Chapter 1 - Crystals

- Materials: crystalline, polycrystalline, amorphous
- Bonds in Solids
- Crystal Structure
- Basis
- Lattice: Bravais Lattice vs. Non Bravais Lattice
- Unit Cell, Primitive Unit Cell, Wigner-Seitz Unit Cell
- Crystal symmetry
- Miller Indices

Why we call it solid-state?

States of the matter:
1. Solid: density $\sim 10^{23}/\text{cm}^3$
   a. Crystal: long range order (lattice + basis)  (Example: Epitaxial silicon and diamond)
   b. Polycrystal: short range order (µm-10µm) (Example: Most metals (Al, Cu) Poly-Si)
   c. Amorphous: no order (Example: Glasses like SiO₂)

![Grain Boundary](image1.png)
(a) Crystalline  (b) Polycrystalline  (c) Amorphous

2. Liquids: no order, takes the shape of the container, still weak bounds; density $\sim 10^{22}/\text{cm}^3$
3. Gases: no order, no bounds between molecules
4. Liquid crystals: atoms mobile, type of long range order  Applications: LCDs
5. Plasma: ionized gas/liquid  (Examples: Sun, Aurora, Lightning, (RIE, Sputtering, PECVD))

Semiconductor devices used in, for example:
- Electronics, e.g., Silicon integrated circuit electronics
- High-performance electronics for, e.g., communications
  e.g., GaAs (gallium arsenide) for cellular communications, satellite receivers

Power electronics
Optoelectronics, e.g.,
- CCD (charge coupled device) cameras
- Photo detectors for, e.g.,
  - Telecommunications, night vision
- Light emitting diodes for, e.g.,
  - Indicators and displays, remote controls
- Semiconductor lasers for, e.g.,
  - Telecommunications, compact disc players
- New kinds of devices, e.g.,
  - Modulators for high-speed telecommunications
  - Wavelength shifters for wavelength division multiplexed networks
  - Two-dimensional arrays of very small devices for optical interconnects ...

<table>
<thead>
<tr>
<th>Table 1: Abbreviated periodic chart of the elements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>H</td>
</tr>
<tr>
<td>Na</td>
</tr>
<tr>
<td>K</td>
</tr>
<tr>
<td>Cu</td>
</tr>
<tr>
<td>Rb</td>
</tr>
<tr>
<td>Cs</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Broad range of usable semiconductor materials:
1. Elemental semiconductors:
   - silicon (Si)
   - germanium (Ge)
   - diamond (C)
2. Binary [two-component] compound semiconductors, e.g.,
   - III-V, e.g., gallium arsenide (GaAs), indium phosphide (InP)
   - IV-VI, e.g., zinc selenide (ZnSe), cadmium telluride (CdTe)
   - IV-V, e.g., silicon carbide (SiC)
   - IV-VI, e.g., lead sulphide (PbS)
3. Alloys,
**3. Metallic Bonding:**
Positive ions plus gas (sea) of electrons. Think of this as the limiting case of ionic bonding in which the negative ions are electrons. (But electrons can't be forced to sit at lattice points from Uncertainty Principle: \(\Delta p \Delta x \geq \hbar/2\) as for electrons m is small so the zero point energy \(\Delta E = \hbar^2/2m\) is very large; the electrons would shake themselves free and are thereby delocalized)

- Bonds are non-directional (high coordination number, high density, malleable and ductile)
- Variable strength
- Free electrons \(\rightarrow\) high electrical conductivity, shiny
  - Electric field associated with incident light makes free electrons at surface move back and forth, re-radiating the light, as a reflected beam)

**4. Van der Waals Bonding:**
Even a neutral atom with a full shell, can, at a given instant, have a dipole moment (i.e. one side of the atom more positive than the other)

This instantaneous dipole will induce a dipole in a neighboring atom, and the resulting dipole-dipole interaction is the origin of the van der Waals bond. Although the original dipole time-averages to zero, the interaction does not – it is always attractive, \(E \sim 1/r^6\)

- Bond is weak (\(\rightarrow\) low melting point, large expansion coefficient)
- Non directional so high coordination number BUT
- Long bond lengths (\(\rightarrow\) low density)
- Examples: Solid inert gases (Argon, Neon), molecular solids (solid Oxygen)

**5. Hydrogen Bonding:**
Hydrogen loses its electron and becomes positively charged particularly easily. Therefore the region of a molecule around a hydrogen atom is often quite positive, and this allows an electrostatic bond to form between it and negative parts of neighboring molecules.

Example: ice – the strength of the hydrogen bond explains the anomalously high melting point of ice.

Central Core Repulsion: Overlap of orbitals rapidly increases the energy of the electrons (Pauli Exclusion Principle forces energies up as soon as electron wavefunctions start to overlap). Present in all cases – need a repulsion to give an equilibrium separation. Characterized mathematically by a high power law, or an exponential – but this is empirical.

### Table 2: Strength of bonds

<table>
<thead>
<tr>
<th>Bond Type</th>
<th>Energy (GPa)</th>
<th>Example of Bond</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covalent</td>
<td>1,000</td>
<td>Diamond</td>
</tr>
<tr>
<td>Ionic</td>
<td>30 - 100</td>
<td>Salt and Ceramics</td>
</tr>
<tr>
<td>Metallic</td>
<td>30 - 150</td>
<td>Metals</td>
</tr>
<tr>
<td>Hydrogen</td>
<td>8</td>
<td>Ice</td>
</tr>
<tr>
<td>Van der Waals</td>
<td>2</td>
<td>Polythene</td>
</tr>
</tbody>
</table>
Crystal Structures:
Translation symmetry: The lattice can be obtained by repetition of a building block called basis.
Bravais lattice: is the set of points defined by \( \vec{R} = n_1 \vec{a}_1 + n_2 \vec{a}_2 + n_3 \vec{a}_3 \) as \( n_i \) is integer.
The shortest possible \( a_i \) gives us primitive vectors.
The volume cell enclosed by the primitive vectors is called the primitive unit cell \( V = \vec{a}_1 \times (\vec{a}_2 \times \vec{a}_3) \).

- **Lattice:**
  \[ \vec{R} = n_1 \vec{a}_1 + n_2 \vec{a}_2 \]

- **Basis:**
  \[ \vec{R} = n_1 \vec{a}_1 - n_2 \vec{a}_2 \]

Crystal structure = Lattice + Basis

- **Wigner-Seitz cell**

- **Primitive unit cells**

- **Non-primitive unit cells**

Figure 2: Lattice, basis, crystal structures, and unit cells

Symmetry operations:
For categorizing crystal systems, symmetry characteristics are to be studied. Many physical properties depend directly on the crystal symmetry. (For biological molecules only rotation and/or translation operations is acceptable. Since, biological protein molecules mainly consist of protein acids, hence, reflection or inversion symmetries are not allowed.)

**N-fold rotation symmetry**, \( C_n \): Rotation by an angle \( \frac{2\pi}{n} \) about an axis through the crystal. (There are restrictions for \( n (n = 1, 2, 3, 4, 6) \), why? Note that \( n = 1 \) corresponds to a rotationally symmetric object (rotation will be about a certain point for 2D or axis for 3D).)

![Different orders of rotational symmetry](image)

- **Inversion center symmetry**, \( I \): Transformation \( r \rightarrow -r \), fixed point is selected as origin (lack of inversion symmetry may lead to piezoelectricity)
- **Plane of symmetry (reflection)**, \( \sigma \): Reflection across a plane (one side of the plane is a mirror image of the lattice on the other side)
- **Rotation-Inversion symmetry (Improper Rotation)**, \( S_n \): Rotation \( C_n \), followed by reflection in the plane normal to the rotation axis.

**One-Dimensional Lattices:**

![1D lattices](image)
Two-Dimensional Lattices:
There are 5 basic classes of 2D Bravais lattices:

1. **Triclinic** (parallelepiped): \(\alpha, \beta, \gamma \neq 90^\circ\)
2. **Monoclinic** (right prism with parallelogram base; here seen from above): \(\alpha \neq 90^\circ, \beta, \gamma = 90^\circ\)
3. **Orthorhombic** (cuboid): \(a \neq b \neq c\)
4. **Tetragonal** (square cuboid): \(a = b < c\)
5. **Rhombohedral** (trigonal trapezohedron): \(a = b = c\), \(\alpha = 60^\circ\)

Figure 5: 2D Bravais lattices: 1 oblique, 3 rectangular, 1 centered rectangular, 4 hexagonal, and 5 square (Ref: Wikipedia)

Three-Dimensional Lattices:
14 Bravais lattices:

<table>
<thead>
<tr>
<th>The 7 lattice systems</th>
<th>The 14 Bravais Lattices</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Triclinic</strong> (parallelepiped)</td>
<td>![Triclinic Lattice]</td>
</tr>
<tr>
<td><strong>Monoclinic</strong> (right prism with parallelogram base; here seen from above)</td>
<td>![Monoclinic Lattice]</td>
</tr>
<tr>
<td><strong>Orthorhombic</strong> (cuboid)</td>
<td>![Orthorhombic Lattice]</td>
</tr>
<tr>
<td><strong>Tetragonal</strong> (square cuboid)</td>
<td>![Tetragonal Lattice]</td>
</tr>
<tr>
<td><strong>Rhombohedral</strong> (trigonal trapezohedron)</td>
<td>![Rhombohedral Lattice]</td>
</tr>
<tr>
<td><strong>Hexagonal</strong> (centered regular hexagon)</td>
<td>![Hexagonal Lattice]</td>
</tr>
<tr>
<td><strong>Cubic</strong> (isometric; cube)</td>
<td>![Cubic Lattice]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Simple</strong></th>
<th><strong>Body-centered</strong></th>
<th><strong>Face-centered</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>![Simple Lattice]</td>
<td>![Body-centered Lattice]</td>
<td>![Face-centered Lattice]</td>
</tr>
</tbody>
</table>
A lattice system is a class of lattices with the same point group. In 3D, there are 7 lattice systems. For convenience a Bravais lattice is depicted by a unit cell which is a factor 1, 2, 3 or 4 larger than the primitive cell. The Bravais lattices were studied by Frankenhaeim (1801-1869), in 1842, who found that there were 15 Bravais lattices. This was corrected to 14 by Bravais in 1848. There are just 14 different ways of arranging points in space lattices such that all the lattice points have exactly the same surrounding. (Why do we not have face-centered tetragonal lattice? (body-centered tetragonal))

Most known semiconductor material (elements, compounds, or alloys) are members of either cubic or hexagonal system.

Specific semiconductor lattices:

<table>
<thead>
<tr>
<th>Semiconductor</th>
<th>Crystal Structure</th>
<th>Lattice Constants [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>Diamond</td>
<td>5.43095</td>
</tr>
<tr>
<td>Ge</td>
<td>Diamond</td>
<td>5.64613</td>
</tr>
<tr>
<td>GaAs</td>
<td>Zincblende</td>
<td>5.6536</td>
</tr>
<tr>
<td>CdS</td>
<td>Zincblende</td>
<td>5.8320</td>
</tr>
<tr>
<td>PbS</td>
<td>Wurtzite</td>
<td>a=4.16, c=6.756</td>
</tr>
</tbody>
</table>

**Diamond:**
Diamond lattice consists of two interpenetrating FCC Bravais lattices, displaced along the body diagonal of the cubic cell by one quarter of the length of the diagonal. Diamond lattice appears in materials such as silicon, germanium, and diamond (C). In this form each atom has four neighboring atoms.

**Zincblende:**
The zincblende is similar to diamond structure except that the two interpenetrating FCC sublattices are of different atoms. Again for each atom there are four neighboring atoms of different kind. GaAs, InP, most III-V, and some II-VI are constructed in this form.

**Wurtzite:**
Named after the mineral wurtzite, it is a crystal structure for many of II-VI such as ZnO, CdS, CdSe, and GaN. As shown in the next Figure it consists of interlocking hexagonal lattices. This structure lacks inversion symmetry; therefore these crystals may show properties such as piezoelectricity.
Miller indices:
That is a method to label distinct planes and direction within a crystal structure. For any given plane, Miller indices can be obtained following these steps:
1. Note where the plane to be indexed intercepts the axes (chosen along unit cell directions). Record result as whole numbers of unit cells in the x, y, and z directions, e.g., 2, 1, 3.
2. Take the reciprocals of these numbers, e.g., 1/2, 1/3, 1/3.
3. Convert to whole numbers with lowest possible values by multiplying by an appropriate integer, e.g., x6 gives 3, 6, 2.
4. Enclose number in parentheses to indicate it is a crystal plane categorization, e.g., (3,6,2).
Planes parallel to a unit cell coordinate axis are viewed as intercepting the axis at infinity, so have an associated Miller Index in that direction of zero, e.g., (100) plane.
Planes intersecting along the negative axis use a bar over the index rather than a negative sign, e.g., $\bar{1}$ rather than -1, e.g., (111).
Groups of equivalent planes, { (100), (010), (001), (100), (010), and (001) all equivalent because rotation about the 3 fold axes of the cube diagonals maps the various faces into one another, making the planes equivalent) are noted in curly brackets, i.e., [100] for the above set of equivalent planes.

Figure 9: Wigner-Seitz unit cell for different Bravais lattices.

Figure 10: Miller indices (a) The (132) plane. (b) The (100) plane. (c) The (111) plane.

Similar procedure can be used to define Miller indices for directions.
Why do we need quantum mechanics?
Not all of the electrons contribute in conductance.

Carrier number = Number of states × filling factor

Quantum Mechanics Fundamentals:
Wave-particle duality is one of the basic ideas of the quantum mechanics, both fundamentally and chronologically. 4 main concepts that lead us to QM are:

1. Black-body radiation: The black-body radiation could not be explained until 1900 Max Planck assumed that the frequency of the emitted light is quantized. He assumed that the light is emitted from oscillators which their energy increases linearly with frequency; the famous $E = h \nu$ relation. This was not a radical proposal - even at that time - since classic oscillators operate similarly.

2. Photoelectric Effect: In 1905 Albert Einstein took Planck’s model and found a solution to another unsolved problem at that time, the photoelectric effect. Light here is absorbed as quanta while in black-body radiation it has been emitted in quanta.
3. Bohr atomic model: To explain Rydberg formula for the spectral lines of atomic hydrogen emission, Bohr assumed that made of positively charged nucleus surrounded by electrons. Electrons travel in circular orbits around the nucleus, similar to the solar system. Bohr model is quantum mechanically modification of the Rutherford model.

![Bohr model](image)

He assumed that the angular momentum is quantized.

\[ J_n = m_e q_n = n \hbar \]

\[ \nu = n \hbar / m_e q_n \]

\[ r_n = \frac{4 \pi \varepsilon_0 (n \hbar)^2}{m_e q_n} \]

\[ E_n = \frac{1}{2} m_e q_n^2 + \frac{1}{2} q^2 / 4 \pi \varepsilon_0 r_n \]

K.E. = \frac{1}{2} m_e q_n^2 = \frac{1}{2} \left( q^2 / 4 \pi \varepsilon_0 r_n \right) \]

P.E. = -q^2 / 4 \pi \varepsilon_0 r_n \quad (P.E. \text{ set } = 0 \text{ at } r = \infty)

\[ E_n = \text{K.E.} + \text{P.E.} = \frac{1}{2} \left( q^2 / 4 \pi \varepsilon_0 r_n \right) \]

\[ E_n = \frac{m_e q_n^4}{2(4\pi \varepsilon_0)^2} = \frac{13.6}{n^2} \text{ eV} \]

4. Wave-particle duality: As photons act as wave and particle, what about electrons? How about other objects?

\[ E = mc^2 = \sqrt{m_e^2 c^4 + p^2 c^2} \]

![Wave-particle duality](image)

\[ \text{photon}(m_0 = 0) \rightarrow h \nu = pc \]

This is called De Broglie's wavelength.

**Uncertainty principle:**

Published by Werner Heisenberg in 1927, the principle means that the position and velocity of a particle cannot both be measured, exactly at the same time (actually pairs of position, energy, and time). This is not about the physical limitations in the measurements. In quantum world, each particle is described by a wave packet. This wave behavior of the particle is reason behind uncertainty principle. The particle is most likely to be found in those places where the undulations of the wave are greatest, or most intense. The more intense the undulations of the associated wave become, however, the more ill defined becomes the wavelength, which in turn determines the momentum of the particle. So a strictly localized wave has an indeterminate wavelength; its associated particle, while having a definite position, has no certain velocity. A particle wave having a well-defined wavelength, on the other hand, is spread out; the associated particle, while having a rather precise velocity, may be almost anywhere. A quite accurate measurement of one observable involves a relatively large uncertainty in the measurement of the other.

