ln(2\pi m) \) is a very small negative number of the order of \(-26\) (since the mass is the mass of one molecule in kilograms.) Again, this is ignorable compared to \( 10^{23} \). So after all that we have for the entropy:

\[
S(N,V,E)/k = N \ln V + \frac{3N}{2} \ln \left( \frac{2\pi m E}{\hbar^2} \right) - \ln(3N/2)! \tag{17.4.24}
\]

Using Stirling’s Approximation (see Section 5.8 on page 5-11) for the factorial and doing a bit of algebra gives us:

\[
S(N,V,E) = Nk \ln \left[ V \left( \frac{4\pi m E}{3Nh^2} \right)^{3/2} \right] + \frac{3Nk}{2} \tag{17.4.25}
\]

This should be our final answer, but if we look closely, there’s a problem. The entropy \( S \) is extensive. So if we double the size of the system (for instance) we double \( N, E, \) and \( V \). Certainly if we double \( N \) and \( E \), the entropy is doubled, since the factor of two cancels out inside the parentheses with \( E/N \) leaving just the doubled \( N \) in both parts of Equation (17.4.25).

But the doubling of \( V \) causes the first term in Equation (17.4.25) to increase by only a factor of the logarithm of 2. This can’t be right.

We discussed this as the second problem in Section 17.3 on page 17-2. We need to include an \( N \) factorial in the denominator in Equation (17.4.22). This adds a term
−N \ln N + N to Equation (17.4.25) on page 17-9 giving us:

\[ S(N, V, E) = Nk \ln \left( \frac{V}{N} \left( \frac{4\pi mE}{3Nh^2} \right)^{3/2} \right) + \frac{5Nk}{2} = Nk \ln \left( \frac{V}{N} \left( \frac{4\pi mE}{3Nh^2} \right)^{3/2} e^{5/2} \right) \]

(17.4.26)
either of which is now our final answer.

The reader should note (by going through the math) that our final answer, Equation (17.4.26) could have been obtained from Equation (17.4.20) on page 17-9 instead of Equation (17.4.21) on page 17-9.

How can this be? Equation (17.4.20) on page 17-9 is for the volume of a 3N dimensional hypersphere while Equation (17.4.21) on page 17-9 is for the surface area of that hypersphere. Surely there is far more “space” in the volume than in the surface?

The answer is that yes, there is more “space” in the volume than in the surface. However the relative increase is negligible. Each minor increase in the radius \( R \) of the hypersphere increases the number of microstates involved by such a huge number that the difference between the volume and the surface just don’t matter.

In fact we’ve demonstrated that in the terms we neglected in going from Equation (17.4.23) on page 17-9 to Equation (17.4.24) on page 17-9. Those neglected terms are the difference and they are vanishingly small.

### 17.5 The Thermodynamics of the Microcanonical Ensemble

Since the value of Omega derived above:

\[ \Omega = \left[ \frac{V}{N} \left( \frac{4\pi mE}{3Nh^2} \right)^{3/2} \right] + \frac{5Nk}{2} = Nk \ln \left( \frac{V}{N} \left( \frac{4\pi mE}{3Nh^2} \right)^{3/2} e^{5/2} \right) \]

(17.5.1)
is identical to that obtained earlier. The thermodynamics are given by equation (5.6.8) on page 5-9 and following equations. The results are repeated below for convenience.

We have:

\[ \left( \frac{\partial S}{\partial E} \right)_{N,V} = \frac{1}{T} \quad \left( \frac{\partial S}{\partial V} \right)_{N,E} = \frac{p}{T} \quad \left( \frac{\partial S}{\partial N} \right)_{V,E} = -\frac{\mu}{T} \]

(17.5.2)

We compute the temperature from Equation (17.4.26) and find that:

\[ \left( \frac{\partial S}{\partial E} \right)_{V,N} = \frac{3Nk}{3E} = \frac{1}{T} \]

(17.5.3)

\(^{18}\)Of course this sort of ad hocery is unsatisfactory. We will do all of this in a much more convenient manner when we talk about semi-classical statistical mechanics.
which not only tells us that the temperature in our ideal gas system can be computed from

$$T = \frac{2E}{3Nk} \quad (17.5.4)$$

but that the energy is given by the familiar equation

$$E = \frac{3}{2}NkT \quad (17.5.5)$$

which serves, as promised, to identify $k$ as Boltzmann’s constant.

It then comes as no surprise that

$$\left(\frac{\partial S}{\partial V}\right)_{N,E} = \frac{Nk}{V} = \frac{p}{T} \quad (17.5.6)$$

resulting in the fairly familiar\(^{19}\)

$$pV = NkT \quad (17.5.7)$$

Finally, the determination of the chemical potential $\mu$ from the appropriate one of Equations (17.5.2) on page 17-10 is fiendishly left as an exercise for the reader.

---

**Example 17.1**

As an exercise let’s calculate\(^{20}\) the entropy of exactly one mole of argon at a temperature of 298.15°C and a pressure of exactly 1 bar using Equation (17.4.26) on page 17-10.

These are not the independent variables of of the microcanonical ensemble, but given the formulas above it is not hard to generate those.

We use the following constants: Avogadro’s number = $6.0221367 \times 10^{23}$ per mole, Boltzmann’s constant = $1.380658 \times 10^{-23}$ Joules/K, and Planck’s constant = $6.626075 \times 10^{-34}$ Joule-sec.

From Equation (17.5.5) we have an energy of 3718.4511 Joules/mol and from Equation (17.5.7) we have a volume of 0.0247897 meters\(^3\).

The atomic mass of argon is 0.039948 kg/mole and, for reference, its entropy under these conditions is 154.8 J/mol-K.

\(^{19}\)If this result is not familiar, the reader is in deep trouble and should switch to reading popular fiction immediately!

\(^{20}\)The alert reader will note that the calculation below carries far more significant figures than necessary. In the “old days” folks carried only the minimum number of significant digits because calculations were done using paper and pencil, three significant digit slide rules, or five place logarithm tables. The result was that accumulated round-off errors often cost the calculation its last digit. Today it is trivial to keep all significant digits (up to the capacity of your calculator) and to do the rounding off to the correct number of significant digits only once, at the very end of the calculation. That way the result is good to ±1 in the last place.
The calculation itself is best taken in stages.

\[
\left(\frac{4\pi mE}{3Nh^2}\right)^{3/2} = \left(\frac{12.566371 \times 6.6325259 \times 10^{-26} \times 3718.4571}{1806641 \times 10^{24} \times 4.3904877 \times 10^{-67}}\right)^{3/2} \\
= \left(\frac{3.0996816 \times 10^{-21}}{7.9320351 \times 10^{-43}}\right)^{3/2} \\
= (3.9078011 \times 10^{21})^{3/2} \\
= 2.4428606 \times 10^{32} \\
V_N = 4.1164293 \times 10^{-26} \\
e^{5/2} = 12.182494 \\
(17.5.8)
\]

We now put the bits together:

\[
S = 8.3145112 \ln (4.1164293 \times 10^{-26} \times 4.1164293 \times 10^{-26} \times 12.182494) \\
= 154.84 \text{ J/mol-K} \\
(17.5.9)
\]

This compares to the experimental value of 154.8. It is exactly this sort of calculation that allowed Sakur and Tetrode to conclude\(^{21}\) that the volume of a microstate was, in fact, \(h^{3N}\).

## 17.6 Discussion

We've solved the microcanonical ensemble for the case of an ideal gas. In doing so we did two integrals, one over the momenta \(p\), the other over the coordinates \(q\).

The coordinate integration was trivial and the integration over the momenta a bit of a mess.

However, unless the potential energy depends on the momenta, the momentum integral Equation (17.4.13) on page 17-8 is always exactly the same. We’ve done it once and we never have to do it again. It’s value is:

\[
I_p = \frac{(2m\pi E/h^2)^{3N/2}}{(3N/2)!} \\
(17.6.1)
\]

The integration over coordinates is, in general, not at all simple. It was simple in the ideal gas case because there was no potential energy. If there is a potential energy, life gets very complex.

First, the potential energy is almost never a simple function of the positions \(q\). In general it depends on the distance between particles. This makes it a function of \(q_i - q_j\), and often a messy one at that.

\(^{21}\)See Section 17.3
So where we had an integral over a hyperspherical shell in phase space for the ideal gas situation, now we have an integral over a very complex shell in hyperspace – at least over the position coordinates.

This is a serious difficulty and in general cannot be done analytically. For this reason the microcanonical ensemble is not often used for hand calculations.
17.7 Appendix: Volume of an $n$-Dimensional Hypersphere

The formula for the volume of an $n$-dimensional hypersphere isn’t obvious. But it can be derived in various ways.

We will use the symbol $V_n$ for the volume of an $n$-dimensional hypersphere and the symbol $A_n$ for the surface area of the same hypersphere.

The results are clearly going to be proportional to the radius $R$ to the $n$th power. Thus we can write:

$$V_n = C_n R^n \quad \text{and} \quad A_n = \frac{dV_n}{dR} = nC_n R^{n-1} \quad (17.7.1)$$

where $C_n$ is a numerical constant independent of $R$.

The quantity $C_n$ can be evaluated by a trick. Consider the integral:

$$\int_{-\infty}^{\infty} e^{-y^2} dy = \pi^{1/2} \quad (17.7.2)$$

which is a standard definite integral. Let $I$ be

$$I = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-y_1^2} e^{-y_2^2} \cdots e^{-y_n^2} dy_1 dy_2 \cdots dy_n \quad (17.7.3)$$

which is Equation (17.7.2) repeated $n$ times:

$$\left[ \int_{-\infty}^{\infty} e^{-y^2} dy \right]^n = \pi^{n/2} \quad (17.7.4)$$

Now $I$ can also be written:

$$I = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-(y_1^2 + y_2^2 + \cdots + y_n^2)} dy_1 dy_2 \cdots dy_n \quad (17.7.5)$$

Here’s the trick: If now I let

$$R^2 = y_1^2 + y_2^2 + \cdots + y_n^2 \quad (17.7.6)$$

then Equation (17.7.5) becomes

$$I = \int_{0}^{\infty} e^{-R^2} dV_n = \int_{0}^{\infty} e^{-R^2} A_n dR = nC_n \int_{0}^{\infty} e^{-R^2} R^{n-1} dR \quad (17.7.7)$$

where we’ve made use of the fact that $dy_1 \cdots dy_n = dV_n$ and the second relation of Equation (17.7.1).\footnote{Equation (17.7.1) is true for hyperspheres. It is not necessarily true for other multidimensional figures. For example the area of a cube is not $dV/dR$, which would be $3R^2$, but $6R^2$ instead. Indeed, the area of an $N$-dimensional hypercube is given by $2dV/dR$.}
This integral is in fact a standard integral. To see that we change variables letting
\( t = R^2 \) so that \( dR = (t^{-1/2}/2)dt \). Then:

\[
I = \frac{n}{2} C_n \int_0^{\infty} e^{-t^{n/2-1}} dt = \frac{n}{2} C_n \left( \frac{n}{2} - 1 \right)! = \left( \frac{n}{2} \right)! C_n = \pi^{n/2}
\] (17.7.8)

where the first integral is the \textit{gamma function}\textsuperscript{23} and the last equality comes from
Equation (17.7.4) on page 17-14.

Thus we have the results:

\[
C_n = \frac{\pi^{n/2}}{(n/2)!}
\] (17.7.9)

\[
V_n = \frac{\pi^{n/2}}{(n/2)!} R^n
\] (17.7.10)

\[
A_n = \frac{n\pi^{n/2}}{(n/2)!} R^{n-1}
\] (17.7.11)

Factorials of non-integers are not usually seen in elementary work. Here when \( n \) is
odd, we will have half-integer factorials. These work like this example:

\[
(5/2)! = (5/2)(3/2)! = (5/2)(3/2)(1/2)! = (5/2)(3/2)(\pi^{1/2})/2 = \frac{15\pi^{1/2}}{8}
\] (17.7.12)

because \((1/2)! = (\pi^{1/2})/2\).

We can see how some of this works out. From these formulas we readily find that:

\[
V_2 = \pi R^2 \quad A_2 = 2\pi R
\]

\[
V_3 = \frac{4\pi}{3} R^3 \quad A_3 = 4\pi R^2
\]

\[
V_4 = \frac{\pi^2}{2} R^4 \quad A_4 = 2\pi^2 R^3
\]

\textsuperscript{23}The gamma function is discussed in section 5.8 on page 5-11.
Chapter 18

The van der Waals Gas

18.1 Introduction

It is now time to consider systems that look somewhat like real gases. We will begin with a “toy” system, that of a van der Waals gas and we will derive van der Waals equation using an approximate and incorrect derivation that works only because of the fortuitous cancellation of errors.

The reason for doing this is to give an insight into the analytical techniques that can be used to study real gases and liquids. There is a limit as to how far these techniques can be pushed but even so they provide great insight into the physical processes that take place in such systems.

Today these systems are mainly studied via computer simulation, either by following the trajectory in phase space of a representative system or by looking at the phase space distribution of representative systems. The former involves integrating the equations of motion, the latter techniques commonly called Monte Carlo methods.

