

Statistical Physics and Thermodynamics

Leon Hostetler

December 15, 2018

Version 0.5

Contents

Preface	iii
1 Introduction	1
1.1 Quantum Theory	1
1.2 Summary: Introduction	12
2 The State	14
2.1 Information and Entropy	14
2.2 Thermodynamic Variables	19
2.3 The Second Law	20
2.4 Equilibrium and Temperature	23
2.5 Variance of Observables	25
2.6 Summary: The State	27
3 Simple Systems	30
3.1 Quantum Harmonic Oscillator	30
3.2 Quantum Rotor	33
3.3 Spin in a Magnetic Field	35
3.4 Paramagnetism	37
3.5 2D Harmonic Oscillator	41
3.6 Summary: Simple Systems	43
4 The Thermodynamic Limit	45
4.1 Homogeneity	45
4.2 Stability	47
4.3 Ensembles	48
4.4 The Laws of Thermodynamics	49
4.5 Summary: The Thermodynamic Limit	53
5 Perfect Gases	54
5.1 Continuum Limit	56
5.2 Classical Limit	58
5.3 Fermi Gas	59
5.4 Bose-Einstein Condensation	62
5.5 Photon Gas	65
5.6 Internal Degrees of Freedom	66
5.7 Summary: Perfect Gas	70
6 Thermodynamic Processes	71
6.1 Summary: Thermodynamic Processes	72
7 Phase Transitions	73
7.1 Summary: Phase Transitions	74

Preface

About These Notes

My class notes can be found at www.leonhostetler.com/classnotes

Please bear in mind that these notes will contain errors. If you find one, please email me at leonhostetler@gmail.com with the name of the class notes, the page on which the error is found, and the nature of the error. If you include your name, I will probably list your name on the thank you page if I decide to compile and sell my notes.

This work is currently licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. That means you are free to copy and distribute this document in whole for noncommercial use, but you are not allowed to distribute derivatives of this document or to copy and distribute it for commercial reasons.

The primary textbook used in this course is the second edition of *Statistical Physics and Thermodynamics: An Introduction to Key Concepts* by Jochen Rau.

Updates

Last Updated: December 15, 2018

Version 0.5: (Dec. 15, 2018) First upload.

Chapter 1

Introduction

NOTE: You should review classical probability and statistics including combinatorics. You should be familiar with the Bernoulli, binomial, Poisson, uniform, and Gaussian distributions. Consult a handbook on mathematics for physicists.

1.1 Quantum Theory

Groups and Vectors

A **group** is a system denoted $[G, \cdot]$ of a set G and an operation \cdot , that fulfills the following requirements:

- The set is *closed* under the operation \cdot . This means that for any pair of $a, b \in G$, then $a \cdot b \in G$.
- The operation is *associative*. That is, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
- There is an identity element $e \in G$ such that $a \cdot e = e \cdot a = a$.
- Every $a \in G$ has an inverse element a^{-1} such that $a^{-1} \cdot a = a \cdot a^{-1} = e$.
- If the operation is commutative, that is, $a \cdot b = b \cdot a$, then it is called an **abelian group**.

A **vector space** over the complex numbers, or a **complex vector space** is a set $V(\mathbb{C})$ of elements called vectors such that:

- There is an operation $+$ such that $[V, +]$ is an abelian group with identity element written “0” (i.e. the zero vector) and an inverse $-x$ for each vector x .
- For any complex numbers $\alpha, \beta \in \mathbb{C}$, and vectors $x, y \in V(\mathbb{C})$, products such as αx are vectors in $V(\mathbb{C})$, and

$$\begin{aligned} - \alpha(\beta x) &= (\alpha\beta)x \\ - 1x &= x \\ - \alpha(x + y) &= \alpha x + \alpha y \\ - (\alpha + \beta)x &= \alpha x + \beta y \end{aligned}$$

We will write vectors using **Dirac notation**. For example, instead of writing x for a vector, we will write it $|x\rangle$. This prevents them from being confused with simple numbers.

We can also have real vector spaces, where $\alpha, \beta \in \mathbb{R}$. Ordinary 3-vectors, for example, are real vectors.

For our purposes, we have a complex vector space that is closed under addition and multiplication by scalar complex numbers. In other words, any vector you can reach by adding together vectors in the space or by multiplying vectors in the space by complex numbers is itself a vector in the space.

Here are a few more useful definitions:

Linear independence : A set of vectors is linearly independent if no member can be written as a linear combination of the others.

Dimension : The dimension of a set of vectors is the maximum number of linearly independent vectors that can co-exist in the space.

Basis : A basis is a set of linearly independent vectors that is big enough to allow any vector to be written as a linear combination of the basis vectors. That is, we can write an arbitrary vector $|a\rangle$ as the linear combination

$$|a\rangle = \sum_i a_i |i\rangle,$$

where the $|i\rangle$ are the basis vectors. The number of basis vectors is the same as the number of dimensions.

Components : The components of an arbitrary vector $|a\rangle$ in a given basis are the factors that multiply the basis vectors. In the example of $|a\rangle$ given above, the a_i are the components.

Inner product : An inner product is a special product of two vectors, $\langle a|b\rangle$. It gives us orthogonality, norms, and allows us to construct orthonormal bases.

Orthonormal basis : With an orthonormal basis, the components are $a_i = \langle i|a\rangle$. The components can then be used to calculate inner products and norms.

Bra : The left side $\langle a|$ of an inner product $\langle a|b\rangle$ is called a “bra”. These exist in a *dual* of the original vector space. While kets can be represented as 1-dimensional column matrices, bras can be represented as 1-dimensional row matrices. Calculating inner products is then just a matter of matrix multiplication.

To describe a physical system, quantum mechanics assigns a complex number, or *amplitude*, to each distinct, available, physical state. Another way to think of it is that quantum mechanics assigns to each such state a pair of real numbers. For example, for a system that contains M distinct available states, we need $2M$ real numbers to assign a pair of real numbers to each state. We can think of the set of $2M$ real numbers as a vector in a $2M$ -dimensional real-valued vector space. Alternatively, we can think of it as a vector in an M -dimensional complex-valued vector space. This abstract vector space is called a **Hilbert space** or a **state space**. Any vector in this space corresponds to a possible quantum-mechanical state of the system, and it can be infinite.

Just as calculus provides the mathematical basis for classical mechanics, linear algebra provides the mathematical basis for quantum mechanics. In quantum mechanics, we use a lot of linear algebra concepts like vectors, matrices, eigenvalues, rotations, etc.

Ordinarily, a position vector might be denoted by $\vec{r}(t)$. With some orthonormal set of basis vectors \vec{e}_j which are predefined in physical terms, the vector can be written as the linear combination

$$\vec{r}(t) = \sum_{j=1}^M r_j(t) \vec{e}_j,$$

and we say that the components are projections onto basis vectors. Or we can write it as the 1-dimensional column matrix

$$\vec{r}(t) = \begin{bmatrix} r_1(t) \\ r_2(t) \\ \vdots \\ r_M(t) \end{bmatrix}.$$

We can also represent these concepts in Dirac notation. It is the same concept, just with new symbols. Now, instead of writing $\vec{r}(t)$, we write $|\psi(t)\rangle$. Instead of $[\vec{r}(t)]^T$, we would write $\langle\psi(t)|$. Instead of the basis vectors \vec{e}_j , we would just write $|j\rangle$. Now, the vector $|\psi(t)\rangle$ as a linear combination of the basis vectors is

$$|\psi(t)\rangle = \sum_{j=1}^M c_j(t) |j\rangle.$$

The inner product $\vec{a} \cdot \vec{b} = \vec{a}^T \vec{b}$ is now simply written as $\langle a|b\rangle$. The components, $r_j(t) = \vec{e}_j \cdot \vec{r}(t)$, are now

$$c_j(t) = \langle j|\psi(t)\rangle.$$

With every vector space V of kets, we get the dual space V' of bras for free. The dual space V' is the space of all linear, complex-valued functions on V . If elements of V are written as $|f\rangle$, then elements of V' are written as $\langle f|$. Then $\langle f|g\rangle$ is a complex number—the value taken by the linear function $\langle f|$ on the vector $|g\rangle$.

If $|i\rangle$ is a basis for V , then $\langle j|$, where $\langle j|i\rangle = \delta_{ji}$, is a basis for V' .

Given a ket

$$|\psi\rangle = \sum_i a_i |i\rangle,$$

then we choose to define

$$\langle\psi| = \sum_j a_j^* \langle j|,$$

so that a normalized state is

$$\langle\psi|\psi\rangle = \sum_{i,j} a_j^* a_i \langle j|i\rangle = \sum_i |a_i|^2 = 1.$$

More generally, if $|\psi\rangle = \sum_i a_i |i\rangle$ and $|\phi\rangle = \sum_j b_j |j\rangle$, then

$$\langle\phi|\psi\rangle = \sum_{i,j} b_j^* a_i \langle j|i\rangle = \sum_i b_i^* a_i = \langle\psi|\phi\rangle^*.$$

Since the vectors in quantum mechanics are complex, the **transpose** operation from linear algebra is replaced by **Hermitian conjugation**. Hermitian conjugation is simply the ordinary transpose operation followed by complex conjugation. So to get a bra from a ket,

$$\langle\psi| = |\psi\rangle^\dagger,$$

where the \dagger denotes Hermitian conjugation. Similarly,

$$\langle b|a\rangle = \langle a|b\rangle^*.$$

In linear algebra, we can write a vector as

$$\vec{r} = \sum_{j=1}^M r_j \vec{e}_j = \sum_{j=1}^M (\vec{e}_j \cdot \vec{r}) \vec{e}_j = \sum_{j=1}^M \vec{e}_j (\vec{e}_j \cdot \vec{r}).$$

In quantum mechanics, this becomes

$$|\psi\rangle = \sum_{j=1}^M c_j |j\rangle = \sum_{j=1}^M \langle j|\psi\rangle |j\rangle = \sum_{j=1}^M |j\rangle \langle j|\psi\rangle = \left[\sum_{j=1}^M |j\rangle \langle j| \right] |\psi\rangle,$$

where

$$\sum_{j=1}^M |j\rangle \langle j| = 1.$$

The operator

$$P_j = |j\rangle \langle j|,$$

is a **projection operator** or a **projector**. We have that

$$P_j |\psi\rangle = |j\rangle \langle j|\psi\rangle = |j\rangle \langle j|\psi\rangle = \langle j|\psi\rangle |j\rangle.$$

Since $P_j |\psi\rangle = \langle j|\psi\rangle |j\rangle$, we say that P_j projects $|\psi\rangle$ onto $|j\rangle$.

In linear algebra, the norm of a vector \vec{r} is

$$\|\vec{r}\| = \sqrt{\vec{r} \cdot \vec{r}}.$$

In quantum mechanics, the norm of $|\psi\rangle$ is

$$\|\psi\| = \sqrt{\langle\psi|\psi\rangle}.$$

In practice, we compute norms in terms of the components along a set of orthogonal unit vectors. We can do this by inserting the identity operator.

$$\begin{aligned} \langle\psi|\psi\rangle &= \langle\psi|I|\psi\rangle = \langle\psi| \left[\sum_{j=1}^M |j\rangle \langle j| \right] |\psi\rangle = \sum_{j=1}^M \langle\psi| \left[|j\rangle \langle j| \right] |\psi\rangle \\ &= \sum_{j=1}^M \langle\psi|j\rangle \langle j|\psi\rangle = \sum_{j=1}^M c_j^* c_j = \sum_{j=1}^M |c_j|^2. \end{aligned}$$

This is equivalent to

$$\vec{r} \cdot \vec{r} = \sum_{j=1}^M r_j^2,$$

from linear algebra.

Operators

In quantum mechanics, an operator is an object that acts on a ket, transforming it into another ket. A state vector, or a vector representing the state of some quantum system, can be specified by giving its components in some physically meaningful basis. An operator can be defined by giving its matrix elements in some physically meaningful basis. Or an operator can be defined implicitly as the solution to some particular equation.

The **outer product**

$$|\psi\rangle \langle\phi|,$$

of two state vectors can be thought of as an operator that maps any ket $|x\rangle$ onto a vector $|\psi\rangle$ with proportionality factor $\langle\phi|x\rangle$.

We will represent operators using capital letters. Suppose A is a generic operator, then an operator is a linear map

$$A|\psi\rangle = |\psi'\rangle.$$

Operators are linear, so if a and b are arbitrary complex numbers, then

$$A(a|\psi_1\rangle + b|\psi_2\rangle) = Aa|\psi_1\rangle + Ab|\psi_2\rangle = aA|\psi_1\rangle + bA|\psi_2\rangle.$$

Given a set of basis vectors, we can represent an operator as a matrix. If we have an M -dimensional subspace with basis vectors $|j\rangle$, then the j th component of a ket $|\psi\rangle$ is $c_j = \langle j|\psi\rangle$. Then for the operator relation $A|\psi\rangle = |\psi'\rangle$, the j th component of the result $|\psi'\rangle$ is

$$\begin{aligned} (|\psi'\rangle)_j &= \langle j|\psi'\rangle = \langle j|A|\psi\rangle = \langle j|A \cdot I|\psi\rangle = \langle j|A \cdot \left[\sum_{k=1}^M |k\rangle \langle k| \right] |\psi\rangle \\ &= \sum_{k=1}^M \langle j|A|k\rangle \langle k|\psi\rangle = \sum_{k=1}^M A_{jk} \langle k|\psi\rangle. \end{aligned}$$

So

$$(|\psi'\rangle)_j = (A|\psi\rangle)_j = A_{jk} (|\psi\rangle)_k,$$

or $c_j = A_{jk} c_k$, where

$$A_{jk} = \langle j|A|k\rangle,$$

are the elements of the matrix representation of operator A in the given basis.

We have already seen the projection operator

$$P_j = |j\rangle \langle j|,$$

which is the outer product of two identical vectors. The sum of projectors onto a set of M orthonormal states in an M -dimensional Hilbert space is the identity operator

$$I = \sum_{j=1}^M P_j = \sum_{j=1}^M |j\rangle \langle j|.$$

If the sum is incomplete, then the resulting operator is the projector onto the subspace spanned by the included vectors. That is, it is the identity operator within that subspace.

$$P_{\text{subsp}} = I_{\text{subsp}} = \sum_{j=l}^m |j\rangle \langle j|,$$

where j and l are in the interval $[1, M]$.

In general, the projection operator onto a state $|\psi\rangle$ is

$$P_\psi = |\psi\rangle \langle \psi|.$$

Projection operators are **idempotent**, meaning

$$P_\psi^2 = P_\psi.$$

A **Hermitian operator** is any operator whose Hermitian conjugate is itself. That is

$$A^\dagger = A.$$

The eigenvalues of Hermitian operators are real, and in quantum mechanics, observables are represented by Hermitian operators.

The eigenvectors are defined only up to a multiplicative constant. For example, if $A|a_i\rangle = a_i|a_i\rangle$, then it is also true that $A(c|a_i\rangle) = a_i(c|a_i\rangle)$, where c is some complex number. Typically, we choose to normalize to 1, so that $\langle a_i|a_i\rangle = 1$.

All eigenvectors corresponding to distinct eigenvalues are orthogonal. If an operator in an M -dimensional Hilbert space has M distinct eigenvalues, i.e. there is no degeneracy, then its eigenvectors form a complete set of orthonormal vectors. That is, they form a complete *basis*. If at least two linearly independent eigenvectors are associated with the same eigenvalue, then the eigenvalue is said to be “degenerate”. The degree of degeneracy of an eigenvalue is the number of linearly independent eigenvectors associated with it.

Remember, if a set of vectors is not orthogonal, we can make them orthogonal by using, for example, the Gram-Schmidt method.

Recall that a basis is a set of orthogonal unit vectors in Hilbert space. Choosing a set of basis vectors is analogous to choosing a coordinate system in 3D space. A basis is a complete set of unit vectors that spans the entire state space.

There are two kinds of basis sets. A **discrete basis** is a set of basis vectors that can be labeled by integers. For example, $|1\rangle, |2\rangle, \dots, |M\rangle$, where M can be finite or infinite. Then the number of basis vectors is either finite or countably infinite. A **continuous basis** is a set of basis vectors labeled by real numbers. For example, $|x\rangle$ where x can be any real number between some upper bound (which could be $+\infty$) and some lower bound (which could be $-\infty$). In a continuous basis, the basis vectors are not countable.

The Principles of QM

The mathematical description of a quantum system rests on several principles:

1. At a fixed time t_0 , the state of a physical system is defined by specifying a ket $|t = t_0\rangle$ belonging to the state space of the system.
2. Every observable, or measurable physical quantity is described by a Hermitian operator A acting in the state space of the system. This operator is called the observable.
3. If we measure a physical property associated with the observable A , the result will always be an eigenvalue of A .

4. When the observable A is measured on a system in the normalized state $|\psi\rangle$, the probability of obtaining the result a is

$$\text{prob}(a|\psi) = \langle P_a \rangle = \langle \psi | P_a | \psi \rangle,$$

where P_a is the projector onto the subspace associated with the eigenvalue a .

5. If the measurement of the observable A on the system in state $|\psi\rangle$ gives the result a , then the state of the system immediately after the measurement is

$$|\psi'\rangle = \frac{P_a |\psi\rangle}{\sqrt{\langle \psi | P_a | \psi \rangle}}.$$

That is, when a state is measured, a definite eigenvalue $P_a |\psi\rangle$ is obtained. Immediately afterward, the state is the projection of the prior state onto the eigenspace associated with the eigenvalue. The square root in the denominator is just a normalization factor. If the eigenvalue is non-degenerate, then we simply have $|\psi'\rangle = P_a |\psi\rangle$.

6. The state $|\psi\rangle$ evolves with time by a unitary transformation. If the state is initially $|\psi(0)\rangle$, then

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle,$$

where $|\psi(t)\rangle$ is the state at time t and $U(t)$ is the time evolution operator. In differential form, the **Schrodinger equation** governs the time evolution as

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H |\psi(t)\rangle.$$

Unless interrupted by a measurement, time evolution is represented by a linear, unitary time evolution operator $U(t)$, where

$$U^\dagger U = U U^\dagger = I.$$

From U we derive the Hamiltonian operator H and the time-dependent Schrodinger equation. For a closed system,

$$U(t) = e^{-\frac{iHt}{\hbar}}.$$

Measurements cause an apparently discontinuous change in the state vector, i.e. it collapses the wave function. After an ideal measurement yielding result a_i , the state is in the corresponding eigenstate $|a_i\rangle$. The best way of preparing systems in a given quantum state is measurement followed by selection of the desired state.

Two-level System

The simplest quantum system is a two-level system. For example, a spin-1/2 particle is a two-level system since it can only have spin up or spin down. Similarly, a photon with two polarization states or an atom with two energy levels are also two-level systems.

A two-level system is described in a two-dimensional Hilbert space. We can use $|0\rangle$ and $|1\rangle$ for our basis vectors, then any pure state is a linear combination of these two basis vectors. For example, if

$$|\psi\rangle = a_0 |0\rangle + a_1 |1\rangle,$$

then $|a_0|^2 + |a_1|^2 = 1$. We can choose a global phase factor such that a_0 is real and positive, then we can rewrite our ket in terms of two angles θ and ϕ

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right) |0\rangle + e^{i\phi} \sin\left(\frac{\theta}{2}\right) |1\rangle,$$

where $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$. This means that there is a one-to-one correspondence between the states of a two-level system and the points on the surface of a sphere of radius 1. This sphere is called the **Bloch sphere**. Pairs of orthogonal states are antipodes on this sphere. For example, we can set $|0\rangle$ to be the north pole of the sphere, then $|1\rangle$ is the south pole. Also, there is a one-to-one correspondence between unitary transformations and rotations on the Bloch sphere.

In a two-level system, every observable can be represented by a 2×2 Hermitian matrix. Any Hermitian matrix can be written as a linear combination of 4 special Hermitian matrices—the unit matrix and the three Pauli matrices.

$$\sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Each of these is Hermitian, so

$$\sigma_i^\dagger = \sigma_i, \quad i = 0, 1, 2, 3.$$

The square of each gives the identity matrix

$$\sigma_i^2 = I, \quad i = 0, 1, 2, 3.$$

The three **Pauli matrices** can be written more compactly as

$$\sigma_j = \begin{bmatrix} \delta_{j3} & \delta_{j1} - i\delta_{j2} \\ \delta_{j1} + i\delta_{j2} & -\delta_{j3} \end{bmatrix},$$

where $j = 1, 2, 3$. Some important identities of the Pauli matrices are

$$\begin{aligned} -i\sigma_1\sigma_2\sigma_3 &= I \\ \det \sigma_j &= -1 \\ \text{tr} [\sigma_j] &= 0, \end{aligned}$$

where, for each, $j = 1, 2, 3$. The commutation relation between the Pauli matrices is given by

$$[\sigma_j, \sigma_k] = 2i\varepsilon_{jkl}\sigma_l.$$

The product of two Pauli matrices is

$$\sigma_j\sigma_k = \delta_{jk}I + i\varepsilon_{jkl}\sigma_l.$$

If our two-level system is the two states of a spin-1/2 particle, then the Pauli matrices represent the observables of the three spin components.

Pure vs. Mixed States

The projector onto a subspace is the operator

$$P = \sum_n |\phi_n\rangle \langle \phi_n|.$$

This operator projects any vector onto a subspace of Hilbert space which is spanned by the orthonormal basis $\{|\phi_n\rangle\}$. The operator P is Hermitian, idempotent, and its trace is the dimension of the subspace onto which it projects.

We are used to quantum mechanical calculations involving pure states. A **pure state** is a state like

$$|\psi\rangle = a_0|0\rangle + a_1|1\rangle + a_2|2\rangle.$$

A pure state like this is in a coherent superposition of the eigenstates $|0\rangle$, $|1\rangle$, and $|2\rangle$, and we see quantum interference effects. Suppose we have a pure state like this, and we want to make a measurement. If we are given the coefficients a_i for the initial state $|\psi(0)\rangle$, then we evolve this state in time until the measurement takes place. Then we use the projector onto eigenstates of the observable to predict the probabilities for different results.

The expectation value of an operator A in the pure state $|\psi\rangle$ gives us the expected average of the results of many measurements of the observable A on a system prepared in state $|\psi\rangle$.

$$\begin{aligned} \langle A \rangle_\psi &= \langle \psi | A | \psi \rangle = \langle \psi | \sum_n |a_n\rangle \langle a_n| A | \psi \rangle = \sum_n \langle \psi | a_n \rangle \langle a_n | A | \psi \rangle \\ &= \sum_n \langle \psi | a_n \rangle a_n \langle a_n | \psi \rangle = \sum_n |\langle \psi | a_n \rangle|^2 a_n = \sum_n \text{prob}(a_n | \psi) a_n. \end{aligned}$$

So the expected value of A is just the weighted average of all possible outcomes.

A pure state can be represented by a projection operator $P_\psi = |\psi\rangle\langle\psi|$.

In terms of a projection operator, the probability of measuring outcome a_i is

$$\begin{aligned}\text{prob}(a_i|\psi) &= \langle\psi|P_i|\psi\rangle = \langle\psi|P_i \cdot I|\psi\rangle = \langle\psi|P_i \cdot \left[\sum_j |j\rangle\langle j|\right]|\psi\rangle \\ &= \sum_j \langle\psi|P_i|j\rangle\langle j|\psi\rangle = \sum_j \langle j|\psi\rangle\langle\psi|P_i|j\rangle = \sum_j \langle j|\left(|\psi\rangle\langle\psi|P_i\right)|j\rangle \\ &= \sum_j \langle j|\left(P_\psi P_i\right)|j\rangle = \text{tr}\left(P_\psi P_i\right).\end{aligned}$$

Similarly, the expectation value of the observable A is obtained in the same manner and is

$$\langle A \rangle_\psi = \sum_i \text{prob}(a_i|\psi)a_i = \langle\psi|A|\psi\rangle = \text{tr}(P_\psi A).$$

What if we don't know the exact initial state of our system? Perhaps we only know the temperature of the system. Perhaps we know the system has to be in the pure state ψ_1 or in the pure state ψ_2 . We don't know in which state it really is—we just know the probability for it to be in one state or the other. This kind of system is in a statistical mixture of states, and it is said to be in a **mixed state**. This is different from a pure state which is not a statistical mixture of pure states, but rather, is in a coherent superposition of eigenstates of whatever observable you're interested in.

Suppose we know with probability p_1 that the system is in state $|\psi_1\rangle$ and with probability p_2 it is in the state $|\psi_2\rangle$. Then our system is in a statistical mixture of the states $|\psi_1\rangle$ and $|\psi_2\rangle$. What is the probability of obtaining some result a_i of a measurement of observable A ? The answer is

$$\begin{aligned}\text{prob}(a_i|\text{mix}(\psi_1, \psi_2)) &= \text{prob}(a_i|\psi_1)\text{prob}(\psi_1) + \text{prob}(a_i|\psi_2)\text{prob}(\psi_2) \\ &= \text{prob}(a_i|\psi_1)p_1 + \text{prob}(a_i|\psi_2)p_2.\end{aligned}$$

Recall that the expectation value of A in a pure state $|\psi\rangle$ is

$$\langle A \rangle_\psi = \sum_i \text{prob}(a_i|\psi)a_i = \langle\psi|A|\psi\rangle = \text{tr}(P_\psi A).$$

For the mixed state which could be in $|\psi_1\rangle$ or in $|\psi_2\rangle$, we have

$$\begin{aligned}\langle A \rangle_{\text{mix}(\psi_1, \psi_2)} &= \sum_i \text{prob}(a_i|\text{mix}(\psi_1, \psi_2))a_i = \sum_i \left[\text{prob}(a_i|\psi_1)p_1 + \text{prob}(a_i|\psi_2)p_2\right]a_i \\ &= \left[\sum_i \text{prob}(a_i|\psi_1)a_i\right]p_1 + \left[\sum_i \text{prob}(a_i|\psi_2)a_i\right]p_2 \\ &= \text{tr}(P_{\psi_1}A)p_1 + \text{tr}(P_{\psi_2}A)p_2 = \text{tr}\left([p_1P_{\psi_1} + p_2P_{\psi_2}]A\right).\end{aligned}$$

We can write our result as

$$\langle A \rangle_{\text{mix}(\psi_1, \psi_2)} = \text{tr}(\rho A),$$

where

$$\rho = p_1P_{\psi_1} + p_2P_{\psi_2}.$$

Density Operator

For the two state mixed state example, we define the **density operator** or **density matrix** as

$$\rho = p_1P_{\psi_1} + p_2P_{\psi_2} = p_1|\psi_1\rangle\langle\psi_1| + p_2|\psi_2\rangle\langle\psi_2|.$$

Remember that, for example, $p_1 = \text{prob}(\psi_1)$.