Mathematically we describe the uncertainty principle as the following

\[ \Delta x \Delta p \geq \hbar \]

or

\[ \Delta p \Delta t \geq \frac{\hbar}{2} \]

A sine wave of wavelength \( \lambda \) implies that the momentum \( p \) is precisely known. But the wavelength and the probability of finding the particle \( \psi \) is spread over all space, it is unknown.

Adding several waves of different wavelength together will produce an interference pattern which begins to localize the wave.

\[ \Delta x > \frac{\hbar}{p} \]

but that process smears the momentum values and makes it more uncertain. This is an inherent and measurable increase in the uncertainty, so when 5x is decreases.

\[ \Delta x \Delta p > \frac{\hbar}{2} \]
Schrödinger Equation:

Electron can behave like plane wave with \( \lambda = h/p \) as \( \Psi \propto \exp (2\pi i x/\lambda) \). We need a wave equation to describe this. Heilmhoit wave equation is as \( \Psi'' = -k^2 \Psi \) where \( k = k_0 = 2\pi/\lambda = p/\hbar \). Rewriting wave equation in terms of \( p \) we will have: \( -\hbar^2 \Psi'' = p^2 \Psi \). By dividing both sides by \( 2m_0 \), where \( m_0 \) is the free electron mass at rest \( m_0 = 9.11 \times 10^{-31} \text{kg} \), we will get:

\[
-\frac{\hbar^2}{2m_0} \Psi'' = \frac{p^2}{2m_0} \Psi
\]

But \( p^2/2m_0 \) is the kinetic energy of electron and in general Total energy \( (E) = \text{Kinetic energy} (K, E) + \text{Potential energy} (V) \)

Hence we could rewrite the above equation as:

\[
-\frac{\hbar^2}{2m_0} \Psi'' = (E - V(r)) \Psi
\]

Or

\[
-\frac{\hbar^2}{2m_0} \Psi'' + V(r) \Psi = E \Psi
\]

This is the time-independent version of the Schrödinger equation. Note that we have not derived Schrödinger equation we merely suggested it as an example. This equation has to be postulated.

\[
E = \sqrt{m_0 c^2 + p^2 c^2} = m_0 c^2 \left[ 1 + p^2 c^2 / 2m_0 + \ldots \right]
\]

\[
E - m_0 c^2 = V + \left( p^2 / 2m_0 \right)
\]

\[
\hbar \omega = \hbar c \omega = V + \left( \hbar c^2 / 2m_0 \right)
\]

\[
h \omega = \left( \hbar c^2 / 2m_0 \right) + V
\]

Assume, \( \Psi(x,t) = A \exp(-i(\omega t - hx)) \)

\[
d \Psi / dt = -i \omega \Psi \text{ and } d^2 \Psi / dt^2 = -\hbar^2 \Psi
\]

\[
\text{ih} \frac{d \Psi}{dt} = \frac{\hbar^2}{2m_0} \frac{d^2 \Psi}{dt^2} = E \Psi
\]
The Schrödinger equation is just such an eigenvalue equation, though up to now we have only thought of it with \( \hat{A} \) as a differential operator. We could also think of it as a matrix, in which case we would find it as some kind of "striped" diagonal matrix, because we only need the values of the most adjacent points in order to calculate local derivatives.

In practice in quantum mechanics, nearly all calculations are done by representing the operators as matrices, though usually the vector elements are not the values of the functions at specific points in space, other "bases" are usually more efficient for representing the state.

Energy operator:

The operator \( \hat{A} \equiv -\frac{\hbar^2}{2m_0} \frac{\partial^2}{\partial r^2} + V(r) \) is known as the Hamiltonian, it is the energy operator. The hat sign is used to distinguish this entity as an operator from a number. Schrödinger equation now can be written as \( \hat{A} \psi = E \psi \). (\( \hat{A} \) is an operator but \( E \) is simply a number, the energy of electron) This equation can be read as:

When \( \hat{A} \) operates on \( \psi \), the result is the total energy \( E \).

Momentum Operator:

We can see for the specific case of a plane wave in the z direction \( \psi \propto e^{i(kx)} \) that

\[
\frac{\hbar}{i} \frac{\partial}{\partial z} \psi = i \hbar k \psi
\]

Generalizing to three dimensions (to handle plane waves propagating in any direction), we could postulate that the momentum operator is

\[
\hat{p} = \frac{\hbar}{i} \frac{\partial}{\partial r}
\]

This postulate works, even for states that are not plane waves is consistent with the energy operator since

\[
\frac{1}{2m_0} \hat{p} \cdot \hat{p} = \frac{\hbar^2}{2m_0} \psi^2
\]

The (vector) momentum is

\[
\hat{p} = \hbar \hat{k}
\]

Note that the relation between momentum \( \hat{p} \) and wavevector \( \hat{k} \) does not involve the mass of the electron, because \( \hat{p} \) and \( \hat{k} \) are related only by a fundamental constant, \( \hbar \) has a more direct physical meaning in terms of physical lengths in the material of interest. We will henceforth talk almost always about \( \hat{k} \) rather than \( \hat{p} \).

Eigenstates:

Starting from Schrödinger equation

\[
\left(-\frac{\hbar^2}{2m_0} \frac{\partial^2}{\partial r^2} + V(r)\right) \psi = E \psi
\]

Only particular set of \( \psi(r) \) will be solutions of this equation, and only particular energies \( E \) correspond to that. The equation is a very specific relationship between the second derivative (curvature \( \psi'' \) of the function) and the difference between the total and potential energies. In linear algebra this is and eigenfunction/eigenvalue problem. Solving this problem will give a set of eigenvalues \( E_n \) with each of which there is associated some eigenfunction \( \psi_n(r) \) that is a solution of the equation (there can be more than one distinct \( \psi_n(r) \) for a given \( E_n \), a situation called degeneracy).

The wavefunction \( \psi(r) \) describes the "state" of the electron. If the wavefunction is one of the eigenfunction \( \psi_n(r) \), the electron is said to be in an eigenstate, with corresponding total energy \( E_n \).

We know from Fourier analysis that any time-dependent function can be broken down into its frequency components, and that the function can equally well be defined by the amplitudes of these components, but sine (and/or cosine) are not the only possible bases.

The eigenfunctions of many (linear) operators form complete sets. For example, the eigenvectors of a matrix form a basis that can describe any vector in the "space" of vectors. It is a postulate of quantum mechanics that the eigenfunctions of the operator corresponding to any physically measurable quantity form a complete set. The eigenfunctions of such an operator are all "orthogonal" to one another.

The eigenvalues of such an operator are all real.

Because of the completeness, any quantum mechanical state (e.g., of an electron) can be described as a linear combination of eigenstates

\[
\psi(r) = \sum \alpha_n \psi_n(r)
\]

where \( \alpha_n \) are numbers that are the "expansion coefficients."

In general, the energy and momentum operators for the electron can have the same eigenstate if the potential energy is uniform. This means that if the system is in an eigenstate of the energy operator, and hence has a definite energy, it can also be in an eigenstate of the momentum operator, and hence can also have a definite momentum. An electron can simultaneously have a definite energy and momentum. (This is not a general property of operators and their associated measurable quantities. position and momentum eigenstates are different. If an electron is at a definite momentum, it does not have a definite position due to Heisenberg uncertainty principle.)

This property of momentum and energy is very useful for us.

Note that "momentum" (\( \hat{k} \)) eigenstates are a natural basis set in crystals for the "envelope" wavefunction. One can calculate band structure by calculating for the energy for each allowed state of momentum. The \( k \) eigenstates and the energy eigenstates are the same.

Time-dependent Schrödinger Equation:

We saw that all particles behave as probability wave. Because of duality of wave and particle all waves also behave as particles! For example electromagnetic waves behave as a particle called photons.

In Photoelectric effect we saw that the maximum energy of photo-electrons are independent of the intensity of light and depends linearly of the frequency. Therefore

\[
E_{\text{photon}} = h \omega
\]

Now, using the wave-particle duality, we try to postulate about the time-dependent Schrodinger Equation. Using energy conservation, we have

\[
E = \frac{p^2}{2m} + V
\]
Introduction to Solid State Physics  

Chapter 2 - Review of Quantum Mechanics

The energy $E$ is related to the angular frequency $\omega$, while the momentum $p$ is related to the wavevector $k$, thus we have:

$$\hbar \omega = \frac{\hbar^2 k^2}{2m} + V$$

Therefore, we speculate that the time-dependent Schrödinger equation should be:

$$i\hbar \frac{\partial \psi}{\partial t} = H \psi$$

or

$$\left( -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial r^2} + V(r) \right) \psi = i\hbar \frac{\partial \psi}{\partial t}$$

For the situation where the system is in an energy eigenstate, we can see by inspection that

$$V(r, t) = \psi(r) \exp\left(-iEt/\hbar\right)$$

is a solution of the time-dependent equation with $\psi(r)$ as a solution of the time-independent equation:

$$\left( -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial r^2} + V(r) \right) \psi(r) = E\psi(r)$$

(i.e., for an energy eigenstate, the time dependence of the wave function is a simple oscillation with frequency proportional to energy.)

For an arbitrary physical state, (not necessarily an eigenstate), if at $t = 0$

$$\psi(r, t) = \sum_m c_m \phi_m(r)$$

Where $\phi_m$ is an eigenstate at the energy $E_m$, i.e.,

$$H\phi_m = E_m \phi_m$$

At an arbitrary time $t$, the state evolves to

$$\psi(r, t) = \sum_m c_m e^{-iE_m t/\hbar} \phi_m(r)$$

Therefore, by solving for the eigenstates of the time-independent Schrödinger equation, we obtain all the information about the time-dependent evolution of a physical state.

Some examples:

**Free Particle:**

Assuming $V = 0$ and considering 1-D for simplicity, Schrödinger equation becomes

$$\frac{d^2 \psi}{dx^2} + \frac{2mE}{\hbar^2} \psi = 0$$

Or like a wave equation as

$$\frac{d^2 \psi}{dx^2} + k^2 \psi = 0$$

Where

$$k = \sqrt{2mE/\hbar^2} \text{ or equivalently } E = \frac{m k^2}{2m}$$

Solution will be in form of

$$\psi(x) = A_+ e^{ikx} + A_- e^{-ikx}$$

Adding the time dependence

$$\psi(x, t) = A_+ e^{i(kx - \frac{Et}{\hbar})} + A_- e^{-i(kx - \frac{Et}{\hbar})}$$

For a free particle:

$$E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m}$$

**Particle in a box:**

Consider a particle in a hypothetically infinite deep potential well. Consider the simple problem of a particle of mass $m$ with spatially varying potential $V(x)$ as in next the Figure

Schrödinger equation will be as:

$$-\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x)$$

where $E$ is the energy of the particle, and $\psi(x)$ is the wavefunction. Consider $V(x)$ as

$$V(x) = \begin{cases} 0 & \text{for } x < L \text{ or } x > L \text{ for } 0 < x < L \end{cases}$$

As $V = \infty$ outside the box there is no possibility of finding particle outside the box, hence $\psi$ must be zero at the walls of the box. For $0 < x < L$:

$$-\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} = E\psi(x)$$

As $\psi(0) = \psi(L) = 0$ solutions to the above equation will be as

$$\psi_s(x) = A_s \sin\left(\frac{\pi x}{L}\right)$$

Where $A_s$ is a normalization constant. Eigenenergies will be as

$$E_s = \frac{n^2 \pi^2 \hbar^2}{2mL^2}$$

$$n = 1, 2, 3, \ldots$$
**Finite potential well:**

Again for simplicity we consider 1-D finite potential well, the time-independent Schrödinger equation can be written as:

\[
\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x)
\]

Now \( V(x) \) is assumed as:

\[
V(x) = \begin{cases} 
U_0 & \text{otherwise} \\
0 & 0 < x < a
\end{cases}
\]

Solution of SE will be as

If \( E > U_0 \):

\[
\frac{d^2\psi(x)}{dx^2} + k^2\psi(x) = 0 \quad \text{where} \quad k = \frac{2m(E-U_0)}{\hbar}
\]

so \( \psi(x) = A_+ e^{-kx} + A_- e^{kx} = A\sin(kx) + B\cos(kx) \)

If \( E < U_0 \):

\[
\frac{d^2\psi(x)}{dx^2} - \alpha^2\psi(x) = 0 \quad \text{where} \quad \alpha = \frac{2mU_0}{\hbar}
\]

so \( \psi(x) = D e^{-\alpha x} + E e^{\alpha x} \)

Needed steps for analytical solution of Schrödinger equation (for \( N \) regions):

Step 1. solve wave equation \( \frac{d^2\psi(x)}{dx^2} + k^2\psi(x) = 0 \) \( \xrightarrow{2N-1} \) \( 2N \) unknowns.

Step 2. Wavefunction at infinity: \( \psi(x = -\infty) = \psi(x = +\infty) = 0 \) \( \xrightarrow{2N-2} \) \( 2N-2 \) unknowns.

Step 3. Continuity of wavefunction: for remaining \( N-1 \) borders between \( N \) regions:

\[
\psi|_{x=x^*} = \psi|_{x=x^*} \quad \text{and} \quad \frac{d\psi}{dx}|_{x=x^*} = \frac{d\psi}{dx}|_{x=x^*} \xrightarrow{2N-2} \text{equations. (DONE!)}
\]

Step 4. find coefficients by setting \( \text{det(matrix)} = 0 \)

Step 5. Normalization \( \int_{-\infty}^{+\infty} |\psi(x)|^2 \ dx = 1 \)

Once \( \psi \) is known, anything else can be computed by its operator as;

\[
p = \int_{-\infty}^{+\infty} \psi^* \left( \frac{\hbar}{i} \frac{d}{dx} \right) \psi \ dx
\]

\[
E = \int_{-\infty}^{+\infty} \psi^* \left[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \right] \psi \ dx
\]
Step 5:

\[
\begin{pmatrix}
1 & 1 & -1 & 0 \\
\sin \kappa a & 0 & 0 & -e^{-\alpha a} \\
\cos \kappa a & -\sin \kappa a & 0 & a e^{-\alpha a}/k_f \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
A \\
B \\
C \\
D
\end{pmatrix}
= 0
\]

\[
\begin{pmatrix}
1 & -1 & 0 & 0 \\
0 & 0 & -e^{-\alpha a} & k_f \\
0 & -e^{-\alpha a} & 0 & A \sin \kappa a \\
0 & 0 & A \sin \kappa a & 0
\end{pmatrix}
= 0
\]

\[
\begin{pmatrix}
C \\
D
\end{pmatrix}
= \begin{pmatrix}
1 & -1 & 0 & 0 \\
0 & 0 & -e^{-\alpha a} & k_f \\
0 & -e^{-\alpha a} & 0 & A \sin \kappa a \\
0 & 0 & A \sin \kappa a & 0
\end{pmatrix}^{-1}
\begin{pmatrix}
0 \\
-k_f \\
-k_f A \sin \kappa a \\
A \sin \kappa a
\end{pmatrix}
\]

\(B, C, \text{ and } A\) is written in terms of \(A\), now \(A\) can be found as:

\[
\int_{-\infty}^{\infty} |\psi|^2 \, dx = 1 \int_{-\infty}^{0} c^2 e^{2ax} \, dx + \int_{0}^{\infty} (A \sin \kappa x + B \cos \kappa x)^2 \, dx + \int_{a}^{\infty} D^2 e^{-2ax} \, dx
\]
Consider atomic cores (nuclei + tightly bound electrons) as shown in the next Figure. One such core gives rise to a potential $U(x)$ because of the Coulomb attraction of the core, with its net charge, for the mobile electrons.

![Figure 2](image)

Two such cores side by side give rise to a different potential, and a large number of cores equally spaced gives rise to a periodic potential $U(x)$ as:

![Figure 3](image)

**Bloch Theorem:**

For a one-dimensional case, Bloch's theorem is as:

If $U(x)$ is periodic such that $U(x) = U(x + a)$

Then $\psi(x + a) = e^{ik}\psi(x)$ or, equivalently, $\psi(x) = e^{ikx}\psi(x)$; $U(x) = U(x + a)$

where $k$ can take on the values: $k = \frac{n\pi}{L}$, while $n = 0, \pm 1, \pm 2, \ldots, \pm N/2$

We know that the crystal is periodic, having the same potential at $x + na$ as it has at $x$ (where $n$ is an integer.) Any charge density must also have the same periodicity (the crystal must look identical in every unit cell.)