The van der Waals equation is:

\[ p + \frac{N^2a}{V^2} \left[ V - Nb \right] = NkT \]  

(18.1.1)

where \( a \) and \( b \) are constants special to each different species of gas. This equation was not derived from any fundamental principle, but was instead proposed as a “corrected” ideal gas equation. The constant \( b \) was to correct for the hard cores of molecules (which were assumed to be spherical) so that the corresponding “ideal” volume was \( V - Nb \). The constant \( a \) corrects for attractive forces. They were assumed to be proportional to the number density in the gas \( N/V \) and to the

\(^1\text{J.D. van der Waals, Sr., Doctoral Dissertation, University of Leiden, 1873.}\)
number of surrounding molecules, also proportional to $N/V$. Since the attractions reduced the ability of a molecule to move and hence exert pressure, the “corrected” pressure was assumed to be $p + (N/V)^2a$.

Van der Waals equation can be developed into a virial expansion

$$p = \frac{NkT}{V - Nb} - \frac{N^2a}{V^2} = \frac{NkT}{V(1 - Nb/V)} - \frac{N^2a}{V^2}$$

which results in

$$pV = NkT \left[ 1 + \frac{Nb}{V} + \left(\frac{Nb}{V}\right)^2 + \cdots \right] - \frac{N^2aV^2}{1}$$

(18.1.2)

The quantity

$$B_2 = b - \frac{a}{kT}$$

(18.1.4)

is known as the second virial coefficient, and

$$B_3 = b^2$$

(18.1.5)

is the third virial coefficient, and so on.

### 18.2 The Approximate Derivation

We shall assume that we have $N$ identical spherical molecules in a volume $V$. We will ignore any possible internal degrees of freedom so in effect we are dealing with a gas of atoms at temperatures too low to produce any electronic excitations.

There are $3N$ cartesian coordinates to be considered and the same number of momentum coordinates. The Hamiltonian for the system is:

$$H(\vec{p}, \vec{q}) = \sum_{j=1}^{N} \frac{p_j^2}{2m} + U(\vec{q})$$

(18.2.1)

and the classical canonical partition function is then:

$$Q(N, V, T) = \frac{1}{h^{3N}N!} \int \exp \left[ -\beta \sum_{j=1}^{N} \frac{p_j^2}{2m} - \beta U(\vec{q}) \right] d\vec{p}d\vec{q}$$

(18.2.2)

The integral over the momenta can be done as usual and gives:

$$Q(N, V, T) = \frac{1}{N!} \left( \frac{2\pi mkT}{\hbar^2} \right)^{3N/2} Z(N, V, T)$$

(18.2.3)

---

2This assumption is implicit in the development of van der Waals equation.
where $Z$ is the configuration integral

$$Z(N, V, T) = \int e^{-\beta U(\vec{q})} d\vec{q}$$  \hspace{1cm} (18.2.4)$$

We will make certain assumptions about the potential energy $U(\vec{q})$. First we’ll assume that it can be written as a sum of terms each depending only on the distance between a pair of molecules, say molecule $i$ and $j$. This is usually called the Assumption of Pairwise Additivity because under this assumption if three molecules are close to each other, the potential is assumed to be the sum of three pair potentials with no contribution from three-body forces.

Second we’ll assume that this pair potential is zero for large separations between a pair of molecules, becomes attractive as the distance between them becomes small, goes through a single minimum at some distance and then increases very rapidly as the molecules get even closer. We will also assume that the potential function goes to zero as the separation increases faster than the inverse third power of the distance.$^3$

The assumption of pairwise additivity means that the potential energy can be written as a sum over the $N(N - 1)/2$ different pairs that can exist in a system of $N$ molecules. Thus

$$U(\vec{q}) = \sum_{j=1}^{N-1} \sum_{i>j}^{N} u(r_{ij})$$  \hspace{1cm} (18.2.5)$$

where $u(r_{ij})$ is the potential energy of interaction between molecules $i$ and $j$.

The configuration integral is then a product. The coordinates of two molecules occur in each of the $N(N - 1)/2$ terms and the coordinates of any particular molecule occur in $N - 1$ different terms of the product. So the configuration integral is not writable as a product of integrals as is the case with the integration over the momenta as in Equation (18.2.2) on page 18-2. The best we can do is write:

$$e^{-\beta U(\vec{q})} = \prod_{N \geq i > j \geq 1} e^{-\beta u(r_{ij})}$$  \hspace{1cm} (18.2.6)$$

When any particular $r_{ij}$ is large, the corresponding $u(r_{ij})$ goes to zero but $\exp[-\beta u(r_{ij})]$ then goes to 1. This is inconvenient since it would be much more useful to deal with something that goes to zero. To make this happen we’ll switch to Mayer $f$ functions, a useful device for this sort of thing. The Mayer $f$ function $f_{ij}$ is defined by:

$$f_{ij} = e^{-\beta u(r_{ij})} - 1$$  \hspace{1cm} (18.2.7)$$

which has the behavior we want. Further since

$$e^{-\beta u(r_{ij})} = 1 + f_{ij}$$

Equation (18.2.6) can now be written as:

$$e^{-\beta U(\vec{q})} = \prod_{N \geq i > j \geq 1} (1 + f_{ij})$$  \hspace{1cm} (18.2.8)$$

$^3$The reasons for this will be discussed below.
We can now expand the product into sums of terms:

\[ e^{-\beta U(q)} = 1 + \sum_{N\geq i>j\geq 1} f_{ij} + \sum_{ij} \sum_{kl} f_{ij} f_{kl} + \cdots \]  

(18.2.9)

Now it is time to make the unjustifiable assumption that was talked about above. That assumption is that it is a good approximation that only the first two terms in the equation above.

With this assumption the configuration integral becomes:

\[ Z(N, V, T) = \int \left[ 1 + \sum_{N\geq i>j\geq 1} f_{ij} \right] d\vec{q}_1 \cdots d\vec{q}_N \]  

(18.2.10)

The integral over 1 is trivial. It leads to a factor of \( V \) (volume) for every three coordinates or a factor of \( V^N \) overall. If there were no potential energy, that is if the \( f_{ij} \) were identically 0, we would be left only with \( V^N \) and we would have recovered the ideal gas law.

But our \( f_{ij} \) are not identically zero. But we do know that all of the \( f_{ij} \) are identical in form since the molecules are identical. So what we have is \( N(N-1)/2 \) identical integrals:

\[ \int f_{ij} d\vec{q}_i d\vec{q}_j \]

Integration over all coordinates but \( i \) and \( j \) is trivial. We get a factor of \( V \) for each of these. There are \( N-2 \) such molecules, so what we have now is

\[ V^{N-2} \int f_{ij} d\vec{q}_i d\vec{q}_j \]

We are left with a pair of molecules. Let us (at least mentally) switch to the coordinates of the center of mass of the pair and to spherical coordinates for the separation and orientation of the pair of molecules.

The center of mass coordinates integrate to another factor of \( V \). And assuming, as we have, that there is no angle dependence in the potential, integration over the relative angular coordinates gives a factor of \( 4\pi \). So if \( r \) is the internuclear distance, the integral has become one over that distance and:

\[ \int f_{ij} d\vec{q}_i d\vec{q}_j = V \int_0^\infty 4\pi f(r) r^2 dr \]  

(18.2.11)

This integral will converge as long as \( f(r_{ij}) \) goes to zero at large \( r \) faster than \( 1/r^3 \), as stipulated above.

Given the restriction on the potential,\(^4\) this integral will have some finite value that

\(^4\)To be technically accurate, the potential has to also have a finite minimum value at \( r = 0 \) as well as going to zero faster than \( 1/r^3 \) as \( r \to \infty \). The condition at \( r = 0 \) is necessary because if the minimum were \( -\infty \), (as would happen, for instance, with a pure gravitational potential) all the molecules would end up together at one point.
we can call $J$. Equation (18.2.11) on page 18-4 will then be
\[ \int f_{ij} d\vec{q}_i d\vec{q}_j = V \int_0^\infty 4\pi f(r) r^2 dr = VJ \] (18.2.12)

There are, as we know, $N(N-1)/2$ such terms and since $N$ is very large this is essentially $N^2/2$. The configuration integral is then:
\[ Z(N, V, T) = V^N \left( 1 + \frac{1}{2} \frac{N^2J}{V} \right) \] (18.2.13)

If the volume per molecule $v = V/N$ is introduced, this becomes
\[ Z(N, V, T) = N^N v^N \left( 1 + \frac{1}{2} \frac{NJ}{v} \right) \] (18.2.14)

Since we are interested in the thermodynamic properties of this gas, we will want to deal with the thermodynamic properties of the gas. What we have overall is:
\[ Q(N, V, T) = \frac{1}{N!} \left( \frac{2\pi mkT}{\hbar^2} \right)^{3N/2} N^N v^N \left( 1 + \frac{1}{2} \frac{NJ}{v} \right) \] (18.2.15)

The pressure is given by
\[ \beta P = \left( \frac{\partial \ln Q}{\partial V} \right) \]
and when one takes the log of Equation (18.2.15) and does the differentiation, one gets:
\[ \beta P = \frac{1}{v} \left[ 1 - \frac{J}{2v} \right] \] (18.2.16)

which looks like van der Waals equation, but isn’t quite yet.

What we want is for $J = (\beta a - b)$, where $a$ and $b$ are the van der Waals constants. The question is: is there any actual potential that can give $J$ in this form?

The answer is yes. Mayer and Mayer demonstrated one such a potential.\(^6\) The potential is:
\[ u(r) = \begin{cases} \infty & 0 \leq r \leq r_o \\ -u_o \left( \frac{r_o}{r} \right)^m & r_o \leq r \leq \infty \end{cases} \] (18.2.17)

where $m$ is a positive integer greater than 3. With this choice the integral we have to do (see Equation (18.2.11) on page 18-4) is:
\[ \int_0^\infty 4\pi f(r) r^2 dr = \int_0^{r_o} 4\pi r^2 dr + \int_{r_o}^\infty 4\pi \left[ e^{-\beta u_o \left( \frac{r_o}{r} \right)^m} - 1 \right] r^2 dr \] (18.2.18)

The first integral on the right is that of a hard sphere and we can define $b$ to be its value:
\[ b = \frac{1}{2} \int_0^{r_o} 4\pi r^2 dr = \frac{2\pi}{3} r_o^3 = 4v_o \] (18.2.19)

\(^5\)This assumes that the molecule is far enough from a wall for the wall not to interfere. Since this will be true for all but a vanishingly small fraction of the molecules, it is a reasonable assumption.

\(^6\)Mayer and Mayer, *Statistical Mechanics*, John Wiley and Sons, 1940, page 267ff
where $v_o$ is the volume of a sphere with a radius of $r_o$, i.e.:

$$v_o = \frac{4\pi}{3} \left( \frac{r_o}{2} \right)^3 = \frac{\pi}{6} r_o^3$$

In the second integral we can expand the exponential in a power series and keep only the first term:

$$e^{-\beta u(r)} - 1 \approx -\beta u(r) = \beta u_o r_o^m r^{-m}$$

and then we can define $a$ to be the value of the second integral

$$\beta a = \frac{1}{2} 4\pi \beta u_o r_o^m \int_{r_o}^{\infty} r^{-(m-2)} dr = \frac{2\pi \beta}{m-3} u_o r_o^3 = \frac{12\beta}{m-3} u_o v_o$$

which not only gives us a value of $a$

$$a = \frac{12}{m-3} u_o v_o$$

but lets us see exactly why the potential has to fall off to zero with $r$ faster than $1/r^3$.

The derivation is, of course, flawed. Two errors were made that luckily cancel out (though we’ve not proven that). A correct, but more complex derivation is possible and will be given.
Chapter 19

Real Gases

19.1 Introduction

Attempts to produce an equation of state suitable for real gases has been over the years both continuous and unavailing. No such universal equation has ever been demonstrated, though all sorts of more or less approximate (and useful) equations have been developed.

The problem lies in the two-phase region below the critical temperature as seen in a $p - V$ diagram for a “typical” real gas. In this region liquid and gas coexist and any isotherm in this region is flat. However, once this region is left, say beyond the point where liquid disappears and only gas remains, the isotherm is nowhere flat.

Thus, while the isotherm itself is continuous, its slope is not, having a finite discontinuity at the boundary of the two-phase region.

It is not really possible to represent this sort of behavior using finite numbers of elementary functions such as powers, trigonometric functions, or logs and exponentials. None of these have a finite region in which their slope is zero. Indeed, none of these have any points at which they show a finite discontinuity either.

Thus one is forced to the conclusion that no equation of state can exist for real gases that contains a finite number of elementary functions alone as components.

Thus modern attempts to produce useful equations of state for real gases focus on providing “realistic” behavior only in special regions such as around the critical point or at high pressures.

There is a way out of this dilemma. What if one considers an infinite number of terms? Perhaps that would at least provide better approximations over a wider range of temperatures and volumes. Infinite series can, for example, approximate
behavior such as a function with a ‘flat’ region. One only need think of the expansion of a square-wave in a Fourier transform.

The idea is simple. Let

$$\beta p = F(N, V, \beta)$$

be the actual (unknown) equation of state for a real gas. Then we can expand this in a power series in the number density $\rho = N/V$ of the gas:

$$\beta p = \rho + B_2 \rho^2 + B_3 \rho^3 + \cdots$$

(19.1.2)

where the $B$’s are known as virial coefficients. These virial coefficients are generally functions of the temperature but not the volume. If the density $\rho$ is allowed to go to zero, what is recovered is:

$$\frac{p}{kT} = \rho = \frac{N}{V}$$

(19.1.3)

which is the ideal gas law.