We can describe our statistical mixed state using the density operator ρ . The probability to obtain result a_n is simply

$$\text{prob}(a_n) = \text{tr}(\rho P_n),$$

where $P_n = |a_n\rangle\langle a_n|$ is the projector onto state $|a_n\rangle$. The proof is as follows:

$$\begin{aligned} \text{prob}(a_n) &= \text{tr}(\rho P_n) = \sum_m \langle m| \left(\rho |a_n\rangle\langle a_n| \right) |m\rangle \\ &= \sum_m \langle m| \left([p_1 P_{\psi_1} + p_2 P_{\psi_2}] |a_n\rangle\langle a_n| \right) |m\rangle \\ &= \sum_m \langle m| \left[p_1 P_{\psi_1} + p_2 P_{\psi_2} \right] |a_n\rangle\langle a_n| m\rangle \\ &= \sum_m \langle a_n| m\rangle \langle m| \left[p_1 P_{\psi_1} + p_2 P_{\psi_2} \right] |a_n\rangle = \langle a_n| \left[p_1 P_{\psi_1} + p_2 P_{\psi_2} \right] |a_n\rangle \\ &= \langle a_n| \left[p_1 |\psi_1\rangle\langle\psi_1| + p_2 |\psi_2\rangle\langle\psi_2| \right] |a_n\rangle = p_1 \langle a_n|\psi_1\rangle\langle\psi_1|a_n\rangle + p_2 \langle a_n|\psi_2\rangle\langle\psi_2|a_n\rangle \\ &= p_1 |\langle a_n|\psi_1\rangle|^2 + p_2 |\langle a_n|\psi_2\rangle|^2 = \text{prob}(a_n|\psi_1) p_1 + \text{prob}(a_n|\psi_2) p_2. \end{aligned}$$

More generally, the density operator for a statistical mixture of states $\{|\psi_j\rangle\}$ with probabilities $\{p_j\}$ is

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|,$$

where the probability weights sum to one

$$\sum_j p_j = 1.$$

The states $|\psi_j\rangle$ are normalized, but they are not necessarily orthogonal.

Recall that for a pure state $|\psi\rangle$, the probability of measuring a result a_i is

$$\text{prob}(a_i|\psi) = \text{tr}(P_\psi P_i),$$

and the expectation value of operator A acting on $|\psi\rangle$ is

$$\langle A \rangle_\psi = \text{tr}(P_\psi A).$$

For a statistical mixture of states, the probability of measuring result a_i is

$$\text{prob}(a_i|\text{mix}) = \text{tr}(\rho P_i),$$

and the expectation value of operator A acting on the mixture is

$$\langle A \rangle_{\text{mix}} = \text{tr}(\rho A).$$

Notice that when going from a pure state to a statistical mixture of states, we replace P_ψ with the density matrix ρ .

Important properties of the density operator include:

- It is Hermitian

$$\rho^\dagger = \rho.$$

- It is positive,

$$\rho \geq 0,$$

meaning that its eigenvalues are nonnegative.

- Its trace is one,

$$\text{tr}(\rho) = 1,$$

since the $|\psi\rangle$ are normalized.

- For a mixed state, ρ is not a projector. For a pure state,

$$\rho^2 = \rho.$$

After measurement, the density matrix becomes

$$\rho \rightarrow \frac{P_i \rho P_i}{\text{tr}(\rho P_i)},$$

known as **Lüder's rule**.

Keep in mind that the density matrix is an operator. In general, it is a complex, time-dependent operator that represents an ensemble of possible states, each with specified probability. The diagonal elements are real, but in general, the off-diagonal elements are complex. For a two-level system, it is a 2×2 matrix.

For two-level systems, the density matrix can always be written in the form

$$\rho = \frac{1}{2} (I + \vec{r} \cdot \vec{\sigma}),$$

where I is the identity matrix, $\vec{\sigma}$ is the vector of Pauli matrices, and \vec{r} is the **Bloch vector**. The Bloch vector gives the position of a point in the Bloch sphere, which has unit radius. In other words, $|\vec{r}| \leq 1$.

For pure states, $|\vec{r}| = 1$, so they are on the surface of the Bloch sphere. For mixed states, $|\vec{r}| \leq 1$, so they are inside the Bloch sphere. Maximally mixed states, $|\vec{r}| = 0$, are at the center of the Bloch sphere. For maximally mixed states, the density matrix ρ is diagonal and the diagonal elements are all the same.

Given the density operator at time zero, its time evolution satisfies

$$\rho(t) = U(t) \rho(0) U^\dagger(t).$$

In differential form, the time evolution of the density operator is given by the **Liouville-von Neumann equation**

$$i\hbar \frac{\partial}{\partial t} \rho(t) = [H(t), \rho(t)],$$

where $H(t)$ is the Hamiltonian, and the brackets denote a commutator.

Tensor Product

Recall that a quantum system with M possible states is described in an M -dimensional complex vector space called a **Hilbert space**. If you have a composite quantum system composed of one quantum system described in a Hilbert space \mathcal{H}_1 and a second quantum system described in a Hilbert space \mathcal{H}_2 then the composite system is described in a Hilbert space that is the **tensor product** of the constituent Hilbert spaces

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2.$$

For example, if you have two spin-1/2 particles a and b described in two-dimensional Hilbert spaces \mathcal{H}_a and \mathcal{H}_b , then the system of the two particles is described in the tensor product space $\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b$, which has dimension $2 \cdot 2 = 4$. The basis states of the composite system are then $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$, where the first number gives the spin of particle a and the second number gives the spin of particle b .

The tensor product is also sometimes called the **direct product**.

For a composite system of n constituents, if the individual wavefunctions are all in pure states, then

$$|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle.$$

States are said to be *separable* if they can be written as a simple tensor product.

Recall that the wave function of the electron in the hydrogen atom, for example, can be written in terms of a space part and an angular momentum part

$$\psi(\vec{r}) = R(r) Y_l^m(\theta, \phi).$$

This wave function does not include the electron's spin. If we apply an orbital angular momentum operator, then that operator acts only on the $Y_l^m(\theta, \phi)$ part of the wave function, and not on the radial part $R(r)$. If the orbital and spin angular momenta are

separable in the Hamiltonian, then including the spin of the electron, we can write

$$\Psi(\vec{r}, s_z, t) = \psi(\vec{r})\xi(s_z, t),$$

where $\xi(s_z, t)$ is the spin part of the wave function. We can also write this in terms of kets as the tensor product

$$|\Psi(t)\rangle = |\psi(t)\rangle \otimes |\xi(t)\rangle.$$

Spin angular momentum acts only on its own vector space, independent of the 3D spatial wave function.

If we can factor a wave function into an energy part and an angular momentum part like

$$|\psi\rangle = |E_n\rangle \otimes |j, m\rangle,$$

then if we apply the Hamiltonian operator \hat{H} , then since \hat{H} acts only on $|E_n\rangle$,

$$\hat{H}(|E_n\rangle \otimes |j, m\rangle) = \hat{H}(|E_n\rangle) \otimes \hat{I}(|j, m\rangle).$$

That is, \hat{H} acts like the identity operator \hat{I} and does nothing on the angular momentum part $|j, m\rangle$.

The tensor product of two matrices A and B is also called the **Kronecker product**. If A is an $m \times n$ matrix and B is a $p \times q$ matrix, then $A \otimes B$ is an $mp \times nq$ **block matrix**

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots \\ a_{21}B & a_{22}B & \\ \vdots & & \ddots \end{bmatrix},$$

where $a_{11}B$ is the top left element of A times the entire matrix B .

In quantum mechanics, discrete linear operators can be represented as square matrices. Suppose we want to know the trace $\text{tr}(A \otimes B)$. As you can see from the block matrix definition of $A \otimes B$, the trace of $A \otimes B$ is a_{11} times the trace of B plus a_{22} times the trace of B , and so on. If A is an $m \times m$ matrix, and B is a $p \times p$ matrix, then

$$\begin{aligned} \text{tr}(A \otimes B) &= a_{11} \text{tr}(B) + a_{22} \text{tr}(B) + \cdots + a_{mm} \text{tr}(B) = \sum_i^m a_{ii} \text{tr}(B) \\ &= \text{tr}(B) \cdot \sum_i^m a_{ii} = \text{tr}(B) \cdot \text{tr}(A) = \text{tr}(A) \cdot \text{tr}(B). \end{aligned}$$

Tip

The density operator ρ of a pure state has exactly one nonzero eigenvalue, and it is equal to 1.

1.2 Summary: Introduction

Skills to Master

- Be comfortable counting combinations and permutations from classical statistics
- Be comfortable calculating the expectation value, variance, and standard deviation of the Bernoulli, Binomial, Poisson, uniform, and Gaussian distributions

A wave function $|\psi\rangle$ can be written as a linear combination of basis vectors $|i\rangle$ idempotent, meaning

$$|\psi\rangle = \sum_i a_i |i\rangle,$$

where the components are

$$a_i = \langle i|\psi\rangle.$$

Then

$$\langle\psi|\psi\rangle = \sum_j a_j^* \langle j|.$$

If $|\psi\rangle = \sum_i a_i |i\rangle$ and $|\phi\rangle = \sum_j b_j |j\rangle$, then

$$\langle\phi|\psi\rangle = \sum_{i,j} b_j^* a_i \langle j|i\rangle = \sum_i b_i^* a_i = \langle\psi|\phi\rangle^*.$$

The norm of $|\psi\rangle$ is

$$\|\psi\| = \sqrt{\langle\psi|\psi\rangle}.$$

We compute this via

$$\langle\psi|\psi\rangle = \sum_{i=1}^M |a_i|^2.$$

Given a set of basis vectors, we can represent an operator A as a matrix with elements

$$A_{jk} = \langle j|A|k\rangle.$$

A Hermitian operator is any operator whose Hermitian conjugate is itself. That is,

$$A^\dagger = A.$$

The eigenvalues of Hermitian operators are real, and in quantum mechanics, observables are represented by Hermitian operators. An operator A acting on an eigenstate $|a_i\rangle$ returns the eigenvalue a_i associated with that eigenstate. That is

$$A|a_i\rangle = a_i|a_i\rangle.$$

The operator

$$P_j = |j\rangle\langle j|,$$

is a projection operator. Projection operators are

$$P_j^2 = P_j.$$

The projector P_j acting on a state $|\psi\rangle$, projects that $|\psi\rangle$ onto $|j\rangle$. That is,

$$P_j|\psi\rangle = |j\rangle\langle j|\psi\rangle = |j\rangle\langle j|\psi\rangle = \langle j|\psi\rangle|j\rangle,$$

where $\langle j|\psi\rangle$ is the component of $|\psi\rangle$ in the direction $|j\rangle$. The sum of projectors onto a set of M orthonormal states in an M -dimensional Hilbert space is the identity operator

$$I = \sum_{j=1}^M P_j = \sum_{j=1}^M |j\rangle\langle j|.$$

The simplest quantum system is a two-level system. For example, a spin-1/2 particle is a two-level system since it can only have spin up or spin down. Every state in a two-level system can be represented as a point on or within the unit Bloch sphere. In a two-level system, every observable can be represented by a 2×2 Hermitian matrix, which can be written as a linear combination of the 2×2 identity matrix and the three Pauli matrices

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Properties of the Pauli matrices include

$$\begin{aligned} \sigma_i^\dagger &= \sigma_i \\ \sigma_i^2 &= I \\ -i\sigma_1\sigma_2\sigma_3 &= I \\ \det \sigma_j &= -1 \\ \text{tr}[\sigma_j] &= 0, \end{aligned}$$

where $i = x, y, z$.

A pure state $|\psi\rangle$, can be written as a linear combination of eigenstates. For example, in a two-level system with eigenstates $|0\rangle$ and $|1\rangle$, a pure state can

be written in the form

$$|\psi\rangle = a_0 |0\rangle + a_1 |1\rangle.$$

Suppose we have a pure state like this, and we want to make a measurement. If we are given the coefficients a_i for the initial state $|\psi(0)\rangle$, then we evolve this state in time until the measurement takes place. Then we use the projector onto eigenstates of the observable to predict the probabilities for different results.

Given a pure state $|\psi\rangle$, the probability of measuring the outcome a_i is

$$\text{prob}(a_i|\psi) = \langle\psi|P_i|\psi\rangle = \text{tr}(P_\psi P_i).$$

The expectation value of the observable A in the pure state $|\psi\rangle$ is the weighted average

$$\langle A \rangle_\psi = \langle\psi|A|\psi\rangle = \sum_i \text{prob}(a_i|\psi)a_i = \text{tr}(P_\psi A).$$

If the system is not known to be in a pure state, but is in some statistical mixture of states $\{|\psi_j\rangle\}$ with corresponding probabilities $\{p_j\}$ (i.e. $p_k = \text{prob}(|\psi_k\rangle)$), then we say the system is in a *mixed* state. The density operator for a statistical mixture of states

$$\rho = \sum_j p_j |\psi_j\rangle \langle\psi_j|,$$

where the probability weights sum to one

$$\sum_j p_j = 1.$$

For a two-level system, the density operator is a 2×2 matrix. The states $|\psi_j\rangle$ are normalized, but they are not necessarily orthogonal. For a statistical mixture of states, the probability of measuring result a_i is

$$\text{prob}(a_i|\{|\psi_j\rangle\}) = \text{tr}(\rho P_i),$$

where $P_i = |a_i\rangle \langle a_i|$. The expectation value of operator A acting on the mixture is

$$\langle A \rangle_{\{|\psi_j\rangle\}} = \text{tr}(\rho A).$$

Important properties of the density operator include:

- It is Hermitian, $\rho^\dagger = \rho$
- It is positive semi-definite, $\rho \geq 0$, meaning that all of its eigenvalues are in $0 \leq \lambda_j \leq 1$.
- Its trace is one, $\text{tr}(\rho) = 1$, since the $|\psi\rangle$ are normalized.
- For a pure state, $\rho^2 = \rho$
- For a pure state, ρ has exactly one nonzero eigenvalue, and it is equal to 1.

For two-level systems, the density matrix can always be written in the form

$$\rho = \frac{1}{2} (I + \vec{r} \cdot \vec{\sigma}),$$

where I is the identity matrix, $\vec{\sigma}$ is the vector of Pauli matrices, and \vec{r} is the **Bloch vector**. The Bloch vector gives the position of a point in the Bloch sphere, which has unit radius. In other words, $|\vec{r}| \leq 1$. For pure states, $|\vec{r}| = 1$, so they are on the surface of the Bloch sphere. For mixed states, $|\vec{r}| < 1$, so they are inside the Bloch sphere. Maximally mixed states, $|\vec{r}| = 0$, are at the center of the Bloch sphere. For maximally mixed states, the density matrix ρ is diagonal and the diagonal elements are all the same.

If A is an $m \times n$ matrix and B is a $p \times q$ matrix, then $A \otimes B$ is an $mp \times nq$ block matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots \\ a_{21}B & a_{22}B & \\ \vdots & & \ddots \end{bmatrix}.$$

An important property is

$$\text{tr}(A \otimes B) = \text{tr}(A) \cdot \text{tr}(B).$$

Chapter 2

The State

2.1 Information and Entropy

In the previous chapter, we found that macroscopic systems are described in terms of mixed states because the details of the system are unknown. The amount of information that is missing is called the **entropy**. The more we know about a system, the smaller its entropy. The entropy of a pure state is zero since we know all there is to know about a pure state.

The entropy of a statement, message, or situation is a measure of the uncertainty in that statement, message, or situation. The **information** associated with an event is the amount of uncertainty that is removed after the outcome of the event becomes known.

Shannon Entropy

Suppose Alice wants to communicate to Bob the results of an experiment. She performs N trials of an experiment, each with d possible results labeled $i = 1, 2, \dots, d$. The trials are independent, so one outcome does not affect subsequent trials. When she performs the N experiments, she gets a sequence of N results, which she wants to communicate to Bob.

Bob knows beforehand the probability distribution $\{p_i\}_{i=1}^d$ for Alice's experiment, and he knows the number of trials N that she will perform. What he doesn't know, prior to Alice's communication, is the specific sequence of outcomes that Alice obtained.

After performing the experiments, Alice wants to communicate the results—the set of results as well as the order in which they occurred—to Bob in the most efficient manner possible. She is restricted to using classical bits, '0' and '1', in her message. We assume here that her communication channel is noiseless, so she doesn't have to worry about redundancy.

For example, suppose Alice's experiment involves throwing a fair dice. Each trial has $d = 6$ possible results, and she performs the experiment $N = 10$ times. Suppose she gets the sequence of results 3, 4, 1, 6, 5, 1, 6, 1, 5, 1, and she wants to communicate these results to Bob. Bob knows what Alice is doing. He knows the probability distribution of each dice throw is $\{p_i\}_{i=1}^d = \{1/6, 1/6, 1/6, 1/6, 1/6, 1/6\}$. He also knows that she performed $N = 10$ trials.

Alice and Bob have agreed beforehand on a coding algorithm. What is the most efficient coding algorithm? That is, what is the shortest string of zeros and ones that communicates the results of Alice's trials to Bob? If I_1 is the minimum number of bits needed to communicate the result of a single trial and I_N is the minimum number of bits needed to communicate the results of N trials, then Alice can definitely communicate the results using NI_1 bits. However, this is not the most efficient way. That is, $I_N \neq NI_1$.

Going back to the dice example, each throw has six possible outcomes. With two bits, Alice can encode $2^2 = 4$ outcomes, so two bits is not sufficient to communicate

the result of a single trial. With three bits, Alice can encode $2^3 = 8$ outcomes. This is sufficient to encode the result of a single trial, so $I_1 = 3$, but there are two wasted “spaces”. If she communicates her ten results separately, she needs a total of $10I_1 = 30$ bits. However, in ten throws, there are $6^{10} = 60,466,176$ possibilities, which is much less than the 2^{30} possibilities that she could represent with 30 bits. That is, there is a lot of wasted information space in her message, and she could, for example, communicate the result of all ten throws at once instead of communicating each result individually, thereby reducing the number of needed bits and getting closer to the I_N minimum number of bits.

As the number of trials N grows large, the average bits needed per trial is

$$I[\{p_i\}] \equiv \lim_{N \rightarrow \infty} \frac{I_N}{N},$$

and this quantity is less than or equal to I_1 . This asymptotic number of bits needed per trial defines the **Shannon entropy**, and it quantifies the **information** contained in the outcome of a single trial. It also quantifies Bob’s initial ignorance or uncertainty of a single trial.

Given N trials of a random experiment, each with d outcomes and having a probability distribution $\{p_i\}$, we obtain an ordered sequence of N outcomes. A sequence is **typical** if the relative frequency of outcome i is equal to its probability. We can assume that all sequences are typical provided that N is large. For a given large N and probability distribution $\{p_i\}$, how many typical sequences are there?

Consider $N = 1002$ trials of a dice roll. Since each of the $d = 6$ outcomes is equally likely, the typical sequence will contain exactly $n_1 = 1002/6 = 167$ ones, $n_2 = 167$ twos, $n_3 = 167$ threes, and so on. How many ordered sequences are there with this property?

In general, for a typical sequence, the frequency of outcome i is equal to its probability. That is,

$$\frac{n_i}{N} = p_i,$$

where n_i is the number of outcomes i in our sequence. We can write this as

$$n_i = Np_i.$$

To find the number of typical sequences, we have to solve a *permutations with repetition* problem. The number of permutations of N items with n_1 identical objects of type 1, n_2 identical objects of type 2, and so on with n_k identical objects of type k is

$$\frac{N!}{n_1!n_2! \cdots n_k!}.$$

In our case, the number of typical sequences is

$$\frac{N!}{(Np_1)!(Np_2)! \cdots (Np_k)!} = \frac{N!}{\prod_i (Np_i)!}.$$

The log base-2 of the number of typical sequences is

$$\log_2 \left(\frac{N!}{\prod_i (Np_i)!} \right) = \log_2 N! - \log_2 \left(\prod_i (Np_i)! \right) = \log_2 N! - \sum_i \log_2 [(Np_i)!].$$

Applying the change of base formula, $\log_2 x = \ln x / \ln 2$ and then Stirling’s approximation $\ln N! \simeq N \ln N - N$ gives us

$$\begin{aligned} \log_2 \left(\frac{N!}{\prod_i (Np_i)!} \right) &= \frac{\ln N!}{\ln 2} - \sum_i \frac{\ln [(Np_i)!]}{\ln 2} \simeq \frac{N \ln N - N}{\ln 2} - \sum_i \frac{Np_i \ln(Np_i) - Np_i}{\ln 2} \\ &= \frac{N \ln N - N}{\ln 2} - \sum_i \frac{Np_i [\ln N + \ln p_i] - Np_i}{\ln 2} \\ &= \frac{N \ln N - N}{\ln 2} - \frac{N \ln N}{\ln 2} \sum_i p_i - \frac{N}{\ln 2} \sum_i p_i \ln p_i + \frac{N}{\ln 2} \sum_i p_i. \end{aligned}$$

But $\sum_i p_i = 1$, so this simplifies to

$$\log_2 \left(\frac{N!}{\prod_i (Np_i)!} \right) \simeq -\frac{N}{\ln 2} \sum_i p_i \ln p_i.$$

Applying the change of base formula again gives us

$$\log_2 \left(\frac{N!}{\prod_i (Np_i)!} \right) \simeq -N \sum_i p_i \log_2 p_i.$$

So we can write the number of typical sequences as

$$K = \frac{N!}{\prod_i (Np_i)!} \simeq 2^{NI},$$

where

$$I = -\sum_i p_i \log_2 p_i.$$

Suppose Alice and Bob compile an ordered list of all K of the typical sequences for the upcoming N experiments. Then after the experiment is performed, the optimal transmission of the results is obtained when Alice simply transmits the position of her specific typical sequence in their previously ordered list of typical sequences. That is, after performing the experiment and obtaining a typical sequence, she finds this typical sequence in their ordered list of K typical sequences and transmits the index number to Bob.

The number of bits required to represent a number K is $\log_2 K$. So the number of bits required to transmit the index associated with Alice's specific sequence is

$$\log_2 2^{NI} = NI.$$

This means the average number of bits required *per trial* is

$$\frac{NI}{N} = I.$$

So for an arbitrary probability distribution $\{p_i\}$,

$$I[\{p_i\}] = -\sum_{i=1}^d p_i \log_2 p_i.$$

This is known as **Shannon's noiseless channel coding theorem**, and it gives the **Shannon entropy**.

Some properties of the Shannon entropy include

- It is non-negative. That is, $I[\{p_i\}] \geq 0$. It is equal to zero if and only if the outcome of the experiment is known beforehand. That is, if the probability distribution is $\{p_i\} = \{0, 0, \dots, 0, 1, 0, 0, \dots\}$.
- It is invariant under permutations of the probabilities $\{p_i\}$. That is, it is unaffected by a relabeling of the probabilities.
- It is additive in the sense that

$$I_{AB} = I_A + I_{B|A}.$$

Meaning, if you perform joint random experiments A and B with possible outcomes $\{a_i\}$ and $\{b_j\}$, and joint outcomes $\{a_i \wedge b_j\}$ then the number of bits required to encode the results of the joint experiment is the number of bits required for A plus the number of bits required to encode B given A

$$I[\{p(a_i \wedge b_j)\}] = I[\{p(a_i)\}] + \sum_i p(a_i) I[\{p(b_j|a_i)\}].$$

Apart from a multiplicative constant, the Shannon entropy is the only function with these three properties of non-negativity, permutation symmetry, and additivity.

The additivity property is easily proved by expanding $p(a_i \wedge b_j) = p(a_i)p(b_j|a_i)$, and then using the fact that $\sum_j p(b_j|a_i) = 1$.

If the two experiments A and B are correlated, then we need less information to specify the outcome. In general,

$$I[\{p(a_i \wedge b_j)\}] \leq I[\{p(a_i)\}] + I[\{p(b_j)\}].$$

The only time there is equality, is if the two experiments A and B are uncorrelated. This is when

$$p(a_i \wedge b_j) = p(a_i)p(b_j),$$

for all i and j .

For a physical system in a mixed state characterized by the density operator $\hat{\rho}$, the Shannon entropy is

$$I[\{p_i\}] = - \sum_i \rho_i \log_2 \rho_i = -\text{tr}(\hat{\rho} \log_2 \hat{\rho}).$$

This is a measure of our ignorance about the mixed state.

Entropy

We now define *the* entropy of statistical mechanics as

$$S[\hat{\rho}] \equiv -k \text{tr}(\hat{\rho} \ln \hat{\rho}),$$

where

$$k \simeq 1.38 \times 10^{-23} \text{ J/K},$$

is **Boltzmann's constant**. Notice that *the* definition of entropy is the same as Shannon entropy multiplied by Boltzmann's constant and with the log base-2 replaced by a natural log. This quantum definition of entropy is sometimes called the **von Neumann entropy** to distinguish it from the non-quantum **Boltzmann entropy** of classical statistical mechanics.

The three properties of Shannon entropy have analogues with von Neumann entropy:

- It is non-negative. That is, $S[\hat{\rho}] \geq 0$. It is equal to zero if and only if the state $\hat{\rho}$ is pure.
- It is invariant under unitary transformations. That is,

$$S[\hat{U} \hat{\rho} \hat{U}^\dagger] = S[\hat{\rho}].$$

- For a bipartite or two-level system composed of subsystems a and b ,

$$S[\hat{\rho}_{ab}] \leq S[\hat{\rho}_a] + S[\hat{\rho}_b].$$

Here, equality holds if and only if the systems are independent or uncorrelated such that $\hat{\rho}_a \otimes \hat{\rho}_b$.