For example, charge density $\rho \propto |\psi^2|$ must be periodic in the same way, hence $|\psi(x + a)|^2 = |\psi(x)|^2$

which means

$$\rho \propto |\psi|^2$$

**Charge density**
where $C$ is a complex number of unit amplitude. (Note that there is no requirement that the
wavefunction itself be periodic with the crystal periodicity, since it is not apparently observable)

Periodic boundary conditions require the wavefunction at $x$ to be the same as the wavefunction at
$x + Na$, hence

$$\psi(x) = \psi(x + Na) = C^n \psi(x)$$

So

$$C^n = 1$$

and so $C$ is one of the $N$ roots of unity, i.e.,

$$C = \exp\left(\frac{2\pi i n}{N}\right) \quad s = 0, 1, 2, \ldots, N - 1$$

Or

$$C = \exp\left(\frac{2\pi i (N - m)}{N}\right) \quad s = 0, 1, 2, \ldots, N - 1$$

Replacing $C$ into equation for $\psi$:

$$\psi(x + a) = e^{i\alpha x} \psi(x)$$

Where

$$k = \frac{2\pi s}{Na}$$

and

$$k = \frac{2\pi n}{Na}$$

Which still gives essentially $N$ states, but now symmetrically disposed about $k = 0$.

Note the allowed $k$ values are evenly spaced by $2\pi / L$, where $L = Na$ is the length of the crystal in
this dimension, regardless of the detailed form of the periodic potential.

Straightforward extension of the 1-D version

$$\psi(x + a) = e^{i\alpha x} \psi(x)$$

Where $a$ is any lattice vector and considering the three crystal basis vector directions, 1, 2, and 3, with lattice constants $a_1, a_2, a_3, N_1, N_2, N_3$, and $k_1 = \frac{2\pi n}{N_1 a_i}$

The total reflectivity is then

$$R = \sum_m r_m e^{i2\pi m a}$$

which diverges when the Bragg condition $ka = m\pi$ is satisfied.

Therefore, no matter how small the reflectivity of each individual scatter is, as long as the Bragg
condition is satisfied, the reflected wave from each scatter coherently interferes with one another. Thus,
Previously we solved Schrödinger's equation in the "well" region: \( 0 < x < a \)
\[
\frac{d^2 \psi_0(x)}{dx^2} + \alpha^2 \psi_0(x) = 0 \quad \text{where} \quad \alpha = \sqrt{2mE/E_0}
\]
In the "barrier" region: \( b < x < a \)
\[
\frac{d^2 \psi_b(x)}{dx^2} + \beta^2 \psi_b(x) = 0 \quad \text{where} \quad \beta = \left\{ \begin{array}{ll}
\beta_+ & (0 < E < E_0) \\
\beta_- & (E > E_0)
\end{array} \right.
\]
Formal solution in well:
\[
\psi_0(x) = A_0 \sin \alpha x + B_0 \cos \alpha x
\]
Formal solution in barrier:
\[
\psi_b(x) = A_b \sin \beta x + B_b \cos \beta x
\]
Boundary conditions:
Continuity requirement:
\[
\left. \frac{d \psi_0(x)}{dx} \right|_{x=0} = \left. \frac{d \psi_b(x)}{dx} \right|_{x=0}
\]
Periodicity requirement:
\[
\psi_0(a) = e^{i(\alpha a)} \psi_0(-a) ; \quad \frac{d \psi_0(a)}{dx} = e^{i(\alpha a)} \frac{d \psi_0(-a)}{dx}
\]
results from boundary conditions
\[
\beta_0 = \beta_+ \quad \alpha A_0 = \beta_+ B_0
\]
\[
A_0 \sin \alpha a = B_0 \cos \beta a
\]
\[
\alpha A_0 \cos \alpha a - \beta_0 B_0 \sin \beta a = e^{i(\alpha a + \beta a)} (\beta_0 \cos \beta a + \beta_0 \sin \beta a)
\]
Matrix vector form:
\[
\begin{pmatrix}
\sin \alpha a + \frac{\alpha + \beta}{\alpha \beta} \sin \beta a \\
\cos \alpha a - \frac{\alpha + \beta}{\alpha \beta} \cos \beta a
\end{pmatrix}
= e^{i(\alpha a + \beta a)}
\begin{pmatrix}
\beta_0 \cos \beta a + \beta_0 \sin \beta a \\
-\cos \alpha a - \frac{\alpha + \beta}{\alpha \beta} \sin \beta a
\end{pmatrix}
\]
\[
[\psi_b(x=a)] = [0]
\]
Taking the determinant of the matrix and setting it equal to zero as usual for linear equation, gives the "characteristic equation" that determines the conditions under which the system has solutions:
\[
\alpha^2 + \beta^2 = 0
\]
\[
\sin \alpha a \cdot \sin \beta b - \cos \alpha a \cdot \cos \beta b = \cos (a + b)
\]
Finding the mathematical solutions to this equation will determine the allowed energies and the wave functions associated with them.
Brillouin zone and Zone Folding

We know that the solutions to the Schrödinger equation for our periodic potential in the crystal are in Bloch form (in one dimension for simplicity):

$$\psi(x) = e^{ikx}u(x)$$

We could recast the Schrödinger equation in a slightly different form by substituting this form of solution into it. Noting that

$$\frac{d\psi}{dx} = \left[\frac{du}{dx} + i\kappa u\right]e^{ikx}$$

and

$$\frac{d^2\psi}{dx^2} = \left[\frac{d^2u}{dx^2} + 2i\kappa \frac{du}{dx} - \kappa^2 u\right]e^{ikx}$$

We should rewrite the Schrödinger equation as:

$$-\frac{\hbar^2}{2m}\left[\frac{d^2}{dx^2} + 2i\kappa \frac{du}{dx} - \kappa^2\right]u + Vu = Eu$$

The detail of this equation does not matter to us for the moment. It is however, another eigenfunction equation; one way of proceeding to solve for the band structure would be to solve this equation for each allowed value of \(k\) to deduce the function \(u\) and the associated energy \(E\).

In fact, solving this equation for any given value of \(k\) will lead to several eigen solutions (actually it will lead to an infinite set of such solutions) with different eigenfunctions, \(u_{k,j}(x)\), and (generally) different eigenvalues for the energy, \(E_{k,j}\), (where \(j\) is an integer that indexes the solutions). This is somewhat different from what we have shown so far, where for each \(k\) we only had one energy solution.

In fact, there is a major simplification possible here that also reconciles these two pictures.

Suppose we know a solution of the above Equation for some particular value of \(k = k_0\)

$$\psi_{k_0} = u_{k_0}(x)e^{ik_0x}$$

Now suppose we want to know something about the solutions for another value of \(k\), specifically the value \(k = k_0 + 2\pi j/a\) (where \(a\) is the lattice constant and \(j\) is any integer). Then, we know that solution must also be in Bloch form, so we can write it as

$$\psi(x) = u(x)e^{i(k_0 + 2\pi j/a)x}$$

and, trivially, we can rewrite this as

$$\psi(x) = \psi(x)e^{ik_0x}$$

Where

$$\psi(x) = u(x)e^{i2\pi jx/a}$$

Note that \(exp\left(\frac{i2\pi jx}{a}\right)\) is also periodic with the lattice periodicity. Hence, \(\psi(x)\) is also periodic with the lattice periodicity (being the product of two functions with the lattice periodicity).
Particle Motion and Effective Mass:
Heisenberg uncertainty principle states for example, $\Delta E \Delta t \geq h$, for the uncertainty $\Delta E$ in the energy and the uncertainty $\Delta t$ in the time; if we know the energy accurately, we cannot accurately define the time at which the system has that energy. Similarly $\Delta p_x \Delta x \geq h$, for the uncertainty in momentum, $\Delta p_x$ in the x momentum and the uncertainty $\Delta x$ in the x position. (If a particle has a specific, accurately defined momentum, we do not know where it is—a monochromatic plane wave extends through all space.)

To understand movement of particles, we need to look at "wave packets", superpositions of waves that look like "pulses" in space. Wave packets are not energy eigenstates, so they change (e.g., move) in time. Specific linear combinations of waves over a range of energies (or, equivalently, in our case, k values or effective momenta $\hbar k$). The narrower the pulse in time, the broader the range of energies required describing it; the narrower the pulse in space, the broader the range of spatial frequencies k or effective momenta $\hbar k$ required to describe it, exactly like temporal or spatial Fourier analysis.

Group Velocity of Wave Packets, effective mass:
Classical wave theory, based on examining the behavior of linear superposition of waves, says the velocity of the center of a wave packet or pulse is the "group velocity"

\[ v_g = \frac{\partial \omega}{\partial k} \]

where $\omega$ is the frequency and $k$ is the wavevector. For example, consider a total wave made up out of a superposition of two waves, both propagating to the right, one at frequency $\omega + \delta \omega / 2$, with a wavevector (angular spatial frequency) $k + \delta k / 2$, and one at a frequency $\omega - \delta \omega / 2$ and a wavevector $k - \delta k / 2$. Then the total wave is

\[ f(t, x) = \sin \left( (\omega + \frac{\delta \omega}{2}) t - \left( k + \frac{\delta k}{2} \right) x \right) + \sin \left( (\omega - \frac{\delta \omega}{2}) t - \left( k - \frac{\delta k}{2} \right) x \right) \]

Since

\[ \sin \alpha + \sin \beta = 2 \sin \left( \frac{\alpha + \beta}{2} \right) \cos \left( \frac{\alpha - \beta}{2} \right) \]

then

\[ f(t, x) = 2 \sin(\omega t - kx) \cos(\delta \omega t / 2 - \delta k x / 2) \]

which can be viewed as an underlying wave $2 \sin(\omega t - kx)$ modulated by an envelope $\cos(\delta \omega t / 2 - \delta k x / 2)$.
Introduction to Solid State Physics

Chapter 4 – Brillouin Zones and Zone Folding

behaves like a mass, and is called the "effective mass", proportional to the inverse of the "curvature" of the band. (Note this agrees with the classical notion of mass and kinetic energy with $E = p^2 / 2m$)

Effective mass is a particularly meaningful and useful concept near minima or maxima (i.e., "edges") in the band structure, which also turns out to where we are most likely to find electrons participating in transport (i.e., transport of charge, or current). There

$$E = E_{\text{edge}} \pm \text{(constant)}(k - k_{\text{edge}})^2$$

So

$$\frac{d^2E}{dk^2} = \text{constant} \quad \text{when } E \text{ near } E_{\text{edge}}$$

parabolic regions in the bands correspond to particular values of effective mass

Note effective mass is a property of the band structure in the crystal. Effective masses can differ very much from the "free electron" mass (they can be smaller or larger, they can even be negative)

$m^*$ is positive near the bottoms of all bands, $m^*$ is negative near the tops of all bands.

Carriers and Current:

Consider a band structure like the Kronig-Penney example; suppose there are two electrons per "atom", and $N$ atoms per band. At low temperature, the lowest two bands would be completely filled, and the third band would be empty. At finite temperature, some of the electrons will have enough thermal energy to be thermally excited into the third band.

Although, an empty band can carry no current – there are no carriers. Less obviously, a full band can carry no current. For every electron with a particular momentum, there is one with exactly the opposite momentum, so there is no net momentum

$$\frac{1}{L} \sum \frac{dE}{dk} n_i = 0$$

For a filled band, $n = 1$ for all states, and $E(k)$ is an even function over $k$. The sum as indicated above goes to zero. Applying a force to the full band of electrons makes no difference to this situation. (The force changes the value of $k$ for each electron as a function of time, however, at each given time, for one electron at a given momentum there is still one with exactly the opposite momentum)

Only partially filled bands can carry current!

For a partially filled band (symmetric in $k$), there will be no net current in the absence of field.

Velocities are equal and opposite for the electrons symmetrically occupying states in the two directions of $k$. With field, the distribution of electrons is skewed by the applied field, giving net current. For nearly full band, conduction is only possible because there are some electrons missing. (can think of the missing (negatively charged) electrons as (positively charged) "holes" in the electron gas. Can think of them as particles with effective masses and energies).

Need to look at everything "upside down". E.g., effective mass associated with the empty states is taken to be the negative of the effective mass $m^*$ deduced previously (gives positive effective masses for holes at the top of a band.)

Reciprocal lattice:

We can easily extend the discussion that we had for 1-d periodic system to 3-d periodic system. For a crystal with primitive lattice vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$, the wavefunction should satisfy the Bloch theorem:

$$\psi(\vec{r} + n \vec{a}) = e^{i \vec{k} \cdot \vec{r}} \psi(\vec{r})$$

where $\psi(\vec{r} + \vec{a}) = \psi(\vec{r})$ for any lattice vector $\vec{a}$.

Similarly, we can show that the energy band has to be periodic in the $k$-space. (i.e., the reciprocal space):

$$E(k + G) = E(k) + \frac{2\pi}{\hbar} \sum \epsilon_i$$

where $\epsilon_i$ is the energy of any lattice vector $a_i$. The set $G$ is the reciprocal lattice for the crystal.

The reciprocal lattice of a Bravais lattice is the set of all vectors $\vec{G}$ such that $e^{i\vec{G} \cdot \vec{R}} = 1$ for all lattice point position vectors $\vec{R}$. This reciprocal lattice of a Bravais lattice is itself a Bravais lattice, and the reciprocal of the reciprocal lattice is the original lattice.

For an infinite three dimensional lattice, defined by its primitive vectors $(\vec{a}_1, \vec{a}_2, \vec{a}_3)$, its reciprocal lattice can be determined by generating its three reciprocal primitive vectors, through the formulas

$$\vec{G}_1 = 2\pi \frac{\vec{a}_2 \times \vec{a}_3}{\vec{a}_1}$$

$$\vec{G}_2 = 2\pi \frac{\vec{a}_3 \times \vec{a}_1}{\vec{a}_2}$$

$$\vec{G}_3 = 2\pi \frac{\vec{a}_1 \times \vec{a}_2}{\vec{a}_3}$$

Or

$$\begin{bmatrix} G_1 \\ G_2 \\ G_3 \end{bmatrix} = 2\pi \begin{bmatrix} a_1 & a_2 & a_3 \\ a_2 & a_3 & a_1 \\ a_3 & a_1 & a_2 \end{bmatrix}^{-1}$$

There are corresponding reciprocal lattice vectors $G = m_1 \vec{G}_1 + m_2 \vec{G}_2 + m_3 \vec{G}_3$ where $m_1, m_2,$ and $m_3$ are integers. The properties associated with any particular point $\vec{k}$ in reciprocal space are the same as
Reciprocal space (also called "k-space") is the space in which the Fourier transform of a spatial function is represented (similarly the frequency domain is the space in which the Fourier transform of a time dependent function is represented). A Fourier transform takes us from "real space" to reciprocal space or vice versa.

\[ F(k) = \int_{-\pi/a}^{\pi/a} f(r) e^{-2\pi i k \cdot r} \, dr \]

A reciprocal lattice is a periodic set of points in this space, and contains the \( k \) points that compose the Fourier transform of a periodic spatial lattice. The Brillouin zone is a volume within this space that contains all the unique k-vectors that represent the periodicity of classical or quantum waves allowed in a periodic structure.

We have seen by example that we can solve the Schrödinger equation, and deduce a dependence of \( E \) on \( k \) that gives us a "band structure" for the crystalline solid. To understand some of the important general properties and great simplifications that are possible in this problem, we need to step back and take a more formal approach for the moment.

Wigner-Seitz unit cell:

A common "standardized" choice of primitive unit cell. The Wigner-Seitz cell is the region round about a lattice point that is closer to that point than to any other point.

The Wigner-Seitz unit cell can be constructed by

(i) drawing lines between the point and (in principle) all other points in the lattice (in practice only the points reasonably close to the one of interest need be considered),

(ii) bisecting each line with a plane perpendicular to the line

(iii) taking the smallest polyhedron formed by these planes about the point

In 3-D have wavevector \( k \). Brillouin zone is volume in k-space, just as repeated zones fill all "space" on the k-axis in one-dimension.
Brillouin zones fill all "k-space" in 3 dimensions.
Brillouin zones are the "unit cells" of a "reciprocal lattice".