The virial coefficients $B_2$, $B_3$, etc., are really thermodynamic functions evaluated at zero density. This can be seen by looking at a Taylor series development of the compressibility factor $Z$

$$Z = \frac{p}{\rho kT}$$

(19.1.4)

around $\rho = 0$:

$$Z = 1 + \left( \frac{\partial Z}{\partial \rho} \right)_{\rho=0} \rho + \frac{1}{2!} \left( \frac{\partial^2 Z}{\partial \rho^2} \right)_{\rho=0} \rho^2 + \cdots + \frac{1}{n!} \left( \frac{\partial^n Z}{\partial \rho^n} \right)_{\rho=0} \rho^n + \cdots$$

(19.1.5)

Since division of Equation (19.1.2) by $\rho$ gives Equation (19.1.5), we can see that the virial coefficients are given by:

$$B_n = \frac{1}{(n-1)!} \left( \frac{\partial^{n-1} Z}{\partial \rho^{n-1}} \right)_{\rho=0}$$

(19.1.6)

So as already said, the virial coefficients are in principle calculable from thermodynamics. However this cannot be done without a theory giving us $Z$ as a function of $\rho$, or experimental data sufficiently accurate to separate out the various derivatives.

What we now set out to do is to find a way to compute the virial coefficients from microscopic properties. This will be done in the next sections.

---

1. This is an expansion in reciprocal powers of the volume. It is also possible to do an expansion in powers of the pressure. The two series are related as almost any Physical Chemistry textbook will show.

2. Which shows that the coefficient of the first term in Equation (19.1.2) must be 1. This is the first virial coefficient $B_1$. Given that it is always 1 for any gas (or that gas would never obey the ideal gas law at low densities, it is always omitted, thus causing virial coefficients to start with the second one and $B_1$ is silently ignored, a fact often causing much concern to undergraduates.

3. Experimental data can be used to find $B_2$ roughly and to find $B_3$ approximately in a few cases. See Dymond and Smith, *The Virial Coefficients of Gases: A Critical Compilation*, Oxford, 1969. Errors in the second virial coefficient are typically of the order of a percent and errors in the third virial coefficient are generally greater.
19.2 Virial Coefficients and Configuration Integrals

We will restrict ourselves to single component systems that possess a virial expansion. The discussion below is long, not because the derivation is difficult, but because we must take many side roads to get to our destination. I will try to point out which discussions are side roads and which are not.

The most convenient approach is via the grand canonical partition function. We have:

$$\Xi(\lambda, V, T) = e^{pV/kT} = \sum_{N=0}^{\infty} Q(N, V, T)\lambda^N = 1 + \sum_{N=1}^{\infty} Q(N, V, T)\lambda^N \quad (19.2.1)$$

where \(\lambda = \exp(\beta\mu)\). This equation is a power series in \(\lambda\) which we have called the absolute activity.

This expansion will be crucial in what follows. However, we must take the first side road here. In the material to come it will be useful to use another activity \(z\) which satisfies:

$$\lim_{\rho \to 0} z = \rho \quad (19.2.2)$$

where \(\rho\) is the number density \(N/V\). So we take the first side road to find out how to do this. Mentally bookmark Equation (19.2.1); we shall return to it below.

We can define an appropriate \(z\) in this way: We have in general

$$\ln \Xi = \beta pV = \ln [1 + Q(1, V, T)\lambda + Q(2, V, T)\lambda^2 + \cdots] \quad (19.2.3)$$

If we let \(\lambda\) go to zero this becomes:

$$\ln \Xi = \beta pV = \ln [1 + Q(1, V, \beta)\lambda] = Q(1, V, \beta)\lambda \quad (19.2.4)$$

since \(\ln(1 + x) = x\) for small \(x\). Since

$$N = \lambda \left( \frac{\partial \ln \Xi}{\partial \lambda} \right) = Q(1, V, T)\lambda = \beta pV \quad (19.2.5)$$

then we can take

$$z = \frac{Q(1, V, T)\lambda}{V} \quad (19.2.6)$$

The quantity \(Q(1, V, \beta)\) is the one-particle canonical partition function. It will occur often so we shall abbreviate it as \(Q_1\). As such it contains all of the internal motions of the molecules. If we are dealing with atoms, \(Q(1, V, \beta) = V/\Lambda^3\) and \(z = (\lambda/\Lambda^3)\), which clearly goes to zero as \(\lambda\) goes to zero. For molecules both rotational and vibrational contributions will be present. For quantum gases \(Q(1, V, T)\) will take other forms. But in any event using \(Q(1, V, T)\) makes the following theory applicable to many different sorts of gases.

\(^4\)In the following I generally follow the treatment given by T.L. Hill, *Introduction to Statistical Thermodynamics*, Dover, New York, 1986, which is a reprint of the second printing of the 1960 edition published by Addison-Wesley.
We now return to Equation (19.2.1) on page 19-3. We can convert it into a power series in $z$ by replacing $\lambda$ with its equivalent $zQ_1/V$ from Equation (19.2.6) on page 19-3:

$$\Xi = e^{pV/kT} = 1 + \sum_{N=1}^{\infty} \left[ \frac{Q(N,V,T)V^N}{Q_1^N} \right] z^N$$  \hspace{1cm} (19.2.7)

If we now define $Z_N(V,T) = Z_N$ to be:

$$\frac{Z_N}{N!} = \frac{Q(N,V,T)V^N}{Q_1^N}$$  \hspace{1cm} (19.2.8)

We can rewrite Equation (19.2.1) on page 19-3 as:

$$\Xi = e^{pV/kT} = 1 + \sum_{N=1}^{\infty} \frac{Z_N}{N!} z^N$$  \hspace{1cm} (19.2.9)

This is always a valid expansion since $e^x = 1 + x + x^2/2! + \cdots$ is an absolutely convergent expansion for any value of $x$, real or imaginary.

In classical mechanics $Z_N$ is simply the configuration integral as can be seen by solving Equation (19.2.8) for $Q(N,V,T)$ and replacing $Q_1$ by $V/\Lambda^3$. This gives:

$$Q(N,V,\beta) = \frac{1}{N!} \left( \frac{V}{\Lambda^3} \right)^N Z_N$$

where the first two factors come from an integration of $\exp(-\beta H)$ over the momentum coordinates (which can always be done for a gas) and the configuration integral part from the integration over the spatial coordinates.

Up to this point we have done nothing but some algebra. Equation (19.2.9) is identical to Equation (19.2.1) on page 19-3 except for changes in variable ($z$ for $\lambda$) and changes in notation.

What we are now going to do is the tricky part of all of this. We are going to expand the grand partition function $\Xi$ in a power series in $z$.

Sadly, this cannot be done directly as the appropriate series do not converge. What one gets is, however the correct result due to cancellation of large terms. We did things just this way in the discussion of the van der Waals equation in a previous Chapter. Now we shall do it correctly.

When we have the appropriate power series expansion in $z$ we will then compare the terms in that series with the terms in Equation (19.2.9). Since the terms in Equation (19.2.9) are in principle known, we can then know the terms in the $pV/kT$ expansion. Those terms will be the virial coefficients while the terms we already have are known through the configuration integral.

Of course we still have to figure out how to integrate the appropriate configuration integrals, but one thing at a time.

We begin by expanding $\exp(\beta pV)$ in a power series. This can always be done since the series for $\exp(x)$ is convergent for all values of $x$. Then:

$$e^{\beta pV} = 1 + (\beta pV) + \frac{1}{2!} (\beta pV)^2 + \cdots + \frac{1}{j!} (\beta pV)^j + \cdots$$  \hspace{1cm} (19.2.10)
Now I assume that $\beta p$ has a convergent power series expansion given by:

$$ \beta p = \sum_{j=1}^{\infty} b_j z^j $$  \hspace{1cm} (19.2.11)

This is a reasonable assumption since $\beta p$ will properly go to zero as $z$ goes to zero. Note that there is no constant term in Equation (19.2.11), a fact that will become important later.

This expansion is then plugged into Equation (19.2.10) on page 19-4 to give:

$$ e^{pV/kT} = 1 + V \sum_{j=1}^{\infty} b_j z^j + \frac{1}{2!} V^2 \left( \sum_{j=1}^{\infty} b_j z^j \right)^2 + \frac{1}{3!} V^3 \left( \sum_{j=1}^{\infty} b_j z^j \right)^3 + \cdots $$

$$ + \frac{1}{m!} V^m \left( \sum_{j=1}^{\infty} b_j z^j \right)^m + \cdots $$ \hspace{1cm} (19.2.12)

What we’d like is an expansion in powers of $z$. This equation is that, but the powers of $z$ are scattered about in many terms. It is not particularly easy to see how those terms are gathered together. It helps to write the equation out in the form:

$$ e^{pV/kT} = 1 + V \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] $$

$$ + \frac{1}{2!} V^2 \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] $$

$$ + \frac{1}{3!} V^3 \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] $$

$$ \times \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] + \cdots $$ \hspace{1cm} (19.2.13)

and then to multiply out the terms to see what they actually are:

$$ e^{pV/kT} = 1 + V \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] $$

$$ + \frac{1}{2!} V^2 \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] $$

$$ + \frac{1}{3!} V^3 \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] $$

$$ \times \left[ b_1 z + b_2 z^2 + b_3 z^3 + \cdots \right] + \cdots $$ \hspace{1cm} (19.2.14)

Finally we collect like powers of $z$. As often happens this can be done fairly easily for the first few terms since low powers of $z$ occur only in the first few products.

$$ e^{pV/kT} = 1 + V b_1 z + \left[ V b_2 + \frac{V^2}{2} b_2^2 \right] z^2 + \left[ V b_3 + 2 \frac{V^2}{2} b_1 b_2 + \frac{V^3}{6} b_1^3 \right] z^3 + \cdots $$ \hspace{1cm} (19.2.15)

It should be noted that this gathering of terms would not be possible if Equation (19.2.11) had a constant term. In that case each term in Equation (19.2.13) beyond the first would contain a single power of $z$. Similar all would contain $z^2$, $z^3$, etc., and the result in Equation (19.2.15) would not exist.

We now need another digression. Again, mentally bookmark Equation (19.2.15) as we will return to it.
It is not easy to see in general what the form of the various terms in Equation (19.2.15) on page 19-5 will be. In fact it is a very difficult. The coefficient of \( z^N \) turns out to be

\[
\sum \prod_{j=1}^N \left( \frac{(Vb_j)^{m_j}}{m_j!} \right) \tag{19.2.16}
\]

where the sum is done over all sets of integers \( m_j \) that satisfy the condition

\[
\sum_{j=1}^N jm_j = N \tag{19.2.17}
\]

There is a rather elegant method for computing these numbers using graphs. These are mathematical structures that can be created for various situations such as this one that enables one to transform a problem in, for example, algebra, to one of counting configurations.\textsuperscript{5} We shall not go into it here. Suffice it to say that what we have done is enough to compute the second and third virial coefficients which is about the practical limit of utility.

We now return to Equation (19.2.15) on page 19-5. What we do next is to compare that equation to Equation (19.2.9) on page 19-4. To see the result let’s write out Equation (19.2.9) on page 19-4

\[
e^{pV/kT} = 1 + Z_1 1! z + Z_2 2! z^2 + Z_3 3! z^3 + \cdots \tag{19.2.18}
\]

And then we shall recall that the power series representation of a given function, here \( \exp (\beta pV) \), in a given set of variables (here \( p, V, T \), and \( z \)) is unique. Thus the right-hand sides of Equations (19.2.18) and (19.2.15) on page 19-5 must be identical. That gives us:

\[
V b_1 = \frac{Z_1}{1!}
\]

\[
V b_2 + \frac{V^2}{2} b_1^2 = \frac{Z_2}{2!}
\]

\[
V b_3 + 2 \frac{V^2}{2} b_1 b_2 + \frac{V^3}{6} b_1^3 = \frac{Z_3}{3!} \tag{19.2.19}
\]

We assume that the \( Z \)’s are known, so we actually want the \( b \)’s in terms of the \( Z \)’s. Since we can start at the top of the list and solve for the \( b \)’s as we go, turning this around isn’t hard. We get:

\[
1! V b_1 = Z_1
\]

\[
2! V b_2 = Z_2 - Z_1^2
\]

\[
3! V b_3 = Z_3 - 3Z_1 Z_2 + 2Z_1^3 \tag{19.2.20}
\]

These simplify a bit since the configuration integral \( Z_1 \) is just \( V \), the volume. That makes \( b_1 = 1 \).