If A is an $n \times n$ matrix, then the exponential of the matrix is defined as the infinite product

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k = I + A + \frac{1}{2} A^2 + \dots.$$

The natural log of a matrix, $\ln B$, is defined to be the matrix A such that $e^A = B$. For example, if $A = 0$ is the $n \times n$ zeros matrix then

$$e^A = I + 0 + 0 + \dots,$$

implying that the natural logarithm of the identity matrix is equal to the zeros matrix.

To take the log of a matrix, just diagonalize it and take the log of the diagonal elements.

Because of the unitary transformation property, to calculate the entropy, we simply calculate

$$S[\hat{\rho}] \equiv -k \sum_j \lambda_j \ln \lambda_j,$$

Tip

The entropy of a pure state is zero.

where the λ_j are the eigenvalues of $\hat{\rho}$. Since $\hat{\rho}$ is positive semi-definite, all of its eigenvalues are in $0 \leq \lambda_j \leq 1$. Since the eigenvalues are always nonnegative, the entropy is always defined. Any zero eigenvalue simply does not contribute to the entropy since $0 \cdot \ln 0 \rightarrow 0$.

In a Hilbert space of dimension d , the maximum entropy is $\ln d$ bits. So for a two-level system, the maximum entropy is $\ln 2$. This is the entropy of a maximally mixed two-level system.

Maximum Entropy Principle

At the quantum level, a state is represented by a density operator $\hat{\rho}$, so we call $\hat{\rho}$ the **microstate** of the system.

The **macrostate** state of a system can be characterized by a set of complete and minimal macroscopic parameters. There are basically three different kind of macroscopic parameters:

Control parameters: These are macroscopic parameters that control the Hamiltonian of a system. An external magnetic field \vec{B} or the volume of a box within which gas particles are constrained, are examples of control parameters. We will denote a generic control parameter by ξ . For a system with multiple control parameters, we let ξ represent the *set* of control parameters.

Sharp constraints: These are hard constraints on the outcomes of measurements. If we counted the number of particles in a box to be N , then we know a later measurement of the number of particles will also give N , so N is an example of a sharp constraint. We will denote a generic sharp constraint by Ξ . For a system with multiple sharp constraints, we let Ξ represent the *set* of sharp constraints.

Expectation values: We might know the expectation value of some macroscopic parameter X . For example, if we uniformly selected our system of particles from some larger system, and we know the values of some macroscopic parameters of the larger system, then we know the expectation values of those parameters for our own system. For a system with multiple expectation values, we let X represent the *set* of expectation values.

A given microstate $\hat{\rho}$ implies a unique macrostate, since given a microstate $\hat{\rho}$, we can calculate macroscopic observables. However, a given macrostate does not imply a unique microstate. In fact, there could be a large number or even infinite microstates which correspond to a single macrostate. Given a specific macrostate, we still want to choose a specific microstate for the purpose of calculations. Which microstate corresponding to a given macrostate should we choose?

According to the **maximum entropy principle**, of all the microstates which are compatible with our specific macrostate, we should choose as representative, the microstate which maximizes the entropy. This corresponds to requiring that the chosen microstate give us the least amount of information.

If the set of parameters describing the macrostate consist only of sharp constraints, then the microstate with the maximum entropy is the totally mixed state selected by the sharp constraints

$$\hat{\rho} = \frac{1}{Z(\Xi)} \hat{I}_{\Xi},$$

where \hat{I}_{Ξ} is the identity operator, and the normalization factor

$$Z(\Xi) = \text{tr}_{\Xi}(I_{\Xi}) = \dim \mathcal{H}_{\Xi},$$

is called the **partition function**. The partition function, in this case, simply counts the number of pure and mutually orthogonal microstates which are compatible with the

macroscopic constraints. If there is more than one sharp constraint, then Ξ can stand for the set of all sharp constraints. The entropy of this state is

$$\begin{aligned} S[\hat{\rho}] &= -k \operatorname{tr}(\hat{\rho} \ln \hat{\rho}) = -k \operatorname{tr} \left(\frac{\hat{I}}{Z} \ln \frac{\hat{I}}{Z} \right) = -k \operatorname{tr} \left(\frac{\hat{I}}{Z} \ln \hat{I} - \frac{\hat{I}}{Z} \ln Z \right) \\ &= -k \operatorname{tr} \left(\frac{\hat{I}}{Z} \ln \hat{I} \right) + k \operatorname{tr} \left(\frac{\hat{I}}{Z} \ln Z \right) = k \frac{\ln Z}{Z} \operatorname{tr}(\hat{I}) = k \ln Z. \end{aligned}$$

Here we used the fact that $Z = \operatorname{tr}(\hat{I})$ is a constant, that $\ln \hat{I} = 0$, that $\operatorname{tr}(cA) = c \operatorname{tr}A$, and that $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.

If in addition to sharp constraints, we also know an expectation value $\langle \hat{X} \rangle$ of operator \hat{X} , then the maximum entropy microstate is

$$\rho = \frac{1}{Z(Y, \Xi)} e^{-Y \hat{X}},$$

where the partition function is now

$$Z(Y, \Xi) = \operatorname{tr}_{\Xi} \left(e^{-Y \hat{X}} \right).$$

This kind of state is called a **Gibbs state**. This microstate is defined on the Hilbert space selected by the sharp constraints Ξ . The parameter Y is a **Lagrange parameter** whose value is adjusted so that ρ yields the correct expectation value of \hat{X} . The entropy of the Gibbs state is

$$S[\rho] = -k \operatorname{tr}(\rho \ln \rho) = k \ln Z + kY \langle X \rangle_{\rho}.$$

More generally, we can have a system with multiple expectation values $X = \{X_i\}$, multiple control parameters $\xi = \{\xi_i\}$, and multiple sharp constraints $\Xi = \{\Xi_i\}$. Furthermore, the expectation values and the sharp constraints may both depend on the control parameters. In that case, the unique state that maximizes the entropy is

$$\rho = \frac{1}{Z(Y, \Xi, \xi)} e^{-\sum_i Y_i X_i(\xi)}, \quad (2.1)$$

where the partition function is

$$Z(Y, \Xi, \xi) = \operatorname{tr}_{\Xi(\xi)} \left(e^{-\sum_i Y_i X_i(\xi)} \right). \quad (2.2)$$

Here, the $Y = \{Y_i\}$ is an entire set of Lagrange parameters, with one for each expectation value. The entropy of the macrostate is

$$S(X, \Xi, \xi) = S[\rho] - k \operatorname{tr}(\rho \ln \rho).$$

By definition of the Gibbs state, this is equal to the largest possible entropy of any compatible microstate ρ'

$$S(X, \Xi, \xi) = \max_{\rho'} S[\rho'].$$

2.2 Thermodynamic Variables

For the Gibbs state, the partition function is a function of its **natural variables** Y , Ξ , and ξ

$$Z(Y, \Xi, \xi) = \operatorname{tr}_{\Xi(\xi)} \left(e^{-\sum_i Y_i X_i(\xi)} \right).$$

If we partially differentiate the natural log of the partition function with respect to a Lagrange parameter, we get the associated expectation value

$$\begin{aligned} \frac{\partial}{\partial Y_i} \ln [Z(Y, \Xi, \xi)] &= \frac{1}{Z} \frac{\partial}{\partial Y_i} \operatorname{tr} \left(e^{-\sum_i Y_i X_i(\xi)} \right) = \frac{1}{Z} \operatorname{tr} \left(-X_i e^{-\sum_i Y_i X_i(\xi)} \right) \\ &= \operatorname{tr} \left(-X_i \frac{1}{Z} e^{-\sum_i Y_i X_i(\xi)} \right) = \operatorname{tr}(-X_i \rho) = -\langle X_i \rangle_{\rho}, \end{aligned}$$

so we say that X is a conjugate variable of Y . Similarly, we define

$$\Upsilon_i \equiv \frac{\partial}{\partial \Xi_i} \ln [Z(Y, \Xi, \xi)]$$

$$v_i \equiv \frac{\partial}{\partial \xi_i} \ln [Z(Y, \Xi, \xi)].$$

So the natural variables are Y, Ξ, ξ , and their associated **conjugate variables** are X, Υ, v . All together, the six variables are the **thermodynamic variables** of the system.

The total differential of $\ln Z$ is

$$d(\ln Z) = -X dY + \Upsilon d\Xi + v d\xi.$$

Notice the conjugate pairs. More generally,

$$d(\ln Z) = - \sum_j \langle X_j \rangle dY_j + \sum_j \Upsilon_j d\Xi_j + \sum_j v_j d\xi_j.$$

The parameters X, Ξ, ξ which originally defined the macrostate, are the natural variables of the macroscopic entropy

$$S(X, \Xi, \xi) = k \ln Z(Y, \Xi, \xi) + k \sum_j Y_j \langle X_j \rangle. \quad (2.3)$$

The Lagrange parameters Y , which are adjusted to yield the correct expectation values, also depend on the natural variables of the macrostate. That is, $Y = Y(X, \Xi, \xi)$.

The total differential of the entropy, in units of k , is

$$d(S/k) = d(\ln Z) + X dY + Y dX,$$

or more generally,

$$d(S/k) = \sum_j Y_j dX_j + \sum_j \Upsilon_j d\Xi_j + \sum_j v_j d\xi_j.$$

This implies the partial derivatives

$$Y_j = \frac{\partial}{\partial X_j} \left(\frac{S}{k} \right)$$

$$\Upsilon_j = \frac{\partial}{\partial \Xi_j} \left(\frac{S}{k} \right)$$

$$v_j = \frac{\partial}{\partial \xi_j} \left(\frac{S}{k} \right).$$

2.3 The Second Law

Now, instead of focusing on states, we focus on the *processes* that connect one state to another. Suppose our initial state has the macroscopic variables X, Ξ, ξ , and some time later, it has the macroscopic variables X', Ξ', ξ' . If systems with the same initial macroscopic variables always end up with the same final macroscopic variables, then the process connecting the two states is called a **reproducible process**. Note that the initial variables X, Ξ, ξ do not have to be the same type as the final variables X', Ξ', ξ' . Reproducibility implies that we can predict the final macrostate from knowledge of the initial macrostate.

Recall that a microstate ρ evolves to ρ' under a unitary transformation. If the initial microstate ρ is compatible with the initial macroscopic variables X, Ξ, ξ , then for a reproducible system, the final microstate ρ' is necessarily compatible with the final macroscopic state X', Ξ', ξ' , even though we will never know exactly what these microstates are.

Since reproducibility holds for any microstate compatible with the macroscopic variables, we are free to choose the initial microscopic state to be the Gibbs state ρ . Then the entropy of the initial macrostate equals the entropy of this Gibbs state $S(X, \Xi, \xi) = S[\rho]$. Remember that of all the possible microstates ρ_i , the Gibbs microstate is the one with the largest entropy. When the system evolves in time, since a unitary transformation

preserves the trace, then the entropy of the microstate remains the same. So

$$S(X, \Xi, \xi) = S[\rho] = S[\rho'].$$

However, this entropy is not necessarily equal to the entropy of the final macrostate. In fact, the entropy of the final macrostate can be greater than or equal to the entropy of the initial macrostate. That is,

$$S(X, \Xi, \xi) \leq S(X', \Xi', \xi').$$

Thus, on the macroscopic level, the entropy cannot decrease when an isolated system undergoes a reproducible process. This is the **second law of thermodynamics**.

Classically, the evolution of a system can be represented by a trajectory in a multi-dimensional phase space of micro-parameters. A point in this phase space represents a pure microstate. A mixed set of microstates would be represented as a probability distribution in this phase space. The micro-parameters are the set of numbers needed to describe a quantum state. Classically, a microstate of a system composed of N molecules has $6N$ micro-parameters—the components of position, $\{q_i\}$, and momentum, $\{p_i\}$, for each molecule. The coordinates vary in time according to canonical equations of motion

$$\begin{aligned}\dot{p}_i &= -\frac{\partial}{\partial q_i} H(q_i, p_i) \\ \dot{q}_i &= \frac{\partial}{\partial p_i} H(q_i, p_i),\end{aligned}$$

for $i = 1, 2, 3, \dots, 3N$. The set of coordinates (q_i, p_i) , which defines a microstate, undergoes continual motion. A representative point carves out a trajectory whose direction at any time t is given by a velocity vector \vec{v} given by the equations of motion.

During its evolution, the system will only pass through *accessible* microstates—the ones that do not violate conservation laws. For example, for an isolated system, the total internal energy must be conserved. The statistical approach is to connect the macroscopic observables (i.e. their averages) to the probability for a certain microstate to appear along the system's trajectory in phase space.

The initial macroscopic parameters impose constraints on the allowed microstates. Classically, these allowed microstates form some volume in phase space. After time evolution, these microstates are mapped to new microstates, so the initial volume is mapped to some new volume in phase space. Since the evolved microstates must be compatible with the final macroscopic data, the evolved volume must be contained in the volume of microstates compatible with the final data. Since the entropy can only increase, the phase space volume associated with macroscopic data can only increase. The phase space volume is a measure of our ignorance about the system's preparation, and its increase mirrors the increase of entropy.

The trajectory of a representative point remains within a limited volume in phase space. This limits the possible values of the coordinates of the point. The requirement of finite energy limits the values of both q_i and p_i through the Hamiltonian $H(q_i, p_i)$. If the energy constraint is sharp, then the trajectory remains in the surface defined by $H(q, p) = E$. If the energy constraint is less sharp, for example, restricted to the interval $[E - \frac{1}{2}\Delta, E + \frac{1}{2}\Delta]$, then the trajectory is restricted to the shell within these limits.

The set of all microstates compatible with a given macrostate forms an ensemble of microstates. At any time t , members of the ensemble will be in all kinds of possible microstates—all of them compatible with the given macrostate. These form a “swarm” of representative points all within the allowed volume. Every member of the ensemble undergoes continual change as each member moves along its own trajectory. We describe the distribution of points in the ensemble by a **density function** $\rho(q, p)$, where q is the entire $3N$ set of position coordinates, and p is the entire $3N$ set of momentum coordinates. The density function quantifies the manner in which the members of the ensemble are distributed over all possible microstates at different instants in time.

At any time t , the number of representative points in the volume element $d\omega =$

$d^{3N}q d^{3N}p$ around the point (q, p) of the phase space is given by the product

$$\rho(q, p; t) d\omega = \rho(q, p; t) d^{3N}q d^{3N}p.$$

Given some physical quantity $f(q, p)$, which may be different for different microstates, its ensemble average is

$$\langle f \rangle = \frac{\int f(q, p) \rho(q, p; t) d\omega}{\int \rho(q, p; t) d\omega},$$

where the integrals go over all of phase space, but $\rho = 0$ in regions where there are incompatible or inaccessible microstates. An ensemble is *stationary* if

$$\frac{\partial \rho}{\partial t} = 0,$$

for all time. Then the ensemble average $\langle f \rangle$ will be independent of time. A system in equilibrium has a stationary ensemble.

Consider an arbitrary volume ω in the relevant region of the phase space, and let the surface enclosing that volume be denoted by σ . Then, the time rate of change of the number of representative points in this volume is $\frac{\partial}{\partial t} \int \rho d\omega$. The net rate at which points flow out of the volume and across the boundary surface σ is $\int \rho \vec{v} d\vec{\sigma}$ where $\vec{v} = (\dot{q}_i, \dot{p}_i)$ is the velocity vector of the representative points in the surface element $d\vec{\sigma}$. Then the divergence theorem implies that

$$\int_{\sigma} \rho \vec{v} d\vec{\sigma} = \int_{\omega} \text{div}(\rho \vec{v}) d\omega,$$

where the divergence is defined as

$$\text{div}(\rho \vec{v}) \equiv \sum_{i=1}^{3N} \left[\frac{\partial}{\partial q_i} (\rho \dot{q}_i) + \frac{\partial}{\partial p_i} (\rho \dot{p}_i) \right].$$

There are no sources or sinks in the phase space, so the total number of representative points must be conserved. This implies that

$$\frac{\partial}{\partial t} \int_{\omega} \rho d\omega = - \int_{\omega} \text{div}(\rho \vec{v}) d\omega,$$

which means

$$\int_{\omega} \left[\frac{\partial \rho}{\partial t} + \text{div}(\rho \vec{v}) \right] d\omega = 0.$$

For an arbitrary volume ω , this implies that

$$\frac{\partial \rho}{\partial t} + \text{div}(\rho \vec{v}) = 0.$$

This is a **continuity equation** for the swarm of representative points in the ensemble. Plugging in the definition for the divergence gives us

$$\begin{aligned} 0 &= \frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N} \left[\frac{\partial}{\partial q_i} (\rho \dot{q}_i) + \frac{\partial}{\partial p_i} (\rho \dot{p}_i) \right] \\ &= \frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N} \left[\frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right] + \rho \sum_{i=1}^{3N} \left[\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right] \\ &= \frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N} \left[\frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right] = \frac{d\rho}{dt}. \end{aligned}$$

The second sum above is zero since the equations of motion tell us that

$$\frac{\partial \dot{q}_i}{\partial q_i} = \frac{\partial^2 H}{\partial q_i \partial p_i} = \frac{\partial^2 H}{\partial p_i \partial q_i} = - \frac{\partial \dot{p}_i}{\partial p_i}.$$

In the final step, we noted that the terms combine to form the total derivative of ρ since $\rho = \rho(q_i, p_i; t)$.

So the total time derivative of the phase space density vanishes along a phase space trajectory. This is **Liouville's theorem**. We can write it in the form

$$\frac{\partial \rho}{\partial t} + [\rho, H] = 0,$$

where

$$[\rho, H] = \sum_{i=1}^{3N} \left[\frac{\partial \rho}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial \rho}{\partial p_i} \frac{\partial H}{\partial q_i} \right] = \sum_{i=1}^{3N} \left[\frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right],$$

is called the **Poisson bracket**.

Liouville's theorem implies that as a representative point moves along its trajectory, the local density of representative points in its immediate vicinity remains constant. That is, the density function behaves like an incompressible liquid.

The second law assumes that

- The system is isolated
- The macroscopic process is reproducible

Whenever these premises are wrong, the second law can be violated. If the system is very small, then a statistical fluctuation could cause a decrease in entropy. However, such a random fluctuation is not reproducible. Memory effects, like hysteresis, can also make a process non-reproducible. However, such memory effects tend to fade away after some characteristic time.

Now we will look at it briefly from an information-theoretic perspective. Reproducibility implies that we can predict the final macrostate from knowledge of the initial macrostate. This prediction cannot contain more information than the data on which it is based, therefore, the information content of the final macrostate cannot be greater than that of the initial macrostate. That is, the information can remain the same or decrease. Since entropy is a measure of our lack of information, this implies that entropy cannot decrease.

The second law implies a fundamental asymmetry in the evolution of macroscopic systems. If a process leads to a final state whose entropy is strictly larger than that of the initial state, then there is no reproducible process leading back to the original macrostate. Such a process is called **irreversible**. A macroscopic process is **reversible** (or **isentropic**) if and only if entropy is conserved.

2.4 Equilibrium and Temperature

In time, any closed system will come into an equilibrium state where all the physical parameters characterizing the system are not changing in time. The process of moving into equilibrium is called the **relaxation process** of the system, and the time it takes is called the **relaxation time**. Equilibrium means that separate parts of the system (subsystems) are also in a state of internal equilibrium and in equilibrium with each other. This means there is no exchange of energy or particles with each other.

There can also be *local* equilibrium. This occurs when subsystems are in states of equilibria, but the subsystems are not in equilibrium with each other. The further the overall system is from total equilibrium, the more macroscopic parameters are needed to characterize the system.

In terms of the density function, in an equilibrium state, the density function is independent of time. So in an equilibrium state, we have a **stationary ensemble**. There are two ways this is possible:

Microcanonical ensemble: The density function is independent of the coordinates.

That is, $\rho(q, p) = \text{constant}$ over the volume of phase space containing points compatible with constraints. Outside of that, $\rho = 0$. This corresponds to an ensemble

of systems which at all times are uniformly distributed over all possible microstates. The ensemble average is then

$$\langle f \rangle = \frac{1}{\omega} \int_{\omega} f(q, p) d\omega,$$

where ω is the volume of the relevant part of phase space. This kind of ensemble is called a “microcanonical ensemble,” with “equal a priori probabilities” for various possible microstates.

Canonical ensemble: More generally, we can have $[\rho, H] = 0$ if we have p and q dependence only on H . That is, $\rho(q, p) = \rho(H)$ with $H = H(q, p)$. The most natural choice for this kind of ensemble is

$$\rho(q, p) = e^{-\frac{H(q, p)}{kT}}.$$

This kind of ensemble is called a “canonical ensemble”.

From now on, we will assume that all processes are reproducible. In equilibrium, the variables are independent of time, i.e. they are constants of motion, so an equilibrium state is a **stationary state**. A system can also be in a nonequilibrium stationary state called a **metastable equilibrium**. However, if perturbed, such a system typically drops into the true equilibrium state.

In terms of entropy, we know from the second law that entropy stays constant or increases. A global maximum of the entropy corresponds to the equilibrium state of the system. A local maximum corresponds to a metastable equilibrium.

For any isolated system, the internal energy U is a constant of the motion. The energy U may be included as a sharp constraint or as an expectation value. Given an energy U , a system can be in three kinds of equilibrium states:

Microcanonical: The energy is given as a sharp constraint.

Canonical: The energy is given as an expectation value, and all other constants of motion are given as sharp constraints. The state is the Gibbs state with the set $\{X_i\}$ being only the Hamiltonian H .

Grand canonical: Energy and particle number N are given as expectation values. The state is a Gibbs state with $\{X_i\} = \{H, N\}$.

In all cases, the internal energy is one of the natural variables of the entropy. Its conjugate variable is the inverse temperature

$$\beta = \frac{\partial}{\partial U} \left(\frac{S}{k} \right).$$

If we let $x = \{X, \Xi, \xi\}$ and $y = \{Y, \Upsilon, v\}$ be the sets of natural variables and their conjugates, then we can write the differential of the total entropy as

$$d(S/k) = \beta dU + y dx.$$

Instead of thinking of the entropy as a function of the energy and the natural variables x , we can also think of the energy as a function of the entropy and the natural variables x . Then

$$dU = \frac{1}{k\beta} dS - \frac{y}{\beta} dx.$$

The last two equations imply that

$$\frac{\partial}{\partial x} \left(\frac{S}{k} \right) = -\beta \frac{\partial U}{\partial x}.$$

We define the **temperature** to be

$$T \equiv \frac{\partial U}{\partial S} = \frac{1}{k\beta}.$$

So with respect to energy, temperature is the conjugate of entropy. If the energy spectrum of the Hamiltonian is bounded below and unbounded above, which is true for real-world systems, then it can be shown that $0 \leq T \leq \infty$.

For a canonical equilibrium state, Eq. (2.1) and Eq. (2.2) with $X = H$ and $Y = \beta$ become

$$\rho = \frac{1}{Z(\beta, \Xi, \xi)} e^{-\beta H(\xi)}$$

$$Z(\beta, \Xi, \xi) = \text{tr}_{\Xi} \left(e^{-\beta H(\xi)} \right).$$

If E_i are the eigenvalues of the Hamiltonian with associated degeneracy factors $\{g_i\}$, then the partition function becomes

$$Z = \sum_i g_i e^{-\beta E_i}. \quad (2.4)$$

The probability that an energy measurement yields the eigenvalue E_i , is

$$P(E_i) = \frac{g_i e^{-\beta E_i}}{Z}.$$

Similarly, we define the **pressure** as the conjugate of the volume V with respect to energy

$$p \equiv -\frac{\partial U}{\partial V},$$

and the **chemical potential** as the conjugate of the particle number N with respect to energy

$$\mu \equiv \frac{\partial U}{\partial N}.$$

2.5 Variance of Observables

If the energy of a system is given as an expectation value rather than as a sharp constraint, then there is a range of possible measurement outcomes.

If the energy is the only macroscopic parameter given as an expectation value, then we are dealing with a canonical state. For the canonical ensemble,

$$\rho = \frac{1}{Z} e^{-\beta H}.$$

Recall that

$$\frac{\partial}{\partial Y_i} \ln Z = -\langle X_i \rangle_{\rho}.$$

In our case, we let $X = H$ and $Y = \beta$, then the expectation value of the energy is

$$U = \langle H \rangle = -\frac{\partial}{\partial \beta} \ln Z = -\frac{1}{Z} \frac{\partial Z}{\partial \beta}.$$

The variance of the energy distribution is

$$\begin{aligned} \text{var}(H) &= \langle H^2 \rangle - \langle H \rangle^2 = \langle H^2 \rangle + \frac{1}{Z} \frac{\partial Z}{\partial \beta} \langle H \rangle - \frac{1}{Z} \frac{\partial Z}{\partial \beta} \text{tr}(\rho H) \\ &= \text{tr} \left(\frac{1}{Z} e^{-\beta H} H^2 \right) + \frac{1}{Z} \frac{\partial Z}{\partial \beta} \text{tr} \left(\frac{1}{Z} e^{-\beta H} H \right) \\ &= \frac{1}{Z} \text{tr} (e^{-\beta H} H^2) + \frac{1}{Z^2} \frac{\partial Z}{\partial \beta} \text{tr} (e^{-\beta H} H) = -\frac{\partial}{\partial \beta} \left[\text{tr} \left(\frac{1}{Z} e^{-\beta H} H \right) \right] \\ &= -\frac{\partial}{\partial \beta} \text{tr}(\rho H) = -\frac{\partial}{\partial \beta} \langle H \rangle = -\left(\frac{\partial U}{\partial \beta} \right)_{\Xi, \xi}, \end{aligned}$$

where the subscript Ξ, ξ in the final line is to explicitly note that the variables Ξ, ξ are held constant while taking the derivative. Since $U = -\frac{\partial}{\partial\beta} \ln Z$, we can also write this result as

$$\text{var}(H) = \frac{\partial^2}{\partial\beta^2} \ln Z.$$

So both the expectation value and the variance of the energy can be obtained by differentiating the logarithm of the partition function with respect to β .