Brillouin zone for a fcc lattice is a bcc Wigner-Seitz unit cell.

Features:
- Valence bands: essentially full of electrons; maximum at \( I' \); three valence bands (heavy hole, light hole and split off hole).
- Conduction bands: essentially empty of electrons:
  - Ge: minimum at \( L \); six equivalent minima.
  - Si: minimum at 0.8(2\( \pi /a \)) along the \( TX \) direction, eight equivalent minima.
  - GaAs: minimum at \( I' \), one minimum.

Notations:
- \( I' \): identifies zone center (k=0)
- \( X' \): denotes zone end along a <100> direction
- \( L' \): denotes zone end along a <111> direction
Band gap energy: (minimum) separation between highest valence band and lowest conduction band.

At room temperature (300K), Ge: 0.663 eV ; Si: 1.125 eV ; GaAs: 1.422 eV (corresponds to a photon wavelength of 871 nm)

Direct and Indirect Bandgap Semiconductors:
GaAs is a "direct bandgap" semiconductor as the maximum in valence band is directly below the minimum in conduction band. This is very important for optical properties. Strong optical transitions are "vertical" on $E$ vs. $k$ diagram because photon has relatively little momentum. This leads to strong optical absorption and emission between band minima and maxima that are directly above one another.

Si and Ge are "indirect bandgap" semiconductors as the maximum in valence band and the minima in conduction band are at different $k$, which can lead to weak optical absorption for photon energies near the bandgap energy and, especially, emission. Photon does not have enough momentum to "connect" the minima at different points in the Brillouin zone. This is why we don't have a silicon laser.

Band Structures of Insulators and Metals:
At low temperature:
Insulator or semiconductor (in pure material): bands either completely full or completely empty of electrons
Metal: one band partially occupied, other bands full (e.g., Sodium)
Semimetal: bands partially occupied because of overlap of bands in energy

Difference between semiconductor and insulator:
Band gap energy is low enough that there is a sufficiently large number of carriers (electrons and/or holes) thermally excited across the bandgap at, e.g., room temperature to give significant conductivity and carrier density and type can be controlled by introduction of small amounts of impurities

Constant Energy Surfaces:
Consider a specific energy, $E = E_0$, just above the energy, $E_k$, of the minimum or minima. In general:
$$E - E_0 = AK_1^2 + BK_2^2 + CK_3^2$$
because the curvature of the band structure is, in general, different in the three different crystal directions $k_i$.

In simple cubic symmetry, at zone center, must have the same behavior in all three directions, so might expect something like
$$E - E_0 = A(k_1^2 + k_2^2 + k_3^2)$$

Effective Mass in 3-D:
In general, effective mass now has to be written as a tensor
$$m^* = \frac{1}{m^*_1 m^*_2 m^*_3} \begin{pmatrix} m^*_1 & m^*_2 & m^*_3 \\ m^*_2 & m^*_3 & m^*_1 \\ m^*_3 & m^*_1 & m^*_2 \end{pmatrix}$$
where, in general
$$\frac{1}{m^*_i} \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial x_i^2}$$

Note: In general that acceleration and force are not colinear, the particle does not necessarily go in the direction you push it!
For conduction bands at zone center in diamond and zinc-blende semiconductors, behavior of band is approximately isotropic and "parabolic".

\[ E - E_c = A(k_x^2 + k_y^2 + k_z^2) \]

So masses are formally

\[ m_{d-x} = m_{d-y} = m_{d-z} = \frac{2A}{\hbar^2} \]

and all "off-diagonal" terms in the effective mass tensor are (approximately) zero. Hence, we can define single scalar electron effective mass, \( m^* \):

\[ E - E_c = \frac{\hbar^2}{2m^*} (k_x^2 + k_y^2 + k_z^2) \]

For which we get back to a simple, scalar-vector relationship, with the particle accelerating in the direction in which it is pushed.

\[ \vec{d} \frac{\vec{p}}{m^*} \]

Near these minima in Ge and Si, the energy behaves approximately as

\[ E - E_c = Ak_x^2 + B(k_y^2 + k_z^2) \]

where \( k_x, k_y, \) and \( k_z \) are along the principal axis. (For Si, \( k_z \) is along the \( <111> \) directions, for Ge, \( k_z \) is along the \( <110> \) directions).

So we can define

\[ m_{11} = \frac{2A}{\hbar^2} \]

\[ m_{22} = m_{33} = \frac{2B}{\hbar^2} \]

Leading to

\[ E - E_c = \frac{\hbar^2}{2m_{11}} k_x^2 + \frac{\hbar^2}{2m_{11}} (k_y^2 + k_z^2) \quad \text{... Ge, Si} \]

where

\[ m_1^* = \frac{\text{Length of the ellipsoid along the axis of revolution}}{\text{Maximum width of the ellipsoid perpendicular to the axis of revolution}} \]

Valence Bands in Diamond and Zinc-Blende Crystals:

Figure 11

One model (Luttinger-Kohn) useful for valence bands in diamond and zinc-blende crystals gives an expression for the top two (heavy and light) hole valence bands

\[ E_h - E = AK^2 \pm \sqrt{B^2K^4 + C^2(K_y^2 + K_z^2)^2} \]

Note this has "cross terms" between the different directions [e.g., \( K_y^2K_z^2 \)] and shows valence band "warping".
Density of States:  
Is needed to understand:
1. Thermal occupation of band (and impurity) states by electrons and/or holes
2. Optical absorption spectra
3. Need to understand explicitly how many states there are available to an electron or hole in a given energy range about some particular energy of interest
4. Density of states (number of possible electron states per unit energy per unit volume)
5. Density of states in k space (a very simple calculation) then deduce the density of states in energy from the (presumed) known relationship between k and E.

Density of States in k Space:
Have already solved this problem, deduced from Bloch theorem that states are spaced by
\[ \frac{2\pi \times 2\pi \times 2\pi}{a \times b \times c} \]

in the three coordinate directions in k space, where a, b, c are the lengths of the crystal in the three directions (note: not the unit cell dimensions - we are following Pierret's notation here, even though it is somewhat inconsistent)

Cyclical boundary conditions: correctly "counts" all the states if we consider states only in one Brillouin zone. In a total range of k\text{max} \times (2\pi/a) \times (2\pi/b) \times (2\pi/c) \times \text{unitcell}.

For example, we will have Nk up states, where Nk is the number of unit cells in the x direction, since Nk \times \text{unitcell} = (2\pi/a).

With a spacing of states of 2\pi/a, the number of states per unit length (in k space is \( \frac{2\pi}{a} \)), hence, density of states per unit volume in k space is \( \frac{abc}{(2\pi)^3} \). But abc is just the volume of the crystal, so the density of states per unit k space volume is

\[ Nk = \frac{abc}{(2\pi)^3} \]

can equally well consider "hard wall" boundary conditions and get the same answer.

actual finite crystal looks like a "box" for the \( \exp(\pm ikx) \) waves (or similar waves in the other crystal directions)

Requiring the waves to be zero at the walls of the box (regardless of the specific form of the unit cell function) requires that the "envelope" be standing waves, so

\[ \psi(x, y, z) = u(x, y, z) \sin kx \sin ky \sin kz \]

where

\[ kx = \frac{n_x \pi}{a} , ky = \frac{n_y \pi}{a} , kz = \frac{n_z \pi}{a}(n_x, n_y, n_z = \pm 1, \pm 2, \pm 3, ...) \]

This would give

\[ Nk = \frac{abc}{(2\pi)^3} = \frac{abc}{(2\pi)^3} \]

negative k values correspond to the same states as positive k values in a standing wave, so should not be counted twice, so should divide by 2 for each dimension, giving

\[ \text{allowed energy states} = \frac{abc}{8\pi^3} \]

Need to correct this result to account for the fact that there are two distinct possible electron states for any given k corresponding to the two different values of electron spin, so we have

\[ \text{allowed electron energy states} = \frac{abc}{4\pi^3} \]

To convert the known density of states in k space to a density of states in energy, we need to invoke some relation between E and k. Most interesting simple case is spherically symmetric "parabolic band" found near some band maxima and minima. There we have

\[ E = \frac{k^2 m}{2} \text{ or } k^2 = \frac{2mE}{\hbar^2} \]

For some effective mass \( m \), where we are presuming for simplicity for the moment that \( E = 0 \) for \( k = 0 \). (We can always change the energy origin later). Now consider a thin spherical shell in k space, the volume of the shell is \( 4\pi k^2 dk \).

In this spherical shell there therefore is a number of \( k \) states
Introduction to Solid State Physics

Chapter 5 – Density of States

(energy states with k between k and k + dk) = \frac{4nk^2dk}{\hbar^2} \frac{abc}{4\pi^2}

Because, by assumption, E \propto k^2 in our parabolic band, this shell also corresponds to all states with energies between E and E + dE, where

\[ dE = \frac{\hbar^2 k^2}{m} \quad \text{or} \quad dk = \frac{1}{\hbar} \sqrt{\frac{m}{2}} \frac{dE}{\sqrt{E}} \]

so the number of states within this energy range is

\[ \langle \text{energy states with } E \text{ between } E \text{ and } E + dE \rangle = \frac{m\sqrt{2mE}}{\pi^2h^3} dE \]

It is conventional to express the density of states as the number of states per unit energy per unit crystal volume, giving

\[ g(E) = \frac{\langle \text{energy states with } E \text{ between } E \text{ and } E + dE \rangle}{\frac{m\sqrt{2mE}}{\pi^2h^3}} \]

Density of States in Energy in Conduction and Valence Bands:

For a conduction band for which the (electron) effective mass \( m^*_e \) is positive

\[ g_c(E) = \frac{m_e\sqrt{2m_eE}}{\pi^2h^3} \quad E \geq E_c \]

For a valence band for which the electron effective mass is negative, we can work instead with a positive hole effective mass \( m^*_h \) to obtain

\[ g_v(E) = \frac{m_h\sqrt{2m_h(E - E_v)}}{\pi^2h^3} \quad E \leq E_v \]

Density of States In GaAs Conduction Band:

In GaAs, lowest conduction band is approximately spherically symmetric near zone center, so we can use a simple effective mass \( m^*_n = m^*_n \) (this mass is very light (~0.07m) which corresponds to steeply curved band, low overall density of states)

Density of States in Si and Ge Conduction Band:

The \( N_e \) minima in conduction band, each approximately "ellipsoidal"

\[ E - E_c = \frac{\hbar^2}{2m^*_n} k^2 + \frac{\hbar^2}{2m^*_n} (k_x^2 + k_y^2) \]

Can be rewrite in form of:

\[ \frac{k_x^2}{\alpha^2} + \frac{k_y^2}{\beta^2} + \frac{k_z^2}{1} = 1 \]

where

\[ \alpha = \frac{2m^*_n(E - E_c)}{\hbar} \quad \beta = \frac{2m^*_n(E - E_c)}{\hbar} \]

Can repeat density of states derivation, now using an ellipsoid in \( k \) space instead of a sphere, with ellipsoid chosen to correspond to a surface of constant energy. Result is similar to "spherical" derivation. total volume of ellipsoids in \( k \) space out to some specific energy is

\[ N_{\text{el}} \left( \frac{\pi n \beta^2}{4} \right) = \frac{4}{\pi} k_{\text{eff}}^3 \]

where \( k_{\text{eff}} \) is the radius of sphere of the same volume.

A single density of state equivalent mass \( m^*_\text{eq} \) can be defined as:

\[ (m^*_\text{eq})^{1/2} = N_{\text{el}}(m^*_n; m^*_h)^{1/2} \]

Giving:

\[ m^*_n = 6^{1/3}(m^*_n)^{2/3} \quad \text{Si} \]
\[ m^*_n = 4^{1/3}(m^*_n)^{2/3} \quad \text{Ge} \]

- Valence Bands in Diamond and Zinc-Blende Semiconductors:

Have two bands, light hole and heavy hole, both with the same energy at \( k = 0 \)

Presuming for simplicity that we can use a simple spherically symmetric band in each case as a first approximation, simply add the densities of states to obtain

\[ g_v(E) = \frac{m_h\sqrt{2m_h(E - E_v)}}{\pi^2h^3} = \frac{m_h\sqrt{2m_h(E - E_v)}}{\pi^2h^3} + \frac{m_h\sqrt{2m_h(E - E_v)}}{\pi^2h^3} \]

or

\[ m^*_v = \left( \frac{(m^*_n)^{3/2} + (m^*_h)^{3/2}}{3/2} \right)^{2/3} \]
• Vibrations in Crystals (Longitudinal/Transverse)
• Dispersion Relation for Elastic Waves (Brillouin Zones)
• Two atoms primitive basis: optical phonons and acoustic phonons
• Quantization of phonon modes

The nuclei (together with the tightly bound valence electrons) can vibrate, allowing "elastic waves" (nuclei connected by "springs" – bonds between atoms). This leads to:
- conduction of sound in the crystal
- conduction of heat
- heat capacity (thermal energy can be stored in the vibrations)
- limitations on conduction of current (electrons or holes see vibrations as imperfections in crystal lattice; nuclei not in their exact lattice positions leads to "scattering" of electrons or holes)
- optical absorption phenomena ("Indirect" optical absorption in "Direct" semiconductors, e.g., Si, Ge)
- difference between low and high frequency dielectric constants

Two key steps in understanding basics of these vibrations are:
1) understand the independent "modes" of oscillation of a regular lattice of masses connected by springs:
   - classes of modes: longitudinal and transverse:
   - dispersion relation between frequency, \( \omega \), and wavevector \( K \) in each class (note: now discussing wavevector of elastic waves in the crystal, not electron wavevector)
2) these vibrations are quantized:
   - just as the photon is the "quantum" for electromagnetic waves (the phonon is the quantum associated with elastic waves in a crystal. energy in any particular mode of oscillation comes in discrete amounts with energy \( E = \hbar \omega \))

Vibrations in Crystals:
Unlike sound waves in, e.g., gases, which have no "shear" strength), vibrations in crystals can be both:
- "longitudinal" (like pressure waves in gases)
- "transverse" (there is a transverse restoring force also)

Nearest neighbor Interaction model (simplest model):
Suppose for simplicity that the force on a given plane of atoms is determined only by its displacement relative to its nearest neighbor planes (note that in general this is not true, though this force is likely the largest single contribution). Suppose also that the force is linearly proportional to those displacements (usual Hooke’s law for springs, good first approximation).

Then, force \( F \), pushing a given plane is the balance of the forces from the two adjacent planes of atoms:
\[
F = C(u_{i+1} - u_i) + C(u_{i-1} - u_i)
\]
where \( C \) is an effective spring constant. Hence the equation of motion of the plane of atoms of mass \( M \) is:
\[
M \frac{d^2 u_i}{dt^2} = C(u_{i+1} + u_{i-1} - 2u_i)
\]

We make an intelligent guess that there are solutions to this equation that are monochromatic oscillations of some frequency \( \omega \), i.e., with some time dependent part of the form \( \exp(\pm i \omega t) \). Hence we have
\[
-M\omega^2 u_i = C(u_{i+1} + u_{i-1} - 2u_i)
\]

We now make the further intelligent guess that these solutions are waves of the form \( u_i = e^{i(Kz + \omega t)} \). With this assumption, we conclude that \( u_{i,\pm} = \frac{1}{2}[u_{i+1} + u_{i-1} - 2u_i] \).

Replacing the above Eq
\[
-M\omega^2 u_i e^{i\omega t} = C[u_{i+1} e^{i\omega t} + e^{i\omega t} (e^{ik} - e^{-ik}) - 2 e^{i\omega t}]
\]

\[
\chi(t) = \int x(t) e^{-i\omega t} dt
\]
\[
\chi(t) = \int x(t - t_0) e^{-i\omega t} dt = \chi(t - t_0) e^{-i\omega t_0}
\]
\[
\chi(t) = x(t) e^{-i\omega t}
\]

Or
\[ M\omega^2 = -C\{e^{iK\alpha} + e^{-iK\alpha} - 2\} = 2C(1 - \cos K\alpha) \]
\[ \omega = \sqrt{\frac{4C}{M} \sin \frac{K\alpha}{2}} \]

Figure 3: The "dispersion relation" between \( \omega \) and \( K \) is clearly periodic.

Note that at \( K = \pm \pi/a \) we have zero slopes at the Brillouin zone edges.