---

\textsuperscript{5}The method is known as Mayer Cluster Theory. An introductory explanation of it is given in T.L. Hill, \textit{An Introduction to Statistical Thermodynamics}, Dover, New York, 1986.
Our last task is to convert all of this to an expansion in powers of the number density \( \rho \) instead of \( z \). Then we will have the virial coefficients \( B_n \) directly. This isn’t hard to do. We recall that
\[
N = z \left( \frac{\partial \ln \Xi}{\partial z} \right) \quad \text{or} \quad \rho = z \left( \frac{\partial p/kT}{\partial z} \right) \tag{19.2.21}
\]
Since
\[
\beta p = \sum_{j=1}^{\infty} b_j z^j \tag{19.2.11}
\]
this results in
\[
\rho = \sum_{j=0}^{\infty} j b_j z^j \tag{19.2.22}
\]
What we want is
\[
z = \rho + a_2 \rho^2 + a_3 \rho^3 + \ldots \tag{19.2.23}
\]
We now substitute Equation (19.2.23) into Equation (19.2.22) and equate coefficients on both sides of the equation. This leads to:
\[
\begin{align*}
a_2 &= -2b_2 \\
a_3 &= -3b_3 - 4a_2 b_2 = -3b_3 + 8b_2^3 \\
\ldots
\end{align*} \tag{19.2.24}
\]
so that in the end we have for the virial coefficients
\[
B_2 = -b_2 \quad B_3 = 4b_2^2 - 2b_3 \quad \text{etc.} \tag{19.2.25}
\]
Putting this all together we have the virial coefficients \( B_n \) given in terms of the expansion coefficients \( b_j \) in Equation (19.2.25). And then we have the expansion coefficients \( B_j \) given in terms of the configuration integrals \( Z_n \) in Equation (19.2.20) on page 19-6. While the configuration integrals are defined in Equation (19.2.8) on page 19-4, what remains to be done is to be more explicit about the potential energy of a system of interacting particles and to show explicit formulas for evaluating the second and third virial coefficients.\(^6\)

### 19.3 Evaluating the Integrals

First we make some assumptions about the potential energies involved in these calculations. We will always assume that the potentials are *pairwise additive*. That means that all potentials, no matter how many particles are involved, can be expressed in terms of sums of potential energies involving two particles.

Thus if three particles, 1, 2, and 3, are interacting, the potential energy \( V(\vec{q}_1, \vec{q}_2, \vec{q}_3) \) is the sum of three potential energies of the form: \( u(\vec{q}_i - \vec{q}_j) \):
\[
V(\vec{q}_1, \vec{q}_2, \vec{q}_3) = u(\vec{q}_2 - \vec{q}_1) + u(\vec{q}_3 - \vec{q}_1) + u(\vec{q}_3 - \vec{q}_2) \tag{19.3.1}
\]
\(^6\)I’ve chosen not to go further because higher virial coefficients are not only hard to measure experimentally, they are also hard to compute numerically.
The same sort of formula will hold true no matter how many particles are involved in an interaction.

We will also assume that the intermolecular potentials have a hard core. That is, the potential energy never drops to negative infinity. If it did so for some value of the intermolecular distance, all the molecules would pile up together with a total energy of minus infinity.

In addition we will assume that the intermolecular potential energy is short ranged. That is, it goes to zero faster than \( 1/r^3 \) and does so over a fairly short distance.

Last, we will make an assumption that makes the calculations much easier, but which is not really necessary: that the intermolecular potentials are spherically symmetric.

What this means is that

\[
V(\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_N) = u(r_{12}) + u(r_{13}) + \ldots + u(r_{1N}) + u(r_{23}) + u(r_{24}) + \ldots + u(r_{2N}) + \ldots + u(r_{N-2,N}) + u(r_{N-1,N})
\]

or, for short

\[
V(\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_N) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} u(r_{ij}) \tag{19.3.2}
\]

There are \( N(N-1)/2 \) terms in these last two equations.

The configuration integral \( Z_1 \) is easy to calculate. With only one particle there can be no potential energy and the integral simply gives \( V \). So we have:

\[
Z_1 = \int_V d\vec{r}_1 = V \tag{19.3.4}
\]

Now from Equation (19.2.25) on page 19-7 we have that \( B_2 = -b_2 \), and from Equation (19.2.20) on page 19-6 we have that \( b_2 = (Z_1^2 - Z_2) / 2V \). Or, using \( Z_1 = V \), get

\[
B_2 = -b_2 = -\frac{Z_2 - Z_1^2}{2V} \tag{19.3.5}
\]

Since \( Z_2 \) is:

\[
Z_2 = \int_{V'} e^{-\beta u(r_{12})} d\vec{r}_1 d\vec{r}_2 \tag{19.3.6}
\]

I can write Equation (19.3.6) as:

\[
B_2 = -\frac{Z_2 - Z_1^2}{2V} = -\frac{1}{2V} \int_{V'} \left[ e^{-\beta u(r_{12})} - 1 \right] d\vec{r}_1 d\vec{r}_2 \tag{19.3.7}
\]

where the one comes from writing the square of the volume \( V^2 \) as:

\[
V^2 = \int_{V'} d\vec{r}_1 d\vec{r}_2 \tag{19.3.8}
\]
In Equation (19.3.7) on page 19-8 we can change the spatial coordinates from the coordinates of each of the two particles to the coordinates of the center of mass of the two particles and the relative spherical coordinates \((r_{12}, \theta, \phi)\) of the two particles.

Integration over the center of mass coordinates is easy. They do not occur in the potential energy so they contribute another factor of \(V\) to the overall integral. As for the relative coordinates, neither of the two orientation angles is present in the potential energy either. They contribute a factor of \(4\pi\). We are then left with:

\[
B_2 = -\frac{1}{2} \int_0^\infty 4\pi r_{12}^2 \left[ e^{-\beta u(r_{12})} - 1 \right] dr_{12} \tag{19.3.9}
\]

For any reasonable interaction potential this integral cannot be done in closed form.\(^7\) However, it is not at all difficult to do either by series expansion or numerical integration. Which method is chosen depends on the actual potential energy \(u(r_{12})\).

But for some very simple potentials the integral can be done.

---

Example 19.1

**Trivial Example:** If there is no potential energy at all, then \(u(r_{12})\) is identically zero, \(\exp(-\beta u(r_{12}))\) is then 1 and \(B_2\) simply becomes 0. To the approximation of just the second virial coefficient\(^8\) the ideal gas law is then obeyed.

We can also rather simply do this example:

---

Example 19.2

**Hard Sphere Gas:** Here the potential is

\[
u(r) = \begin{cases} 
\infty & 0 \leq r \leq \sigma \\
0 & \sigma < r < \infty 
\end{cases}
\]

where \(\sigma\) is the distance of closest approach. \(B_2\) is then:

\[
B_2 = -2\pi \int_0^\sigma (-1) r_{12}^2 dr_{12} - 2\pi \int_\sigma^\infty (1 - 1) r_{12}^2 dr_{12} = \frac{2}{3} \pi \sigma^3 \tag{19.3.10}
\]

where the integral from \(\sigma\) to \(\infty\) contributes nothing to the total.

Since \(\sigma\) is the distance of closest approach of molecules 1 and 2 it is the distance between the centers of those molecules and hence twice the radius of one of them.

---

\(^7\)Yes, I know that this seems to be a big disappointment after all this work, but the situation is not at all hopeless.

\(^8\)Actually, all the virial coefficients will be zero in this case.
If the radius of a single molecule is \( a \), then \( \sigma = 2a \) and we see that \( B_2 \) is four times the volume of a single molecule.

We can also examine the square well potential.

**Example 19.3**

**Square Well Gas:** Here the potential is

\[
    u(r) = \begin{cases} 
    \infty & 0 \leq r \leq \sigma \\
    -\epsilon & \sigma < r \leq w\sigma \\
    0 & w\sigma < r < \infty 
    \end{cases}
\]

where again \( \sigma \) measures the distance of closest approach and \( w \) measures the size of the attractive region in multiples of the distance of closest approach.\(^9\) The depth of the potential well is \(-\epsilon\) so that normally \( \epsilon \) is a positive number.

The integral for \( B_2 \) breaks up into three integrals each covering one of the regions in Equation (19.3.11).

\[
    B_2 = -2\pi \int_0^\sigma (-1)r_{12}^2 dr_{12} - 2\pi \int_{\sigma}^{w\sigma} [e^{\beta\epsilon} - 1] r_{12}^2 dr_{12} - 2\pi \int_{w\sigma}^{\infty} (1 - 1)r_{12}^2 dr_{12} 
\]

\[
    = 2\pi \sigma^3 - 2\pi \sigma^3 [e^{\beta\epsilon} - 1] \left[ \sigma^3 (w - 1) \right] 
\]

\[
    = \frac{2}{3} \pi \sigma^3 \left[ 1 - (e^{\beta\epsilon} - 1) (w - 1) \right] 
\]

(19.3.11)

As checks we note that if \( \epsilon \) is zero or if \( w \) is 1 this reduces to the hard-sphere gas.

This gas also exhibits a **Boyle Temperature** \( T_B \) at which the second virial coefficient vanishes. At this temperature the gas behaves in an “ideal” manner. Any combination of \( w \) and \( \beta\epsilon \) that satisfies the symmetric formula:

\[
    we^{\beta\epsilon} - e^{\beta\epsilon} - w = 0 
\]

(19.3.12)

will cause the second virial coefficient of the Square-Well potential gas to vanish. The only restriction is that \( w \) must be greater than 1 for the potential to make physical sense.

This second virial coefficient is similar in behavior to actually observed second virial coefficients in that it is negative at low temperatures and positive at high ones. The crossover point is, of course, the Boyle temperature. Further, at very high temperatures \( B_2 \) becomes constant, a behavior again observed in experiment. The limiting value for \( B_2 \) can easily be seen to be the hard-sphere value.

\(^9\)Since \( \sigma \) is in fact the diameter of one of our necessarily spherical molecules, \( w \) is also the size in diameters.
19.4 The Third Virial Coefficient

While the second virial coefficient is not too hard to evaluate, the third is another story. Indeed, as the order of the virial coefficients increases, the difficulty in evaluating them increases very fast.

But first we must obtain as simple an expression as possible for the third virial coefficient $B_3$. It is given by:

$$B_3 = 4b_2^2 - 2b_3$$

which looks simple enough.

Before going ahead it is good to collect the appropriate formulas in one place: We have:

$$Z_1 = \int dr_1$$
$$Z_2 = \int\int e^{-\beta u(r_{12})} dr_1 dr_2$$
$$Z_3 = \int\int\int e^{-\beta u(r_{12}) - \beta u(r_{13}) - \beta u(r_{23})} dr_1 dr_2 dr_3$$

where the $r$'s represent the three coordinates of a given particle, and

$$1!Vb_1 = Z_1$$
$$2!Vb_2 = Z_2 - Z_1^2$$
$$3!Vb_3 = Z_3 - 3Z_1Z_2 + 2Z_1^3$$

It is helpful to shorten the notation slightly by writing

$$x_{ij} = e^{-\beta u(r_{ij})}$$

With these things in mind we can write

$$3!Vb_3 = \int\int\int [x_{12}x_{13}x_{23} - 3(1)x_{12} + 2(1)^3] dr_1 dr_2 dr_3$$

where the first term is $Z_3$, the second is three times $Z_1$ times $Z_2$, and the third is 2 times $Z_1^2$.

We can make this look more symmetric by realizing that integrating over $x_{12}$ gives exactly the same result as integrating over $x_{13}$ or $x_{23}$. So we can rewrite Equation (19.4.5) as:

$$3!Vb_3 = \int\int\int [x_{12}x_{13}x_{23} - x_{12} - x_{13} - x_{23} - 2] dr_1 dr_2 dr_3$$

A glance at the definition of $B_3$ above shows that twice $b_3$ must be subtracted from four times $b_2^2$. It is convenient to get a simple expression for $b_2^2$ in the following way:
We note that
\[ 2!Vb_2^2 = \iiint (x_{12} - 1)(x_{13} - 1)dr_1dr_2dr_3 \]
\[ = V \left[ \int (x_{12} - 1)dr_{12} \right] \times \left[ \int (x_{13} - 1)dr_{13} \right] \quad (19.4.7) \]

Since the equation for \( b_3 \) is multiplied by three already and that for \( b_2^2 \) above is already multiplied by two, we end up using the form
\[ 3Vb_3 = 12Vb_2^2 - 6Vb_3 \]

Putting this all together gives:
\[ 3Vb_3 = \iiint ((x_{12} - 1)(x_{13} - 1) + (x_{12} - 1)(x_{23} - 1) + (x_{13} - 1)(x_{33} - 1) \]
\[ - [x_{12}x_{13}x_{23} - x_{12} - x_{13} - x_{23} - 2]dr_1dr_2dr_3 \quad (19.4.8) \]

Multiplying and canceling the terms that cancel out,\(^{10}\) we end up with:
\[ 3Vb_3 = -\iiint (x_{12} - 1)(x_{13} - 1)(x_{23} - 1)dr_1dr_2dr_3 \quad (19.4.9) \]

which really isn’t so bad.\(^{11}\)

If we switch from \( dr_1dr_2dr_3 \) to coordinates relative to particle 1, i.e. \( dr_1dr_{12}dr_{13} \) we can then integrate over particle 1 and get another factor of the volume \( V \). We then have:
\[ B_3 = -\frac{1}{3} \iiint (x_{12} - 1)(x_{13} - 1)(x_{23} - 1)dr_{12}dr_{13} \quad (19.4.10) \]

This is very difficult to integrate. To do so one must not only express \( \bar{r}_{23} \) in terms of \( \bar{r}_{12} \) and \( \bar{r}_{13} \) (actually, their components)\(^{12}\) but in addition then integrate the exponentials containing the potential energy. This is especially difficult for the term arising from \( u(r_{23}) \).