Since $\text{var}(H) = -\left(\frac{\partial U}{\partial\beta}\right)_{\Xi, \xi}$ and variance is always non-negative, this implies that

$$\left(\frac{\partial U}{\partial\beta}\right)_{\Xi, \xi} \leq 0.$$

This implies that

$$\left(\frac{\partial\beta}{\partial U}\right)_{\Xi, \xi} \leq 0.$$

That is, for canonical states, the internal energy is a monotonically decreasing function of the inverse temperature. This means we can say that the temperature is a monotonically increasing function of the energy

$$\left(\frac{\partial T}{\partial U}\right)_{\Xi, \xi} \geq 0.$$

We can generalize to expectation values and variances of arbitrary sets of observables

$$\text{var}(X_i) = \frac{\partial^2}{\partial Y_i^2} \ln Z(Y, \Xi, \xi) = -\left(\frac{\partial X_i}{\partial Y_i}\right)_{Y_k, \Xi, \xi},$$

where $k \neq i$. Defined in this way, the variance is always non-negative and only equal to zero when the state is an eigenstate of X_i .

2.6 Summary: The State

Skills to Master

-

Entropy

For a mixed state, our ignorance—the amount of information that is missing—is called the *entropy*. The entropy of a statement, message, or situation is a measure of the uncertainty in that statement, message, or situation. The **information** associated with an event is the amount of uncertainty that is removed after the outcome of the event becomes known.

Given N trials of a random experiment, each with d outcomes and having a probability distribution $\{p_i\}$, we obtain an ordered sequence of N outcomes. A sequence is *typical* if the relative frequency of outcome i is equal to its probability. We can assume that all sequences are typical provided that N is large. What is the entropy of this ordered sequence. That is, what is the minimum number of bits needed to losslessly transmit the sequence given that N , d , and $\{p_i\}$ are known by the receiver?

For an arbitrary probability distribution $\{p_i\}$, the average number of bits required, *per trial*, to transmit the sequence is given by the Shannon entropy

$$I[\{p_i\}] = -\sum_{i=1}^d p_i \log_2 p_i.$$

For a physical system in a mixed state characterized by the density operator $\hat{\rho}$, the Shannon entropy is

$$I[\{p_i\}] = -\sum_i \rho_i \log_2 \rho_i = -\text{tr}(\hat{\rho} \log_2 \hat{\rho}).$$

The entropy of statistical mechanics is defined as

$$S[\hat{\rho}] \equiv -k \text{tr}(\hat{\rho} \ln \hat{\rho}),$$

where

$$k \simeq 1.38 \times 10^{-23} \text{ J/K},$$

is Boltzmann's constant.

Properties of the entropy include

- It is non-negative. That is, $S[\hat{\rho}] \geq 0$.
- For a pure state, $S[\hat{\rho}] = 0$.

- The entropy of a maximally mixed system is $\ln d$ bits, where d is the dimension of the Hilbert space. For example, for a two-level system, the maximally mixed state has an entropy of $\ln 2$ bits.

To calculate the entropy, we simply calculate

$$S[\hat{\rho}] \equiv -k \sum_j \lambda_j \ln \lambda_j,$$

where the λ_j are the eigenvalues of $\hat{\rho}$. Since the eigenvalues are always nonnegative, the entropy is always defined. Any zero eigenvalue simply does not contribute to the entropy.

At the quantum level, a state is represented by a density operator $\hat{\rho}$, so we call $\hat{\rho}$ the microstate of the system. The macrostate state of a system can be characterized by a set of complete and minimal macroscopic parameters of three different types:

Control parameters (ξ): Macroscopic parameters like an external magnetic field or the volume of a box which control the Hamiltonian of a system.

Sharp constraints (Ξ): Hard constraints, like conserved particle number, on the outcomes of measurements.

Expectation values ($\langle X \rangle$): If we know the expectation value of some macroscopic parameter X .

A given microstate $\hat{\rho}$ implies a unique macrostate, but a given macrostate does not imply a unique microstate. Given a specific macrostate, we want to choose a representative microstate for the purpose of calculations. According to the **maximum entropy principle**, of all the microstates which are compatible with our specific macrostate, we should choose as representative the microstate which maximizes the entropy.

If the set of parameters describing the macrostate consists only of sharp constraints, then the microstate

with the maximum entropy

$$\hat{\rho} = \frac{1}{Z(\Xi)} \hat{I}_{\Xi}, \quad Z(\Xi) = \text{tr}_{\Xi}(I_{\Xi}) = \dim \mathcal{H}_{\Xi},$$

where \hat{I}_{Ξ} is the identity operator. The entropy of this state is

$$S[\hat{\rho}] = k \ln Z.$$

If the macroscopic data do not include any expectation values, the maximum entropy microstate is the totally mixed state in the Hilbert space selected by the sharp constraints. Otherwise, the maximum entropy microstate is a Gibbs state, which features Lagrange parameters.

If in addition to sharp constraints, we also know an expectation value $\langle \hat{X} \rangle$ of operator \hat{X} , then the maximum entropy microstate is the Gibbs state

$$\rho = \frac{1}{Z(Y, \Xi)} e^{-Y \hat{X}}, \quad Z(Y, \Xi) = \text{tr}_{\Xi}(e^{-Y \hat{X}}).$$

More generally, we write the partition function as

$$Z(Y, \Xi, \xi) = \text{tr}_{\Xi(\xi)}(e^{-\sum_i Y_i X_i(\xi)}).$$

The parameter Y is a Lagrange parameter whose value is adjusted so that ρ yields the correct expectation value of \hat{X} . The entropy of the Gibbs state is

$$S[\rho] = -k \text{tr}(\rho \ln \rho) = k \ln Z + kY \langle X \rangle_{\rho}.$$

The thermodynamic variables of a system are the three (sets of) natural variables Y, Ξ, ξ , and their associated conjugate variables X, Υ, v defined as

$$\begin{aligned} X_i &\equiv -\frac{\partial}{\partial Y_i} \ln [Z(Y, \Xi, \xi)] \\ \Upsilon_i &\equiv \frac{\partial}{\partial \Xi_i} \ln [Z(Y, \Xi, \xi)] \\ v_i &\equiv \frac{\partial}{\partial \xi_i} \ln [Z(Y, \Xi, \xi)]. \end{aligned}$$

The total differential of $\ln Z$ is

$$d(\ln Z) = -\sum_j \langle X_j \rangle dY_j + \sum_j \Upsilon_j d\Xi_j + \sum_j v_j d\xi_j.$$

The parameters X, Ξ, ξ which originally defined the macrostate, are the natural variables of the macroscopic entropy

$$S(X, \Xi, \xi) = k \ln Z(Y, \Xi, \xi) + k \sum_j Y_j \langle X_j \rangle.$$

The Lagrange parameters Y , which are adjusted to yield the correct expectation values, also depend on

the natural variables of the macrostate. That is, $Y = Y(X, \Xi, \xi)$. The total differential of the entropy is

$$d(S/k) = \sum_j Y_j dX_j + \sum_j \Upsilon_j d\Xi_j + \sum_j v_j d\xi_j.$$

This implies the partial derivatives

$$\begin{aligned} Y_j &= \frac{\partial}{\partial X_j} \left(\frac{S}{k} \right) \\ \Upsilon_j &= \frac{\partial}{\partial \Xi_j} \left(\frac{S}{k} \right) \\ v_j &= \frac{\partial}{\partial \xi_j} \left(\frac{S}{k} \right). \end{aligned}$$

For any isolated system, the internal energy U is a constant of the motion. The energy U may be included as a sharp constraint or as an expectation value. Given an energy U , a system can be in three kinds of equilibrium states:

Microcanonical: The energy is given as a sharp constraint.

Canonical: The energy is given as an expectation value, and all other constants of motion are given as sharp constraints. The state is the Gibbs state with the set $\{X_i\}$ being only the Hamiltonian H .

Grand canonical: Energy and particle number N are given as expectation values. The state is a Gibbs state with $\{X_i\} = \{H, N\}$.

In all cases, the internal energy is one of the natural variables of the entropy. Its conjugate variable is the inverse temperature

$$\beta = \frac{\partial}{\partial U} \left(\frac{S}{k} \right).$$

We also have the relation

$$\frac{\partial}{\partial x} \left(\frac{S}{k} \right) = -\beta \frac{\partial U}{\partial x},$$

where $x = \{X, \Xi, \xi\}$ is a natural variable.

With respect to energy, temperature is the conjugate of entropy.

$$T \equiv \frac{\partial U}{\partial S} = \frac{1}{k\beta}.$$

If the energy spectrum of the Hamiltonian is bounded below and unbounded above, which is true for real-world systems, then $0 \leq T \leq \infty$. For a canonical

system, the temperature is a monotonically increasing function of the energy.

For a canonical system,

$$\rho = \frac{1}{Z(\beta, \Xi, \xi)} e^{-\beta H(\xi)}$$

$$Z(\beta, \Xi, \xi) = \text{tr}_{\Xi} \left(e^{-\beta H(\xi)} \right) = \sum_i g_i e^{-\beta E_i},$$

where E_i are the eigenvalues of the Hamiltonian with associated degeneracy factors $\{g_i\}$. The probability that an energy measurement yields the eigenvalue E_i , is

$$P(E_i) = \frac{g_i e^{-\beta E_i}}{Z}.$$

The energy is

$$U = \langle H \rangle = -\frac{\partial}{\partial \beta} \ln Z = -\frac{1}{Z} \frac{\partial Z}{\partial \beta}.$$

The variance of the energy is

$$\text{var}(H) = -\left(\frac{\partial U}{\partial \beta} \right)_{\Xi, \xi} = \frac{\partial^2}{\partial \beta^2} \ln Z.$$

More generally,

$$\text{var}(X_i) = \frac{\partial^2}{\partial Y_i^2} \ln Z(Y, \Xi, \xi) = -\left(\frac{\partial X_i}{\partial Y_i} \right)_{Y_{k \neq i}, \Xi, \xi},$$

where $k \neq i$.

Pressure is the conjugate of the volume V with respect to energy

$$p \equiv -\frac{\partial U}{\partial V},$$

and the chemical potential is the conjugate of the particle number N with respect to energy

$$\mu \equiv \frac{\partial U}{\partial N}.$$

Some useful properties of the trace include

$$\text{tr}(cA) = c \text{tr}A$$

$$\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$$

$$\text{tr}(AB) = \text{tr}(BA).$$

Chapter 3

Simple Systems

Our investigations in the previous chapter were all very abstract—using ideas like sharp constraints, expectation values, and control parameters. Now we will look at several specific examples with concrete variables.

For simple systems, our goal is typically to

1. calculate the energy spectrum
2. calculate the *canonical* partition function
3. relate the energy, entropy, and temperature
4. calculate the variance of the energy

3.1 Quantum Harmonic Oscillator

For a quantum harmonic oscillator, the Hamiltonian is

$$H = \left(N + \frac{1}{2}\right) \hbar\omega.$$

and the energy levels are given by

$$E_n = \hbar\omega \left(n + \frac{1}{2}\right), \quad n = 0, 1, 2, \dots$$

By Eq. (2.4), we get the canonical partition function

$$Z(\beta) = \sum_{n=0}^{\infty} e^{-\beta E_n} = \sum_{n=0}^{\infty} e^{-\beta\hbar\omega(n+\frac{1}{2})} = e^{-\beta\hbar\omega/2} \sum_{n=0}^{\infty} (e^{-\beta\hbar\omega})^n.$$

We now have a geometric series, which has the sum

$$Z(\beta) = e^{-\beta\hbar\omega/2} \cdot \frac{1}{1 - e^{-\beta\hbar\omega}} = \frac{1}{e^{\beta\hbar\omega/2} - e^{-\beta\hbar\omega/2}} = \frac{1}{2 \sinh\left(\frac{\beta\hbar\omega}{2}\right)}.$$

Differentiating gives us the energy

$$U(\beta) = -\frac{\partial}{\partial\beta} \ln Z = -\frac{1}{Z} \frac{\partial Z}{\partial\beta} = \frac{\hbar\omega}{2 \tanh\left(\frac{\beta\hbar\omega}{2}\right)}.$$

Inverting this gives us

$$\beta(U) = \frac{2}{\hbar\omega} \tanh^{-1}\left(\frac{\hbar\omega}{2U}\right).$$

The energy is bounded below ($U \geq \hbar\omega/2$) and unbounded above, so the inverse temperature is positive, $\beta > 0$. This implies that the temperature is always positive $T > 0$.

From Eq. (2.3), with $X = U$ and $Y = \beta$, we get

$$\frac{S(\beta)}{k} = \ln Z + \beta U = -\ln\left(2 \sinh\left[\frac{\beta\hbar\omega}{2}\right]\right) + \beta U.$$

From the above equation for $\beta(U)$, we have that

$$\frac{\beta\hbar\omega}{2} = \tanh^{-1}\left(\frac{\hbar\omega}{2U}\right).$$

Taking the sinh of both sides, we can apply the identity

$$\sinh\left[\tanh^{-1}\left(\frac{1}{x}\right)\right] = \frac{1}{\sqrt{x^2-1}}, \quad x > 1,$$

with $x = 2U/(\hbar\omega)$ to get

$$\sinh\left(\frac{\beta\hbar\omega}{2}\right) = \frac{1}{\sqrt{\left(\frac{2U}{\hbar\omega}\right)^2 - 1}}.$$

Plugging this into $S(\beta)/k$ and simplifying gives us the entropy in terms of the energy

$$\begin{aligned} \frac{S(U)}{k} &= -\ln\left(2\sinh\left[\frac{\beta\hbar\omega}{2}\right]\right) + \beta U = -\ln\left(\frac{2}{\sqrt{\left(\frac{2U}{\hbar\omega}\right)^2 - 1}}\right) + \frac{2U}{\hbar\omega} \tanh^{-1}\left(\frac{\hbar\omega}{2U}\right) \\ &= -\ln 2 + \frac{1}{2} \ln\left[\left(\frac{2U}{\hbar\omega}\right)^2 - 1\right] + \frac{2U}{\hbar\omega} \tanh^{-1}\left(\frac{\hbar\omega}{2U}\right). \end{aligned}$$

What we did here in going from $\ln Z(\beta)$ to $S(U)/k$ is a **Legendre transformation**.

The general steps are:

1. Start from $\ln Z$ as a function of β
2. Differentiate to get $U(\beta)$
3. Invert the result to get $\beta(U)$
4. Plug this into $S(U)/k = \ln Z + \beta U$

Next, we check that $S(U)/k$ has the right properties as derived in the previous chapter:

- We have $S \geq 0$ for all allowed values of U ($U \geq 0$)
- We have $S = 0$ only when we are in the ground state (a pure state) with $U = \hbar\omega/2$
- The entropy increases monotonically with energy, or $\frac{\partial S}{\partial U} > 0$, that is, $\beta > 0$
- The entropy function is “concave down,” meaning $\frac{\partial^2 S}{\partial U^2} < 0$. So the inverse temperature goes down as the energy goes up.
- The first derivative is equal to the inverse temperature, $\frac{\partial S/k}{\partial U} = \beta$

In the low temperature limit, $kT \ll \hbar\omega$, which is equivalent to $\beta\hbar\omega \gg 1$. This implies that $e^{-\beta\hbar\omega} \ll 1$, and so

$$Z = \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}} \approx e^{-\frac{\beta\hbar\omega}{2}}.$$

So

$$\ln Z \approx -\frac{\beta\hbar\omega}{2},$$

and

$$U = -\frac{\partial}{\partial \beta} \ln Z \approx \frac{\hbar\omega}{2}.$$

That is, the internal energy approaches the energy of the ground state in the low temperature limit.

In the high temperature limit, $kt \gg \hbar\omega$, which is equivalent to $\beta\hbar\omega \ll 1$. This implies that $\sinh(\beta\hbar\omega/2) \approx \beta\hbar\omega/2$.

$$Z(\beta) = \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}} = \frac{1}{2\sinh\left(\frac{\beta\hbar\omega}{2}\right)} \approx \frac{1}{\beta\hbar\omega}.$$

So

$$\ln Z = \ln\left(\frac{1}{\beta\hbar\omega}\right),$$

and

$$U = -\frac{\partial}{\partial \beta} \ln Z \approx \beta \hbar \omega \cdot \frac{1}{\beta^2 \hbar \omega} = \frac{1}{\beta} = kT.$$

So in the high temperature limit, the energy scales linearly with temperature, and there is no more dependence on \hbar . In the high temperature limit, the thermal energy scale kT is much larger than the differences between the discrete energy levels. That is, at high temperatures, the quantum nature of the oscillator stops playing a role. In this limit,

$$\frac{S(\beta)}{k} = \ln Z + \beta U \approx \ln \left(\frac{1}{\beta \hbar \omega} \right) + 1 \approx \ln(kT) - \ln(\hbar \omega) \approx \ln(kT) \approx \ln U.$$

The variance of the Hamiltonian can easily be calculated as

$$\text{var}(H) = \frac{\partial^2}{\partial \beta^2} \ln Z = \frac{(\hbar \omega)^2}{4 \sinh^2 \left(\frac{\beta \hbar \omega}{2} \right)}.$$

In the low temperature limit,

$$\text{var}(H) \approx 0, \quad kT \ll \hbar \omega.$$

So for very low temperatures, the energy fluctuations approach zero. This is to be expected since the ground state is pure. In the high temperature limit,

$$\text{var}(H) \approx \frac{1}{\beta^2} = (kT)^2, \quad kT \gg \hbar \omega.$$

So for very high temperatures, the energy fluctuations are of the order of the energy itself.

Example 3.1.1

Suppose an isotropic harmonic oscillator with Hamiltonian

$$H = (N_x + N_y + 1) \hbar \omega,$$

is initially in the state

$$\rho = \frac{7}{8} |00\rangle \langle 00| + \frac{1}{16} |10\rangle \langle 10| + \frac{1}{16} |01\rangle \langle 01|.$$

Calculate $\langle H \rangle$, S , and the occupation probabilities for the states $|00\rangle$, $|10\rangle$, and $|01\rangle$. Also find the probabilities of measuring different values of the energy.

In general,

$$\langle H \rangle_\rho = \text{tr}(\rho H) = \sum_i \langle i | H \rho | i \rangle.$$

Note that $\text{tr}(\rho H) = \text{tr}(H \rho)$ due to the cyclic property of the trace. In our case, the obvious set of basis vectors is the eigenvectors of H , $|i\rangle = |n_x n_y\rangle$ where $n_x, n_y = 0, 1, 2, \dots$. Recall that

$$\langle n_x n_y | n'_x n'_y \rangle = \delta_{n_x n'_x} \delta_{n_y n'_y},$$

so in our case, there are only three nonzero terms in the sum

$$\begin{aligned} \langle H \rangle_\rho &= \sum_{n_x, n_y} \langle n_x n_y | H \rho | n_x n_y \rangle \\ &= \sum_{n_x, n_y} \langle n_x n_y | (N_x + N_y + 1) \hbar \omega \\ &\quad \times \left[\frac{7}{8} |00\rangle \langle 00| + \frac{1}{16} |10\rangle \langle 10| + \frac{1}{16} |01\rangle \langle 01| \right] |n_x n_y\rangle \\ &= \frac{7}{8} \langle 00 | H | 00 \rangle \langle 00 | 00 \rangle + \frac{1}{16} \langle 10 | H | 10 \rangle \langle 10 | 10 \rangle + \frac{1}{16} \langle 01 | H | 01 \rangle \langle 01 | 01 \rangle \\ &= \frac{7}{8} \langle 00 | \hbar \omega | 00 \rangle + \frac{1}{16} \langle 10 | 2\hbar \omega | 10 \rangle + \frac{1}{16} \langle 01 | 2\hbar \omega | 01 \rangle = \frac{9}{8} \hbar \omega. \end{aligned}$$

For the entropy, we get

$$S/k = \text{tr}(\rho \ln \rho) = - \sum_j \lambda_j \ln \lambda_j = -\frac{7}{8} \ln \left(\frac{7}{8}\right) - \frac{1}{16} \ln \left(\frac{1}{16}\right) - \frac{1}{16} \ln \left(\frac{1}{16}\right).$$

The occupation probability for the state $|00\rangle$ is the probability of finding the harmonic oscillator in this state. In general, recall that the probability of measuring a_i is

$$P(a_i|\rho) = \text{tr}(\rho P_i) = \text{tr}(\rho |i\rangle \langle i|) = \sum_j \langle j|\rho|i\rangle \langle i|j\rangle.$$

In our case, the probability of finding it in the ground state is

$$P(|00\rangle) = \sum_{n_x, n_y} \langle n_x, n_y|\rho|00\rangle \langle 00|n_x n_y\rangle.$$

The only surviving terms are where $n_x n_y = 00$, which simply extracts the coefficient of $|00\rangle \langle 00|$ from ρ

$$P(|00\rangle) = \langle 00|\rho|00\rangle = \frac{7}{8}.$$

Similarly,

$$P(|10\rangle) = \frac{1}{16}$$

$$P(|01\rangle) = \frac{1}{16}.$$

Our energy levels are

$$E_{n_x n_y} = (n_x + n_y + 1) \hbar\omega,$$

So the possible energies in this case are $\hbar\omega$ with probability equal to the probability of the system being in $|00\rangle$

$$P(E = \hbar\omega) = \frac{7}{8},$$

and $2\hbar\omega$ with probability equal to the probability of the system being in $|10\rangle$ or $|01\rangle$

$$P(E = 2\hbar\omega) = \frac{1}{16} + \frac{1}{16} = \frac{1}{8}.$$

If we measure the energy to be $\hbar\omega$, then the system can only be in the ground state $|00\rangle$, so after the measurement it is in a pure state $\rho = |00\rangle \langle 00|$ with $S = 0$. If we measure the energy to be $2\hbar\omega$, then after the measurement it is in a mixed state $\rho = \frac{1}{2} |10\rangle \langle 10| + \frac{1}{2} |01\rangle \langle 01|$, since it could be in either $|10\rangle$ or $|01\rangle$, with $S \neq 0$.

3.2 Quantum Rotor

For the quantum rotor, we have only rotational degrees of freedom. The Hamiltonian is

$$H = \frac{1}{2I} \vec{L}^2,$$

where I is the moment of inertia and \vec{L} is the angular momentum operator. The eigenvalues of \vec{L}^2 are $\hbar^2 l(l+1)$ with $l = 0, 1, 2, \dots$, so the energy spectrum is given by

$$E_l = \frac{\hbar^2}{2I} l(l+1), \quad l = 0, 1, 2, \dots$$

However, now our energy levels are degenerate with a degeneracy factor

$$g_l = 2l + 1.$$

The canonical partition function is then

$$Z(\beta) = \sum_{l=0}^{\infty} g_l e^{-\beta E_l} = \sum_{l=0}^{\infty} (2l+1) \exp\left(-\frac{\beta \hbar^2}{2I} l(l+1)\right).$$

This sum doesn't have a nice analytical form, but we can study the low and high temperature limits.

In the low temperature limit, $kT \ll \hbar^2/I$, which implies $\beta \hbar^2/I \gg 1$, and this implies that

$$e^{-\frac{\beta \hbar^2}{I}} \ll 1.$$

Because of this, the terms in the sum quickly become very small, so we can approximate it well by taking only the first two terms

$$Z(\beta) \approx 1 + 3e^{-\frac{\beta \hbar^2}{I}}.$$

With $x = 3e^{-\frac{\beta \hbar^2}{I}} \ll 1$, when we take the natural log, we use the approximation $\ln(1+x) \approx x$ to get

$$\ln Z(\beta) \approx 3e^{-\frac{\beta \hbar^2}{I}}.$$

Taking the partial derivative with respect to β gives us

$$U(\beta) = -\frac{\partial}{\partial \beta} \ln Z = -\frac{3\hbar^2}{I} 3e^{-\frac{\beta \hbar^2}{I}}.$$

Inverting this gives us

$$\beta(U) \approx -\frac{I}{\hbar^2} \ln\left(\frac{UI}{3\hbar^2}\right).$$

Getting the entropy as a function of its natural variable U ,

$$\frac{S(U)}{k} = \ln Z(\beta(U)) + \beta(U)U \approx -\frac{UI}{\hbar^2} \left[1 - \ln\left(\frac{UI}{3\hbar^2}\right)\right].$$

In the limit $T \rightarrow 0$ or $\beta \rightarrow \infty$, this shows that $U \rightarrow 0$ and $S \rightarrow 0$. So for $T \rightarrow 0$, the rotor drops to the non-degenerate pure ground state where it has zero entropy as expected.

In the high temperature limit, $kT \gg \hbar^2/I$, which implies $\beta \hbar^2/I \ll 1$. We now introduce the dimensionless variable

$$x_l = \sqrt{\frac{\beta \hbar^2}{I}} \left(l + \frac{1}{2}\right),$$

then its square is

$$x_l^2 = \frac{\beta \hbar^2}{I} l(l+1) + \frac{\beta \hbar^2}{4I}.$$

Its incremental value is

$$\Delta x = x_{l+1} - x_l = \sqrt{\frac{\beta \hbar^2}{I}}.$$

We can now rewrite the partition function as

$$Z(\beta) = \frac{2I}{\beta \hbar^2} e^{-\frac{\beta \hbar^2}{8I}} \sum_{l=0}^{\infty} x_l e^{-\frac{x_l^2}{2}} \Delta x.$$

For $kT \gg 1$, the Δx become very small, so we can approximate the sum with an integral

$$Z(\beta) \approx \frac{2I}{\beta \hbar^2} e^{-\frac{\beta \hbar^2}{8I}} \int_{x_0}^{\infty} x e^{-\frac{x^2}{2}} dx = \frac{2I}{\beta \hbar^2}.$$

The internal energy is

$$U(\beta) = -\frac{\partial}{\partial \beta} \ln Z = -\frac{1}{Z} \frac{\partial Z}{\partial \beta} \approx -\frac{1}{Z} \frac{\partial}{\partial \beta} \frac{2I}{\beta \hbar^2} = \frac{1}{\beta} = kT.$$

The entropy is

$$\begin{aligned}\frac{S(U)}{k} &= \ln Z + \beta U \approx \ln \left(\frac{2I}{\beta \hbar^2} \right) + 1 = \ln \left(\frac{2UI}{\hbar^2} \right) + 1 \\ &= \ln U + \ln \left(\frac{2I}{\hbar^2} \right) + 1 \sim \ln U \sim \ln T.\end{aligned}$$

So, like with the harmonic oscillator, at high temperatures, the thermal energy scale is much larger than the energy level spacing. Similarly, there is no dependence on \hbar at high temperatures, and the energy scales logarithmically with temperature or energy. This is a general feature. Whenever Z scales as some power of the inverse temperature, i.e. $Z \sim \beta^{-s}$ for $s > 0$, then $S \sim \ln U \sim \ln T$.