Brillouin Zones for Elastic Waves, Allowed Values of \( K \):

Note that we need only consider \( K \) values inside the first Brillouin zone \(-\pi/a \leq K \leq \pi/a\)

The waves under consideration are those of the positions of the discrete atoms. The wave in the region between the atoms has no meaning; it does not correspond to the displacement of anything. Hence, adding on \( 2\pi/a \) to the value of \( K \) makes no difference to the wave.

Note that, if we impose periodic boundary conditions on the elastic wave, we obtain the same allowed values of \( K \) as we encountered for electron waves, i.e., values spaced by \( 2\pi/Na \), where \( N \) is the number of atomic planes in the direction of interest. We can also regard this as simply defining a Fourier basis set for describing any elastic wave in the crystal – to describe the displacement of the \( N \) atomic planes, we only need \( N \) different (i.e., linearly independent) functions with a total of \( N \) coefficients, hence it is reasonable to make the choice that we consider only these discrete values of \( K \).

Group Velocity:
Transfer of elastic energy will take place at the group velocity
\[ v_g = \frac{d\omega}{dK} \]

this is the velocity at which pulses of elastic displacement will propagate through the crystal.

Generalizing to three dimensions gives
\[ v_g = \nabla_K \omega(K) \]

For the simple dispersion relation of the "nearest neighbor" model, we have
\[ v_g = \frac{C\alpha^2}{M} \sin \frac{K\alpha}{2} \]

which is zero at the edges of the Brillouin zone.

For long wavelengths (i.e., \( K \) very small), can expand the cosine in Eq to obtain
\[ \omega = \sqrt{\frac{4C}{M} K} \]

which gives dispersionless propagation with a particular velocity. The "velocity of sound" \( a\sqrt{C/M} \)

Optical Phonons:
So far, have considered only one atom per unit cell (all atoms identical); with two atoms per unit cell, we can have additional, distinct kind of "mode" of vibration.

The two different atoms vibrating in opposite directions called an "optical" mode (though it is not a mechanical vibration, not an electromagnetic one. (If the two atoms had opposite net charges, then an electromagnetic field would pull them in opposite directions, thus exciting this kind of vibration). Other kind of mode (different kinds of atoms oscillating in the same direction as their different neighbors) known as an "acoustic" mode, because it corresponds to conventional sound waves for small \( K \).

Now need two different wave "coordinates": \( u_x \) (for one kind of atom), and \( v_x \) (for the other one) (note: can still have both longitudinal and transverse oscillations)

Still presume, for simplicity, that forces are only from nearest neighbors. Now, have two coupled equations since forces are, by assumption, from atoms of the other kind. Using same arguments as before now have:
\[ M_1 \frac{d^2u_1}{dt^2} = C(u_{x+1} + u_{x-1} - 2u_x) \]
\[ M_2 \frac{d^2v_x}{dt^2} = C(u_{x+1} + u_{x-1} - 2v_x) \]

Proposing wave equation as:
\[ u = v e^{i\omega K a} e^{-i\omega t} \quad \text{and} \quad v = u e^{i\omega K a} e^{-i\omega t} \]

So

\[ -M_1 \omega^2 u = C u \left(1 + e^{-i\alpha K a}\right) - 2 Cu \]
\[ -M_2 \omega^2 v = C v \left(1 + e^{i\alpha K a}\right) - 2 Cv \]

In matrix form

\[
\begin{pmatrix}
2C - M_1 \omega^2 & -C \left(1 + e^{-i\alpha K a}\right) \\
-C \left(1 + e^{i\alpha K a}\right) & 2C - M_2 \omega^2
\end{pmatrix}
\begin{bmatrix}
u \\ v
\end{bmatrix} = 0
\]

To have a solution the determinant should be zero

\[
\begin{vmatrix}
2C - M_1 \omega^2 & -C \left(1 + e^{-i\alpha K a}\right) \\
-C \left(1 + e^{i\alpha K a}\right) & 2C - M_2 \omega^2
\end{vmatrix} = 0
\]

which gives

\[ M_1 M_2 \omega^4 - 2C(M_1 + M_2) \omega^2 + 2C(1 - \cos \alpha K a) = 0 \]

quadratic – gives two solutions \( \omega \) for each \( K \):

1. one corresponding to adjacent pairs of atoms oscillating in the same direction. (acoustic mode – lower frequency solution)

\[ \omega^2 = \frac{C/2}{M_1 + M_2} K^2 a^2 \quad \text{near} \ K = 0 \]

\[ \omega^2 = \frac{2C}{M_1} \quad \text{at} \ K = \pm \pi/a \]

2. one corresponding to adjacent pairs of atoms oscillating in the opposite direction. (optical mode – higher frequency solution \( K \), has finite frequency even at zero)

\[ \omega^2 = 2C \left(\frac{1}{M_1} + \frac{1}{M_2}\right) \quad \text{near} \ K = 0 \]

\[ \omega^2 = \frac{2C}{M_2} \quad \text{at} \ K = \pm \pi/a \]

Note that at \( K = 0 \) we can see for the optical branch \( u/v = -M_2/M_1 \), so there is no net motion of the center of mass.

Quantization of Harmonic Oscillator:

In a simple "harmonic oscillator", i.e., some system that oscillates sinusoidally in time and space, we can write the total energy of the oscillator as (classically)

\[ E = \frac{p^2}{2m} + \frac{1}{2} C x^2 \quad (*) \]

where \( M \) is the oscillating mass and \( C \) is the "spring constant".

Note that this is a sum of kinetic and potential energies. This is the form of potential energy we get if the restoring force is proportional to the displacement \( x \), as it is in a simple spring.

In considering a harmonic oscillator in general in quantum mechanics, whether we are talking about electrons, atoms, or some other entities, we will usually assert that we can replace the classical momentum by the momentum operator. As before, we will assert that this operator, operating on the state of the system (e.g., the wave function), will give us the value of the momentum if we are in an eigenstate. Taking (in one dimension)
\[ p = -i\hbar \frac{\partial}{\partial x} \]

as usual will give us an eigenvalue/eigenfunction equation instead of \( E \):

\[ \frac{\hbar^2}{2M} \frac{\partial^2}{\partial x^2} + Cx \phi(x) = E \phi(x) \]

We can now solve this equation for the allowed values of \( E \), and the associated wavefunctions \( \phi(x) \).

If we were simply dealing with a mass \( M \) on a spring, for example, \( |\phi(x)|^2 \) would tell us the probability of finding the mass in the vicinity of the position \( x \).

The solutions to this mathematical equation are well known. In particular, the allowed energies are

\[ E_n = \left(n + \frac{1}{2}\right)\hbar\omega \quad (***) \]

Where \( n = 0, 1, 2, \ldots \) and \( \omega = \sqrt{C/M} \). **Note that the allowed values of energy are quantized and equally spaced.**

The classical vibration modes of a crystal have energies that obey the same kind of equation as (**), though the "coordinate" \( x \) is not the physical displacement of one atom, but rather the amplitude of an entire mode or wave. Similarly, the "mass" \( M \) and the "spring constant" \( C \) are not the mass of one atom or the spring constant of one "bond", but are rather properties of the entire mode.

We can, however, go about quantizing the results in the same way, and so we find the allowed energies in a given mode are quantized just as in Eq. (**). This is analogous to the situation with electromagnetic modes. We can think of the units \( \hbar\omega \) as corresponding to particles that here we call phonons (by analogy with photons). Changing from one eigenstate to another therefore corresponds to emission or absorption of phonons. Optical phonons typically have energies of \( \sim 10s \) of meV. Phonons can also be viewed as having a "crystal" momentum \( \hbar K \) if they have a wavevector \( K \).
Semi-conductors (and insulators) tend to be substantially transparent for photon energies below the band gap energy (between the highest occupied valence band and the lowest unoccupied conduction band) and relatively strongly absorbing for photon energies above the band gap energy.

Two broad classes of semi-conductors as far as optical properties are concerned are:

1. Direct gap semi-conductors
   - Strong optical absorption and emission above the bandgap energy
   - Very abrupt onset of optical absorption

2. Indirect gap semi-conductors
   - Relatively weaker optical absorption and emission above the bandgap energy
   - Relatively smooth onset of optical absorption

Direct Gap Optical Absorption:

Optical wavelength $\lambda$ is very long compared to the separation between atoms $a$ (Typical wavelength of interest $\sim 1 \mu m$; Lattice constant of GaAs $\sim 0.5 nm$). Therefore optical wavevector $k_{opt} = 2\pi/a$ is very small compared to the size of the Brillouin zone ($\sim 2\pi/a$). Therefore transitions between a valence band state and a conduction band state are essentially vertical if they are to conserve momentum $hk$.

Hence, simple model of direct optical absorption: absorption of a photon raises an electron in a particular $k$ state in the valence band vertically to the state of the same $k$ in the conduction band (so-called vertical transitions).

Presume that, when we get the energy of the photon exactly correct for the separation between an occupied valence band state and an empty conduction band state, there is a certain probability per unit time that we will make the transition, absorbing a photon in the process, i.e., we would have a transition rate

$$ W = A \delta(E_v(k) - E_g(k) - \hbar\omega) $$

Here $A$ is a constant representing the strength of the absorption corresponding to this transition, $E_v(k)$ the energy in the conduction band and $E_g(k)$ the energy in the valence band corresponding to the wavevector $k$, and $\hbar\omega$ is the photon energy.

The $\delta()$ function is a mathematical abstraction of an absorption "line" that is convenient for subsequent algebra. It is infinitely narrow and infinitely high, but has unit area. More physically, we should choose some spectral line with a finite width and height to corresponding to the transition, but though it gives the same answer in the limit, it is mathematically clumsy to take such finite widths. The total transition rate is sum of the transition rates for all of the different possible $k$ values. 

Presuming that this transition rate is the same for all states (approximately true near band minima and maxima), we obtain a relation of the form:

$$ W_{TOT} = A \sum_k \delta(E_v(k) - E_g(k) - \hbar\omega) $$

This formula tells us how the transition rate, and therefore the absorption coefficients for the incident light, varies with the frequency $\omega$ of the incident light. We need to change the summation to an integral to evaluate the mathematical answer, so, using the density of states we formally rewrite (considering unit volume)

$$ \sum_k \approx \int dq g(q)d^3q $$

where $g(q)$ is the density of states in $k$-space. Change variables in the integral to energy $E_j = E_v(k) - E_g(k)$. Assuming parabolic bands, can define $E_j$ as

$$ E_j = E_v(k) - E_g(k) = \frac{\hbar^2 k^2}{2m_e} \left( \frac{1}{m_e} + \frac{1}{m_h} \right) + E_g = \frac{\hbar^2 k^2}{2\mu_{eff}} + E_g $$

$m_e$ is valence band effective mass, and reduced effective mass, $\mu_{eff}$ is defined as

$$ \frac{1}{\mu_{eff}} = \frac{1}{m_e} + \frac{1}{m_h} $$

hence, define "joint density of states" $g_j(E_j)$, and write

$$ g_j(E_j)dE_j = g(q)d^3q $$

leading, for $E_j \geq E_g$, to

$$ g_j(E_j) = \frac{1}{2\pi^2} \left( \frac{2\mu_{eff}}{\hbar^2} \right)^{3/2} (E_j - E_g)^{1/2} $$

hence

$$ W_{TOT} = A \sum_k \delta(E_v(k) - E_g(k) - \hbar\omega) = A \int_{E_g}^{\infty} \frac{1}{2\pi^2} \left( \frac{2\mu_{eff}}{\hbar^2} \right)^{3/2} (E_j - E_g)^{1/2} dE_j $$

i.e. for $\hbar\omega \geq E_g$,

$$ W_{TOT} \propto (\hbar\omega - E_g)^{1/2} $$

Optical absorption length:

Since $\alpha$ is the probability of absorption of a photon per unit length, if we start out with $n_p$ photons crossing a unit area, then after a short distance $dz$, $n_p \propto dz$ photons will have been absorbed, so
\[
\frac{dn_p}{dx} = -n_p \alpha \\
n_p(x) = n_p(0)e^{-\alpha x}
\]

or, since intensity \( I \) is proportional to the number of photons crossing a unit area per second

\( I(x) = I(0)e^{-\alpha x} \)

consequently, after a distance \( 1/\alpha \), the intensity has fallen to \( 1/e \) of its initial value, hence \( 1/\alpha \) is the "absorption length." \( \alpha \) is mostly commonly quoted in units of \( \text{cm}^{-1} \). Typical value for \( \alpha \) just above the bandgap energy is \( \sim 10^4 \text{ cm}^{-1} \) in a direct gap semiconductor, corresponding to an absorption length of \( \sim 1 \mu \text{m} \).

---

**Figure 2:** Optical absorption of GaAs at various temperatures

Spectrum of direct gap semiconductor: has abrupt absorption onset near bandgap \( E_g \), has smooth rise above the bandgap energy, and the peak occurs below the bandgap of the bulk crystal. Also have clear appearance of a relatively strong peak near to the bandgap energy, especially at low temperatures, absorption does not have smoothly rising curve predicted by non-excitonic model – instead almost step-like rise, followed by a much "straighter" increase with increasing photon energy, (explanation: excitonic effects)

**Indirect gap optical absorption:**
Indirect gap material, like silicon, smallest energy separation for electrons in the conduction band and holes in the valence band occurs for very different electron and hole momentum. In this case, optical absorption requires participation of a phonon to conserve the overall momentum.

"two-step" process, with two different types possible

Process I:
- electron transition from the valence band absorbing a photon, followed by
- electron transition within conduction band, emitting or absorbing a phonon

Process II:
- hole transition from conduction band to valence band, followed by
- hole transition within valence band, emitting or absorbing a phonon

---

**Figure 3:**

Note: The energy is not conserved in intermediate stages of "two-step" process necessarily, though energy must be conserved overall. Photon energy usually is substantially less than bandgap energy (phonon energy is also typically small (e.g., 10's of meV))

Analysis straightforward, but more complex than simple direct gap derivation: form of result, approximately

\[
\alpha \propto B_{\text{abs}}(\hbar \omega + E_{\text{phonon}} - E_g)^2 + B_{\text{em}}(\hbar \omega - E_{\text{phonon}} - E_g)^2
\]

where \( E_{\text{phonon}} \) is the phonon energy, and the factors \( B_{\text{abs}} \) and \( B_{\text{em}} \) correspond to absorption and emission of a phonon, respectively. (For the case of absorption of a phonon, process depends on temperature since there needs to be a thermal population of phonons present from which one can be absorbed)

---

**Figure 4:** Absorption spectrum of Ge (plotted so that quadratic increase of absorption with energy gives a straight line)
Indirect absorption spectra do not have the abruptly rising character of the direct absorption spectrum (generally significantly weaker than direct absorption.)
Introduction to Solid State Physics

Chapter 8 – Statistical/Thermal Physics

- Basic ideas of thermal physics
- Binary model systems
- Probability, average value
- Thermal equilibrium
- Concept of entropy and temperature
- Laws of thermodynamics
- Boltzmann factor
- Partition function
- Free energy
- Chemical potential

Motivation for us to study this in a solid state physics fundamentals course, e.g.,

a. Transport (conduction) in semiconductors is based on thermal populations of charged particles, because conductivity depends on the thermal distributions
b. Concept of doping in semiconductors depends on thermal ionization of dopants (to deduce number of electrons or holes in material, need to know statistical mechanics of dopants and band levels)
c. Operation of semiconductor lasers and light-emitting diodes (totally dependent on precise form of thermal distributions of electrons and holes in semiconductor bands)
d. General Motivation, to understand some basic concepts like temperature, entropy, second law of thermodynamics that influence and limit all physical devices

We will introduce concepts of temperature, and entropy in their modern perspective that a system is to be found in the statistically most likely condition (not from the thermodynamic perspective of heat engines, deduced before the statistical nature of entropy was understood). Then we will extend and use these concepts to look at important classes of systems such as ideal “gasses”, including: 1. Fermi gas (represents fermions such as electrons and holes), 2. Bose gas (represents bosons such as photons and phonons). Afterward we will apply these concepts to understand key concepts in semiconductor statistics

Basic idea for statistical physics:
For a system with a large number of particles there are:

- Macroscopic properties (Total energy, temperature, volume, magnetization)
- Microscopic descriptions (Quantum states of many-particle system, or, equivalently, the occupation of single particle orbitals)

A given set of macroscopic properties can correspond to a large number of microscopic states. The basic question of statistical physics is: how can we deduce macroscopic properties of a given material system from its microscopic description? It turns out that the answer can be found in counting the number of microscopic states that correspond to a given set of macroscopic properties.