## 19.5 The Lennard-Jones Potential

A very common intermolecular pairwise potential energy is one given by Lennard-Jones.\(^{13}\) It is given by:
\[ u(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] = \epsilon \left[ \left( \frac{r_o}{r} \right)^{12} - 2 \left( \frac{r_o}{r} \right)^6 \right] \quad (19.5.1) \]

\(^{10}\) Though students (and even faculty) sometimes end up canceling terms that really don’t in messy expressions. This is a cause of great angst and wailing.

\(^{11}\) One cannot look at Equation (19.4.9) without thinking that such a beautiful symmetric result cannot accidentally be the result of all that manipulation that preceded it. There must be another way to get it that emphasizes the symmetry. And there is. The method is called Mayer Cluster Theory and it is a way of getting directly to the virial coefficients in terms as simple as they can be. Sadly the theory itself is a bit complex. It is not included here because we (and most people) do not need to go past the third virial coefficient and because it is a subject worthy of an appendix all its own.

\(^{12}\) Readers are invited to try this for themselves.

\(^{13}\) Who is, in fact, a single person.
The potential starts at $\infty$ at $r = 0$, falls to 0 at $r = \sigma$, reaches a minimum of $-\epsilon$ at $r = r_a$ and then rises to zero as $r \to \infty$. The relationship between $\sigma$ and $r_a$ is $r_a = \frac{1}{6}\sigma$. Either form in Equation (19.5.1) on page 19-12 may be used and both forms can be found in the literature.

With this form it turns out that $B_2$ is negative at low temperatures, rises to positive values as the temperature is increased, and then reaches a maximum and falls off slightly as the temperature is increased further. This agrees with what is observed experimentally.

The second virial coefficient for the Lennard-Jones potential cannot be obtained in a simple form. It however can be obtained exactly in a series representation that converges fairly rapidly.

The second virial coefficient is given by:

$$B_2 = -2\pi \int_0^{\infty} \left[ e^{-\beta u(r)} - 1 \right] r^2 dr \quad (19.5.2)$$

We begin by obtaining a modified form for the second virial coefficient. We integrate by parts. Thus we have

$$\int_a^b U dV = UV \bigg|_a^b - \int_a^b V dU \quad (19.5.3)$$

and we choose

$$U = e^{-\beta u} - 1 \quad \text{and} \quad dV = r^2 dr \quad (19.5.4)$$

then

$$dU = -\beta \left( \frac{du}{dr} \right) e^{-\beta u} dr \quad \text{and} \quad V = \frac{r^3}{3} \quad (19.5.5)$$

Now if $u(r) \to 0$ faster than $r^{-3}$ as $r \to \infty$, then $UV \to 0$ as $r \to \infty$. Since $UV$ is zero at $r = 0$, the term $UV$ in Equation (19.5.5) is then zero at both limits and hence vanishes. What is left is:

$$B_2 = -\frac{2}{3} \pi \beta \int_0^{\infty} \left( \frac{du}{dr} \right) e^{-\beta u} r^3 dr \quad (19.5.6)$$

With Equation (19.5.1) on page 19-12 (in the $\sigma$ form inserted into this we have:

$$B_2 = -\frac{8 \beta \pi \epsilon}{3 \sigma} \int_0^{\infty} \left[ -12 \left( \frac{\sigma}{r} \right)^{13} + 6 \left( \frac{\sigma}{r} \right)^7 \right] \exp \left[ -4\beta \epsilon \left( \frac{\sigma}{r} \right)^{12} \right] \exp \left[ 4\beta \epsilon \left( \frac{\sigma}{r} \right)^6 \right] r^2 dr \quad (19.5.7)$$

It is exceptionally convenient to introduce new variables and combination of constants. In particular we will let

$$x = r/\sigma, \quad \theta = 1/\beta \epsilon, \quad \text{and} \quad \alpha = 4/\theta \quad (19.5.8)$$

---

14In the following I use the development presented by Donald Rapp, *Statistical Mechanics*, Holt, Rinehart, and Winston, 1972. I’ve also worked it out in detail, something that is rather hard to find in most textbooks.
Further, the quantity
\[ B^*_2 = \frac{B_2}{\frac{2}{3} \pi \sigma^3} \]  
(19.5.9)
is known as the reduced second virial coefficient and is in common use in the literature. With these substitutions we get:
\[ B^*_2 = -\alpha \int_0^\infty \left[ -12x^{-10} + 6x^{-4} \right] e^{-\alpha x^{-12}} e^{\alpha x^{-6}} dx \]  
(19.5.10)
The integral is evaluated by expanding the second exponential in a power series thusly:
\[ e^{\alpha x^{-6}} = \sum_{n=0}^{\infty} \frac{1}{n!} \alpha^n x^{-6n} \]  
(19.5.11)
We now insert this into Equation (19.5.10) in place of the appropriate exponential to get
\[ B^*_2 = -\alpha \int_0^\infty \left[ -12x^{-10} + 6x^{-4} \right] e^{-\alpha x^{-12}} \sum_{n=0}^{\infty} \frac{1}{n!} \alpha^n x^{-6n} dx \]  
(19.5.12)
We can interchange the integration and the summation and then break the result into two integrals:
\[ B^*_2 = 12 \sum_{n=0}^{\infty} \frac{1}{n!} \alpha^{n+1} \int_0^\infty x^{-6n-10} e^{-\alpha x^{-12}} dx \]
\[ -6 \sum_{n=0}^{\infty} \frac{1}{n!} \alpha^{n+1} \int_0^\infty x^{-6n-4} e^{-\alpha x^{-12}} dx \]  
(19.5.13)
Now we need a small digression. The integrals in Equation (19.5.13) can be done in terms of the Gamma function \( \Gamma(z) \). This is why we’ve been going through all these contortions and why we’ve expanded one exponential and not the other. The Gamma function is a well-known function that is fully tabulated in many places and is closely related to the factorial function.

What we have above is an integral that can be written generally as:
\[ I = \int_0^\infty r^{-\gamma} e^{-c/r^b} dr \]  
(19.5.14)
The Gamma Function is defined to be
\[ \Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \]  
(19.5.15)
where the real part of \( z \) must be positive. We need to get the integral in Equation (19.5.14) into the form of Equation (19.5.15). This isn’t hard. We take the blunt approach and let
\[ t = cr^{-b} \text{ thus } r = \left( \frac{c}{t} \right)^{1/b} \text{ and } dr = -\frac{1}{b} \left( \frac{c}{t} \right)^{1/b} t^{-1} dt \]  
(19.5.16)
Substituting we get:
\[
I = - \int_{\infty}^{0} \left( \frac{c}{t} \right)^{a/b} e^{-t} \frac{1}{b} \left( \frac{c}{t} \right)^{1/b} \frac{1}{t} dt
\]  
(19.5.17)

which, after some algebra, simplifies to
\[
I = \frac{1}{b} c^{(1-a)/b} \int_{0}^{\infty} t^{(a-1)/b-1} e^{-t} dt
\]  
(19.5.18)

where the integral is now in the form of Equation (19.5.14) on page 19-14 if we let 
\[ z = \frac{(a-1)}{b}. \] 
Thus
\[
I = \frac{1}{b} c^{(1-a)/b} \Gamma \left( \frac{a-1}{b} \right)
\]  
(19.5.19)

With the digression over we can evaluate the integrals in Equation (19.5.13) on page 19-14:
\[
\int_{0}^{\infty} x^{-6n-10} e^{-ax^{-12}} dx = \frac{1}{12} \alpha^{-\frac{(6n+9)}{12}} \Gamma \left( \frac{6n+9}{12} \right)
\]  
(19.5.20)
\[
\int_{0}^{\infty} x^{-6n-4} e^{-ax^{-12}} dx = \frac{1}{12} \alpha^{-\frac{(6n+3)}{12}} \Gamma \left( \frac{6n+3}{12} \right)
\]  
(19.5.21)

and insert the results into Equation (19.5.13) on page 19-14:
\[
B^{*}_{2} = 12 \sum_{n=0}^{\infty} \frac{1}{n!} \alpha^{n+1} \frac{1}{12} \alpha^{-(2n+3)/4} \Gamma \left( \frac{2n+3}{4} \right)
\]
\[ -6 \sum_{n=0}^{\infty} \frac{1}{n!} \alpha^{n+1} \frac{1}{12} \alpha^{-(2n+1)/4} \Gamma \left( \frac{2n+1}{4} \right)
\]  
(19.5.22)

This simplifies slightly:
\[
B^{*}_{2} = \sum_{n=0}^{\infty} \frac{1}{n!} \alpha^{(2n+1)/4} \Gamma \left( \frac{2n+3}{4} \right) - \sum_{n=0}^{\infty} \frac{1}{n!} \frac{1}{2} \alpha^{(2n+3)/4} \Gamma \left( \frac{2n+1}{4} \right)
\]  
(19.5.23)

What I am now trying to do is to combine the two sums into one to give us a more compact formula. To do this I need to manipulate Equation (19.5.23) a bit. We play with each sum separately. First I’ll split out the first \((n = 0)\) term from the first sum:
\[
B^{*}_{2} = \alpha^{1/4} \Gamma \left( \frac{3}{4} \right) + \sum_{n=1}^{\infty} \frac{1}{n!} \alpha^{(2n+1)/4} \Gamma \left( \frac{2n+3}{4} \right)
\]
\[ -\sum_{n=0}^{\infty} \frac{1}{n!} \frac{1}{2} \alpha^{(2n+3)/4} \Gamma \left( \frac{2n+1}{4} \right)
\]  
(19.5.24)

and then I’ll re-index the second sum by setting \(k = n + 1\):
\[
B^{*}_{2} = \alpha^{1/4} \Gamma \left( \frac{3}{4} \right) + \sum_{n=1}^{\infty} \frac{1}{n!} \alpha^{(2n+1)/4} \Gamma \left( \frac{2n+3}{4} \right)
\]
\[ -\sum_{k=1}^{\infty} \frac{k}{k!} \frac{1}{2} \alpha^{(2k+1)/4} \Gamma \left( \frac{2k-1}{4} \right)
\]  
(19.5.25)
Now, realizing that $k$ can be renamed $n$ we can factor out the terms in $\alpha$ and $n!$ to get:

$$B_2^* = \alpha^{1/4} \Gamma\left(\frac{3}{4}\right) + \sum_{n=1}^{\infty} \frac{1}{n!} \alpha^{(2n+1)/4} \left[ \Gamma\left(\frac{2n+3}{4}\right) - \frac{n}{2} \Gamma\left(\frac{2n-1}{4}\right) \right]$$  \hspace{1cm} (19.5.26)

The term in the square brackets can be simplified if we use a property of the gamma function

$$\Gamma(z + 1) = z\Gamma(z)$$  \hspace{1cm} (19.5.27)

which is easy to prove using Equation (19.5.15) on page 19-14. With this we can write:

$$\Gamma\left(\frac{2n+3}{4}\right) = \Gamma\left(\frac{2n-1}{4} + 1\right) = \frac{2n-1}{4} \Gamma\left(\frac{2n-1}{4}\right)$$  \hspace{1cm} (19.5.28)

With this the term in the square brackets in Equation (19.5.26) becomes

$$\Gamma\left(\frac{2n-1}{4}\right) \left[ \frac{2n-1}{4} - \frac{n}{2} \right] = -\frac{1}{4} \Gamma\left(\frac{2n-1}{4}\right)$$

and so we succeed in putting the two sums together:

$$B_2^* = \alpha^{1/4} \Gamma\left(\frac{3}{4}\right) - \sum_{n=1}^{\infty} \frac{1}{4n!} \alpha^{(2n+1)/4} \Gamma\left(\frac{2n-1}{4}\right)$$  \hspace{1cm} (19.5.29)

We can even simplify this by noting that

$$\Gamma\left(\frac{3}{4}\right) = -\frac{1}{4} \Gamma\left(-\frac{1}{4}\right)$$

and if this is used in the first term, we note that it fits the pattern of the sum and is, in fact, unsurprisingly, the zeroth term of it. Thus our final result is the amazingly simple:

$$B_2^* = -\sum_{n=0}^{\infty} \frac{1}{4n!} \left(\frac{4}{\theta}\right)^{(2n+1)/4} \Gamma\left(\frac{2n-1}{4}\right)$$  \hspace{1cm} (19.5.30)

where I’ve substituted $4/\theta$ for $\alpha$ since the explicit temperature dependence is good to know.\(^{15}\)

The result converges quite rapidly for values of $\theta$ of about 4 or more. As $\theta$ gets smaller, more and more terms are needed. And while it looks as if the second virial coefficient is always negative, in fact the first term in Equation (19.5.30) is positive, and at high temperatures that term dominates the others.

Evaluating this series is not a problem on a computer. Given any system interacting via Lennard-Jones potentials, Equation (19.5.30) will provide the second virial coefficient if both $\epsilon$ and $\sigma$ are known to any desired degree of accuracy.

\(^{15}\)Recall that $\theta = kT/\epsilon$. 
Chapter 20

Distribution Functions

There is yet another approach to the entire problem of interacting subsystems. This the the distribution function technique.

A distribution function is a function that gives the chance that one particle is at one particular location in the system while another is at a second location, a third at yet another location, and so on. It turns out that if we knew the distribution function for all the particles in a system we should, in principle, be able to determine all the thermodynamic properties of that system.

Of course, except in trivial cases (perfect crystals for one) knowledge of such a thing is not possible.