At the high and low temperature limits, the variance of H is the same for the rotor as it is for the harmonic oscillator.

3.3 Spin in a Magnetic Field

Now we consider a spin-1/2 particle in a magnetic field. The magnetic field features as a control parameter. A spin-1/2 particle in a magnetic field has a magnetic moment. For an electron, that magnetic moment is called the **Bohr magneton**

$$\mu_B = \frac{e\hbar}{2m},$$

where e is the elementary charge, and m is the mass of the electron. The associated observable is

$$\vec{M} = -\mu_B \vec{\sigma} = -\frac{e\hbar}{2m} \vec{\sigma},$$

where $\vec{\sigma}$ is the vector of Pauli matrices. The sign is negative because the electron's charge is $-e$. We will now systematically analyze this system.

Step 1: We construct the canonical partition function.

We will assume the particle's position is fixed, so that its only degrees of freedom are restricted to the spin degrees of freedom. Without loss of generality, we can take the external magnetic field to be in the z direction. Then the Hamiltonian of our particle is

$$H(B) = -\vec{B} \cdot \vec{M} = \mu_B B \sigma_z.$$

The B in the Hamiltonian functions as a control parameter. The Pauli matrix σ_z has eigenvalues ± 1 , so the energy spectrum consists of only the two levels

$$E_{\pm} = \pm \mu_B B.$$

So the canonical partition function is

$$Z(\beta, B) = \sum_j g_j e^{-\beta E_j} = e^{-\beta \mu_B B} + e^{\beta \mu_B B} = 2 \cosh(\beta \mu_B B).$$

Step 2: We get U by differentiating $\ln Z$ with respect to β .

The energy is

$$U(\beta, B) = -\frac{\partial}{\partial \beta} \ln Z = -\frac{1}{Z} \frac{\partial Z}{\partial \beta} = -\mu_B B \tanh(\beta \mu_B B).$$

Step 3: We invert to get β as a function of the natural variables U and B .

This is easily done to get

$$\beta(U, B) = -\frac{1}{\mu_B B} \tanh^{-1} \left(\frac{U}{\mu_B B} \right).$$

Step 4: Obtain the entropy as a function of the natural variables U and B .

We start with

$$\frac{S(U, B)}{k} = \ln Z + \beta U,$$

where Z is a function of β and B , and β is a function of U .

In this case, we start by using the identity

$$\cosh x = \frac{1}{\sqrt{1 - \tanh^2 x}},$$

to rewrite the partition function as

$$Z(\beta, B) = \frac{2}{\sqrt{1 - \tanh^2(\beta\mu_B B)}}.$$

Next, we solve the equation for $\beta(U, B)$ for the quantity $\beta\mu_B B$, and plug it in to get

$$\begin{aligned} Z(\beta(U), B) &= \frac{2}{\sqrt{1 - \tanh^2\left(-\tanh^{-1}\left(\frac{U}{\mu_B B}\right)\right)}} = \frac{2}{\sqrt{1 - \left(-\frac{U}{\mu_B B}\right)^2}} \\ &= \frac{2}{\sqrt{1 - u^2}} = \frac{2}{\sqrt{(1+u)(1-u)}} = \frac{1}{\sqrt{\frac{1+u}{2} \frac{1-u}{2}}}, \end{aligned}$$

where $u \equiv U/(\mu_B B)$ is the energy in units of $\mu_B B$. Taking the natural log gives us

$$\ln Z = -\frac{1}{2} \left[\ln\left(\frac{1+u}{2}\right) + \ln\left(\frac{1-u}{2}\right) \right].$$

So plugging $\ln Z$, $\beta(U)$, and U into the general form for the entropy given earlier, we get

$$\frac{S(U, B)}{k} = -\frac{1}{2} \left[\ln\left(\frac{1+u}{2}\right) + \ln\left(\frac{1-u}{2}\right) \right] - u \tanh^{-1}(u).$$

Using the identity

$$\tanh^{-1} x = \frac{1}{2} \ln\left(\frac{1+x}{1-x}\right),$$

allows us to write the entropy more compactly as

$$\frac{S(U, B)}{k} = - \left[\frac{1+u}{2} \ln\left(\frac{1+u}{2}\right) + \frac{1-u}{2} \ln\left(\frac{1-u}{2}\right) \right].$$

Since $-\mu_B B < U < \mu_B B$, we know that $-1 < u < 1$. The entropy is zero at $u = \pm 1$, corresponding to the pure eigenstates, and it is positive between -1 and 1 . In this case, the entropy is not a monotonic function of the energy. This implies that the temperature could be negative. This occurs because the energy spectrum is bounded above and below as opposed to the harmonic oscillator and rotor examples in which the energy spectrum was unbounded above.

Step 5: Find the low ($kT \ll \mu_B B$) and high ($kT \gg \mu_B B$) temperature limits.

In the low temperature limit, $kT \ll \mu_B B$, which can also be written as $\beta\mu_B B \gg 1$. In this limit, $e^{-\beta\mu_B B} \rightarrow 0$, so the partition function becomes

$$Z \approx e^{\beta\mu_B B}.$$

This implies that

$$\ln Z \approx \beta\mu_B B,$$

and the internal energy becomes

$$U = -\frac{\partial}{\partial \beta} \ln Z \approx -\mu_B B.$$

So in the low temperature limit, the energy approaches that of the lowest energy pure state. In this limit, the energy goes to zero

$$S = \ln Z + \beta U \approx \beta\mu_B B + \beta(-\mu_B B) = 0.$$

In the high temperature limit, we have $kT \gg \mu_B B$ or $\beta\mu_B B \ll 1$. Then the partition function becomes

$$Z \approx 2 + (\beta\mu_B B)^2,$$

since

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots$$

Then

$$\ln Z \approx \ln 2 + \frac{1}{2} (\beta \mu_B B)^2,$$

since

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots.$$

The internal energy is then

$$U = -\frac{\partial}{\partial \beta} \approx -(\mu_B B)^2 \beta.$$

Apparently, $U \rightarrow 0$ from below as $T \rightarrow \infty$. It does not approach the upper of the two energy levels as one might have expected. In the high temperature limit, the entropy becomes

$$S/k = \ln Z + \beta U \approx \ln 2 - \frac{1}{2} \left(\frac{U}{\mu_B B} \right)^2.$$

Notice that in the high temperature limit, the energy and entropy still depend on the Planck constant (via μ_B). Since the energy spectrum is bounded above, even in the high temperature limit, the quantum nature cannot be neglected.

Step 6: Calculate the variance of the energy.

The variance is calculated as

$$\begin{aligned} \text{var}(H) &= \frac{\partial^2}{\partial \beta^2} \ln Z(\beta) = \frac{\partial^2}{\partial \beta^2} \ln (2 \cosh(\beta \mu_B B)) = \mu_B B \frac{\partial}{\partial \beta} \tanh(\beta \mu_B B) \\ &= \mu_B B [1 - \tanh^2(\beta \mu_B B)] \mu_B B = \left[\frac{\mu_B B}{\cosh(\beta \mu_B B)} \right]^2. \end{aligned}$$

In the low temperature limit,

$$\cosh(\beta \mu_B B) \approx \frac{1}{2} e^{\beta \mu_B B} \rightarrow \infty,$$

so

$$\text{var}(H) \rightarrow 0,$$

in the low temperature limit. So in this limit, the energy fluctuations approach zero.

In the high temperature limit,

$$\cosh(\beta \mu_B B) \approx 1,$$

so

$$\text{var}(H) \rightarrow (\mu_B B)^2.$$

So in this limit, the fluctuations $\sqrt{\text{var}(H)} = \mu_B B$ are of the order of the range of the energy spectrum itself.

3.4 Paramagnetism

A paramagnetic material is one in which the magnetic dipoles (formed by the atomic electrons) align parallel with an externally applied field forming an overall induced magnetic dipole moment. This is as opposed to a diamagnetic material in which the magnetic moments align antiparallel to the external field and a ferromagnetic field in which the induced overall dipole moment may persist even after the external field is turned off.

A simple model for a paramagnet is a rigid lattice formed by N non-interacting spins. This lattice is placed in a uniform external magnetic field $\vec{B} = B \hat{z}$. In this model, the spins interact with the external magnetic field, but they do not interact with each other. Now, N is a sharp constraint, B is a control parameter, and $U = \langle H \rangle$ is an expectation value.

Recall that the Hamiltonian of a single spin system is $H(B) = \mu_B B \sigma_z$. For the paramagnet, we have N of these spin systems. Since they do not interact with each other,

the Hamiltonian for this model is simply the sum of the Hamiltonians of the individual spin systems

$$H(B) = \sum_{i=1}^{\infty} H_i(B) = \mu_B B \sum_{i=1}^{\infty} \sigma_{z,i},$$

where Pauli matrix $\sigma_{z,i}$ is of the i th spin system. In reality, the above means

$$H(B) = (H_1 \otimes I \otimes \cdots \otimes I) + (I \otimes H_2 \otimes \cdots \otimes I) + \cdots + (I \otimes \cdots \otimes H_N).$$

That is, the operator H_i acts only on the i th spin system, and not on the others. This implies that the exponential of the total Hamiltonian factorize as

$$e^{-\beta H} = e^{-\beta H_1} \otimes e^{-\beta H_2} \otimes \cdots \otimes e^{-\beta H_N}.$$

Recall that the partition function has the form $Z = \text{tr}(e^{-\beta H})$, and that the trace has the property $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$. For our model, this implies that

$$\begin{aligned} Z(\beta, N, B) &= \text{tr} [e^{-\beta H_1} \otimes e^{-\beta H_2} \otimes \cdots \otimes e^{-\beta H_N}] \\ &= \text{tr} [e^{-\beta H_1}] \text{tr} [e^{-\beta H_2}] \cdots \text{tr} [e^{-\beta H_N}] = Z_1 Z_2 \cdots Z_N. \end{aligned}$$

That is, the partition function of our model is just the product of the partition functions of the individual spin systems. For a single spin system, the partition function was $Z_i = 2 \cosh(\beta \mu_B B)$, so for our paramagnet, the partition function is simply

$$Z(\beta, N, B) = [2 \cosh(\beta \mu_B B)]^N.$$

The factorization of Z implies a number of things:

- The statistical operator of the paramagnet factorizes into the tensor product of the individual statistical operators

$$\rho = \frac{1}{Z} e^{-\beta H} = \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_N.$$

This means the individual magnetic moments are statistically independent—there are no correlations.

- The logarithm of the partition function is additive

$$\ln Z = \ln Z_1 + \ln Z_2 + \cdots + \ln Z_N.$$

For the paramagnet,

$$\ln Z(\beta, N, B) = N [2 \cosh(\beta \mu_B B)].$$

- The energy is additive

$$U = -\frac{\partial}{\partial \beta} \ln Z = U_1 + U_2 + \cdots + U_N.$$

In our case, the energy of a single spin does not depend on its location in the lattice. Therefore, the energies of the single spins are all the same, and the total energy is just N times the energy of a single spin

$$U(\beta, N, B) = -N \mu_B B \tanh(\beta \mu_B B).$$

- The entropy is additive

$$\frac{S}{k} = \ln Z + \beta U = \frac{S_1}{k} + \frac{S_2}{k} + \cdots + \frac{S_N}{k}.$$

This satisfies a general property—whenever systems are uncorrelated, the overall entropy is the sum of the individual entropies.

The entropy of our paramagnet is

$$\frac{S(\beta, N, B)}{k} = \ln Z + \beta U = N \ln [2 \cosh(\beta \mu_B B)] - \beta N \mu_B B \tanh(\beta \mu_B B).$$

We can also invert the energy to write β in terms of U , B , and N as

$$\beta(U, N, B) = -\frac{1}{\mu_B B} \tanh^{-1} \left(\frac{U}{N \mu_B B} \right).$$

Then the entropy can be written as

$$\frac{S(U, N, B)}{k} = \ln Z + \beta U = N \ln [2 \cosh(\beta \mu_B B)] - \frac{U}{\mu_B B} \tanh^{-1} \left(\frac{U}{N \mu_B B} \right).$$

Using the identity

$$\cosh x = \frac{1}{\sqrt{1 - \tanh^2 x}},$$

and letting

$$u = \frac{U}{N \mu_B B},$$

we can write the entropy as

$$\begin{aligned} \frac{S(U, N, B)}{k} &= N \ln \left[\frac{2}{\sqrt{1 - \tanh^2(\tanh^{-1} u)}} \right] - N u \tanh^{-1} u \\ &= N \ln \left[\frac{2}{\sqrt{1 - u^2}} \right] - \frac{N u}{2} \ln \left(\frac{1 + u}{1 - u} \right) \\ &= N \ln \left[\frac{1}{\sqrt{\frac{1+u}{2} \cdot \frac{1-u}{2}}} \right] - \frac{N u}{2} \ln \left(\frac{1 + u}{1 - u} \right) \\ &= -N \left[\frac{1 + u}{2} \ln \left(\frac{1 + u}{2} \right) + \frac{1 - u}{2} \ln \left(\frac{1 - u}{2} \right) \right]. \end{aligned}$$

Differentiating this with respect to the energy gives us

$$\begin{aligned} \frac{\partial(S/k)}{\partial U} &= -N \left[\frac{1}{2} \ln \left(\frac{1 + u}{2} \right) - \frac{1}{2} \ln \left(\frac{1 - u}{2} \right) \right] \frac{\partial u}{\partial U} = -N \left[\frac{1}{2} \ln \left(\frac{1 + u}{1 - u} \right) \right] \frac{\partial u}{\partial U} \\ &= -N \tanh^{-1} u \frac{\partial u}{\partial U} = -\frac{1}{\mu_B B} \tanh^{-1} \left(\frac{U}{N \mu_B B} \right) = \beta, \end{aligned}$$

as expected.

The natural variables of U are S , B , and N . In the high temperature limit $kT \gg \mu_B B$ or $\beta \mu_B B \ll 1$, which implies that $u = U/N \mu_B B$ is small. After expanding all of the logarithms in $S(U, N, B)/k$, we can use the approximation

$$\ln(1 \pm u) \simeq \pm u - \frac{1}{2} u^2,$$

to get

$$\frac{S(U, N, B)}{k} \simeq N \left[\ln 2 - \frac{1}{2} \left(\frac{U}{N \mu_B B} \right)^2 \right].$$

Inverting this gives us the energy in the high temperature limit as a function of its natural variables

$$U(S, B, N) \simeq N \mu_B B \sqrt{2 \ln 2 - \frac{2S}{kN}}.$$

Next, we want to calculate the total magnetic moment M . With respect to the entropy, the conjugate pair is $(B, \beta M)$. That is,

$$\frac{\partial(S/k)}{\partial B} = \beta M.$$

Recall from the previous chapter, that for a control parameter ξ_i ,

$$v_i = \frac{\partial(S/k)}{\partial \xi_i} = \frac{\partial}{\partial \xi_i} \ln Z(Y, \Xi, \xi).$$

Here, $B = \xi$ is a control parameter, so

$$\frac{\partial(S/k)}{\partial B} = \frac{\partial}{\partial B} \ln Z,$$

and therefore,

$$\beta M = \frac{\partial}{\partial B} \ln Z = \frac{\partial}{\partial B} N \ln [2 \cosh(\beta \mu_B B)],$$

which gives us the total magnetic moment

$$M = N \mu_B \tanh(\beta \mu_B B).$$

Plugging in $\beta = -(1/\mu_B B) \tanh^{-1}(U/N\mu_B B)$, gives us

$$U = -BM.$$

With respect to the energy, we have the conjugate pair $(B, -M)$, so

$$\frac{\partial U}{\partial B} = -M.$$

Differentiating

$$U = -\frac{\partial}{\partial \beta} \ln Z = -N \mu_B B \tanh(\beta \mu_B B),$$

with respect to B gives us again

$$M = N \mu_B \tanh(\beta \mu_B B).$$

For an individual spin, the energy levels are

$$E_{\pm} = \pm \mu_B B,$$

and the partition function is

$$Z = 2 \cosh(\beta \mu_B B).$$

Recall that the probability of finding a system at energy E_i is

$$P(E_i|\rho) = \frac{g_i e^{-\beta E_i}}{Z}.$$

So the probability that a specific spin's magnetic moment is aligned (+) or anti-aligned (-) with the external magnetic field, as a function of temperature, is

$$P_{\pm} = \frac{e^{\pm \frac{\mu_B B}{kT}}}{2 \cosh\left(\frac{\mu_B B}{kT}\right)}.$$

From $U = -N \mu_B B \tanh(\beta \mu_B B)$, we have that

$$u = -\tanh(\beta \mu_B B),$$

where $u = U/N\mu_B B$. Then,

$$\frac{1+u}{2} = \frac{e^{-\frac{\mu_B B}{kT}}}{2 \cosh\left(\frac{\mu_B B}{kT}\right)} = P_-,$$

and

$$\frac{1-u}{2} = \frac{e^{\frac{\mu_B B}{kT}}}{2 \cosh\left(\frac{\mu_B B}{kT}\right)} = P_+.$$

This allows us to write the entropy in terms of the probabilities as

$$S/k = -N \left[\frac{1+u}{2} \ln \left(\frac{1+u}{2} \right) + \frac{1-u}{2} \ln \left(\frac{1-u}{2} \right) \right] = -N [P_- \ln(P_-) + P_+ \ln(P_+)].$$

So the entropy per spin is

$$\frac{S}{N} = -k [P_- \ln(P_-) + P_+ \ln(P_+)].$$

We know that

$$P_{\pm} = \frac{1 \mp u}{2} = \frac{1 \mp \frac{U}{N\mu_B B}}{2} = \frac{N\mu_B B \mp U}{2N\mu_B B}.$$

But $U = -BM$, so we can write the probability for a spin's magnetic moment to be aligned with the external magnetic field as

$$P_{\pm} = \frac{1 \pm m}{2}.$$

where

$$m = \frac{M}{N\mu_B},$$

is the average magnetic moment per elementary spin (in units of the Bohr magneton).

The entropy per elementary magnet is

$$\frac{S}{N} = -k [P_- \ln(P_-) + P_+ \ln(P_+)] = -k \left[\frac{1-m}{2} \ln \left(\frac{1-m}{2} \right) + \frac{1+m}{2} \ln \left(\frac{1+m}{2} \right) \right].$$

We can also write this in terms of T and B if we plug $P_{\pm} = e^{\pm \frac{\mu_B B}{kT}} / 2 \cosh(\mu_B B / (kT))$ into the first equation.

Plugging $\beta = -(1/\mu_B B) \tanh^{-1}(U/N\mu_B B)$ into $T = 1/(k\beta)$, substituting in $U = -BM$ and then using $m = M/N\mu_B$, we find that

$$T(m, B) = \frac{\mu_B B}{k \tanh^{-1}(m)}.$$

3.5 2D Harmonic Oscillator

Consider a 2-dimensional isotropic harmonic oscillator, whose Hamiltonian is simply the sum of the Hamiltonians of two 1D oscillators

$$H = (N_x + N_y + 1) \hbar\omega.$$

If the energy is given as an expectation value, then the maximally mixed state is the Gibbs state with

$$\rho = \frac{1}{Z} e^{-\beta H} = \frac{1}{Z} e^{-\beta(N_x + N_y + 1)\hbar\omega}.$$

The energy levels are

$$E_{n_x n_y} = (n_x + n_y + 1) \hbar\omega,$$

where $n_x, n_y = 0, 1, 2, \dots$

The partition function is

$$Z = \sum_j g_j e^{-\beta E_j} = \sum_{n_x} \sum_{n_y} e^{-\beta(n_x + n_y + 1)\hbar\omega} = e^{-\beta\hbar\omega} \sum_{n_x} (e^{-\beta\hbar\omega})^{n_x} \sum_{n_y} (e^{-\beta\hbar\omega})^{n_y}.$$

This is simply the square of the partition function for a 1D harmonic oscillator, so

$$Z = \frac{1}{4 \sinh^2 \left(\frac{\hbar\omega}{2kT} \right)}.$$

The energy is then

$$U = -\frac{\partial}{\partial \beta} \ln Z = \frac{\hbar\omega}{\tanh \left(\frac{\hbar\omega}{2kT} \right)}.$$

The entropy is

$$S/k = \ln Z + \beta U = -2 \ln \left[2 \sinh \left(\frac{\hbar\omega}{2kT} \right) \right] + \frac{\hbar\omega}{kT \tanh \left(\frac{\hbar\omega}{2kT} \right)}.$$

Example 3.5.1

Calculate the occupation probability for the ground state of the 2D harmonic oscillator as a function of temperature.

In the ground state, $n_x = n_y = 0$, and so the ground state energy is $E_{00} = \hbar\omega$.

The occupation probability is calculated as

$$P(E_{00}|\rho) = \frac{ge^{-\beta E_{00}}}{Z}.$$

The degeneracy of the ground state is zero, so $g = 1$. Plugging in $E_{00} = \hbar\omega$, $\beta = 1/kT$ and Z gives us

$$P(E_{00}|\rho) = 4e^{-\frac{\hbar\omega}{kT}} \sinh^2\left(\frac{\hbar\omega}{2kT}\right).$$

Now, consider an anisotropic harmonic oscillator with frequency $\omega/2$ in the x direction and 2ω in the y direction. The Hamiltonian is now

$$H = \left(N_x + \frac{1}{2}\right) \frac{\hbar\omega}{2} + \left(N_y + \frac{1}{2}\right) 2\hbar\omega = \left(\frac{1}{2}N_x + 2N_y + \frac{5}{4}\right) \hbar\omega.$$

The partition function again factors into two geometric series

$$Z = e^{-\frac{5}{4}\beta\hbar\omega} \sum_{n_x} \left(e^{-\frac{1}{2}\beta\hbar\omega}\right)^{n_x} \sum_{n_y} \left(e^{-2\beta\hbar\omega}\right)^{n_y}.$$

Using the formula for the sum of a geometric series and then simplifying, we get

$$Z = \frac{1}{4 \sinh\left(\frac{\hbar\omega}{4kT}\right) \sinh\left(\frac{\hbar\omega}{kT}\right)}.$$

The internal energy is

$$U = -\frac{\partial}{\partial\beta} \ln Z = \frac{\hbar\omega}{4 \tanh\left(\frac{\hbar\omega}{4kT}\right)} + \frac{\hbar\omega}{\tanh\left(\frac{\hbar\omega}{kT}\right)}.$$

3.6 Summary: Simple Systems

Skills to Master

•

Simple Systems

Typically, we will be given the Hamiltonian H for a simple canonical system, and have to calculate various quantities.

1. Obtain the energy levels E_i of the system. These are the eigenvalues of H .
2. Calculate the partition function

$$Z = \sum_i g_i e^{-\beta E_i},$$

where $\beta = 1/kT$ and g_i is the degeneracy of the i th energy level. To calculate the probability that an energy measurement yields the eigenvalue E_i , use

$$P(E_i) = \frac{g_i e^{-\beta E_i}}{Z}.$$

3. Obtain the energy of the system

$$U = -\frac{\partial}{\partial \beta} \ln Z = -\frac{1}{Z} \frac{\partial Z}{\partial \beta}.$$

4. After obtaining U , solve it for β to get β in terms of its natural variables.
5. Calculate the entropy of the system as a function of its natural variables by plugging β into

$$S/k = \ln Z + \beta U.$$

6. Calculate the variance of the energy (or Hamiltonian) as

$$\text{var}(H) = \frac{\partial^2}{\partial \beta^2} \ln Z.$$

7. Check the above quantities in the low temperature limit $kT \ll \hbar\omega$ or $\beta\hbar\omega \gg 1$.
8. Check the above quantities in the high temperature limit $kT \gg \hbar\omega$ or $\beta\hbar\omega \ll 1$.

You may also be given the Hamiltonian H of a system that is in a specified initial state

$$\rho = \sum_j p_j |\psi_j\rangle \langle \psi_j|.$$

When ρ is written in this way, the coefficient p_j is the probability for the system to be in the j th state $|j\rangle$.

The probability that a measurement yields a given energy E' is the sum of the coefficients of the states with that energy. The entropy is

$$S/k = \text{tr}(\rho \ln \rho) = -\sum_j \lambda_j \ln \lambda_j,$$

where the λ_j are the eigenvalues of ρ . The energy expectation value is

$$\langle H \rangle_\rho = \text{tr}(\rho H) = \sum_i \langle i | H \rho | i \rangle.$$

When choosing the basis vectors $|i\rangle$, use the eigenvectors of the Hamiltonian.

Composite Systems

The following applies to composite systems of simple systems in which the simple systems do not interact with each other:

1. The Hamiltonian is sum of the Hamiltonians of the simple systems.

$$H = H_1 + H_2 + H_3 + \dots$$

2. The partition function is the product of the partition functions of the simple systems

$$Z = Z_1 Z_2 Z_3 \dots$$

3. The statistical operator is the tensor product of the statistical operators of the simple systems

$$\rho = \rho_1 \otimes \rho_2 \otimes \rho_3 \otimes \dots$$

4. The energy is additive

$$U = U_1 + U_2 + U_3 \dots$$

5. The entropy is additive

$$\frac{S}{k} = \frac{S_1}{k} + \frac{S_2}{k} + \frac{S_3}{k} + \dots$$

Misc.