Binary Model Systems:
A simple example of system that can have large number of “particles” is a system that in it each “particle” only allowed being in one of two states: “up” or “down”, or equivalently, “1” or “0”, “full” or “empty.” This binary model is typically choose to analyze “spin” system, magnets that are either (spin) “up” or (spin) “down”.

Microscopic descriptions: spin orientation on each lattice site.

Macroscopic descriptions: total magnetization.

e.g., consider 10 such particles, This figure is one particular possible state of those 10 particles

\[ T_1 T_2 T_3 T_4 T_5 T_6 T_7 T_8 T_9 T_{10} \]

We could generate all possible states of the system of N such binary particles by “multiplying” out the expression

\[ (1 + \tilde{I}_1)(1 + \tilde{I}_2)(1 + \tilde{I}_3) \ldots (1 + \tilde{I}_N) \]

\[ (*) \]

This product of terms is called a “generating function” for the system. The “multiplication rule” is as

\[ (I_1 + \tilde{I}_1)(I_2 + \tilde{I}_2) = \tilde{T}_1 T_2 + \tilde{T}_2 T_1 + \tilde{T}_1 \tilde{T}_2 + \tilde{T}_2 \tilde{T}_1 \]

The above statement reads in words: The possible states of a system [with a particle 1 that can be ("particle 1 up" OR "particle 1 down") AND (with a particle 2 that can be) ("particle 2 up" OR "particle 2 down") ARE ("particle 1 up" AND "particle 2 up") OR ("particle 1 up" AND "particle 2 down") OR ("particle 1 down" AND "particle 2 up") OR ("particle 1 down" AND "particle 2 down")]

We can use mathematical expressions with multiplications and additions since they obey the same algebra as the logical expressions. This is a neat algebraic trick that enables us to evaluate the various possible states of multiple particle systems

Consider we apply this to find total magnetization of binary magnet system. Assigning a magnet a magnetic moment of +1 if it is “up” and −1 if it is “down” gives a total magnetic moment M that is the sum of the individual magnetic moments. For N magnets, the possible values of M are

\[ M = N, N - 2, N - 4, \ldots, -N + 2, -N \]

In a magnetic field, the energy of a given state of a system will depend on this total magnetic moment. We would like to know the multiplicities associated with each of these different magnetic moment values, i.e., how many states will there be of a given energy in the presence of a magnetic field E, i.e.,

there is only one state with M = N, i.e., TT TT \ldots TT TT

but there are N states with M = N – 2, i.e., TT TT \ldots TT TT TT TT \ldots TT TT TT TT, etc.

Enumeration of States: For simplicity in subsequent algebra, we introduce notation, \( N_s \) as number of magnets up, and \( N_{\bar{s}} \) as number of magnets down. Assume total number of magnets \( N \) is even for simplicity. Define “spin excess”, \( s \), as

\[ N_s = N - s \]

\[ N_{\bar{s}} = \frac{N}{2} + s \]

When considering total magnetizations or total energies, we don't care which magnets are which. Though we do still need to count all of the possible states to get the multiplicities correctly so we can drop
the site labels (subscripts) as long as we still count all the possible states. The “generating function” can be written as:

\[(1 + x^N)^N\]

Like for \( N = 2 \) we have:

\[(1 + x^2)^2 = 1 + 2 \cdot 1 + 4 x^2 \]

This correctly counts the number of states with one magnet up and one magnet down (2), though no longer cares about which magnet is which. We will refer to the set of states with the same total magnetization as a “class” of states.

For arbitrary \( N \), we know the binomial expansion

\[(x + y)^N = \sum_{t=0}^{N} \binom{N}{t} x^{N-t} y^t\]

Or by changing notation

\[(x + y)^N = \sum_{s=-N/2}^{N/2} \frac{N!}{(N/2 + s)! (N/2 - s)!} x^{N/2 + s} y^{N/2 - s}\]

Or we can rewrite \( (1 + x^N)^N \) as:

\[(1 + x^N)^N = \sum_{s=-N/2}^{N/2} \frac{N!}{(N/2 + s)! (N/2 - s)!} x^{N/2 + s} y^{N/2 - s}\]

The coefficient

\[g(N, s) = \frac{N!}{(N/2 + s)! (N/2 - s)!} = \frac{N!}{2^{N/2} \Gamma(N/2 + 1) \Gamma(N/2 - 1)!}\]

is the number of states having \( N_u \) magnets up and \( N_d \) down, or equivalently, it is the multiplicity of the class of states with \( N_u \) magnets up and \( N_d \) down, we can call \( g(N, s) \) the multiplicity function.

![Figure 1: Multiplicity function for 10 magnets](image)

Usually, systems of interest to us will have many elements. What happens to multiplicity functions for simple binary systems as we go to very large numbers of elements? Factorials are difficult to work out for large numbers. Solution is to use Stirling’s approximation for factorials when \( N \) is large. For large numbers, more convenient to work with logarithms (entropy defined that way for example). Stirling’s approximation in logarithmic form

\[
\log N! \approx \frac{1}{2} \log 2 \pi + \left( N + \frac{1}{2} \right) \log N - N
\]

Note: we will use a notation that “log” refers to the natural logarithm (bas e).

Gaussian Approximation to Binomial Coefficients: it can be shown that using Stirling’s approximation, for our binary model system with magnets with two possible states

\[g(N, s) \approx g(N, 0) e^{-s^2/N}\]

where

\[g(N, 0) = \left( \frac{2}{\pi N} \right)^{1/2} 2^{N}\]

Note that the Gaussian function drops to 1/e of its peak value when

\[s/N = (1/2N)^{1/2}\]

Quantity \( (1/2N)^{1/2} \) is a reasonable measure of the fractional width of the distribution. For large \( N \) this fractional width gets very narrow. (e.g., fractional width is \( \approx 10^{-11} \) for \( N = 10^{22} \).

We could say: nearly all states of this binary model system have substantially equal numbers of magnets up and down or the multiplicity function is very sharp for large numbers.

To analyze any system statistically, we need to be able to count states. It is particularly important for thermal physics to know number of states of a system at a given energy (or within some range of energy about the energy of interest). For example:

- density of states in semiconductor bands gives number of states for a single electron in the vicinity of some energy of interest.
- states of one electron in a hydrogen atom can have “degeneracy,” means several quantum states with identical energy. For example of what is known as “multiplicity” in thermal physics, number of different states with the same energy. Coupled states of two electron atom (Lithium). This is more complicated than hydrogen (simple example of multi-particle system).

We will have to be able to deal with system that could have very large numbers of particles, yet still be able to count the states at or near to some energy.

**Average Values:**

Average value of function \( f(x) \) when probability is \( P(s) \) of the value of \( x \) being \( s \), is

\[
\bar{f} = \sum_s f(s) P(s)
\]

where by definition for a probability

\[\sum_s P(s) = 1\]

Consider, for example, the system of \( N \) binary magnets, for the moment with no field applied. Then all the states have the same energy, and all are equally accessible, so each state of the system should have exactly the same probability. The total number of possible states of the magnets is \( 2^N \) (each magnet can be either up or down, so there are two ways of choosing the state of each of the \( N \) magnets, and hence \( 2^N \) states altogether). So, the probability of a given value of \( s \) for the case of \( N \) magnets is

\[P(N, s) = g(N, s)/2^N\]
Suppose now we want to know the mean square deviation of \( s \) from zero. Then formally, our function is \( f(s) = s^2 \), so we have

\[
\langle s^2 \rangle = \frac{(2\pi N)^{3/2} 2^{N}}{2N} \int_{-\infty}^{\infty} dx x^2 e^{-x^2 / N} = \frac{2(2\pi N)^{3/2} N^{1/2}}{2N} \int_{-\infty}^{\infty} dx x^2 e^{-x^2}
\]

\[
\langle s^2 \rangle = N \quad ; 
\langle (2s)^2 \rangle = N
\]

Quantity \( \langle (2s)^2 \rangle \) is the mean square spin excess. Root mean square spin exce

\[
\langle (2s)^2 \rangle^{1/2} = \sqrt{N}
\]

The fractional fluctuation in \( 2s \) is

\[
F \equiv \frac{\langle (2s)^2 \rangle^{1/2}}{N} = \frac{1}{\sqrt{N}}
\]

Hence, if we have \( 10^{28} \) particles, the fluctuation \( F \) is \( \sim 10^{-16} \).

**Fundamental Assumption:**

The fundamental assumption of thermal physics is: a closed system is equally likely to be in any of the quantum states accessible to it. A closed system will have constant energy, constant number of particles, constant volume, constant values of all external parameters that may influence the system, including gravitational, electric, and magnetic fields. A quantum state is accessible if its properties are compatible with the physical specification of the system (e.g., energy range, number of particles). The macroscopic property of the system is determined by the most probable configuration.

Since all accessible quantum states are equally likely, the probability of being in some state \( q \) that is one of \( g \) accessible states is

\[
P(q) = \frac{1}{g}
\]

If some physical quantity has a value \( X(q) \) when it is in the state \( q \), then the so-called ensemble average value is

\[
\langle X \rangle = \sum_{q} X(q) P(q)
\]

which in this case reduces to

\[
\langle X \rangle = \frac{1}{g} \sum_{q} X(q)
\]

A given system may well be in a state \( q \), so for that system the value of our physical quantity will be \( X(q) \), but we generally do not know in advance which of the various accessible states the system will be in. We need a concept that tells us what "on the average" will be the value of \( X \), so we imagine a set (an ensemble) of \( g \) identical and separate replicas of the system, each of which is in a different one of the (equally accessible) states. This is one example of what is called an ensemble average.

**Thermal Contact:**

Bringing two systems together so they can freely exchange energy

---

**Most Probable Configuration:**

Consider, for example, two spin systems (binary magnet systems), one with spin excess \( 2s_1 \), the second with spin excess \( 2s_2 \). A configuration of this system is the set of all states with particular values of \( s_1 \) and \( s_2 \). The multiplicity of this configuration is the product of the multiplicities of the individual systems

\[
g_1(N_1, s_1) \cdot g_2(N_2, s - s_1)
\]

where \( s = s_1 + s_2 \). Since for any particular state of system 1, system 2 can be in any of the \( g_2(N_2, s - s_1) \) states, the total multiplicity of the combined system with \( N = N_1 + N_2 \) particles (but with \( N_1 \) and \( N_2 \) fixed because of the impenetrable wall between the systems) is

\[
g(N, s) = \sum_{s_1} g_1(N_1, s_1) \cdot g_2(N_2, s - s_1)
\]

There will be some value of \( s_2 \), which we will call \( s_1^* \), that corresponds to the configuration with the highest multiplicity, i.e., the most probable configuration for which \( g_1(N_1, s_1^*) \cdot g_2(N_2, s - s_1) \) is the largest. We know that, for a large system, the configurations for which \( s_1 \) is close to \( s_1^* \) totally dominate the number of accessible states. Nearly all of the accessible states of the system correspond to situations in which \( s_1 \) is close to \( s_1^* \), hence the properties of the system tend in practice to be the properties of those configurations with \( s_1 \) close to \( s_1^* \).

We expect to find the system in the most probable configuration (or those with very similar \( s_1^* \) after we have given sufficient time for energy to exchange back and forth between the systems. This concept of allowing the system sufficient time for energy exchange is the concept we know as thermal equilibrium. The configuration the system has in thermal equilibrium is the most probable configuration or one very similar to it.

**Example - Two Spin Systems in Thermal Contact:**

We presume we have two, initially separate, systems, one with spin excess \( 2s_1 \), the other with spin excess \( 2s_2 \). We will presume we are applying a constant magnetic field, \( B \), to these two systems. The energy of a system with spin excess \( 2s_0 \) is (for a field parallel to the magnet directions)

\[
U(s_0) = -2mBs_0
\]

where \( m \) is the "magnetic moment" of one magnet (the minus sign comes from the sign convention for magnetic moments and field directions). The total energy of the system is
Chapter 8 – Statistical/Thermal Physics 1

**Introduction to Solid State Physics**

\[ U(s) = U_1(s_1) + U_2(s_2) = -2mB(s_1 + s_2) = -2mBs \]

In our overall system, when we bring the two systems into thermal contact, we will conserve energy overall, so the total spin excess \( 2s \) will be conserved (in our simple system, we are allowing no other “degrees of freedom” (e.g., kinetic energy) in the problem that could have energy)

Now we want to understand for our given set of conditions:

i. given total energy \( U \) (and hence given total spin excess \( 2s \))
ii. given numbers of magnets \( N_1 \) and \( N_2 \) (these remain fixed because we do not allow magnets themselves to pass between the two systems)

what is the most likely configuration if we allow energy to pass between the systems. i.e., if we allow spin excess to pass between the systems (if we allow \( s_1 \) to change while keeping \( s \) constant, or equivalently if we allowed \( s_2 \) to change while keeping \( s \) constant)

If we count the possible states for the two spin systems considered together, the multiplicity of a given configuration is explicitly

\[ g_1(N_1, s_1) \cdot g_2(N_2, s_2) = g_1(0)g_2(0)e^{\frac{-2s^2}{N_1} - \frac{2s^2}{N_2}} \]

Or in terms of \( s = s_1 + s_2 \)

\[ g_1(N_1, s_1) \cdot g_2(N_2, s - s_1) = g_1(0)g_2(0)e^{\frac{-2s^2}{N_1} - \frac{2(s-s_1)^2}{N_2}} \]

It will be more convenient to work with the log of the multiplicity – this makes no real difference since the number is maximized so also is its logarithm. Hence we have

\[ \log(g_1(N_1, s_1)g_2(N_2, s - s_1)) = \log(g_1(0)g_2(0)) - \frac{2s^2}{N_1} - \frac{2(s-s_1)^2}{N_2} \]

Differentiating to find the maximum, we have at a maximum

\[ \frac{d}{ds}(\log(g_1(N_1, s_1)g_2(N_2, s - s_1))) = \frac{-4s}{N_1} + \frac{4(s-s_1)}{N_2} \]

(As a check, we can evaluate the second derivative, which is \(-4(1/N_1 + 1/N_2)\) which is negative, confirming that we have a maximum rather than a minimum)

Hence we find that the multiplicity is maximum for the condition

\[ \frac{s_1}{N_1} = \frac{s - s_1}{N_2} \]

So

\[ \frac{s_1}{N_1} = \frac{s_2}{N_2} = \frac{s}{N} \]

In our constant magnetic field, the energy is simply proportional to the value of \( s \) for the system in question, so these ratios are essentially the average energy per magnet. We are finding that in the configuration with the largest multiplicity (and hence the most likely single configuration), the average energy per magnet is the same in each part of the system, and for the system overall, this is hinting at a larger truth, to which we will return below

With \( \frac{s_1}{N_1} \) and \( \frac{s_2}{N_2} \) denoting the values of \( s_1 \) and \( s_2 \) at the maximum multiplicity (i.e., in the most probably configuration), we can rewrite

\[ \frac{s_1}{N_1} = \frac{s_2}{N_2} = \frac{s}{N} \]

Hence at our most probably configuration

\[ (g_1g_2)_{\text{max}} = g_1(s_1)g_2(s-s_1) = g_1(0)g_2(0)e^{-2s^2/N} \]

Now let us investigate how sharp this maximum is. Consider a small deviation \( \delta \) from the maximum, so that \( s_1 = s_1 + \delta; s_2 = s_2 - \delta \)

hence,

\[ g_1(N_1, s_1 + \delta)g_2(N_2, s_2 - \delta) = g_1(0)g_2(0)e^{\frac{-2s^2}{N_1} + \frac{4\delta^2}{N_1} + \frac{4\delta^2}{N_2} - \frac{28^2}{N_1} - \frac{28^2}{N_2}} \]

If either of \( N_1 \) or \( N_2 \) is large, this is a very sharp distribution e.g., for \( N_1 = N_2 = 10^{12} \) and \( \delta = 10^{-12} \), i.e., \( \delta/N_1 = 10^{-10} \), then \( 28^2/N_1 = 200 \), and the product \( g_1g_2 \) is reduced to \( e^{-440} = 10^{-174} \) of its maximum value! Note, however, we can still see substantial relative fluctuations in very small systems, even when thermally connected to very large systems (important for reliability of decisions for very small transistors)

**Thermal Equilibrium:**

Let us generalize to any two systems in thermal contact, with constant total energy \( U = U_1 + U_2 \). The multiplicity of the combined system will be

\[ g(N, U) = \sum g_1(N_1, U_1, g_2(N_2, U - U_1) \]

where we are summing over all configurations for which \( U_1 < U \) and \( g_1 \) and \( g_2 \) are the multiplicities of the individual systems with numbers of “particles” \( N_1 \) and \( N_2 \) respectively. In general, if we make small (actually infinitesimal) changes in quantities such as \( U_1 \) and \( U_2 \), we will make small changes in the quantity \( g \) that depends on \( U_1 \) and \( U_2 \).