We seek instead a simplification of this scheme. The one we will pursue is the common situation where we can decompose the system potential energy into sums of spherically symmetric pair interaction energies.

This manner of thought leads to the definition of distribution functions in terms of the potential energies in a system.

And this approach has one major advantage: the real distribution functions with the actual naturally-occurring potential energies are experimentally observable with “real” experiments such as x-ray and neutron diffraction. And actual distribution functions for systems with assumed potentials of any sort are observable as the results of computer experiments (the first of which go back about half a century.)

This approach also suffers from a major disadvantage: it is difficult to calculate distribution functions directly from the potential energies.
20.1 Defining Distribution Functions

We will consider systems having fixed $N$, $V$, and $T$. The bulk number density is $ho = N/V$ and the positions of particle $i$ in this system will be given by the vector $\vec{r}_i$.

Be warned. The seeming complexity of what follows comes from the paucity of symbols available. Here we have followed common practice which is commonly thought to be unnecessarily complex.

To be more formal we ask for the chance that we will find particle 1 at a position $\vec{r}_1$ in an infinitesimal volume element $d\vec{r}_1$, particle 2 at a position $\vec{r}_2$ in an infinitesimal volume element $d\vec{r}_2$, and so forth up to and including finding particle $N$ at a position $\vec{r}_N$ in an infinitesimal volume element $d\vec{r}_N$. The density for this is given by:

$$P_N(\vec{r}_1,\ldots,\vec{r}_N) = \frac{e^{-\beta U_N}}{Z_N} \tag{20.1.1}$$

where $Z_N$ is the configurational integral

$$Z_N = \int \cdots \int e^{-\beta U_N} d\vec{r}_1 \cdots d\vec{r}_N \tag{20.1.2}$$

and $U_N$ is the the potential energy resulting from having particle 1 at $\vec{r}_1$, etc.

This density is obviously normalized:

$$\int \cdots \int P_N(\vec{r}_1,\ldots,\vec{r}_N) d\vec{r}_1 \cdots d\vec{r}_N = \frac{1}{Z_N} \int \cdots \int e^{-\beta U_N} d\vec{r}_1 \cdots d\vec{r}_N = 1 \tag{20.1.3}$$

which, since $P_N \geq 0$ always, suffices to make $P_N$ a legitimate probability.

We can also define a reduced probability density $P_n$, $1 \leq n \leq N$ (note the lower case $n$):

$$P_n(\vec{r}_1,\ldots,\vec{r}_n) = \int \cdots \int P_N(\vec{r}_{n+1},\ldots,\vec{r}_N) d\vec{r}_{n+1} \cdots d\vec{r}_N \tag{20.1.4}$$

which, from Equation (20.1.3) can be seen also to be normalized to 1.

These are specific probability distributions dealing with the chance of finding particular particles in particular places. Much more important to us$^1$ is the generic distribution function (often called simply the distribution function) that some particle is at $\vec{r}_1$ and that another is at $\vec{r}_2$, etc., no matter what the identities of the particles. Note the subtle difference between this and the specific distribution functions which deal with a particular particle being at $\vec{r}_1$.

Since there are $N!$ different ways to distribute $N$ particles among $N$ positions, there will be $N!$ different $P_N$’s that will give a particular generic distribution. Thus it is natural to define the $N$th order distribution function by:

$$\rho_N(\vec{r}_1,\ldots,\vec{r}_N) = N! P_N(\vec{r}_1,\ldots,\vec{r}_N) \tag{20.1.5}$$

---

$^1$Particularly since particles are often indistinguishable!
since we expect all the $P_N$’s to have the same value.

The distribution function $\rho_N$ is not a probability since it is trivial to show that:

$$\int \cdots \int \rho_N(\vec{r}_1, \ldots, \vec{r}_N)\vec{r}_1 \ldots \vec{r}_N = N!$$  \hspace{1cm} (20.1.6)$$

and not 1.

Just as we have a reduced probability density in Equation (20.1.4) on page 20-2 we can also define a reduced generic distribution function $\rho_n$:

$$\rho_n(\vec{r}_1, \ldots, \vec{r}_n) = \frac{N!}{(N-n)!} P_n(\vec{r}_1, \ldots, \vec{r}_n)$$  \hspace{1cm} (20.1.7)$$

where the $(N-n)!$ in the denominator is due to the fact that the first $n$ particles are fixed in specific conditions.

The simplest of these reduced generic distribution functions is $\rho_1$. This answers the question “is there a particle at $\vec{r}_1$ given that all $N-1$ particles are somewhere in the volume $V$?” A moment’s reflection will show that $\rho_1$ must be a constant since all points in the fluid are equivalent.\(^2\) Thus we have:

$$\int \rho_1(\vec{r}_1)d\vec{r}_1 = \rho_1 \int d\vec{r}_1 = \rho_1 V = N$$  \hspace{1cm} (20.1.8)$$

where the $N$ arises from setting $n = 1$ in Equation (20.1.7) and then making use of the normalization of $P_1$. From this it is easy to see that:

$$\rho_1 = \frac{N}{V} = \rho \quad \text{(fluids only)}$$  \hspace{1cm} (20.1.9)$$

or, in other words, $\rho_1$ is the bulk density of the fluid. This is the answer to the question above, the chance of finding a particle at $\vec{r}_1$ averaged over all other particles is just $\rho_1 = N/V$.

The next distribution function, $\rho_2$ answers the question “what is the chance that there is a particle at $\vec{r}_1$ and a second particle at $\vec{r}_2$?” We can write an integral analogous to Equation (20.1.8):

$$\iint \rho_2(\vec{r}_1, \vec{r}_2)d\vec{r}_1d\vec{r}_2 = N(N-1)$$  \hspace{1cm} (20.1.10)$$

That’s not too useful as it stands, but we can relate it to the bulk density $\rho$ with a small digression.

**Small Digression:** If there was no potential energy at all in this system, that is, if we had a system of $N$ non-interacting point particles, then the probability of finding particle 1 at $\vec{r}_1$ would be $d\vec{r}_1/V$, independent of what any other particle was doing. For $n$ such particles (out of $N$ particles total) we would have:

$$\frac{d\vec{r}_1}{V} \frac{d\vec{r}_2}{V} \ldots \frac{d\vec{r}_n}{V} = \frac{d\vec{r}_1 \ldots d\vec{r}_n}{V^n} = P_n(\vec{r}_1, \ldots, d\vec{r}_n)d\vec{r}_1 \ldots d\vec{r}_n$$  \hspace{1cm} (20.1.11)$$

\(^2\)This is not true in a solid!
and so
\[ P_n(\vec{r}_1, \ldots, d\vec{r}_n) = \frac{1}{V^n} \]  
(20.1.12)
for a fluid of independent particles. By definition (see Equation (20.1.7) we can then write:
\[ \rho_n = \frac{N!}{(N-n)!} P_n = \frac{N!}{(N-n)!} \frac{1}{V^n} = \frac{N!}{N^n(N-n)!} \rho^n \]  
(20.1.13)
From this we can immediately see that, again for a fluid of independent particles:
\[ \rho_1(\vec{r}_1) = \rho = N/V \]  
(20.1.14)
\[ \rho_2(\vec{r}_1, \vec{r}_2) = \rho^2 \frac{N(N-1)}{N^2} = \rho^2 \left( 1 - \frac{1}{N} \right) \]  
(20.1.15)

End of Small Digression.

But in our system the particles interact with a potential \( U_N \). It is still true that \( \rho_1 = \rho \), which we proved above (Equation (20.1.9)). But the relation Equation (20.1.15) only holds for systems with no interaction potential.

We make the connection by introducing what is called the correlation function \( g_n(\vec{r}_1, \ldots, \vec{r}_n) \) defined this way:
\[ \rho_n(\vec{r}_1, \ldots, \vec{r}_n) = g_n(\vec{r}_1, \ldots, \vec{r}_n) \rho^n \]  
(20.1.16)
In a system of independent particles we get from Equation (20.1.13):
\[ g_n(\vec{r}_1, \ldots, \vec{r}_n) = \frac{N!}{N^n(N-n)!} \]  
(20.1.17)
which, for \( n \ll N \) is essentially 1 since
\[ \frac{N!}{N^n(N-n)!} = \frac{N}{N} \frac{N-1}{N} \cdots \frac{N-n+1}{N} = 1 \times \left( 1 - \frac{1}{N} \right) \times \cdots \times \left( 1 - \frac{n}{N} \right) = 1 + O \left( \frac{1}{N} \right) \]  
(20.1.18)
since on multiplying out the right-hand side the first term is \( 1^n = 1 \) and the second \( n/N \) and further terms are smaller. Since by hypothesis \( n \ll N \), this term (and all following terms) are essentially zero.

From Equations (20.1.7) and (20.1.16) it follows that
\[ g_n \rho^n = \frac{N!}{(N-n)!} P_n = \frac{1}{Z_N} \int \cdots \int e^{-\beta U_N} d\vec{r}_{n+1} \ldots d\vec{r}_N \]  
(20.1.19)
and so
\[ g_n(\vec{r}_1, \ldots, \vec{r}_n) = \frac{V^n N!}{N^n(N-n)!} \frac{1}{Z_N} \int \cdots \int e^{-\beta U_N} d\vec{r}_{n+1} \ldots d\vec{r}_N \]  
(20.1.20)
Note the difference between this and Equation (20.1.17).
The correlation function satisfies:

\[ \int \cdots \int g_n(\vec{r}_1, \ldots, \vec{r}_n) d\vec{r}_1 \ldots d\vec{r}_n = \frac{V^n N!}{N^n (N-n)!} \]  

(20.1.21)

or

\[ \frac{1}{V^n} \int \cdots \int g_n(\vec{r}_1, \ldots, \vec{r}_n) d\vec{r}_1 \ldots d\vec{r}_n = \frac{N!}{N^n (N-n)!} = 1 + O\left(\frac{1}{N}\right) \]  

(20.1.22)

for \( n \ll N \), showing that under that condition it can be considered normalized.

The correlation function of interest to us here is \( g_2 \). This is given by:

\[ g_2(\vec{r}_1, \vec{r}_2) = \frac{N(N-1)}{\rho^2} \frac{1}{Z_N} \int \cdots \int e^{-\beta U_N} d\vec{r}_3 \ldots d\vec{r}_N \]  

(20.1.23)

and it normalizes to

\[ \frac{1}{V^2} \int \int g_2(\vec{r}_1, \vec{r}_2) d\vec{r}_1 d\vec{r}_2 = \frac{N(N-1)}{\rho^2} = 1 - \frac{1}{N} \]  

(20.1.24)

We can make two changes in this. The first is to change to the coordinates of the center of mass of particles 1 and 2 and to the relative separation between them. The second is to recall that the interparticle potentials are assumed to be spherically symmetric. Then we can instantly integrate over the two angles of the interparticle orientation. This gives:

\[ \frac{1}{V} \int_V g(r) 4\pi r^2 dr = 1 - \frac{1}{N} \]  

(20.1.25)

where \( r \) has been written for \( r_{12} \). The factor of \( V \) appears because of the integration over \( r_{12} \).

The function \( g_2(r) = g(r) \) is often called the {\it radial distribution function}. It has the following significance. If we place one specific particle at \( \vec{r}_1 \), then the probability of observing a second particle a distance \( \vec{r}_{12} \) away is

\[ \text{const} \times g_2(\vec{r}_1, \vec{r}_{12}) d\vec{r}_1 d\vec{r}_{12} \]

The constant can be evaluated by integrating the expression above over \( d\vec{r}_{12} \) to get

\[ \text{const} \times \int g_2(\vec{r}_{12}) 4\pi r_{12}^2 d\vec{r}_{12} = (N-1) \]  

(20.1.26)

since the integration over the volume must catch all the particles except the particle at \( \vec{r}_1 \). Thus the left-hand side is equal to \( N - 1 \). But if we compare Equation (20.1.26) to Equation (20.1.25) we see that the constant must be the number density \( \rho = N/V \).

Thus the probability \( P(r) dr \) of observing a second particle at a distance \( r \) inside a spherical shell of thickness \( dr \) is

\[ P(r) dr = \rho g(r) 4\pi r^2 dr \]  

(20.1.27)

and the name {\it radial distribution function} for \( g_2 \) is justified.
20.2 Thermodynamic Functions

With the radial distribution function properly defined, and with the assumption that the potential energy is made up of sums over spherically symmetric pair potentials, it is possible to relate all the thermodynamic properties of a system to the radial distribution function.

Though we are working in the \([N,V,T]\) representation of an ensemble, the canonical ensemble, whose characteristic function is the Helmholtz free energy \(A\), it is not easy to present a simple expression for \(A\) in terms of the radial distribution function. This is unfortunate since if we had such an \(A\), we could derive all the other thermodynamic functions from it.

Instead we shall derive formulas for the energy \(E\) the pressure \(p\) and the chemical potential \(\mu\) based upon the radial distribution function. These derivations are done in some detail because they lend themselves to the computation of thermodynamic properties in computer simulations of fluid systems.