For a quantum harmonic oscillator, the Hamiltonian is

$$H = \left(N + \frac{1}{2} \right) \hbar\omega.$$

For a quantum rotor, the Hamiltonian is

$$H = \frac{1}{2I} \vec{L}^2,$$

where I is the moment of inertia and \vec{L} is the angular momentum operator. The eigenvalues of \vec{L}^2 are $\hbar^2 l(l+1)$ with $l = 0, 1, 2, \dots$. For a simple spin system, the Hamiltonian is

$$H(B) = -\vec{B} \cdot \vec{M} = \mu_B B \sigma_z,$$

with eigenvalues ± 1 .

For the paramagnet, we can calculate the total magnetic moment M . With respect to the entropy, the conjugate pair is $(B, \beta M)$. That is,

$$\frac{\partial(S/k)}{\partial B} = \beta M.$$

Here, $B = \xi$ is a control parameter, so we also have the relation

$$\frac{\partial(S/k)}{\partial B} = \frac{\partial}{\partial B} \ln Z,$$

With respect to the energy, we have the conjugate pair $(B, -M)$, so

$$\frac{\partial U}{\partial B} = -M.$$

For the entropy, remember that

- The quantity S/k is unitless
- $S \geq 0$ for all allowed values of U
- We have $S = 0$ only when we are in the ground state (a pure state)
- Whenever Z scales as some power of the inverse temperature, i.e. $Z \sim \beta^{-s}$ for $s > 0$, then $S \sim \ln U \sim \ln T$.
- The first derivative is equal to the inverse temperature, $\frac{\partial S/k}{\partial U} = \beta$

For the temperature, remember that

1. If the energy spectrum is bounded below and not from above (e.g. harmonic oscillator) and the state is canonical, then the temperature is always non-negative, finite, and increases monotonically with the energy.
2. If the energy spectrum is bounded above and below (as in the simple spin system), then the temperature can be negative.

Useful formulae:

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$$

$$\sinh x = \frac{e^x - e^{-x}}{2}$$

$$\cosh x = \frac{e^x + e^{-x}}{2}$$

$$\cosh^2 x - \sinh^2 x = 1$$

$$\cosh x = \frac{1}{\sqrt{1 - \tanh^2 x}}$$

$$\tanh^{-1} x = \frac{1}{2} \ln \left(\frac{1+x}{1-x} \right)$$

If x is small, then

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

Chapter 4

The Thermodynamic Limit

For the model of a paramagnet in the previous chapter, the relative size of fluctuations scales as $1/\sqrt{N}$. That is, in the macroscopic limit where N is large, it no longer matters whether the energy is given as an expectation value or as a sharp constraint. In general, a system composed of N identical constituents, becomes macroscopic as N becomes large.

Macroscopic systems are typically homogeneous and thermodynamically stable.

4.1 Homogeneity

A system is homogeneous if all samples taken from the system exhibit the same thermodynamics properties regardless of the location, the size, or the shape of the sample. That is, every sample of a homogeneous macroscopic system is thermodynamically representative of the system as a whole.

Note, homogeneity does depend on the length scale. At the scale of a meter or two, the ocean may be homogeneous, but it is not homogeneous on the very large scale. A small sample drawn from the bottom of the Marianna trench will not have the same thermodynamics properties as a small sample drawn from near the surface. Homogeneity also breaks down on the very small scale. A single droplet of water is an inhomogeneous system due to surface effects of a water droplet. Finally, a system, like the ocean, is nonhomogeneous at a phase transition—such as between ice and liquid near the poles, even though each may be homogeneous within a phase.

Suppose the total volume of a macroscopic system is V , and the volume of a small sample is some fraction f , where $0 < f \leq 1$, of the total volume. Then the volume of the sample is fV . Then the macroscopic system is homogeneous if any two samples of volume fV drawn from the macroscopic system have the same energy fU , regardless of the shape or location of the sample. Since the internal energy of a sample must be independent of its shape, this implies that it is independent of the size and shape of the sample's interface with the surrounding material. In other words, there is no surface contribution to the internal energy of a sample, and this implies that we can neglect the interactions of a sample with its surroundings.

Since the properties of a sample are independent of position, the energy density of the macroscopic system must be uniform, and a sample that occupies a fraction f of the total volume of the system with total energy U , must have energy fU . We can make the same argument for the other additive variables like particle number N . All of these variables that scale with the volume of the system are called **extensive variables**.

Since all samples of a homogeneous system must have the same thermodynamic variables, the entropy for two samples with the same volume must be the same. This implies that there are no surface contributions to the entropy. If we partition a homogeneous system into multiple pieces of the same volume, then each piece must have the same entropy. This is true whether the pieces stay together or are separated. This implies that

there are no statistical correlations between different samples in a homogeneous system. However, removing statistical correlations causes an increase in entropy, so if the entropy does not change when the partitions are moved apart, there must not have been any statistical correlations between those regions to begin with. In other words, for a system to be homogeneous, the statistical correlations between different regions must be negligible.

Since the entropy of the whole is the sum of the entropies of its pieces, the entropy is also an extensive variable. To calculate the entropy of a sample, we can either calculate the total entropy and then multiply it by its fraction f , or we can calculate the entropy of the sample. Both approaches must give the same value, which implies the relations

$$f \frac{\partial}{\partial U} S(U, N, V)/k = \frac{\partial}{\partial(fU)} S(fU, fN, fV)/k = \frac{\partial}{\partial(fU)} fS(U, N, V)/k.$$

This implies that the temperature of the sample is the same as the temperature of the system. That is, temperature is an **intensive variable**.

Given a conjugate pair of thermodynamic variables, one is extensive and the other is intensive. For example, U and β are conjugate pairs with respect to the entropy, and U is extensive and β is intensive. Similarly, S and T are conjugate variables with respect to the energy, and S is extensive and T is intensive. In general, the conjugate, with respect to the entropy or energy, of an extensive quantity is intensive and vice versa.

Above, we compared the entropy of a sample in two different ways to conclude that temperature is intensive. We can do the same thing with the energy. Doing so implies the relation

$$U = \frac{\partial U}{\partial S} S + \frac{\partial U}{\partial V} V + \frac{\partial U}{\partial N} N = TS - pV + \mu N.$$

This is the **Euler relation** for homogeneous systems.

The total differential of the internal energy, obtained from the fact that U is a function of S , V , and N , is

$$dU = \frac{\partial U}{\partial S} dS + \frac{\partial U}{\partial V} dV + \frac{\partial U}{\partial N} dN = T dS - p dV + \mu dN.$$

However, the total differential of the internal energy as obtained from the Euler relation is

$$dU = T dS - p dV + \mu dN + S dT - V dp + N d\mu.$$

Comparing these two gives us the **Gibbs-Duhem relation**

$$S dT - V dp + N d\mu = 0.$$

With this, we can write any of the three intensive variables T , p , and μ in terms of the other two.

Intensive variables contain no information about the volume of the macroscopic system, so we always need at least one extensive variable to fully characterize a macroscopic system.

Solving the Euler relation for the entropy gives us

$$S = k\beta U + k\beta pV - k\beta \mu N.$$

We can write this relation as

$$S = \frac{\partial S}{\partial U} U + \frac{\partial S}{\partial V} V + \frac{\partial S}{\partial N} N,$$

and we can generalize it to

$$S = \sum_i \frac{\partial S}{\partial x_i} x_i,$$

where the x_i are arbitrary extensive variables. Taking the total differential of this gives us

$$dS = \sum_i \frac{\partial S}{\partial x_i} dx_i + \sum_{ij} \frac{\partial^2 S}{\partial x_i \partial x_j} x_i dx_j.$$

However, the total differential of the entropy is

$$dS = \sum_i \frac{\partial S}{\partial x_i} dx_i,$$

so by comparison, we must have

$$\sum_{ij} \frac{\partial^2 S}{\partial x_i \partial x_j} x_i dx_j = 0,$$

which implies

$$\sum_i \frac{\partial^2 S}{\partial x_i \partial x_j} x_i = 0.$$

The matrix of the second partial derivatives of the entropy is called the **Hessian** of the entropy.

$$\begin{bmatrix} \frac{\partial^2 S}{\partial x_1^2} & \frac{\partial^2 S}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 S}{\partial x_2 \partial x_1} & \frac{\partial^2 S}{\partial x_1^2} & \\ \vdots & & \ddots \end{bmatrix}.$$

The sum given above implies that the vector

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix},$$

is an eigenvector of the Hessian, and its associated eigenvalue is 0.

4.2 Stability

For a macroscopic system to be stable, different regions of the system must not spontaneously separate into regions with different thermodynamic properties. A homogeneous and stable system must be in the unique state which has the highest possible entropy that is compatible with the macroscopic variables.

Since a homogeneous state has a strictly larger entropy than an inhomogeneous state, this implies that the entropy as a function of U or V or N is strictly concave (down)

$$\frac{\partial^2 S}{\partial U^2} < 0, \quad \frac{\partial^2 S}{\partial V^2} < 0, \quad \frac{\partial^2 S}{\partial N^2} < 0.$$

We can generalize this further to

$$\begin{bmatrix} \frac{\partial^2 S}{\partial U^2} & \frac{\partial^2 S}{\partial U \partial N} \\ \frac{\partial^2 S}{\partial U \partial N} & \frac{\partial^2 S}{\partial N^2} \end{bmatrix} < 0, \quad \begin{bmatrix} \frac{\partial^2 S}{\partial U^2} & \frac{\partial^2 S}{\partial U \partial V} \\ \frac{\partial^2 S}{\partial U \partial V} & \frac{\partial^2 S}{\partial V^2} \end{bmatrix} < 0, \quad \begin{bmatrix} \frac{\partial^2 S}{\partial N^2} & \frac{\partial^2 S}{\partial N \partial V} \\ \frac{\partial^2 S}{\partial N \partial V} & \frac{\partial^2 S}{\partial V^2} \end{bmatrix} < 0.$$

That is, the three Hessian “submatrices” are negative definite. A necessary and sufficient condition for this is that the determinants are strictly positive

$$\begin{vmatrix} \frac{\partial^2 S}{\partial U^2} & \frac{\partial^2 S}{\partial U \partial N} \\ \frac{\partial^2 S}{\partial U \partial N} & \frac{\partial^2 S}{\partial N^2} \end{vmatrix} > 0, \quad \begin{vmatrix} \frac{\partial^2 S}{\partial U^2} & \frac{\partial^2 S}{\partial U \partial V} \\ \frac{\partial^2 S}{\partial U \partial V} & \frac{\partial^2 S}{\partial V^2} \end{vmatrix} > 0, \quad \begin{vmatrix} \frac{\partial^2 S}{\partial N^2} & \frac{\partial^2 S}{\partial N \partial V} \\ \frac{\partial^2 S}{\partial N \partial V} & \frac{\partial^2 S}{\partial V^2} \end{vmatrix} > 0.$$

For the full 3×3 Hessian, we require

$$\begin{vmatrix} \frac{\partial^2 S}{\partial U^2} & \frac{\partial^2 S}{\partial U \partial N} & \frac{\partial^2 S}{\partial U \partial V} \\ \frac{\partial^2 S}{\partial U \partial N} & \frac{\partial^2 S}{\partial N^2} & \frac{\partial^2 S}{\partial N \partial V} \\ \frac{\partial^2 S}{\partial U \partial V} & \frac{\partial^2 S}{\partial N \partial V} & \frac{\partial^2 S}{\partial V^2} \end{vmatrix} = 0.$$

4.3 Ensembles

Recall that there are three kinds of thermodynamic parameters: control parameters, expectation values, and sharp constraints. There are also three common types of equilibrium states or **ensembles** determined by the type of parameters in our system. For systems whose macrostate is determined by the three thermodynamic variables U , N , and V , following are the different ensembles:

Microcanonical: U and N are sharp constraints, V is a control parameter, and there are no expectation values.

Canonical: U is an expectation value, N is a sharp constraint, and V is a control parameter.

Grand Canonical: U and N are both expectation values, there are no sharp constraints, and V is a control parameter.

In the thermodynamic limit, a system is large, homogeneous, and stable. The three ensembles defined above are not in general equivalent. However, in the thermodynamic limit, they are equivalent. That is, for any system, in the thermodynamic limit,

$$S_M = S_C = S_G.$$

That is, we can treat it as a microcanonical system and calculate the entropy S_M , or treat it as a canonical system and calculate the entropy S_C , or treat it as a grand canonical system and calculate the entropy S_G , and we will find that all three entropies are the same in the thermodynamic limit. Since all other thermodynamic variables can be calculated from the entropy, all three ensembles must be completely equivalent in the thermodynamic limit.

For example, let's consider the paramagnet of the previous chapter as a microcanonical ensemble whose energy U is given as a sharp constraint with some precision ΔU . That is, the energy of the paramagnet is in the interval $[U - \Delta U, U + \Delta U]$. We want to calculate the entropy of the system.

There are a total of N elementary magnets, each of which has two possible orientations—up or down. So there are a total of 2^N possible microstates. We want to find the number of microstates whose energies are compatible with our sharp constraint.

Suppose at energy U , there are N_+ elementary magnets with spin up, and N_- with spin down. Then the energy of this microstate is

$$U = N_+ \mu_B B - N_- \mu_B B.$$

Since there are N total magnets, we know that $N = N_+ + N_-$, so

$$U = N_+ \mu_B B - (N - N_+) \mu_B B.$$

This allows us to solve for N_+ and N_- :

$$N_{\pm} = \frac{N}{2} \pm \frac{U}{2\mu_B B}.$$

If we define

$$u = \frac{U}{N\mu_B B},$$

as before, we can write this as

$$N_{\pm} = \frac{N}{2} (1 \pm u).$$

Suppose at energy $U + \Delta U$, there are $N_+ + n$ up spins. Then there are $N - N_+ - n$ down spins and the energy is

$$U + \Delta U = (N_+ + n) \mu_B B - (N - N_+ - n) \mu_B B = U + 2n \mu_B B.$$

This implies that

$$n = \frac{\Delta U}{2\mu_B B}.$$

There are n energy levels in $(U, U + \Delta U]$, so there are a total of $2n + 1$ compatible energy levels in $[U - \Delta U, U + \Delta U]$.

At energy U , there are N total magnets and N_+ with spin up and N_- with spin down. Calculating the number of microstates that are compatible with this is a matter of computing a permutation with repetition. The number of compatible microstates at energy U is

$$\frac{N!}{N_+!N_-!}.$$

If we increase the number of up spins by one (and therefore decrease the number of down spins by one), we have a degeneracy of

$$\frac{N!}{(N_+ + 1)!(N_- - 1)!},$$

and so on. So the total number of compatible microstates for energies $[U - \Delta U, U + \Delta U]$ is the sum

$$Z_M = \sum_{i=-n}^n \frac{N!}{(N_+ + i)!(N_- - i)!}.$$

If we assume that the number $(2n + 1)$ of compatible energy levels is much smaller than the total number of energy levels for the paramagnet, then we can approximate this sum as

$$Z_M \simeq (2n + 1) \frac{N!}{N_+!N_-!}.$$

This is our partition function.

The microcanonical entropy is calculated as

$$\frac{S}{k} = \ln Z_M.$$

In our case, we use the approximation $\ln M! \simeq M \ln M - M$ to approximate the logarithm

$$\ln Z_M = \ln(2n + 1) - \ln(N_+!) - \ln(N_-!).$$

After doing that and plugging in n , N_+ , and N_- , and performing some algebra, we find that

$$\frac{S}{k} = \ln \left(\frac{\Delta U}{\mu_B B} + 1 \right) + S_C,$$

where

$$S_C = -N \left\{ \frac{1+u}{2} \ln \left(\frac{1+u}{2} \right) + \frac{1-u}{2} \ln \left(\frac{1-u}{2} \right) \right\},$$

is the canonical entropy for the paramagnet. So we see that if $\Delta U = 0$, then the microcanonical entropy equals the canonical entropy.

4.4 The Laws of Thermodynamics

In the thermodynamic limit, all macroscopic states and processes obey the four laws of thermodynamics. These laws govern equilibrium (zeroth law), energy (first law), entropy (second law), and the ground state (third law).

Zeroth Law

The zeroth law of thermodynamics deals with the equilibrium of isolated systems. An isolated system is in equilibrium if and only if its macrostate is completely characterized by the values of the system's constants of motion.

Two systems, which are individually in equilibrium, are defined to be in equilibrium with each other if upon contact with each other, their macroscopic properties do not change. Here, "contact" is defined to be a weak coupling between the two systems that

allows for the exchange of extensive resources like energy or particles. So two systems are in equilibrium with each other if the exchange of extensive resources is possible, but no net exchange takes place.

Formally, the zeroth law states that *two systems are in equilibrium with each other if and only if all the intensive thermodynamic variables which are conjugate to the resources being exchanged between the two systems have the same values in both systems*. For example:

- If energy exchange is possible between the two systems, then they are in equilibrium if and only if they have the same temperature.
- If particle exchange is possible between the two systems, then they are in equilibrium if and only if they have the same chemical potential.
- If volume exchange is possible between the two systems (i.e. the systems are separated by a movable wall), then they are in equilibrium if and only if they have the same pressure.

Suppose we have two systems A and B which are individually in equilibrium. Suppose energy exchange is allowed between the two systems, but all other extensive resources (like particle number and volume) are fixed. We will consider two cases:

Case 1: We assume that A is hotter than B ($T_A > T_B$). Since energy exchange between the two systems is allowed, U_A and U_B can change, but $U_A + U_B$ is constant. Let dU be the energy that flows from the hotter body to the colder body, then $U_A \rightarrow U_A - dU$ and $U_B \rightarrow U_B + dU$. Recall the relation

$$d(S/k) = \beta dU + y dx.$$

This tells us that the entropies of the two systems change as

$$dS_A = -\frac{1}{T_A} dU, \quad dS_B = \frac{1}{T_B} dU,$$

so the total entropy change is

$$dS = dS_A + dS_B = \left(1 - \frac{T_B}{T_A}\right) dS_B.$$

But since $T_A > T_B$ and $dS_B > 0$, this implies that $dS > 0$. That is, the total entropy is increased by the transfer of energy from the hotter body to the colder body. This means the total system before the internal transfer was not the maximum entropy state, and so it was not an equilibrium state. In other words, systems A and B were not in equilibrium with each other.

Case 2: Now suppose that A and B had the same temperatures from the outset ($T_A = T_B$). Then

$$dS = dS_A + dS_B = \left(1 - \frac{T_B}{T_A}\right) dS_B = 0.$$

So the infinitesimal flow of energy does not alter the total entropy. This means the total entropy was already at a maximum, and so A and B must be in equilibrium with each other.

The zeroth law implies an equivalence relation for equilibrium. If systems A and B are in equilibrium with each other, and B and C are in equilibrium with each other, then A and C are in equilibrium with each other. This relation allows an experimental approach to measuring an intensive variable—for example, temperature. We can measure the temperature of a system A using a reference system R of known temperature by bringing the two systems into contact with each other provided that one system is much larger than the other. If R is much larger than A , then the temperature of A will rise to that of R and the temperature change of R will be negligible. In this case, R is called a **heat bath** or **thermostat**. If A is much larger than R , then the temperature of R will become the same as the temperature of A , and the temperature change of A is negligible. In this case, R is called a **thermometer**.

First Law

The first law is concerned with energy and its conservation, and it consists of two parts:

- The internal energy of a system is a *system property*
- The internal energy of an isolated system is conserved

If a system is composed of several parts which are in thermal contact with each other, i.e. they are allowed to exchange energy with each other, then it is the total energy that is conserved.

In the thermodynamic limit, the energy needed to establish or break thermal contact between parts of a larger system is small compared to the total energy of the system. That is, the contribution to the total energy by the interfaces between the parts of a system is negligible. This implies that the total energy of a system (with or without thermal contact between the parts) is just the sum of the energies of the parts.

Second Law

In a previous section, we discussed the second law of thermodynamics, which states that for a reproducible macroscopic process, the entropy of an isolated system cannot decrease. We required a distinction between thermodynamic variables given as expectation values and those given as sharp constraints. Entropy and reproducibility of a process depended on the choice of ensemble. Furthermore, we noted the possibility that statistical fluctuations are exempt from the second law and could temporarily decrease the entropy.

In the thermodynamic limit, the different ensembles become equivalent, and statistical fluctuations become negligible. Furthermore, we can assume that all macroscopic processes are reproducible. Therefore, in the thermodynamic limit, we can drop the caveats noted above, and the second law becomes simply the statement that *the entropy of an isolated system cannot decrease*.

The entropy is a **system property** or **state property**. This means it is a unique feature of the system's macrostate, and its value does not depend on how that macrostate was achieved or on what ensemble was used. For a system composed of multiple parts, it is the total entropy of the system that cannot decrease. The entropy of individual parts may decrease.

Third Law

The third law of thermodynamics, also called **Nernst's theorem**, deals with the properties of the ground state of a macroscopic system. It is the only one of the four laws of thermodynamics whose origin is purely quantum mechanical.

By definition, a system is in its ground state if and only if its temperature is exactly zero, $T = 0$. If the ground state is a unique pure state, i.e., it is not degenerate, then the entropy is also exactly zero, $S = 0$. If the ground state is degenerate with a degeneracy factor g_0 , then the entropy of the ground state at zero temperature is

$$S = k \ln g_0.$$

Empirically, we find that in all real systems, the multiplicity of the ground state grows more slowly than exponentially with the number of constituents. This implies that

$$\lim_{N \rightarrow \infty} \frac{S(T=0)}{N} = \lim_{N \rightarrow \infty} \frac{\ln g_0}{N} = 0.$$

More generally, the third law states that *the entropy per constituent tends to zero as the temperature approaches zero*

$$\boxed{\lim_{N \rightarrow \infty} \frac{S}{N} = 0, \quad \text{as } T \rightarrow 0.}$$

The third law holds for all real systems, but it may be violated by model systems, as it is in the simple model of a paramagnet analyzed in a previous section. The third law may also be formulated in terms of the entropy density S/V in the limit of large volume

$$\lim_{V \rightarrow \infty} \frac{S}{V} = 0, \quad \text{as } T \rightarrow 0.$$

The two formulations are equivalent provided that both N and V are taken to infinity in the macroscopic limit, and their ratio (i.e. the particle density) remains fixed. When the macroscopic variables include N and not V (as in spin systems), then the third law must be formulated in terms of N . When the macroscopic variables include V and not N (as in a photon gas), then the third law must be formulated in terms of V .

One implication of the third law is that it is impossible to create and maintain any macroscopic system at $T = 0$.

4.5 Summary: The Thermodynamic Limit

Skills to Master

-

Chapter 5

Perfect Gases

When many identical particles move freely in space and there is no interaction between the particles, then we have what is called a **perfect gas**. The study of perfect gases began with the kinetic theory of gases developed by Maxwell and Boltzmann. Here, we will treat a perfect gas as a quantum system. If the particles have no internal structure, and their only property is mass, then we are dealing with a **monoatomic gas**. If the particles have internal structure with internal dynamics such as rotations and vibrations, then we are dealing with a **molecular gas**. We will begin our study of perfect gases by assuming they are monoatomic.

Since the perfect gas is a quantum mechanical system, its particles are indistinguishable as a matter of principle. Overall, a many-particle state is written in the form $|n_1 n_2 \dots\rangle$ where n_i is the number of particles in the single-particle state $|i\rangle$. The set $\{|i\rangle\}$ of single-particle states forms an orthonormal basis. The many-particle basis states span a Hilbert space called the **Fock space**.

According to the **spin-statistics theorem**,

- Particles with integer spin (e.g. photons) are called **bosons**. For bosons, multiparticle states are symmetric under the exchange of identical particles. This implies that there's no limitation on the number of particles allowed in a single-particle state. That is, a single-particle state $|i\rangle$ can have any number (from 0 to ∞) of particles.
- Particles with half-integer spin (e.g. electrons) are called **fermions**. For fermions, multiparticle states are antisymmetric under the exchange of identical particles. Then the Pauli exclusion principle limits the number of particles that can occupy a single particle state $|i\rangle$ to 0 or 1.

Since there is no interaction between the particles in a perfect gas, the total energy of the gas is just the sum of the single-particle energies. The Hamiltonian of a perfect gas is just

$$\hat{H} = \sum_i E_i \hat{n}_i,$$

where E_i is the energy of the single particle level i , and \hat{n}_i , the **occupation number operator**, counts the number of particles in level i . For bosons, \hat{n}_i has eigenvalues $0, 1, \dots, \infty$. For fermions, it has eigenvalues 0 and 1. The total number of particles in the gas is just

$$\hat{N} = \sum_i \hat{n}_i.$$

There is no interaction *between* the particles of a perfect gas, but there may be interactions between the particles and an external potential (such as the walls of a box). Any such external potential is accounted for in the individual energy levels E_i , and so it doesn't change the overall Hamiltonian \hat{H} .