This small change is called the differential, and can be written for this specific case as

\[ dg = \left( \frac{\partial g_1}{\partial U_1} \right)_{N_1} g_2 dU_1 + \left( \frac{\partial g_2}{\partial U_2} \right)_{N_2} dU_2 \]

Remember that the notation \( \left( \frac{\partial g_1}{\partial U_1} \right)_{N_1} \) means the derivative with respect to \( U_1 \) with \( N_1 \) held constant. In our particular situation here, where total energy is conserved, we know that \( dU_1 = -dU_2 \). If the system is to be in the state with the largest multiplicity, then \( g(N, U) \) should be at a maximum as far as the choice of \( U_1 \) is concerned. Therefore, if we were to make an infinitesimal small change in \( U_1 \), there should be no change in \( g(N, U) \)

Hence we can write that the differential of \( g \) should be zero

\[ dg = \left( \frac{\partial g_1}{\partial U_1} \right)_{N_1} g_2 dU_1 + \left( \frac{\partial g_2}{\partial U_2} \right)_{N_2} dU_1 \]

From which we can conclude

\[ \frac{1}{g_1} \left( \frac{\partial g_1}{\partial U_1} \right)_{N_1} = \frac{1}{g_2} \left( \frac{\partial g_2}{\partial U_2} \right)_{N_2} \]

equivalently

\[ \left( \frac{\partial \log g_1}{\partial U_1} \right)_{N_1} = \left( \frac{\partial \log g_2}{\partial U_2} \right)_{N_2} \]
Now let us define a quantity, which we will call the "entropy" as \( s(N, U) = \log g(N, U) \). Then we have, for the configuration with the largest multiplicity (the most probably configuration)

\[
\frac{\partial s}{\partial U}\bigg|_{N_1} = \frac{\partial s}{\partial U}\bigg|_{N_2}
\]

This is the condition for thermal equilibrium for two systems in thermal contact. We could restate this as saying "the rate of change of entropy with energy is the same for all systems in thermal equilibrium with each other" (at least for the case here of conserved numbers of particles in each system).

Temperature:

We are used to the idea that when systems in thermal equilibrium with one another that the temperature is the same, \( T_1 = T_2 \). Hence we can now identify the quantities in previous Eq with temperature. To accord with our conventional understanding of temperature, we use the reciprocal of these quantities, and also use a constant to get the temperature into the unit with which we are familiar, i.e.,

\[
\frac{1}{T} = k_B \left( \frac{\partial s}{\partial U}\right)_N
\]

where \( k_B = 1.381 \times 10^{-23} \) joules/Kelvin, is the Boltzmann constant. The Boltzmann constant is only there because of our system of units. It is more convenient to work with "fundamental temperature", \( \tau \), which we can define as

\[
\frac{1}{\tau} = \left( \frac{\partial s}{\partial U}\right)_N
\]

or \( \tau = k_B T \). Note that the fundamental temperature has dimensions of energy, which is the real unit of temperature—other units (Kelvin, Celsius, Fahrenheit, Rankine) are technically redundant even if practically convenient (300K is \( \approx 4.14 \times 10^{13} \) joules).

In classical thermodynamics

\[
\frac{1}{\tau} = \left( \frac{\partial S}{\partial U}\right)_N
\]

where the (thermodynamic) entropy \( S \) corresponds with our statistical "fundamental" entropy through

\[
S = k_B \sigma
\]

Again, units for entropy are technically unnecessary since it is really a pure number, being the logarithm of a number.

Example of bringing simple systems into thermal contact:

Systems \( S_1 \) and \( S_2 \) each have 2 magnets, both in same magnetic field. Magnet "up" has energy +1, magnet "down" has energy -1. Initial conditions: \( S_1 \) has energy \( U_1 = 2 \), \( S_2 \) has energy \( U_2 = -2 \).

Entropy and heat flow:

Suppose we allow a small amount of energy \( dU \) to flow from body 1 to body 2, then the change in entropy overall will be

\[
\Delta s = \left( \frac{\partial s}{\partial U}\right)_{N_1}(-dU) + \left( \frac{\partial s}{\partial U}\right)_{N_2}(dU) = \Delta U \left( \frac{1}{\tau_1} - \frac{1}{\tau_2} \right)
\]

hence entropy increases if energy ("heat") flows from a hot body to a colder one.

Numerical example of heat flow:

Consider an object at 350 K, placed in thermal contact object at 290 K. What is the change in entropy when 0.1 J of energy has been transferred? (Presume large objects (large thermal mass) so that there has been negligible change in temperature during the transfer.)

\[
\Delta s_1 = \frac{-0.1}{350} K = -2.86 \times 10^{-4} \text{ JK}^{-1}
\]

\[
\Delta s_2 = \frac{0.1}{290} K = 3.45 \times 10^{-4} \text{ JK}^{-1}
\]

\[
\Delta s = \Delta s_1 + \Delta s_2 = 0.59 \times 10^{-4} \text{ JK}^{-1}
\]

Entropy has increased in flowing heat from a hot body to a colder one. In fundamental entropy "units", the change in entropy is

\[
\Delta \sigma = 0.59 \times 10^{-4}
\]

i.e., the number of states accessible to the system has increased by a factor \( e^{0.59 \times 10^{-4}} \)

Law of increase of entropy:

Proof that: total entropy always increases when two systems are brought into thermal contact. If total energy \( U = U_i + U_j \) is a constant, the total multiplicity (i.e., the total number of accessible states) after the systems are in thermal contact, is

\[
g(U) = \sum_{u_i} g_1(u_i) g_2(u - u_i)
\]
Note we sum over all possible values of energy of first system (or equivalently, all possible values of energy of second system, or equivalently all possible ways in which energy can be divided between the two systems). The sum, contains the original multiplicity $g_1(U_{10})g_2(U - U_{10})$ of the total system before the two parts were brought into thermal contact (where $U_{10}$ is the initial energy of system 1); but the sum also contains various other terms, all of which are positive. Hence $g(U) > g_1(U_{10})g_2(U - U_{10})$

Therefore the total entropy has increased.

Figure 3: Ways of increasing entropy

Relation between entropy and information:

In information theory, the "entropy" of a random variable is defined as

$$H(X) = - \sum_i p_i \log p_i$$

where the sum is over all possible values of the random variable $X$. For example, could be one alphabet character that could take on any of the 26 values. Information theory has to deal with the possibility that the different "states" (values of $X$) could have different probabilities ("e" might be more likely than "q").

In a physical system, all the accessible states are equally likely. The total number of accessible states is $g = \exp(\sigma)$ and so, for any of the physical states $i$

$$p_i = \frac{1}{g} = e^{-\sigma}$$

and so, using the information theory definition of entropy we have

$$H(X) = - \sum_i e^{-\sigma} \log e^{-\sigma} = \sigma \sum_i e^{-\sigma} = \sigma$$

Hence the information theory definition of entropy is consistent with the thermal definition.

"Units" of information:

The only practical difference between information entropy and thermal entropy is that information is usually measured in "bits", which is based on taking logarithms to the base 2, whereas in thermal entropy logarithms are usually taken to the base $e$. This makes no fundamental difference. We could choose to measure thermal entropy in "bits" as well.

If we use logarithms to the base $e$ to measure information, then the information is being measured in "nits".

\[ \text{Init} = \log_2 e \text{ bits} \approx 1.44 \text{ bits} \]

In information theory, the most efficient way (i.e., fewest sent characters) to send is to code the information being sent so that all characters (i.e., all states) are equally likely. The information entropy is essentially the average number of bits that need to be sent to send a "character" given the best possible coding. If a (thermal or information) system has entropy $\sigma$, I need to give $\sigma$ bits $(\log_2 e) \sigma$ bits to specify which state the system is in.

Laws of thermodynamics:

Traditionally, thermodynamics is based on four postulates (though with the statistical treatment given here essentially contains them all).

**Zeroth law:**

If two systems are in thermal equilibrium with a third system, they must be in thermal equilibrium with each other. This follows automatically from our statistical definition of the most likely condition after thermal contact. If

$$\left( \frac{\partial \log g_1}{\partial U_1} \right)_{N_1} = \left( \frac{\partial \log g_2}{\partial U_2} \right)_{N_2}$$

and

$$\left( \frac{\partial \log g_3}{\partial U_3} \right)_{N_3} = \left( \frac{\partial \log g_4}{\partial U_4} \right)_{N_4}$$

Then

$$\left( \frac{\partial \log g_1}{\partial N_1} \right)_{U_1} = \left( \frac{\partial \log g_2}{\partial N_2} \right)_{U_2}$$

Or if $T_1 = T_2$ and $T_3 = T_4$ then $T_1 = T_3$.

**First law:**

Common form: Heat is a form of energy (and energy overall is conserved)

**Second law:**

Statistical form: entropy is increased when a constraint internal to a closed system is removed

Common form: if a closed system is in a configuration that is not the equilibrium configuration, the most probable consequence will be that the entropy of the system will increase monotonically in successive instants of time

**Thermodynamic (Kelvin-Planck) form:** It is impossible for any cyclic process to occur whose sole effect is the extraction of heat from a reservoir and the performance of an equivalent amount of work

**Clausius form:** A transformation whose only result is to transfer heat from a body at a given temperature to a body at a higher temperature is impossible

**Third law:**

Common form: The entropy of a system approaches a constant value as the temperature approaches zero

Nernst form: At absolute zero, the entropy difference disappears between all those configurations of a system which are in internal thermal equilibrium

Principle of unattainability of absolute zero: It is impossible to accomplish a process as a result of which the temperature of a body is reduced to $T = 0 \ K$

The third law is obvious from the statistical approach as long as the system has a definite multiplicity in its ground state (and hence a definite entropy)

Calculating properties as a function of temperature:
Boltzmann factor: 
We want to know the probability that the system $S$ is in some particular quantum state given that we have a particular temperature $\tau$ for the total system. Consider two possible states of the system $S$, state 1 and state 2, with associated energies $E_1$ and $E_2$. The probability that the system $S$ is in state 1 is proportional to the multiplicity of the reservoir $R$ when it has energy $U_R = E_1$. This multiplicity is simply the number of ways the total system can exist in which system $S$ is in state 1. So, with a similar argument for state 2, we have:

$$P_1 = \text{Multiplicity of } R \text{ at energy } U_R = E_1$$
$$P_2 = \text{Multiplicity of } R \text{ at energy } U_R = E_2$$

where $P_1$ (or $P_2$) is the probability of $S$ being in state 1 (or state 2).

Note: We are not calculating the probabilities that the system $S$ has energy $E_1$ or energy $E_2$, we are calculating probabilities of the system being in a specific state of energy $E_1$ or energy $E_2$.

Now we know that

$$\text{Multiplicity of } R \text{ at energy } U_R = E_1 \Rightarrow e^{\sigma(U_R - E_1)}$$

and so we have

$$P_1 = e^{\sigma(U_R - E_1)}$$
$$P_2 = e^{\sigma(U_R - E_2)}$$

By expanding $e^{\sigma(U_R - E)}$ about the point $U_R$

$$e^{\sigma(U_R - E)} = e^{\sigma(U_R)} - \sigma(U_R - E)$$

A term of the form $e^{-\sigma E/\tau}$ is called a Boltzmann factor. It tells us the relative occupation probability of two quantum states separated by energy $\varepsilon$ in thermal equilibrium.

Partition function:
The Boltzmann factor also can be viewed as telling us the probability of occupation of a state relative to the occupation of some, possibly hypothetical, state at energy 0, in which case, we can view the total system $R + S$ as a closed system, total energy $U_0 = U_R + U_S$ is constant, e.g., if system $S$ has energy $\varepsilon$, then the reservoir has energy $U_R = \varepsilon - \varepsilon$. 

Boltzmann factor as an “unnormalized” occupation probability. We can normalize the probabilities by dividing the sum of all such “unnormalized” probabilities. That sum is called the “partition function”

$$Z(\varepsilon) = \sum_{\varepsilon_q} e^{-\varepsilon_q / \tau}$$

where the sum is over all possible states $q$ of the system $S$.

Hence the (normalized) probability of finding the system in a particular state $q$ of energy $\varepsilon_q$ is

$$P_q(\varepsilon_q) = \frac{e^{-\varepsilon_q / \tau}}{Z}$$

Note that the partition function has some useful properties. For example, the ensemble average energy, denoted by $\langle \varepsilon \rangle$, of the system is

$$\langle \varepsilon \rangle = \sum_{\varepsilon_q} \varepsilon_q P_q(\varepsilon_q) = \frac{1}{Z} \sum_{\varepsilon_q} \varepsilon_q e^{-\varepsilon_q / \tau} = \tau \frac{d}{d \tau} \ln Z$$

It is a useful expression for solving many thermodynamic problems, below we will see an example.

Example: Calculating average energy in a two-state system
We treat a system of one particle with two states, one of energy 0 and one of energy $E$, the particle is in thermal contact with a reservoir at temperature $\tau$, we want to find the energy of the system as a function of $\tau$. The partition function for the two-state system of the particle is:

$$Z = e^{-E/\tau} + e^{-E/\tau} = 1 + e^{-E/\tau}$$

The average energy is:

$$\langle \varepsilon \rangle = \frac{1}{Z} \sum_{\varepsilon_q} \varepsilon_q P_q(\varepsilon_q) = \tau \frac{d}{d \tau} \ln Z$$

Helmholtz free energy:
Prior to considering systems in thermal contact with a reservoir, we have been dealing with closed systems that have fixed energy, $U$, and particle number, $N$. (We have also implicitly assumed all other parameters, such as volume and magnetic field are fixed.) Under these conditions, entropy was the quantity that was an extremum (a maximum in this case) in thermal equilibrium. In considering systems in contact with a thermal reservoir, instead of fixing the energy of the system of interest, we will be fixing the temperature. For the moment, we will still consider the total particle number to be fixed, as well as other parameters such as total volume and magnetic field.)

What quantity now will be an extremum in thermal equilibrium?
Answer: the Helmholtz free energy $F = U - TS$
It is at a minimum in equilibrium at constant temperature $\tau$ and volume $V$.

Proof that Helmholtz energy is minimized at constant $\tau, V, N$:
Suppose we consider a small transfer of energy from the reservoir $R$ to the system $S$ at constant temperature and constant volume. The change in the Helmholtz free energy in the system $S$ will be the differential

$$dF = dU_S - T dS_T$$

where we have used the constancy of the temperature to eliminate the term $\sigma \partial T$ from the differential. But

$$\frac{1}{T} = \frac{\delta S}{\delta U}$$
Under these conditions, showing that is an extremum. Further arguments can prove this extremum is a minimum rather than a maximum (though we omit these here). Hence we know that $F$ is at a minimum in a system at constant temperature and constant volume.

Minimum property of free energy of a paramagnetic system:

Consider the model system considered before, with $N_1$ magnets up and $N_2$ down, with $N = N_1 + N_2$, and spin excess $2s = N_1 - N_2$, and magnetic moment (per magnet) $\tau$. We could evaluate the entropy of this system to be (using Stirling's approximation, simplified somewhat further for very large $N$)

$$s(N) \approx -\left(\frac{N}{2} + s\right) \log\left(\frac{1 + s}{N}\right) - \left(\frac{N}{2} - s\right) \log\left(\frac{1 - s}{N}\right)$$

The energy of this system in a magnetic field $B$ is $U = -2s\mu B$, so the Helmholtz free energy is

$$F = U - TS = -2s\mu B + \frac{(N_1 + s)}{N} \log\left(\frac{1 + s}{N}\right) + \frac{(N_2 - s)}{N} \log\left(\frac{1 - s}{N}\right)$$

Now let us find the configuration (i.e., the value of $s$) for which $F$ is extremum (minimum). To do this, we differentiate w.r.t. $s$ while holding $\tau$ and $N$ constant, giving

$$\left(\frac{\partial F}{\partial s}\right)_{N, \tau} = 0 = -2\mu B + \tau \log\frac{N + 2s}{N - 2s}$$

Hence the thermal equilibrium value of $2s$, which we denote by ($2s_0$), is the value that satisfies the above equation, i.e.,

$$\frac{N + (2s_0)}{N - (2s_0)} = e^{2\mu B/\tau}$$

or, after some rearrangement

$$(2s_0) = N \left( e^{2\mu B/\tau} - 1 \right) = N \log \frac{mB}{\tau}$$

Hence the total magnetization and total energy (which are proportional to $2s_0$) follow a hyperbolic tangent function as a function of the ratio $mB/\tau$.