20.2.1 Thermodynamic Functions: the Energy

Since we are working in the canonical ensemble, we know that the energy \(E\) is given by:

\[
E = -\left(\frac{\partial \ln Q(N,V,T)}{\partial \beta}\right) = \frac{3}{2} N kT + kT^2 \left(\frac{\partial \ln Z_n}{\partial T}\right)_{N,V} = \frac{3}{2} N kT + \langle U \rangle \tag{20.2.1}
\]

The first term, \(3NkT/2\) comes from the momentum part of the partition function and is the proper result for an ideal gas. The remainder is from the configuration integral. The term \(\langle U \rangle\) is called the mean potential energy and is given by:

\[
\langle U \rangle = \frac{1}{Z_N} \int \cdots \int U e^{-\beta U} d\bar{r}_1 \cdots d\bar{r}_N \tag{20.2.2}
\]

We assume that the potential \(U_N\) is made up of pairwise additive contributions. Then it is the sum of \(N(N-1)/2\) terms of identical forms (but different integration arguments), all of which give the same result when integrated. So we need only calculate one such integral which we will assume is the one over \(u(r)\), where \(r\) stands for \(r_{12}\). The mean potential energy is then:

\[
\langle U \rangle = \frac{N(N-1)}{2Z_N} \int \cdots \int u(r)e^{-\beta U} d\bar{r}_1 \cdots d\bar{r}_N \tag{20.2.3}
\]

Moving \(Z_N\) inside the integral (it is just a constant)

\[
\langle U \rangle = \frac{N(N-1)}{2} \int \int u(r) \left[ \frac{\int \cdots \int e^{-\beta U} d\bar{r}_3 \cdots d\bar{r}_N}{Z_N} \right] d\bar{r}_1 d\bar{r}_2 \tag{20.2.4}
\]

and now recognizing that if we move the \(N(N-1)\) term inside as well, the quantity in the square brackets is just the distribution function, we get with no difficulty

\[
\langle U \rangle = \frac{1}{2} \int \int u(r)\rho_2(\bar{r}_1 \bar{r}_2) d\bar{r}_1 d\bar{r}_2 \tag{20.2.5}
\]
and now introducing the radial distribution function \( g(r) \)

\[
(U) = \frac{N(N-1)}{2V} \int_0^\infty u(r)g(r)4\pi r^2 dr
\]  

(20.2.6)

Replacing \( N(N-1) \) with \( N^2 \), which for large \( N \) introduces no error we then end up with

\[
\frac{E}{NkT} = \frac{3}{2} + \frac{\rho}{2kT} \int_0^\infty g(r)4\pi r^2 dr
\]  

(20.2.7)

Equation (20.2.7) is easily adapted to computer simulation work because the pair correlation function is easily computed.

### 20.2.2 Thermodynamic Functions: the Pressure

For large \( V \) we know from thermodynamics that the pressure is independent of the shape of the container. So we will assume that the container is a cube.

Now the pressure is given, in the canonical ensemble, by:

\[
p = \left( \frac{\partial \ln Q(N,V,T)}{\partial V} \right)_{N,T} = kT \left( \frac{\partial \ln Z_N}{\partial V} \right)_{N,T}
\]  

(20.2.8)

where the last term arises because the volume dependence is in the configuration integral and not in the momenta at all.

Now \( Z_N \) is an integral over the volume of the container of the system. Hence the limits depend on the volume. And we have to differentiate with respect to the volume. We’ve already seen an example of how to handle this when the Debye function was discussed. Here we will see another way.

The configuration integral is given by

\[
Z_N = \int_0^{V^{1/3}} \cdots \int_0^{V^{1/3}} e^{-\beta U} dx_1 dy_1 dz_1 \cdots dx_N dy_N dz_N
\]  

(20.2.9)

We need to differentiate Equation (20.2.9) with respect to \( V \). The new trick is to change variables so as to get rid of the \( V \) dependence in the limit. We do this by using a new coordinate equal to the old one multiplied by \( V^{1/3} \) so that

\[
x_i' = V^{1/3}x_i, \quad y_i' = V^{1/3}y_i, \quad z_i' = V^{1/3}z_i \quad \text{etc.}
\]  

(20.2.10)

Then the configuration integral Equation (20.2.9) becomes

\[
Z_N = V^N \int_0^1 \cdots \int_0^1 e^{-\beta U} dx_1' dy_1' dz_1' \cdots dx_N' dy_N' dz_N'
\]  

(20.2.11)

Everything has to be transformed to the new coordinates, even \( U \):

\[
U = \sum_{1 \leq i < j \leq N} u(r_{ij})
\]  

(20.2.12)
and
\[ r_{ij} = \left[ (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 \right]^{1/2} = V^{1/3} \left[ (x'_i - x'_j)^2 + (y'_i - y'_j)^2 + (z'_i - z'_j)^2 \right]^{1/2} \] (20.2.13)

Now we can simply differentiate under the integral sign to get
\[
\left( \frac{\partial Z_N}{\partial V} \right)_{N,T} = NV^{N-1} \int_0^1 \cdots \int_0^1 e^{-\beta U} \, dx'_1 \, dy'_1 \, dz'_1 \cdots dx'_N \, dy'_N \, dz'_N - \frac{VN}{kT} \int_0^V \cdots \int_0^V \left( \frac{\partial U}{\partial V} \right) e^{-\beta U} \, dx'_1 \, dy'_1 \, dz'_1 \cdots dx'_N \, dy'_N \, dz'_N \] (20.2.14)

We can now transform back to the original variables and then integrate, remembering that we get \( N(N-1)/2 \) identical terms. The result is:
\[
\left( \frac{\partial \ln Z_N}{\partial V} \right) = \frac{N}{V} - \frac{1}{6VkT} \int_0^V \cdots \int_0^V r_{12} \, dr_{12} \, \rho_2(r_1, r_2) \, dr_1 \, dr_2 \] (20.2.15)

where we’ve switched back to \( \ln Z_N \). This now must be put into Equation (20.2.8) from which we get:
\[
\frac{p}{kT} = \rho - \frac{2\pi \rho^2}{6kT} \int_0^\infty ru'(r) g_j(r) \, 4\pi r^2 \, dr \] (20.2.16)

where we’ve switched to \( r \) for \( r_{12} \) and \( u' \) stands for differentiation with respect to \( r \).

This is known as the pressure equation and it is often used to compute the pressure when doing computer calculations on various models.\(^3\)

This equation has a certain satisfactory look to it. In an ideal system \( P/kT \) would simply equal \( \rho \). If the pair potential is zero in this equation, that’s what we get as the integral disappears. But in general it does not. However it shows up as a sort of “correction” term to the ideal case.

Indeed we can push the similarity a bit. We write the correlation coefficient \( g \) as a power series in the number density:
\[
g(r, \rho, T) = g_0(r, T) + \rho g_1(r, T) + \rho^2 g_2(r, T) + \cdots \] (20.2.17)

If we then substitute this into Equation (20.2.16) we get
\[
\frac{p}{kT} = \rho - \frac{2\pi \rho^2}{3kT} \sum_{j=0}^\infty \rho^j \int_0^\infty r^3 u'(r) g_j(r, T) \, dr \] (20.2.18)

If this is now compared to the virial expansion for a real gas in powers of the density we see that:
\[
B_{j+2} = -\frac{2\pi}{3kT} \int_0^\infty r^3 u'(r) g_j(r, T) \, dr \] (20.2.19)

and indeed it can be shown that at low density:
\[
g_0(r, T) = e^{-\beta u(r)} \] (20.2.20)

\(^3\)This isn’t the only equation that can be used for computing the pressure.
20.2.3 Thermodynamic Functions: the Chemical Potential

We introduce here another technique often used in theoretical work. It works like this: The chemical potential is the change in (Helmholtz) free energy with the number of particles. Usually in thermodynamics we use moles and assume that the number of moles is a continuous variable.

Here we will make use of this to discover the chemical potential as being the difference between two Helmholtz free energies, one with \( N \) particles and one with \( N - 1 \) particles. In fact we are going to show a way to add a particle to the system by having it fade in when we want it and fade out when we don’t.

We do this by means of a coupling parameter \( \xi \) which can range from 0 to 1. We pick one particle to be the “ghost” particle. Here we will make it particle 1. Everything to do with particle 1 is multiplied by \( \xi \). In particular, the interaction of particle 1 with all other particles depends on \( \xi \). When \( \xi = 1 \), particle 1 is there, fully interacting with all other particles. When \( \xi = 0 \), particle 1 is a point particle not interacting with any other particle. Indeed, it can pass through any other particle. It is a ghost particle.

The point of this is that we can change the number of particles in the system at will by changing the coupling parameter.

To be definite, if \( u(r_{ij}) \) is the pair potential acting between particles \( i \) and \( j \), then we must replace it everywhere where one of its subscripts is 1 by \( \xi u(r_{1j}) \). So as particle 1 fades out, all such terms vanish and its interactions with other particles vanish as well.

We begin by noting that the chemical potential is given by:

\[
\mu = \left( \frac{\partial A}{\partial N} \right)_{V,T} = A(N,V,T) - A(N-1,V,T) \tag{20.2.21}
\]

The Helmholtz free energy is given by:

\[
-\beta A = \ln Z_N - \ln N! - N \ln \Lambda^3 \tag{20.2.22}
\]

Inserting this into Equation (20.2.21):

\[
-\beta \mu = \ln \frac{Z_N}{Z_{N-1}} - \ln N - N \ln \Lambda^3 \tag{20.2.23}
\]

Now remember our ghost particle, particle 1. When \( \xi = 1 \),

\[
Z_N(\xi = 1) = Z_N \tag{20.2.24}
\]

but when \( \xi = 0 \)

\[
Z_N(\xi = 0) = V Z_{N-1} \tag{20.2.25}
\]

The factor of \( V \) comes from particle 1. Particle 1 in this case is a ghost, just like an ideal gas particle. And integration over \( dr_1 \) thus gives a factor of \( V \).

With this we have

\[
\ln \frac{Z_N}{Z_{N-1}} = \ln \frac{Z_N(\xi = 1)}{Z_N(\xi = 0)} + \ln V \tag{20.2.26}
\]
and Equation (20.2.23) becomes
\[
-\beta \mu = \ln \frac{Z_N(\xi = 1)}{Z_N(\xi = 0)} + \ln V - \ln N - \ln \Lambda^3
\]  
(20.2.27)

Now something tricky is done. Consider that
\[
\ln \frac{Z_N(\xi = 1)}{Z_N(\xi = 0)} = \int_0^1 \left( \frac{\partial \ln Z_N}{\partial \xi} \right) d\xi.
\]  
(20.2.28)

How is this possible? Because the derivative \( \partial Z_N / \partial \xi \) is the rate at which \( Z \) is changing as one particle’s interactions are “turned on”. The effect of that change is given by the integral in Equation (20.2.28):

Then:
\[
-\beta \mu = \ln \frac{Z_N}{Z_{N-1}} = \ln V + \int_0^1 \left( \frac{\partial \ln Z_N}{\partial \xi} \right) d\xi
\]  
(20.2.29)

Assuming spherically symmetric pair potentials as usual, the total potential is
\[
U(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N, \xi) = \sum_{2 \leq j \leq N} \xi u(r_{1j}) + \sum_{2 \leq i < j \leq N} u(r_{ij})
\]  
(20.2.30)

Putting this into the configuration integral gives:
\[
Z_N = \int \cdots \int \exp \left[ -\beta \sum_{2 \leq j \leq N} \xi u(r_{1j}) \right] \exp \left[ -\beta \sum_{2 \leq i < j \leq N} u(r_{ij}) \right] d\vec{r}_1 \ldots d\vec{r}_N
\]  
(20.2.31)

and if we now differentiate this with respect to \( \xi \) we get:
\[
\left( \frac{\partial Z_N}{\partial \xi} \right) = -\beta \int \cdots \int \left[ \sum_{2 \leq j \leq N} u(r_{1j}) \right] \exp \left[ -\beta \sum_{2 \leq i < j \leq N} u(r_{ij}) \right] d\vec{r}_1 \ldots d\vec{r}_N
\]  
(20.2.32)

If we interchange the sum and the integrals we find that we have \( N - 1 \) identical integrals:
\[
\left( \frac{\partial \ln Z_N}{\partial \xi} \right) = -(N-1)\beta \int \int u(r_{12}) \exp \left[ -\beta \sum_{2 \leq i < j \leq N} u(r_{ij}) \right] \frac{Z_N}{Z_N} d\vec{r}_1 d\vec{r}_2
\]  
(20.2.33)

As with Equation (20.2.4) this is
\[
\left( \frac{\partial \ln Z_N}{\partial \xi} \right) = -N\beta \int \int u(r_{12}) \rho_2(\vec{r}_1, \vec{r}_2) d\vec{r}_1 d\vec{r}_2
\]  
(20.2.34)

and since \( \rho_2 = \rho g(r) \)
\[
\left( \frac{\partial \ln Z_N}{\partial \xi} \right) = -\beta \rho \int_0^\infty u(r) g(r; \xi) 4\pi r^2 dr
\]  
(20.2.35)

Finally, putting this into Equation (20.2.29), we get our final answer
\[
\beta \mu = \ln \rho \Lambda^3 + \rho \beta \int_0^1 \int_0^\infty u(r) g(r; \xi) 4\pi r^2 dr d\xi
\]  
(20.2.36)
20.3 Further Results

We have not at all exhausted the discussion of distribution functions in statistical thermodynamics. But here all we will do is give a brief overview of other developments.

In general the consideration of distribution functions gives rise to integral equations that are difficult to solve.