For a perfect gas, we are given the macroscopic information: internal energy U , total particle number N , and volume V . In the thermodynamic limit, all ensembles are equivalent, so we are free to choose the one that makes the calculations the easiest—the grand canonical ensemble. In this ensemble, the partition function is

$$Z(\beta, \alpha, V) = \text{Tr} \left[e^{-\sum_i (\beta E_i - \alpha) \hat{n}_i} \right],$$

where

$$\alpha = \beta\mu.$$

On the right side of Z , the volume V is encoded in the individual energy levels E_i via the external potential. The trace can be evaluated over the Fock space as

$$Z(\beta, \alpha, V) = \sum_{n_1, n_2, \dots} \langle n_1 n_2 \dots | e^{-\sum_i (\beta E_i - \alpha) \hat{n}_i} | n_1 n_2 \dots \rangle,$$

where the sum goes over all the n_i . For each sum, it goes over all of the allowed values for that n_i (i.e. $0, 1, \dots, \infty$ for bosons and 0 and 1 for fermions). The number operator \hat{n}_i simply counts the number of particles in state $|i\rangle$, so we can easily evaluate this trace. The matrix elements are

$$\begin{aligned} \langle n_1 n_2 \dots | e^{-\sum_i (\beta E_i - \alpha) \hat{n}_i} | n_1 n_2 \dots \rangle &= \langle n_1 n_2 \dots | n_1 n_2 \dots \rangle e^{-\sum_i (\beta E_i - \alpha) n_i} \\ &= e^{-\sum_i (\beta E_i - \alpha) n_i} = e^{-(\beta E_1 - \alpha) n_1 - (\beta E_2 - \alpha) n_2 \dots} \\ &= e^{-(\beta E_1 - \alpha) n_1} e^{-(\beta E_2 - \alpha) n_2} \dots \\ &= \prod_i \left[e^{-(\beta E_i - \alpha)} \right]^{n_i} = \prod_i z_i^{n_i}, \end{aligned}$$

where

$$z_i = e^{-(\beta E_i - \alpha)} = e^{-\beta(E_i - \mu)}.$$

So the partition function becomes

$$Z(\beta, \alpha, V) = \sum_{n_1, n_2, \dots} \prod_i z_i^{n_i} = \sum_{n_1} z_1^{n_1} \sum_{n_2} z_2^{n_2} \dots = \prod_i \sum_n z_i^n.$$

For bosons, the sum goes over $n = 0, 1, \dots, \infty$. This yields an infinite geometric series with sum $(1 - z_i)^{-1}$, so for bosons,

$$Z(\beta, \alpha, V) = \prod_i (1 - z_i)^{-1}.$$

For fermions, the sum goes over two values $n = 0, 1$. This yields the sum $(1 + z_i)$, so for fermions

$$Z(\beta, \alpha, V) = \prod_i (1 + z_i).$$

Both cases can be combined as

$$Z(\beta, \alpha, V) = \prod_i (1 \pm z_i)^{\pm 1},$$

where the plus sign is taken for fermions and the minus sign is taken for bosons.

The logarithm of the partition function is then

$$\ln Z(\beta, \alpha, V) = \pm \sum_i \ln(1 \pm z_i).$$

For the grand canonical ensemble, we know that the total particle number is given by

$$N = \frac{\partial}{\partial \alpha} \ln Z(\beta, \alpha, V).$$

In our case, this gives us

$$N = \frac{\partial}{\partial \alpha} \left[\pm \sum_i \ln(1 \pm z_i) \right] = \pm \sum_i \frac{1}{1 \pm z_i} \left(\pm \frac{\partial z_i}{\partial \alpha} \right) = \sum_i \frac{z_i}{1 \pm z_i} = \sum_i \frac{1}{1/z_i \pm 1}.$$

Here, we used the fact that $\frac{\partial z_i}{\partial \alpha} = z_i$. Plugging in the definition of z_i gives us

$$N = \sum_i \frac{1}{e^{(\beta E_i - \alpha)} \pm 1} = \sum_i \frac{1}{e^{\beta(E_i - \mu)} \pm 1}.$$

This can be interpreted as the sum of the expected occupation numbers of the single-particle state i , so

$$\langle n_i \rangle = \frac{1}{e^{(\beta E_i - \alpha)} \pm 1} = \frac{1}{e^{\beta(E_i - \mu)} \pm 1}.$$

5.1 Continuum Limit

Suppose our particles are confined in a fixed volume. We can model this in quantum mechanics using the infinite square well. In 1D, a non-relativistic particle in a square well has the energy levels

$$E_n = \frac{\hbar^2 \pi^2 n^2}{2mL^2}, \quad n = 1, 2, 3, \dots,$$

where m is the mass of the particle, and L is the width of the potential well. For a 3D potential well (cube), the energy levels are

$$E_{\vec{n}} = \frac{1}{2m} \left(\frac{\hbar \pi}{L} \right)^2 (n_x^2 + n_y^2 + n_z^2),$$

where $n_x, n_y, n_z = 1, 2, 3, \dots$, and $\vec{n} = (n_x, n_y, n_z)$. We now have degeneracy—there are many states available for each energy of a single particle.

In the large volume limit ($L \rightarrow \infty$), the spacing between the energy levels goes to zero, and the energy levels effectively form a continuum. Then the sum over the discrete energy levels in the partition function, can be approximated by an integral

$$\sum_{\vec{n}} \rightarrow \int_{n_x, n_y, n_z > 0} d^3 \vec{n}.$$

Since the partition function is only a function of the energy, and the energy is a function only of the n 's squared, then the partition function depends only on the length of the index vector

$$n = |\vec{n}| = \sqrt{n_x^2 + n_y^2 + n_z^2} = \frac{L}{\hbar \pi} \sqrt{2mE_{\vec{n}}}.$$

We can use spherical coordinates and integrate over the angles. Since $n_x, n_y, n_z > 0$, we only integrate over one-eighth of the full solid angle (i.e. only over the octant in which all three coordinates are positive)

$$\int_{n_x, n_y, n_z > 0} d^3 \vec{n} = \int dn_x dn_y dn_z = \int_0^n \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} n'^2 \sin \theta d\theta d\phi dn' = \frac{\pi}{2} \int_0^n n'^2 dn'.$$

Plugging in

$$n^2 = \left(\frac{L}{\hbar \pi} \right)^2 2mE,$$

and

$$dn = \frac{mL}{\hbar \pi \sqrt{2mE}} dE,$$

from the derivative $\frac{dn}{dE}$ gives us

$$\frac{\pi}{2} \int_0^n n'^2 dn' = \frac{(2m)^{\frac{3}{2}} L^3}{(2\pi)^2 \hbar^3} \int \sqrt{E} dE.$$

So we can replace the discrete sum over energy levels in the partition function by an integral over the energy

$$\sum_i \rightarrow \int \mathcal{D}(E) dE,$$

This is the **continuum limit**. Here

$$\mathcal{D}(E) = \frac{gV}{(2\pi)^2} \left(\frac{\sqrt{2m}}{\hbar} \right)^3 \sqrt{E}, \quad E \geq 0,$$

is the **density of states** or **level density**, where $V = L^3$ is the volume, and g is the degeneracy factor. Note, $\mathcal{D}(E) = 0$ for $E < 0$.

The total number of energy levels below some given energy E is the integral of the level density

$$\mathcal{N}(E) = \int_{-\infty}^E \mathcal{D}(E') dE'.$$

We can also think of the level density as the derivative of this quantity: $\mathcal{D}(E) = \mathcal{N}'(E)$. For non-relativistic particles in a 3D box, we get

$$\mathcal{N}(E) = \frac{2gV}{3(2\pi)^2} \left(\frac{\sqrt{2m}}{\hbar} \right)^3 E^{\frac{3}{2}}, \quad E \geq 0.$$

Recall that

$$\ln Z(\beta, \alpha, V) = \pm \sum_i \ln(1 \pm z_i).$$

In the continuum limit, this is now

$$\ln Z(\beta, \alpha, V) = \pm \int_{-\infty}^{\infty} \mathcal{D}(E) \ln(1 \pm z_i) dE = \pm \int_{-\infty}^{\infty} \mathcal{D}(E) \ln(1 \pm e^{-\beta E + \alpha}) dE.$$

We can integrate this by parts ($\int u dv = uv - \int v du$) with

$$\begin{aligned} u &= \ln(1 \pm e^{-\beta E + \alpha}) \\ dv &= \mathcal{D}(E) dE \\ du &= \mp \frac{\beta e^{-\beta E + \alpha}}{1 \pm e^{-\beta E + \alpha}} dE = \mp \frac{\beta}{e^{\beta E - \alpha} \pm 1} dE \\ v &= \mathcal{N}(E). \end{aligned}$$

The uv term goes away because $\ln(1 \pm e^{-\beta E + \alpha})$ evaluated at $E \rightarrow \infty$ and $\mathcal{N}(E)$ evaluated at $E \rightarrow -\infty$ are both zero. So

$$\ln Z(\beta, \alpha, V) = \beta \int_{-\infty}^{\infty} \frac{\mathcal{N}(E)}{e^{\beta E - \alpha} \pm 1} dE.$$

So this gives us an alternative way of calculating the logarithm of the grand canonical partition function.

The factor

$$f(E) = \frac{1}{e^{\beta E - \alpha} \pm 1},$$

in the integrand is called the **occupation factor**, and it has the same for as $\langle n_i \rangle$.

The total number of particles is

$$N = \frac{\partial}{\partial \alpha} \ln Z(\beta, \alpha, V) = \int_{-\infty}^{\infty} \mathcal{D}(E) f(E) dE.$$

This is the continuum limit of the earlier equation

$$N = \sum_i \frac{1}{e^{(\beta E_i - \alpha)} \pm 1}.$$

Then for non-relativistic particles in 3D, the total number of particles is

$$N \propto \left(\frac{\sqrt{mkT}}{\hbar} \right)^3 V \int_0^{\infty} \frac{\sqrt{x}}{e^{x - \alpha} \pm 1} dx, \quad (5.1)$$

where $x = \beta E = E/kT$.

5.2 Classical Limit

The classical limit occurs when quantum effects can be neglected. This occurs when the wave functions of the particles barely overlap. In other words, the typical distance between particles is much larger than their typical wavelength λ . For a gas of N particles confined in a volume V , the volume available per particle is $\sim V/N$, which implies an average distance between particles of $\sim (V/N)^{1/3}$. From the de Broglie wavelength of a particle $\lambda = \hbar/p$ with $p = \sqrt{mE}$, and a typical energy scale of $E \sim kT$, we get that a perfect gas can be treated classically if

$$\sqrt[3]{\frac{V}{N}} \gg \frac{\hbar}{\sqrt{mkT}}. \quad (5.2)$$

This means we can treat a gas classically when the density is low or the temperature is high. On the other hand, quantum effects become significant when the density is high or the temperature is low. At room temperature, a typical gas can be treated classically.

Comparing Eq. (5.1) and (5.2) implies that for a gas to be treated classically, we need

$$\int_0^\infty \frac{\sqrt{x}}{e^{x-\alpha} \pm 1} dx \ll 1.$$

This is guaranteed only if $e^{-\alpha} \gg 1$. Then

$$e^{-\alpha} \gg 1 \implies e^{x-\alpha} \gg 1 \implies e^{x-\alpha} \pm 1 \rightarrow e^{x-\alpha}.$$

That is, in the classical limit, there is no longer a need to distinguish between bosons and fermions since $e^{x-\alpha} \pm 1 \rightarrow e^{x-\alpha}$. Now, we are able to evaluate the integral in the log of the partition function:

$$\begin{aligned} \ln Z(\beta, \alpha, V) &= \beta \int_{-\infty}^{\infty} \frac{\mathcal{N}(E)}{e^{\beta E - \alpha} \pm 1} dE \propto \beta V \int_0^\infty \frac{E^{3/2}}{e^{\beta E - \alpha} \pm 1} dE \\ &\propto \beta^{-3/2} V \int_0^\infty \frac{x^{3/2}}{e^{x-\alpha} \pm 1} dx, \end{aligned}$$

where $x = \beta E$. Here, we changed the lower limit of integration from $-\infty$ to 0 because $\mathcal{N}(E) = 0$ for $E \leq 0$. In the classical limit, $e^{x-\alpha} \pm 1 \rightarrow e^{x-\alpha}$, so our integral becomes

$$\ln Z(\beta, \alpha, V) \propto e^\alpha \beta^{-3/2} V \int_0^\infty x^{3/2} e^{-x} dx.$$

The integral is now just some number, so we conclude that in the classical limit,

$$\ln Z(\beta, \alpha, V) \propto e^\alpha \beta^{-3/2} V. \quad (5.3)$$

Differentiating this gives us the particle number

$$N = \frac{\partial}{\partial \alpha} \ln Z(\beta, \alpha, V) = \ln Z(\beta, \alpha, V).$$

Recall that pressure is defined as

$$p = -\frac{\partial U}{\partial V},$$

and in general

$$\frac{\partial(S/k)}{\partial x} = -\beta \frac{\partial U}{\partial x}.$$

So with $x = V$, we have that

$$p = -\frac{\partial U}{\partial V} = \frac{1}{\beta} \frac{\partial(S/k)}{\partial V} = \frac{1}{\beta} \frac{\partial}{\partial V} \ln Z(\beta, \alpha, V).$$

This is a general result, and is true not just for a perfect gas.

A perfect gas is assumed to be homogeneous, which implies that the logarithm of the grand canonical partition function must be extensive, and so it must scale linearly with volume. This means the derivative with respect to the volume is equivalent to dividing

by V . That is,

$$p = \frac{1}{\beta} \frac{\partial}{\partial V} \ln Z(\beta, \alpha, V) = \frac{1}{\beta V} \ln Z(\beta, \alpha, V).$$

Comparing this with Eq. (5.3) implies that

$$e^\alpha \propto p\beta^{\frac{5}{2}}.$$

This quantity is called the **fugacity**.

Comparing $p = (1/\beta V) \ln Z$ with $N = \ln Z$ gives us the classical **ideal gas law**

$$\boxed{pV = NkT.}$$

The internal energy is

$$U = -\frac{\partial}{\partial \beta} \ln Z(\beta, \alpha, V) = \frac{3}{2\beta} \ln Z(\beta, \alpha, V).$$

Combining this with N and the ideal gas law, implies

$$U = \frac{3}{2}NkT = \frac{3}{2}pV.$$

5.3 Fermi Gas

A Fermi gas is a perfect gas consisting of fermions. A Fermi gas is a good approximation for several physical systems including electrons in a metal or the neutrons inside of a neutron star. Many of the important physical properties of a Fermi gas are determined by a small fraction of the total particles—those whose energies are close to the edge of the Fermi distribution.

Recall that for a perfect gas,

$$\langle n_i \rangle = \frac{1}{e^{(\beta E_i - \alpha)} \pm 1} = \frac{1}{e^{\beta(E_i - \mu)} \pm 1}.$$

For fermions, we use the plus sign in the formula, so

$$\langle n_i \rangle_f = \frac{1}{e^{(\beta E_i - \alpha)} + 1} = \frac{1}{e^{\beta(E_i - \mu)} + 1}.$$

In the low temperature limit, $T \rightarrow 0$ or equivalently, $\beta \rightarrow \infty$, we have a discontinuous jump at $E_i = \mu$. In this limit,

$$\langle n_i \rangle \rightarrow \begin{cases} 1 & \text{for } E_i < \mu \\ 0 & \text{for } E_i > \mu. \end{cases}$$

At $T = 0$, $\langle n_i \rangle$ becomes a step function as shown in Fig. (5.1). Then

- If μ is given, then N is the number of energy levels below μ
- If N is given (i.e. in the canonical and microcanonical ensembles), then μ must lie somewhere between the N th and the $(N + 1)$ th energy levels

At nonzero temperature, the discontinuity becomes smeared out as shown in Fig. (5.2), and μ becomes the energy at which the expected occupation is 1/2. Now, if N is given, we can no longer state with certainty that μ lies between the N th and the $(N + 1)$ th energy levels. The chemical potential $\mu(T)$ is now a function of the temperature, and at high temperatures, it may even lie below the ground state of the system.

When the temperature goes to zero, we can approximate the occupation factor by a step function,

$$f(E) \sim \tilde{f}(E) = \begin{cases} 1 & \text{for } E < \mu \\ 0 & \text{for } E > \mu. \end{cases}$$

Here, the tilde on the second f indicates that it is the step function approximation. We can use this to simplify the grand canonical partition function to get

$$\ln \tilde{Z}(\beta, \alpha, V) = \int_{-\infty}^a \mathcal{N}(x/\beta) dx,$$

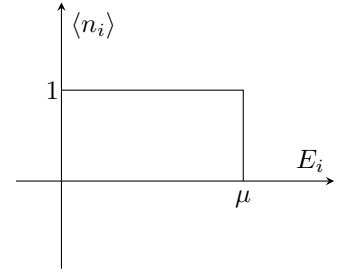


Figure 5.1: Particle number as a step function.

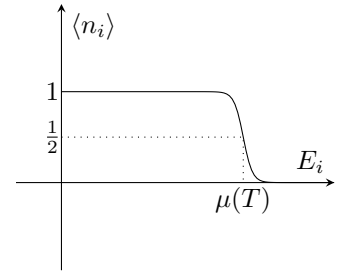


Figure 5.2: Particle number as a “smeared” step function.

where $x = \beta E$. Previously, we found that for non-relativistic particles, the number of levels scales as $\mathcal{N}(x/\beta) \propto x^{\frac{3}{2}}$. So

$$\ln \tilde{Z}(\beta, \alpha, V) = \frac{2\alpha}{5} \mathcal{N}(x/\beta) \propto \alpha^{\frac{5}{2}} \beta^{-\frac{3}{2}}.$$

In the low temperature limit (i.e. the step-function approximation), all levels up to the chemical potential are filled with one particle, and all levels above the chemical potential are empty. This means the total number of particles equals the number of energy levels up to the chemical potential

$$\tilde{N} = \frac{\partial}{\partial \alpha} \ln \tilde{Z}(\beta, \alpha, V) = \mathcal{N}(\alpha/\beta) = \mathcal{N}(\mu).$$

At $T = 0$, this result is exact. That is, $\tilde{N} = N$, and the total number of particles equals the number of states below the **Fermi energy** E_F

$$N = N(E_F),$$

where E_F is the chemical potential at zero temperature, which is also called the Fermi energy. At higher temperatures, the result is only an approximation $\tilde{N} \approx N$.

Since a perfect Fermi gas is homogeneous, we can use the general formula for the pressure of a homogeneous perfect gas

$$p = \frac{1}{\beta} \frac{\partial}{\partial V} \ln Z(\beta, \alpha, V) = \frac{1}{\beta V} \ln Z(\beta, \alpha, V).$$

Plugging in $\ln Z$ from the step-function approximation gives us

$$\tilde{p} = \frac{1}{\beta V} \ln \tilde{Z}(\beta, \alpha, V) = \frac{2\alpha \tilde{N}}{5\beta V}.$$

At $T = 0$, this result is exact. That is, $\tilde{p} = p$ and $\tilde{N} = N$. And since at $T = 0$, $\mu = \alpha/\beta = E_F$, we have that at $T = 0$

$$pV = \frac{2}{5} N E_F.$$

This is completely different from the classical ideal gas law $pV = NkT$. At $T = 0$, this implies that a Fermi gas has a non-zero positive **degeneracy pressure**

$$p = \frac{2NE_F}{5V}.$$

This is a direct consequence of the **Pauli exclusion principle**, which tells us that the particles cannot all settle into the ground state. Some particles must occupy higher energy states even at zero temperature. Associated with these higher energy states is non-zero momenta, which gives us the non-zero pressure. Recall that

$$p = - \left(\frac{\partial U}{\partial V} \right)_N.$$

This relation implies that if we compress a degenerate Fermi gas, reducing the wavelengths of all the wavefunctions then the energies must increase.

It is the electron degeneracy pressure in white dwarf stars that resists the gravitational collapse of the star. Similarly, in neutron stars, the neutron degeneracy pressure resists the gravitational collapse of the star. Note, the electron degeneracy pressure in white dwarfs has nothing to do with the electrostatic repulsion between electrons or from the thermal motion of the electrons. Rather, this degeneracy pressure derives from the antisymmetrization requirement for the wave functions of identical fermions.

We can also show that the internal energy of a Fermi gas will never decrease to zero, even at zero temperature.

$$\tilde{U} = - \frac{\partial}{\partial \beta} \ln \tilde{Z}(\beta, \alpha, V) = \frac{3}{2\beta} \ln \tilde{Z}(\beta, \alpha, V).$$

Using

$$\ln \tilde{Z}(\beta, \alpha, V) = \frac{2\alpha}{5} \tilde{N},$$

gives us

$$\tilde{U} = \frac{3\alpha\tilde{N}}{5\beta}.$$

At $T = 0$, this result becomes exact and

$$U = \frac{3}{5}NE_F.$$

This is different from the classical result

$$U = \frac{3}{2}NkT,$$

which implies that $U = 0$ when $T = 0$. However, if we combine the result for U with the previous result for p , then the relationship between U and p

$$U = \frac{3}{2}pV,$$

agrees with the classical result.

Recall that for particles confined in a box,

$$E_{\vec{n}} = \frac{1}{2m} \left(\frac{\hbar\pi}{L} \right)^2 (n_x^2 + n_y^2 + n_z^2).$$

For fermions at $T = 0$, there is a maximum energy—the Fermi energy—beyond which the states in the system are not occupied. That is, there is a maximum value n_{max} of $n = |\vec{n}|$. The Fermi energy corresponds to the surface of the sphere in momentum space with radius $n_{max} = n_F$

$$E_F = \frac{1}{2m} \left(\frac{\hbar\pi}{L} \right)^2 n_F^2 = \frac{1}{2m} \left(\frac{\hbar\pi}{V^{1/3}} \right)^2 n_F^2.$$

In the continuum limit, we showed that the number of levels with energy below E is

$$\mathcal{N}(E) = \frac{2gV}{3(2\pi)^2} \left(\frac{\sqrt{2m}}{\hbar} \right)^3 E^{3/2}, \quad E \geq 0.$$

For electrons, we have degeneracy $g = 2$ since there are two spin orientations. Solving for E gives us

$$E_F = \frac{\hbar^2}{2m} \left(3\pi^2 \frac{N}{V} \right)^{2/3},$$

where $N = \mathcal{N}(E_F)$. Note that E_F is an intensive variable that depends on the ratio N/V . The boundary separating the occupied and unoccupied states in momentum space is called the **Fermi surface**. Now at $T = 0$, we have that

$$U = \frac{3}{5}NE_F = \frac{3}{5} \frac{\hbar^2}{2m} \left(\frac{3\pi^2}{V} \right)^{2/3} N^{5/3} = \frac{\hbar^2}{10\pi^2 m} (3\pi^2 N)^{5/3} V^{-2/3}.$$

For the pressure,

$$p = - \left(\frac{\partial U}{\partial V} \right)_N = \frac{2}{3} \frac{3}{5} \frac{\hbar^2}{10\pi^2 m} (3\pi^2 N)^{5/3} V^{-5/3} = \frac{2}{3} \frac{U}{V}.$$

The **Fermi temperature** is the temperature at which the thermal energy equals the Fermi energy, $kT_F = E_F$:

$$T_F = \frac{\hbar^2}{2mk} (3\pi^2)^{2/3} \left(\frac{N}{V} \right)^{2/3}.$$

If the actual temperature is less than the Fermi temperature, then a system can be regarded as “cold” and most of the electrons will be in the lowest accessible state. When T rises above zero, we expect the system to behave as a perturbed version of the $T = 0$ system provided that the temperature is sufficiently low; $T \ll T_F$.

Example 5.3.1

Estimate the Fermi temperature of copper.

For copper, we have a density of 8.94 g/cm^3 and an atomic weight of 63.546 g/mol , so the moles of copper atoms per unit volume is

$$8.94 \frac{\text{g}}{\text{cm}^3} \cdot \left(\frac{100 \text{ cm}}{\text{m}} \right)^3 \cdot \frac{1}{63.546 \frac{\text{g}}{\text{mol}}} = 140\,699 \frac{\text{mol}}{\text{m}^3}.$$

Using Avogadro's number, this implies that the number of copper atoms per unit volume is

$$140\,699 \frac{\text{mol}}{\text{m}^3} \cdot 6.022 \times 10^{23} \frac{\text{atoms}}{\text{mol}} = 8.4729 \times 10^{28} \frac{\text{atoms}}{\text{m}^3}.$$

If we assume one conduction electron per atom, then the electrons per cubic meter is

$$\frac{N}{V} = 8.4729 \times 10^{28} \frac{\text{electrons}}{\text{m}^3}.$$

For electrons, $m = 5.11 \times 10^{-5} \text{ eV}/c^2$. We can write Planck's constant as $\hbar = 6.582 \times 10^{-16} \text{ eVs}$ and Boltzmann's constant as $k = 8.617 \times 10^{-5} \text{ eV/K}$, then

$$\frac{\hbar^2}{mk} = \frac{(6.582 \times 10^{-16} \text{ eVs})^2}{5.11 \times 10^{-5} \frac{\text{eV}}{c^2} \cdot 8.617 \times 10^{-5} \frac{\text{eV}}{\text{K}}}.$$

Plugging in $c = 2.997 \times 10^8 \text{ m/s}$ and simplifying gives us

$$\frac{\hbar^2}{mk} = 8.837 \text{ m}^2\text{K}.$$

Plugging these results into the formula for the Fermi temperature gives us

$$T_F = 8.16 \times 10^4 \text{ K}.$$

Remember, that the Fermi energy is the energy of the most energetic particles in a gas at $T = 0$. At this temperature, since the Fermi energy is nonzero, these particles are still moving with a lot of kinetic energy. It is only when the temperature of the gas exceeds the Fermi temperature, i.e. when $T > T_F = E_F/k$, that the particles begin moving significantly faster than at $T = 0$. This might be confusing if you think of temperature as a quantification of the motion of the particles, and that at absolute zero there is no motion at all—an idea still taught in chemistry. However, in statistical mechanics, $T = 0$ does not mean the particles are motionless. Rather, it means that the thermal motion of the particles is at a minimum. Classically, the particles would be motionless, but we are using a quantum-mechanical approach, and the uncertainty principle tells us that they cannot be motionless even at $T = 0$.

5.4 Bose-Einstein Condensation

A **Bose gas** is a perfect gas of bosons. Recall that for a perfect gas, the occupation number is

$$\langle n_i \rangle = \frac{1}{e^{(\beta E_i - \alpha)} \pm 1} = \frac{1}{e^{\beta(E_i - \mu)} \pm 1}.$$

For bosons, we take the negative sign, so

$$\langle n_i \rangle_b = \frac{1}{e^{(\beta E_i - \alpha)} - 1} = \frac{1}{e^{\beta(E_i - \mu)} - 1}.$$

If μ is fixed while $T \rightarrow 0$ (i.e. $\beta \rightarrow \infty$), then $\langle n_i \rangle \rightarrow 0$. That is, all the higher energy levels depopulate. If the particle number N is held fixed and $T \rightarrow 0$, then μ is strictly less than all energy levels including the ground state, for which $\langle n_0 \rangle \rightarrow N$. That is, all particles drop into the ground state, and all other levels depopulate. This is the phenomenon of **Bose-Einstein condensation**. As a Bose gas is cooled down, eventually the ground

state acquires a macroscopic occupation number. That is, a macroscopic fraction of the particles will drop into the ground state.

If we set our energy scale such that the ground state energy is $E_0 = 0$, then the occupation number becomes

$$N_0 = \frac{1}{e^{-\alpha} - 1}.$$

Solving this for α gives us

$$\alpha = -\ln\left(1 + \frac{1}{N_0}\right) < 0.$$

As $T \rightarrow 0$, the occupation of the ground state becomes large $N_0 \gg 1$. So we can expand the logarithm using

$$\ln(1+x) \approx x - \frac{1}{2}x^2,$$

to get

$$\alpha \approx -\frac{1}{N_0}.$$

As $T \rightarrow 0$, $\alpha \rightarrow -1/N \approx 0$. This implies that $\mu = \alpha/\beta$ approaches zero from below. So as $T \rightarrow 0$, the ground state occupation N_0 approaches the total number of particles N , and the chemical potential μ approaches the ground state energy from below.

The formula for the total particle number

$$N = \frac{\partial}{\partial \alpha} \ln Z(\beta, \alpha, V) = \int_{-\infty}^{\infty} D(E) f(E) dE,$$

was derived in the continuum limit, and assumed a level density $D(E) \propto \sqrt{E} = 0$ for $E \leq 0$. That is, this formula leaves out the occupation of the ground state. This is okay at high temperatures, but not at low temperatures when we want to know the ground state occupation. So really, this formula gives the number of particles outside the ground state

$$N - N_0 = \frac{\partial}{\partial \alpha} \ln \bar{Z}(\beta, \alpha, V),$$

where \bar{Z} indicates that this is the Z from the continuum derivation. For bosons,

$$\ln \bar{Z}(\beta, \alpha, V) = \beta \int_{-\infty}^{\infty} \frac{\mathcal{N}(E)}{e^{\beta E - \alpha} - 1} dE,$$

with the number of levels given by

$$\mathcal{N}(E) = \frac{2gV}{3(2\pi)^2} \left(\frac{\sqrt{2m}}{\hbar}\right)^3 E^{\frac{3}{2}}.$$

Plugging \mathcal{N} into $\ln \bar{Z}$, substituting $x = \beta E$, and noting that the integrand is only defined for $x \geq 0$, we get that

$$\ln \bar{Z}(\beta, \alpha, V) = \frac{2g}{3(2\pi)^2} \left(\frac{\sqrt{2m}}{\hbar}\right)^3 \beta^{-\frac{3}{2}} V F(\alpha),$$

where

$$F(\alpha) = \int_0^{\infty} \frac{x^{\frac{3}{2}}}{e^{x-\alpha} - 1} dx.$$

At low temperature, α is small and we can Taylor expand $F(\alpha)$ around 0 as

$$F(\alpha) \approx F(0) + F'(0)\alpha = F(0) \left[1 + \frac{F'(0)}{F(0)}\alpha\right] = F(0) [1 + (\ln F)'(0)\alpha].$$

For $F(0)$, we integrate by parts (integrate $x^{3/2}$ and differentiate $1/(e^x - 1)$) to get

$$\begin{aligned} F(0) &= \int_0^\infty \frac{x^{3/2}}{e^x - 1} dx = \frac{2x^{5/2}}{5(e^x - 1)} \Big|_0^\infty + \frac{2}{5} \int_0^\infty \frac{x^{5/2} e^x}{(e^x - 1)^2} dx \\ &= \frac{2}{5} \int_0^\infty \frac{x^{5/2}}{(e^{x/2} - e^{-x/2})^2} dx = \frac{1}{4} \frac{2}{5} \int_0^\infty \frac{x^{7/2-1}}{\sinh^2(x/2)} dx. \end{aligned}$$

Now we use the identity

$$\int_0^\infty \frac{x^{t-1}}{\sinh^2(x/2)} dx = 4\Gamma(t)\zeta(t-1).$$

with $t = 7/2$ to get

$$F(0) = \frac{1}{4} \frac{2}{5} 4\Gamma\left(\frac{7}{2}\right) \zeta\left(\frac{5}{2}\right).$$

This simplifies to

$$F(0) = \frac{3\sqrt{\pi}}{4} \zeta\left(\frac{5}{2}\right).$$

Differentiating $F(\alpha)$ with respect to α , and then plugging in zero gives us

$$\begin{aligned} F'(0) &= \int_0^\infty \frac{x^{3/2} e^x}{(e^x - 1)^2} dx = \int_0^\infty \frac{x^{3/2}}{(e^{x/2} - e^{-x/2})^2} dx \\ &= \frac{1}{4} \int_0^\infty \frac{x^{3/2}}{\sinh^2(x/2)} dx = \frac{3\sqrt{\pi}}{4} \zeta\left(\frac{3}{2}\right). \end{aligned}$$

So our approximation becomes

$$F(\alpha) \approx \frac{3\sqrt{\pi}}{4} \zeta\left(\frac{5}{2}\right) \left[1 + \frac{\zeta\left(\frac{3}{2}\right)}{\zeta\left(\frac{5}{2}\right)} \alpha \right].$$

Putting it all together,

$$\ln \bar{Z}(\beta, \alpha, V) = g \left(\frac{m}{2\pi\hbar^2} \right)^{3/2} \zeta\left(\frac{5}{2}\right) \left[1 + \frac{\zeta\left(\frac{3}{2}\right)}{\zeta\left(\frac{5}{2}\right)} \alpha \right] \beta^{-3/2} V.$$

Plugging this into the formula for particle number gives us

$$N - N_0 = \frac{\partial}{\partial \alpha} \ln \bar{Z}(\beta, \alpha, V) = g \left(\frac{m}{2\pi\hbar^2} \right)^{3/2} \zeta\left(\frac{3}{2}\right) \beta^{-3/2} V.$$

So the number of bosons outside of the ground state, $N - N_0$, scales as $T^{3/2}$. This becomes inaccurate at higher temperatures, and breaks down completely when $N - N_0 = N$, because at that point, the ground state has been completely emptied. The temperature at which this occurs is called the **Bose-Einstein temperature**. Setting $N - N_0 = N$ and solving for T , gives us

$$T_{BE} = \frac{2\pi\hbar^2}{km} \left(\frac{N}{\zeta\left(\frac{3}{2}\right) gV} \right)^{2/3}.$$

When going from higher to lower temperature, T_{BE} is the temperature at which macroscopic occupation of the ground state sets in. Recall that g is a possible degeneracy factor. For example, we could have $g = (2s + 1)$ for massive spin- s particles.

For example, for ${}^4\text{He}$ with spin-0 (i.e. $g = 1$), and atomic mass 6.664×10^{-27} kg, mass density $M/V \approx 0.14$ g/cm³, and particle density $N/V = M/(mV) \approx 2.1 \times 10^{28}$ m⁻³, this formula gives us $T_{BE} \approx 3.1$ K. This is within the correct order of magnitude of the actual value $T_{BE} = 2.17$ K despite ${}^4\text{He}$ not being a perfect gas.

One interesting property of Bose-Einstein condensates is that bosons in the ground state all have zero energy and hence zero momentum, and do not contribute to the pressure. If N_0 is large, then $\alpha \sim 1/N_0 \approx 0$, and we can neglect the second term in $\ln \bar{Z}$

and write

$$\ln \bar{Z}(\beta, \alpha, V) \simeq g \left(\frac{m}{2\pi\hbar^2} \right)^{\frac{3}{2}} \zeta \left(\frac{5}{2} \right) \beta^{-\frac{3}{2}} V.$$

Then the pressure is

$$p = -\frac{\partial U}{\partial V} = \frac{1}{\beta} \frac{\partial(S/k)}{\partial V} = \frac{1}{\beta} \frac{\partial}{\partial V} \ln Z \simeq g \left(\frac{m}{2\pi\hbar^2} \right)^{\frac{3}{2}} \zeta \left(\frac{5}{2} \right) \beta^{-\frac{5}{2}}.$$

So $p \propto T^{\frac{5}{2}}$. The pressure of the Bose-Einstein condensate, which depends only on those bosons still outside the ground state, is a function only of the temperature. It does not depend on α or V . So

$$\left(\frac{\partial p}{\partial V} \right)_T = 0.$$

Since this derivative goes to zero, the derivative of the volume with respect to the pressure must diverge to infinity. This tells us that the condensate reacts extremely strongly to small changes in pressure. A very small change in pressure will cause a very large change in volume. This instability implies that Bose condensation involves a phase transition.

5.5 Photon Gas

Photons have spin-1, so they are bosons. A photon gas, must therefore be modelled as a Bose gas.

Earlier, we derived the level density for non-relativistic gas particles. Since photons are relativistic, this formula cannot be used. Instead, we start with the fact that the energy of a photon is

$$E = pc.$$

For photons in an infinite cube well, this implies that

$$E_{\vec{n}} = \frac{\hbar\pi c}{L} |\vec{n}|,$$

where $n_x, n_y, n_z = 1, \dots, \infty$. So

$$|\vec{n}| = \frac{L}{\hbar\pi c} E_{\vec{n}}.$$

Again, to obtain the continuum limit, the discrete sum over the energy levels can be converted into an integral over \vec{n} . This is accomplished again by integrating over one octant in spherical coordinates. We find that (after including the degeneracy $g = 2$ for the two polarizations of photons) the level degeneracy for a photon gas is

$$\mathcal{D}(E) = \frac{V}{\pi^2 \hbar^3 c^3} E^2.$$

Integrating this with respect to E gives us the number of photon states up to a given energy

$$\mathcal{N}(E) = \frac{V}{3\pi^2 \hbar^3 c^3} E^3.$$

In general, for bosons,

$$\ln Z(\beta, \alpha, V) = \beta \int_{-\infty}^{\infty} \frac{\mathcal{N}(E)}{e^{\beta E - \alpha} - 1} dE.$$

For photons, $\alpha = 0$, so

$$\ln Z(\beta, V) = \beta \int_{-\infty}^{\infty} \frac{\mathcal{N}(E)}{e^{\beta E} - 1} dE = \frac{\beta V}{3\pi^2 \hbar^3 c^3} \int_{-\infty}^{\infty} \frac{E^3}{e^{\beta E} - 1} dE$$

We can make the substitution $x = \beta E$, and since we are only integrating over positive energies, our integral becomes

$$\ln Z(\beta, V) = \frac{V\beta^{-3}}{3\pi^2 \hbar^3 c^3} \int_0^{\infty} \frac{x^3}{e^x - 1} dx.$$

From the Riemann zeta function identity

$$\int_0^{\infty} \frac{x^{t-1}}{e^x - 1} dx = \zeta(t)\Gamma(t),$$

this implies that

$$\ln Z(\beta, V) = \frac{V\beta^{-3}}{3\pi^2\hbar^3c^3}\zeta(4)\Gamma(4).$$

Given that $\zeta(4) = \pi^4/90$ and $\Gamma(4) = 6$, this gives us

$$\ln Z(\beta, V) = \frac{\pi^2}{45\hbar^3c^3}V\beta^{-3}.$$

5.6 Internal Degrees of Freedom

In our previous treatments of perfect gases, we assumed that the gas particles had no internal structure. This is adequate for monoatomic gases, but what about diatomic gases? A simple diatomic gas particle, has the same translational degree of freedom as a monoatomic gas particle, but in addition to that, it may have rotational and vibrational degrees of freedom. These internal dynamics may affect the thermodynamic properties of the gas.

These “internal” degrees of freedom for a single gas particle can be characterized by an internal Hamiltonian H_{int} , which implies an internal partition function

$$\zeta(\beta) = \text{tr} \left(e^{-\beta\hat{H}_{int}} \right).$$

This is for a single gas particle.

We will assume that there are a fixed number N of gas molecules, so we use the canonical ensemble. The translational degrees of freedom are described in an N -particle Hilbert space. The internal degrees of freedom are described by an *internal* Hilbert space for each of the N particles.

If we assume that the gas is dilute enough so that internal degrees of freedom are of different molecules are independent and that the internal degrees of freedom are independent of the translational degrees of freedom, then the partition function factorizes are

$$Z(\beta, N, V) = Z_t(\beta, N, V)\zeta(\beta)^N,$$

where $Z(\beta, N, V)$ is the full canonical partition function for the entire gas, and $Z_t(\beta, N, V)$ is the canonical partition function for a monoatomic gas. That is, Z_t is the partition function for the translational part of a multi-atomic gas.

If the two atoms in a diatomic molecule are the same, then we have to take into account the indistinguishability of quantum particles. Here, we will assume the two atoms in a molecule are different elements.

A diatomic molecule has two internal degrees of freedom—rotation about its center of mass and oscillations or vibrations along the axis between the two atoms. Since the two motions are independent, we can factorize the internal partition function into a rotational part and an oscillation part

$$\zeta(\beta) = \zeta_{rot}(\beta)\zeta_{osc}(\beta),$$

where the rotational part is given by the partition function of the quantum rotor,

$$\zeta_{rot}(\beta) = \sum_{l=0}^{\infty} (2l+1)e^{-\beta\hbar^2 l(l+1)/2\Theta},$$

and the oscillation part is given by the partition function of the harmonic oscillator

$$\zeta_{osc}(\beta) = \frac{1}{2 \sinh\left(\frac{\beta\hbar\omega}{2}\right)}.$$

The factorization of the total partition function for the gas, implies that the internal energy of the gas is

$$U = U_t + NU_{rot} + NU_{osc},$$

where U_t is the translational internal energy, or the internal energy of a monoatomic gas. If the gas is sufficiently dilute, and the temperature is high enough to justify the classical limit, then

$$U_t = \frac{3}{2}kT.$$

From the low and high temperature limits for the quantum rotor and the harmonic oscillator, we know that

$$NU_{rot}(T) \simeq \begin{cases} 0 & \text{for } kT \ll \hbar^2/\Theta \\ NkT & \text{for } kT \gg \hbar^2/\Theta \end{cases},$$

and

$$NU_{osc}(T) \simeq \begin{cases} \frac{N\hbar\omega}{2} & \text{for } kT \ll \hbar\omega \\ NkT & \text{for } kT \gg \hbar\omega \end{cases}.$$

This implies that the energy approaches some fixed lower bound at low temperatures and grows linearly with temperature at high temperatures.

The **specific heat** of a material (or gas) is the amount of heat per unit mass required to raise the temperature by one degree Kelvin. At constant volume, the specific heat can be calculated as

$$C_V = \frac{T}{N} \left(\frac{\partial S}{\partial T} \right)_{V,N} = \frac{1}{N} \left(\frac{\partial U}{\partial S} \frac{\partial S}{\partial T} \right)_{V,N} = \frac{1}{N} \left(\frac{\partial U}{\partial T} \right)_{V,N}.$$

It is the sum of contributions from the translational and internal degrees of freedom

$$C_V = C_{V,t} + C_{V,rot} + C_{V,osc}.$$

This means the specific heat of a diatomic gas is always higher than that of a monoatomic gas. Recall that for a monoatomic gas,

$$C_{V,t} = \frac{3}{2}k.$$

In the low temperature limit, both $C_{V,rot}$ and $C_{V,osc}$ approach zero, and in the high temperature limit, they both approach Boltzmann's constant k . So the overall specific heat, C_V , looks like a double step function.

Now let us consider a dilute solution composed of a large number \tilde{N} of solvent particles (such as water molecules), and a relatively small number N of solute particles (such as dissolved CO_2 particles). The macroscopic variables describing a dilute solution are the total energy U , the volume V , and N and \tilde{N} .

We will assume the solution is sufficiently dilute so that the effect of the solute particles on the solvent particles can be neglected. We will also assume that the solute particles are sufficiently separated so that no interaction occurs between them. That is, the internal dynamics of different solute particles are independent. Then, the interaction of a solute particle with the solvent particles surrounding it can be treated as a form of internal degrees of freedom. In short, the solute particles can be treated as a perfect gas with additional degrees of freedom coming from their interactions with the solvent particles.

The effective potential created by the cloud of solvent particles surrounding a solute particle may depend on the temperature and the density of the solvent. If the solution is homogeneous, then we can apply the Gibbs-Duhem relation. Then the four intensive variables which are conjugate to U , V , N , and \tilde{N} are not independent of each other. Any one of them can be expressed in terms of the other three. For example, the density of the solvent (an intensive variable) can be expressed as a function of β (conjugate variable of U), $\tilde{\alpha}$ (conjugate variable of \tilde{N}), and the density of the solute. Assuming the solution is

highly dilute, then the solute density can be neglected, and so the density of the solvent can be expressed in terms of β (or temperature) and $\tilde{\alpha}$. This tells us that we can write H_{int} as a function of β and $\tilde{\alpha}$. So we can do the same for the internal partition function

$$\zeta(\beta, \tilde{\alpha}) = \text{tr} \left(e^{-\beta \hat{H}_{int}(\beta, \tilde{\alpha})} \right),$$

and the overall partition function

$$Z(\beta, \tilde{\alpha}, N, V) = Z_t(\beta, N, V) \zeta(\beta, \tilde{\alpha})^N.$$

To study the overall solution, it is more convenient to use the grand canonical ensemble, and we can obtain the grand canonical partition function from the canonical partition function by doing a discrete Laplace transform

$$\begin{aligned} Z_G(\beta, \alpha, V) &= \sum_{N'=0}^{\infty} e^{\alpha N'} \text{tr}_{N'} \left[e^{-\beta \hat{H}} \right] = \sum_{N'=0}^{\infty} e^{\alpha N'} Z_C(\beta, N', V) \\ &= \sum_{N'=0}^{\infty} e^{\alpha N'} Z_C(\beta, \tilde{\alpha}, N', V) = \sum_{N'=0}^{\infty} e^{\alpha N'} Z_t(\beta, N', V) \zeta(\beta, \tilde{\alpha})^{N'} \\ &= \sum_{N'=0}^{\infty} e^{\{\alpha + \ln \zeta(\beta, \tilde{\alpha})\} N'} Z_t(\beta, N', V) = Z_{G,t}(\beta, [\alpha + \ln \zeta(\beta, \tilde{\alpha})], V). \end{aligned}$$

So the grand canonical partition function of the solute is related to that of a monoatomic gas, simply by a shift of the variable α to $\alpha + \ln \zeta(\beta, \tilde{\alpha})$. Other than that shift, the solute particles behave like a monoatomic gas.

Earlier we showed that the log of the grand canonical partition function of a monoatomic gas is proportional to e^α in the classical limit. For the dilute solute particles, this becomes

$$e^{\alpha + \ln \zeta(\beta, \tilde{\alpha})} = \zeta(\beta, \tilde{\alpha}) e^\alpha.$$

This implies that the log of the grand canonical partition function of the solute is related to the same thing for the monoatomic gas by a factor of ζ

$$\ln Z_G(\beta, \tilde{\alpha}, \alpha, V) = \zeta(\beta, \tilde{\alpha}) \ln Z_t(\beta, \alpha, V).$$

So the overall log of the partition function for the solution is

$$\ln Z(\beta, \tilde{\alpha}, \alpha, V) = \ln \tilde{Z}(\beta, \tilde{\alpha}, V) + \zeta(\beta, \tilde{\alpha}) \ln Z_t(\beta, \alpha, V),$$

where \tilde{Z} is the partition function of the solvent independent of α (solute).

We can get the number of dissolved particles by differentiating

$$N = \frac{\partial}{\partial \alpha} \ln Z(\beta, \tilde{\alpha}, \alpha, V) = \zeta(\beta, \tilde{\alpha}) \frac{\partial}{\partial \alpha} \ln Z_t(\beta, \alpha, V).$$

In the classical limit of the grand canonical partition function, $\ln Z(\beta, \alpha, V) \propto e^\alpha \beta^{-3/2} V$. This implies that

$$\frac{\partial}{\partial \alpha} \ln Z_t(\beta, \alpha, V) = \ln Z_t(\beta, \alpha, V),$$

and so

$$\boxed{N = \zeta(\beta, \tilde{\alpha}) \ln Z_t(\beta, \alpha, V).}$$

We can get the pressure via

$$\begin{aligned} p_{tot} &= -\frac{\partial U}{\partial V} = \frac{1}{\beta} \frac{\partial(S/k)}{\partial V} = \frac{1}{\beta} \frac{\partial}{\partial V} \ln Z(\beta, \tilde{\alpha}, \alpha, V) \\ &= \frac{1}{\beta} \frac{\partial}{\partial V} \ln \tilde{Z}(\beta, \tilde{\alpha}, V) + \frac{1}{\beta} \frac{\partial}{\partial V} \zeta(\beta, \tilde{\alpha}) \ln Z_t(\beta, \alpha, V). \end{aligned}$$

This demonstrates that the total pressure of the solution has two contributions—one from the solvent and one from the solute

$$\boxed{p_{tot} = \tilde{p} + p,}$$

where the pressure of the pure solvent is given by

$$\tilde{p} = \frac{1}{\beta} \frac{\partial}{\partial V} \ln \tilde{Z}(\beta, \tilde{\alpha}, V),$$

and the pressure of the solute is given by

$$p = \zeta(\beta, \tilde{\alpha}) \frac{1}{\beta} \frac{\partial}{\partial V} \ln Z_t(\beta, \alpha, V).$$

Since the solvent and solute are both homogeneous, the logarithm of the grand canonical partition function must be extensive. This implies that it must scale linearly with the volume. But this implies that the derivative with respect to V is equivalent to dividing by V , and so

$$p = \zeta(\beta, \tilde{\alpha}) \frac{1}{\beta V} \ln Z_t(\beta, \alpha, V) = \frac{1}{\beta V} N.$$

This also implies that the solute satisfies the ideal gas law

$$p = \frac{NkT}{V}.$$

In summary, for a dilute solution, in addition to the pressure exerted by the pure solvent, there is also an extra pressure exerted by the solute, which acts as if it were a classical perfect gas. This extra pressure is called **osmotic pressure**.

5.7 Summary: Perfect Gas

Skills to Master

-

The grand canonical density operator is

$$\hat{\rho} = \frac{1}{Z} \exp \left(- \sum_j (\beta E_j - \alpha) \hat{n}_j \right).$$

The grand canonical entropy is

$$\frac{S}{k} = \ln Z - \alpha N + \beta U.$$

Chapter 6

Thermodynamic Processes

6.1 Summary: Thermodynamic Processes

Skills to Master

-

Chapter 7

Phase Transitions

7.1 Summary: Phase Transitions

Skills to Master

-

Index

- Abelian group, 12
- Basis, 13
- Bayes' theorem, 5
- Bernoulli distribution, 6
- Binomial distribution, 7
- Binomial theorem, 8
- Bloch vector, 20
- Block matrix, 22
- Bohr magneton, 48
- Boltzmann entropy, 30
- Boltzmann's constant, 30
- Bose gas, 75
- Bose-Einstein condensate, 76
- Bose-Einstein condensation, 75
- Bose-Einstein temperature, 77
- Bra, 13
- Canonical, 37
- Canonical ensemble, 37
- Central limit theorem, 11
- Classical limit, 71
- Combinations, 2
- Component, 13
- Conjugate variables, 33
- Continuity equation, 35
- Continuous basis, 16
- Continuum limit, 69, 70
- Control parameter, 31
- Degeneracy pressure, 73
- Density matrix, 19
- Density of states, 70
- Density operator, 19
- Dimension, 13
- Direct product, 21
- Discrete basis, 16
- Ensembles, 61
- Entropy, 27, 30
- Equilibrium, 36
- Euler relation, 59
- Expectation value, 31
- Extensive variable, 58
- Fermi energy, 73
- Fermi gas, 72
- Fermi surface, 74
- Fermi temperature, 74
- First law, 64
- Fock space, 67
- Fugacity, 72
- Gaussian distribution, 10
- Gibbs state, 32
- Gibbs-Duhem relation, 59
- Grand canonical, 37
- Group, 12
- Harmonic oscillator, 43, 54
- Heat bath, 63
- Hermitian conjugate, 14
- Hermitian operator, 16
- Hilbert space, 13
- Homogeneity, 58
- Idempotent, 15
- Information, 27
- Inner product, 13
- Intensive variable, 59
- Irreversible process, 36
- Isentropic process, 36
- Kronecker product, 22
- Lagrange parameter, 32
- Law of large numbers, 11
- Law of total probability, 4
- Laws of Thermodynamics, 62
- Legendre transformation, 44
- Level density, 70
- Linear independence, 13
- Liouville's theorem, 36
- Liouville-von Neumann equation, 21
- Lüder's rule, 20
- Macrostate, 31
- Marginalization rule, 4
- Maximum entropy principle, 31
- Mean, 6
- Median, 6
- Metastable equilibrium, 37
- Microcanonical, 37
- Microcanonical ensemble, 37
- Microstate, 31
- Mixed state, 19
- Mixed states, 18
- Mode, 6
- Molecular gas, 67
- Monoatomic gas, 67
- Multiplication principle, 1
- Natural variables, 33
- Nernst's theorem, 64
- Normal distribution, 10
- Occupation number operator, 67
- Operators, 15
- Orthonormal basis, 13
- Osmotic pressure, 82
- Outer product, 15
- Paramagnetism, 50
- Pauli exclusion principle, 73
- Pauli matrix, 17
- Perfect gas, 67
- Permutation, 1
- Photon gas, 78
- Poisson bracket, 36
- Poisson distribution, 9
- Product rule, 5
- Projection operator, 14
- Projector, 14
- Pure state, 18
- Quantum harmonic oscillator, 43
- Quantum rotor, 46
- Relative frequency, 8
- Reproducible process, 33
- Reversible process, 36
- Rule of product, 1
- Schrodinger equation, 17

- Second law, 64
- Second law of thermodynamics, 33
- Shannon entropy, 27, 29
- Shannon's noiseless channel coding theorem, 29
- Sharp constraint, 31
- Specific heat, 80
- Spin system, 48
- Stability, 60
- State property, 64
- State space, 13
- System property, 64
- Temperature, 36, 37
- Tensor product, 21
- Thermodynamic variables, 32, 33
- Thermometer, 63
- Thermostat, 63
- Third law, 64
- Transpose, 14
- Two-level system, 17
- Typical sequence, 28
- Uniform distribution, 10
- Variance, 6, 38
- Vector, 12
- Vector spaces, 12
- Von Neumann entropy, 30
- Zeroth law, 62