One particular approach has led to some very useful results. Back in 1914 Ornstein and Zernike\textsuperscript{4} considered the function

\[ h(r) = g(r) - 1 \]  

where \( r \) stands for \( r_{12} \) here and in what follows.

This \( h \) function is useful just as the analogous functions \( f_{ij} = \exp(-\beta u(r_{ij} - 1) \) were useful in discussing non-ideal gases. It measures the influence of particle 1 on particle 2 when their separation is \( r \).

We follow Ornstein and Zernike and divide the correlation function into two parts: the \textbf{direct correlation function} representing the correlation between two particles 1 and 2 in a system containing \( N \) particles in total; and the \textbf{indirect correlation function} representing the correlation between particles 1 and 2 by means of the influence of particle 1 on another particle \( i \) and the influence of that particle on particle 2. This indirect part can be called \( c(r) \) where \( r \) stands for the coordinates of the particles.

Thus they wrote:

\[ h(r_{12}) = c(r_{12}) + \rho \int c(r_{12})h(r_{23})dr_3. \] (20.3.2)

This is the Ornstein-Zernike equation. There is, of course, some handwaving to this but it does turn out to be useful. We can consider Equation (20.3.2) to be a definition of \( h(r) \) or, perhaps more usefully, a relationship between \( h(r) \) and \( c(r) \).

If we had a formula for \( c(r) \) in terms of \( h(r) \) or even \( g(r) \) we would have an integral formulation for \( h(r) \).

But we don’t.

What we can do is approximate the “direct correlation” by assuming that

\[ c(r) = g_{\text{total}}(r) - g_{\text{indirect}}. \] (20.3.3)

except that we have no real idea as to how to split things up.

Percus and Yevick worked roughly as follows: they invoked a \textit{potential of mean force} \( w \) which is a potential energy that is a function of \( r \) whose derivative with respect

\textsuperscript{4}This paper is reprinted in H. Frish and J. L. Lebowitz, \textit{The Equilibrium Theory of Classical Fluids}, Benjamin, 1964.
to \( r \) is the average (mean) force between pairs of molecules.\(^5\) We then assume that the total correlation function \( g(r) \) can be written as:

\[
g_{\text{total}}(r) = e^{-\beta w(r)} \tag{20.3.4}
\]

The indirect correlation function is then:

\[
g_{\text{indirect}}(r) = e^{-\beta [w(r) - u(r)]]} \tag{20.3.5}
\]

In these terms \( c(r) \) is given by:

\[
c(r) = e^{-\beta w(r)} - e^{-\beta [w(r) - u(r)]} \tag{20.3.6}
\]

Now to make things look neat (and less comprehensible) we define

\[
y(r) = e^{\beta u(r)} g(r) \tag{20.3.7}
\]

then

\[
c(r) = g(r) - y(r) = e^{-\beta u(r)} y(r) - y(r) = f(r) y(r) \tag{20.3.8}
\]

where

\[
f(r) = e^{-\beta u(r)} - 1 \tag{20.3.9}
\]

If this is now placed into the Ornstein-Zernike Equation, (20.3.2) on page 20-11, we get after some manipulation:

\[
y(r_{12}) = 1 + \rho \int_0^{\infty} f(r_{13}) y(r_{13}) h(r_{23}) dr_3 \tag{20.3.10}
\]

which is known as the **Percus-Yevick Equation**.

### 20.3.1 Percus-Yevick Hard Spheres

It turns out that the Percus-Yevick equation can be solved analytically for hard spheres. One gets two equations of state from this, one by expanding the pressure equation, Equation (20.2.16) on page 20-8, and the other by expanding the compressibility equation

\[
\frac{1}{\beta} \left( \frac{\rho}{P} \right) = 1 + \rho \int [g(r) - 1] dr, \tag{20.3.11}
\]

which we did not derive.

The two equations are given in terms of \( y \) (a different \( y \))

\[
y = \frac{\pi \rho \sigma^3}{6}, \tag{20.3.12}
\]

where \( \rho \) is the number density \( N/V \), and \( \sigma \) is the diameter of the hard spheres. The equations are:

\[
\frac{\beta P}{\rho} = 1 + 2y + 3y^2 \quad \frac{(1 - y)^2}{(1 - y)^2}, \tag{20.3.13}
\]

\(^5\)This is a common concept that can often be used to simplify many problems in physics and statistical thermodynamics.
from the pressure equation and
\[
\frac{\beta p}{\rho} = \frac{1 + y + y^2}{(1 - y)^3}, \tag{20.3.14}
\]
from the compressibility equation.

Equation (20.3.14) gives reasonable results up through the fifth virial coefficient for the hard sphere gas. The pressure equation is not quite as accurate.
Appendix A

History

The field of statistical mechanics (and what, at the elementary level is almost the same thing, statistical thermodynamics) is a field with a long history. It is also a field that has undergone major changes in the past two decades.

It has also undergone major changes in the past. Since the late 19th century, its very language has changed as well. Thus old books and papers are hard to read today as both the terminology and the notation are strange to us. Anyone interested in the development of classical statistical mechanics needs to know not only this language but that the field grew primarily out of the study of the kinetic theory of gases and was developed into statistical mechanics first by James Clerk Maxwell, Ludwig Boltzmann, and then by J. Willard Gibbs. Maxwell’s work consisted primarily in the discovery of the Maxwell-Boltzmann distribution law. A major review of Boltzmann’s work has been written by Paul and Tatiana Ehrenfest which served to make the field a standard part of both theoretical chemistry and theoretical physics.

While I know of no satisfactory written history of statistical mechanics, an exposition of Boltzmann’s work is given in The Conceptual Foundations of the

---

1 Even the now standard symbols of thermodynamics have changed. Gibbs, for instance, used η for the entropy.

2 Tatiana Ehrenfest’s role in the development of statistical mechanics is often played down. In the preface to the English translation of their article that I reference below, Tatiana Ehrenfest-Afanassjewa modestly claims that “The great task of collecting the literature and of organizing the Encyklopädie article was done by Paul Ehrenfest. My contribution consisted only in discussing with him all the problems involved and I feel that I succeeded in clarifying some concepts that were often incorrectly used.” This was written in 1959. Paul Ehrenfest had died in 1933, a suicide as his mentor Boltzmann had been in 1906. Tatiana Ehrenfest-Afanassjewa’s modesty was well-known during Paul Ehrenfest’s lifetime. It is not so well known today. She did much more than “discussing with him all the problems involved...”.

Statistical Approach in Mechanics by Paul and Tatiana Ehrenfest, originally an article in the *Encyklopädie Der Mathematischen Wissenschaften*, Leipzig, 1912 where it is No. 6 of Volume VI:2:II. It is currently available under the English title given above as translated by Michael J. Moravesik in a Dover Phoenix Edition published by Dover Publications, Mineola, 2002. This is not an easy book for novices but it repays close study. It also contains a bibliography of the literature up to 1912 that is no longer well known (or sadly, easily available.)

Gibbs wrote extensively on Statistical Mechanics. His book, *Elementary Principles in Statistical Mechanics* was first published in 1902 and is kept in print today by Dover Publications. It is much more accessible to students than the Encyclopedia Article by the Ehrenfests. The article by the Ehrenfests mentioned above contains a review of the work by Gibbs on this subject\(^4\) that is interesting in the contrasts between the two views of statistical mechanics.

Development of the field continued for the next 60 years with the advent of quantum mechanics providing a major creative impulse.

The problem of the theory of heat capacities is an example. It is easy to show classically that the constant volume heat capacity of monatomic solids is \(3R\) per mole for all temperatures, but this disagrees with experiments which show that the heat capacity falls to zero as the temperature goes to absolute zero.

A satisfactory theory for the heat capacity of solids was worked out early in the 20th century by application of quantization to the problem.\(^5\)

In the 20’s and 30’s of the last century the quantum approach to statistical mechanics was used very successfully to provide answers to a number of other problems. Very readable books written in that period are those by Fowler and Guggenheim\(^6\) and by Tolman.\(^7\)

However, the systems solved were primarily ideal or idealized problems. More difficult ones such as the behavior of real gases were not attacked until the late 30’s when Meyer and Meyer, introduced a formalism for determining the behavior of non-ideal gases.\(^8\) But while the formalism is very pretty, it was not suitable for actual calculations much beyond the second virial coefficient.

Textbooks such as that written by T.L. Hill\(^9\) made the field accessible to the less mathematically sophisticated students of chemistry as opposed to those in physics and applied mathematics. As a result the theories and method of thinking of statistical mechanics became part of the armament of every physical and biological

---

\(^4\)It was Gibbs’s last major scientific publication.


\(^8\)Statistical Mechanics by Mayer and Mayer, Wiley, 1940. The theory is developed in Chapter 13. As is the case with Tatiana Ehrenfest, it is usually the husband, in this case Joseph Mayer, that is remembered. In fact, his wife Maria Goeppert-Meyer (who won the Nobel Prize for the development of this shell theory of the atomic nucleus) was an equal contributor.

\(^9\)An *Introduction to Statistical Thermodynamics* by Terrel L. Hill, Addison-Wesley, 1960. This text, as of this writing, is still in print as a Dover Publications reprint of 1986.
However by this time the calculations that could be done with pencil and paper had mostly been done. Theories such as the Mayer Cluster Theory mentioned above could not be used for calculations beyond the simple because of the computational complexity required. Pure theory could be further extended, but without calculated results to compare to experiment, such extensions were dry and sterile.

Several signposts to the future did occur. For example in 1957 B.J. Alder and T.E. Wainwright programmed an early digital computer (an IBM 704) to simulate a gas of hard spheres. In doing so they used systems of 32 and 108 particles moving under classical equations of motion.\(^1\) This computation could not then (and cannot now) be done by hand. The labor involved would be prohibitive.

But it was clear that even a simple problem involving hard spheres with no other interaction potential between them required huge amounts of computer time. And while computers were improving in speed, they remained relatively slow until the 1990s.\(^1\)

Today, all that has changed. Very fast very large memory multiple CPU machines are readily available. Computations only dreamed about a few years ago are now feasible. And as a result statistical mechanics is once again a thriving research area.


\(^{11}\)The Author notes that an early IBM machine, the IBM 650, could multiply two ten digit numbers together in a millisecond. Today’s single core machines alone operate in the range of a billion such multiplications a second, an increase in speed by a factor of a million.
Appendix B

Fundamental and Derived Physical Constants

The physical constants used in this work are given below. Those above the horizontal line are considered ‘fundamental’. Those below are ‘derived’ from fundamental constants. The listing is not exhaustive. Some constants used in the text are not given here, but are derivable from those that are.

The numbers in parentheses are the uncertainties in the last two digits given.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed of light</td>
<td>(c)</td>
<td>299792458 (exact)</td>
<td>m/s</td>
</tr>
<tr>
<td>Electric Constant(^1)</td>
<td>(\varepsilon_0)</td>
<td>8.854187817 \times 10^{-12} (exact)</td>
<td>F/m</td>
</tr>
<tr>
<td>Planck’s constant</td>
<td>(h)</td>
<td>6.62606896(33) \times 10^{-34}</td>
<td>Js</td>
</tr>
<tr>
<td>Electron charge</td>
<td>(e)</td>
<td>1.602176487(40) \times 10^{-19}</td>
<td>C</td>
</tr>
<tr>
<td>Electron mass</td>
<td>(m_e)</td>
<td>9.10938215(45) \times 10^{-31}</td>
<td>kg</td>
</tr>
<tr>
<td>Proton mass</td>
<td>(m_p)</td>
<td>1.672621637(83) \times 10^{-27}</td>
<td>kg</td>
</tr>
<tr>
<td>Avogadro’s number</td>
<td>(N_o)</td>
<td>6.02214179(30) \times 10^{23}</td>
<td>1/mol</td>
</tr>
<tr>
<td>Boltzmann’s constant</td>
<td>(k)</td>
<td>1.3806504(24) \times 10^{-23}</td>
<td>J/K</td>
</tr>
<tr>
<td>Faraday constant</td>
<td>(\mathcal{F})</td>
<td>96485.3399(24)</td>
<td>C/mol</td>
</tr>
<tr>
<td>Gas constant</td>
<td>(R)</td>
<td>8.314472(15)</td>
<td>J/mol-K</td>
</tr>
<tr>
<td>Atomic mass unit</td>
<td>(u)</td>
<td>1.660538782(83) \times 10^{-27}</td>
<td>kg</td>
</tr>
</tbody>
</table>

The constants given above were taken from the National Institute of Science and Technology (NIST) web pages at:

\(^1\)The electric constant \(\varepsilon_0\) and the magnetic constant \(\mu_0\) are related by the relation \(\varepsilon_0\mu_0 = 1/c^2\) where \(c\) is the speed of light and \(\mu_0\) is defined to be exactly \(4\pi \times 10^{-7} \text{N} \cdot \text{A}^{-2}\)
http://physics.nist.gov/cuu/Constants/

These derive from Peter J. Mohr and Barry N. Taylor, CODATA Recommended Values of the Fundamental Physical Constants, 2006.
Appendix C

List of Works Consulted

Abramowitz and Stegun, eds, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, AMS Series 55, 1964. This edition, with corrections, as of this writing, is in print from Dover Publications.


Fowler, R. and E.A. Guggenheim, *Statistical Thermodynamics*, Cambridge,
APPENDIX C. LIST OF WORKS CONSULTED

1939.