Relative importance of energy and entropy:

We have therefore solved the problem of the total magnetization of this system as a function of temperature. Note that, as temperature rises, it is more difficult to obtain nearly complete magnetization of the system in one direction or the other. At low temperatures, the minimization of the energy dominates in minimizing $F$. As temperature rises, the influence of entropy becomes more important in minimizing $F$ - the system no longer merely wants to remain in its lowest energy state since that has very low entropy.

Chemical Potential:

We have so far considered systems in which the number of particles is fixed within the system, but if we remove some barrier between two systems with different numbers of particles, we may expect particles to flow one way or another ("diffusion") even if the two systems are at the same temperature. Just as temperature governs the flow of energy from one system to another, chemical potential governs the flow of particles. In equilibrium, the (total) chemical potentials of the two systems will be equal.

Diffusive and thermal equilibrium:

In this problem, the total volume is fixed as is the temperature (and total number of particles), so the total Helmholtz free energy will be a minimum in equilibrium. Total Helmholtz free energy

$$F = F_1 + F_2 = U_1 + U_2 + \tau (Q_1 - Q_2)$$

for systems 1 and 2 together. Now suppose we consider infinitesimal changes in the numbers of particles in the systems. Then the change in $F$ is

$$dF = \left(\frac{\partial F_1}{\partial N_1}\right) dN_1 + \left(\frac{\partial F_2}{\partial N_2}\right) dN_2$$

But the total number of particles is conserved, i.e., $N = N_1 + N_2 = $ constant, so in diffusive equilibrium, losses of particles by one system corresponds to gains by the other, or $\delta N_1 = -\delta N_2$

Hence

$$dF = \left(\frac{\partial F_1}{\partial N_1}\right) dN_1 + \left(\frac{\partial F_2}{\partial N_2}\right) dN_2$$

so that at equilibrium

$$\left(\frac{\partial F_1}{\partial N_1}\right) = \left(\frac{\partial F_2}{\partial N_2}\right)$$

Hence we have found a quantity that must be the same in the two systems if they are in diffusive equilibrium at a given temperature (and a given volume). This quantity is the chemical potential $\mu$, i.e.,

$$\mu(T, V, N) \equiv \frac{\partial F}{\partial N}$$

so that, in diffusive equilibrium

$$\mu_1 = \mu_2$$

Direction of particle flow:

Rewriting Eq. (*), but considering a condition away from diffusive equilibrium so that $dF$ is not zero, we see that

$$dF = (\mu_1 - \mu_2) dN_1$$
Hence, if \( \mu_1 > \mu_2 \), a positive \( dN \) leads to an increase in \( F \). Hence, since \( F \) is a minimum at equilibrium, with \( \mu_1 > \mu_2 \), we have to see a decrease in \( N \) to approach equilibrium, i.e., particles flow from high chemical potential to low chemical potential.

Chemical potentials for several species:

If we have several species of particles, each species has its own chemical potential. For species \( j \),

\[
\mu_j = \frac{\partial F}{\partial N_j} \bigg|_{N_1, N_2, \ldots, N_{j-1}}
\]

where in the differentiation the numbers of all particles are held constant except for the species \( j \).
Chemical Potential and Entropy:

So far, we have defined chemical potential in terms of the change of Helmholtz free energy $F$ with respect to particle number $N$. There are other equivalent definitions possible, and we need one of these for a subsequent derivation, so we derive it here. The relation we will derive is

$$\mu(U, V, N) = -\left(\frac{\partial \sigma}{\partial N}\right)_{U, V}$$

Note the similarity of this relation to the definition of temperature as:

$$\frac{1}{T} = \left(\frac{\partial S}{\partial N}\right)_{U, V}$$

Just as we originally did for temperature, we are currently considering entropy as a function of $U$, $V$, and $N$. Hence we can write the differential

$$d\sigma = \frac{\partial \sigma}{\partial U} dU + \frac{\partial \sigma}{\partial V} dV + \left(\frac{\partial \sigma}{\partial N}\right)_{U, V} dN$$

which is the change in entropy for arbitrary changes $dU$, $dV$, and $dN$.

We are interested in the situation where the volume is held constant, so we set $dV = 0$.

Now we make the restriction that we will choose a combination of $dU$, $dV$, and $dN$ such that the temperature change $\delta T$ is zero, and we denote this restricted combination by $(\delta \sigma)_T$, $(\delta U)_T$, and $(\delta N)_T$.

Then $d\sigma = 0$ when

$$(\delta \sigma)_T = \left(\frac{\partial \sigma}{\partial U}\right)_T (\delta U)_T + \left(\frac{\partial \sigma}{\partial N}\right)_{T, U} (\delta N)_T$$

Dividing by $(\delta N)_T$,

$$\frac{(\delta \sigma)_T}{(\delta N)_T} = \left(\frac{\partial \sigma}{\partial U}\right)_T (\delta U)_T + \left(\frac{\partial \sigma}{\partial N}\right)_{T, U}$$

The ratio $(\delta \sigma)/\delta N$ is $(\delta \sigma)/\delta N$ at $(\delta U)/\delta N$ at $(\delta T)/\delta N$, all at constant volume. Hence, using $T = \left(\frac{\partial \sigma}{\partial U}\right)_T$, and $(\delta \sigma)/\delta N$ at $(\delta U)/\delta N$, we have

$$\left(\frac{\partial \sigma}{\partial N}\right)_{T, U} = \left(\frac{\partial U}{\partial N}\right)_{T, U} + \left(\frac{\partial \sigma}{\partial U}\right)_T$$

i.e.,

$$\left(\frac{\partial \sigma}{\partial N}\right)_{T, U} = -\left(\frac{\partial U}{\partial N}\right)_{T, U}$$

But we know from our original definition of chemical potential that

$$\mu = \left(\frac{\partial \sigma}{\partial N}\right)_{T, U}$$

Hence we have

$$\mu(U, V, N) = -\left(\frac{\partial \sigma}{\partial N}\right)_{U, V}$$

Chemical potential and potential energy change:

The two most important applications of the chemical potential for semiconductor devices are in understanding what happens to the "gas" of electrons and/or holes when we join dissimilar materials and when we apply voltages. In both cases, we have situations where one part of the structure has one value of energy for the lowest state, and the other part has another value of energy for the lowest state. This difference in energies for the lowest state in different parts of the whole structure is a potential energy difference (i.e., a difference in energy associated only with position).

If we change the overall energy of a system of $N$ particles, e.g., by lifting it up in a gravitational field, or changing the electrical potential seen by a set of charged particles, we change the chemical potential. Here we will calculate just what this change is. Suppose for example, we change the energy per particle by an amount $q\Delta V$ by applying change in voltage $\Delta V$ to a set of particles each of which has charge $q$ (e.g., we have the set of particles in an electric field, and we lift them "up."

The change in energy of the set of particles is therefore $\Delta U = Nq\Delta V$. Simply making this change in potential energy does not in itself affect the entropy or the temperature. For example, lifting a balloon up (slowly so it does not get heated by air friction!) will not affect its temperature, and the entropy of the gas molecules inside the balloon is not changed — there are still the same number of states available for the molecules inside the balloon. This change in potential energy does; however, change the Helmholtz free energy through the energy term $U$, i.e., the total change in free energy is

$$\Delta F = \Delta U = Nq\Delta V$$

(where we have noted that the entropy and temperature are not changed in this process). Before we made this change in Helmholtz free energy, we had an initial Helmholtz free energy for our system of $F_0$, so we had a chemical potential

$$\mu = \left(\frac{\partial F_0}{\partial N}\right)_{T, U}$$

After this change as in $\Delta F$, we have a new chemical potential

$$\mu = \left(\frac{\partial F}{\partial N}\right)_{T, U} = \left(\frac{\partial F_0}{\partial N}\right)_{T, U} + \left(\frac{\partial \Delta F}{\partial N}\right)_{T, U}$$

This has a very simple interpretation: changes in potential energy (per particle) add an equal amount to the chemical potential.

Note this means that the actual value of chemical potential depends on the energy origin chosen, but this does not matter in practice since we are only concerned about differences in chemical potential since it is those differences that drive movement of particles.

Note that it is still true that the chemical potential is constant throughout the system in thermal equilibrium, despite any changes in potential energy we may make in different parts of the system.

Internal and Total Chemical Potential:

Sometimes for convenience a distinction is made between:

1. "Internal" chemical potential, $\mu_{int}$: chemical potential neglecting externally imposed changes in the energy of all particles of a given species, e.g., neglecting gravitationally, electrostatically, or
magnetically induced energy changes, would include all effects from, e.g., temperature, volume, number of particles \( \mu \).

2. "external chemical potential", \( \mu_{\text{ext}} \): the contribution to chemical potential arising solely from externally imposed changes in the energy of all particles of a given species, e.g., all the potential energies (gravitational, electrostatic, magnetic).

With this approach, we have a total chemical potential, \( \mu_{\text{tot}} = \mu_{\text{int}} + \mu_{\text{ext}} \).

**The total chemical potential is the quantity that is the same throughout at system at thermal equilibrium.**

**Semiconductor Heterostructures and Chemical Potential:**

The most important situation for this course is that in which two different semiconductor materials are joined. The materials could be different in the chemical composition of the material, e.g., GaAs and Al,Ga,As or the doping type or concentration of the material (e.g., p-n junction).

Suppose we take two different, large pieces of semiconductor material and join them together. We do not at the moment know what happens in the region near to where they are joined, but we can presume that, far enough away from this junction, the materials will look the same as they did before.

![Figure 1: Chemical potentials and discontinuous joined materials](image)

In a p-type material (far away from the junction), the chemical potential is near the valence band edge, and in an n-type material (far away from the junction) it is near the conduction band edge (we will deduce this explicitly later). We have not yet deduced what happens near the junction, but we do know that the chemical potential must be constant throughout the system if it is in thermal equilibrium. This is the starting point for the further analysis of semiconductor diodes.

**Chemical potentials and applied voltage:**

Note we are also assuming for the moment that there is no applied voltage on the structure. If we apply a voltage, we will raise the chemical potential at one side compared to that at the other, which (certainly for forward bias) can cause flow of particles from one material to the other. (This is the normal forward current of a diode (so-called "diffusion current").) However, the flow of particles across the junction is (in a steady state situation) balanced by the flow of particles into or out of the materials through the electrical wire that connects them. Hence with an applied voltage the diode is not in thermal equilibrium, though it may be in a steady state situation with a steady current.

**Gibbs factor and Gibbs sum:**

To derive one of the most useful relations for semiconductor physics, the Fermi-Dirac distribution function, and another important relation for semiconductor physics and optics, the Bose-Einstein distribution, we need first to discuss the Gibbs factor and Gibbs sum. The Gibbs factor is the extension of the Boltzmann factor to the case where particles can also be exchanged with the reservoir. The Gibbs sum is the extension of the partition function to include states with different numbers of particles.

Note: The ensemble of systems (or set of states) that are counted in the partition function is sometimes called the canonical ensemble. The ensemble of systems (or set of states) that are counted in the Gibbs sum is sometimes called the grand canonical ensemble.

Consider now a system \( S \) in thermal and diffusive contact with a large reservoir \( R \), i.e., in addition to being able to transfer energy between the system and the reservoir, we can also now transfer particles. Now the entire closed system of reservoir and system \( S \) has \( N \) identical particles, and energy \( U \). When the system \( S \) has \( N \) particles, the reservoir has \( N' \) particles. When the system \( S \) has energy \( \varepsilon \), the reservoir has energy \( U_0 - \varepsilon \).

Just as for the derivation of the Boltzmann factor, we consider the system \( S \) to be in a particular state in which it has energy \( \varepsilon \) and \( N \) particles.

State 1 is a state in which system has energy \( \varepsilon \) and \( N_1 \) particles.

State 2 is a state in which system has energy \( \varepsilon_2 \) and \( N_2 \) particles.

The ratio of the probabilities of state 1 or state 2 is

\[
\frac{P(N_1, \varepsilon_1)}{P(N_2, \varepsilon_2)} = g(N_0 - N, U_0 - \varepsilon_1) g(N_0 - N, U_0 - \varepsilon_2)
\]

Now we know that, by definition, the multiplicity \( g \) is just the exponential of the entropy, \( \sigma \), i.e.,

\[
b \exp \left( \frac{\sigma(N_0 - N, U_0 - \varepsilon_1)}{N} \right) = e^{\sigma(N_0 - N, U_0 - \varepsilon_2)}
\]

so

\[
\frac{P(N_1, \varepsilon_1)}{P(N_2, \varepsilon_2)} = e^{\sigma(N_0 - N, U_0 - \varepsilon_1)} = e^{\sigma(N_0 - N, U_0 - \varepsilon_2)}
\]

where

\[
\Delta \sigma = \sigma(N_0 - N, U_0 - (\varepsilon_1)) - \sigma(N_0 - N, U_0 - (\varepsilon_2))
\]

Just as for the Boltzmann factor derivation, we expand \( \sigma(N_0 - N, U_0 - \varepsilon) \) in a Taylor series about the values \( N_0 \) and \( U_0 \), with the difference that this time we need to expand in two variables, \( N \) and \( \varepsilon \), to obtain

\[
\sigma(N_0 - N, U_0 - \varepsilon) = \sigma(N_0, U_0) - N \left( \frac{\partial \sigma}{\partial N} \right) \varepsilon_0 - \varepsilon \left( \frac{\partial \sigma}{\partial \varepsilon} \right)_N \varepsilon_0 + ...
\]

On the assumption that the reservoir is very large, we may neglect all higher order terms in the expansion, just as in the Boltzmann factor derivation, to obtain

\[
\Delta \sigma \approx -(N_1 - N_2) \left( \frac{\partial \sigma}{\partial N} \right) \varepsilon_0 - (\varepsilon_1 - \varepsilon_2) \left( \frac{\partial \sigma}{\partial \varepsilon} \right)_N \varepsilon_0
\]

and so, using the relations...
\[
\frac{1}{T} \equiv \left( \frac{\partial \sigma}{\partial H_0} \right)_{N_0} \quad \text{and} \quad -\frac{\mu}{T} \equiv \left( \frac{\partial \sigma}{\partial N_0} \right)_{H_0}
\]

we have

\[
\Delta \sigma = \frac{(N_1 - N_2)\mu - (\varepsilon_1 - \varepsilon_2)}{T}
\]

Hence

\[
P(N_1, \varepsilon_1) = \frac{e^{(N_1\mu - \varepsilon_1)/T}}{Z}
\]

A factor of the form \( e^{((N\mu - \varepsilon)/T)} \) is called a Gibbs factor, and represents an unnormalized probability that the system \( S \) can be in a state of energy \( \varepsilon \) and particle number \( N \).

To normalize the probabilities, we need to sum over all the possible unnormalized probabilities. This sum is now over all possible numbers of particles in \( S \), and all possible states of \( S \) for each possible number of particles, i.e., we construct the “Gibbs sum” (or grand partition function) \( Z \).

\[
Z(\mu, T) = \sum_{\mu=0}^{\infty} \sum_{N=0}^{\infty} \exp \left( \frac{N\mu - \varepsilon_\mu}{T} \right) = \sum_{\mu=0}^{\infty} \exp \left( \frac{N\mu - \varepsilon_\mu}{T} \right)
\]

where \( \sum_{\mu=0}^{\infty} \sum_{N=0}^{\infty} \exp \left( \frac{N\mu - \varepsilon_\mu}{T} \right) \) means the sum over all possible numbers of particles in \( S \), and all possible states of \( S \) for each possible number of particles. Note that the state with zero particles (\( N = 0 \)) is included in this sum. Hence, the absolute probability of finding the system \( S \) in a state with \( N_1 \) particles and energy \( \varepsilon_1 \) is

\[
P(N_1, \varepsilon_1) = \frac{1}{Z} e^{(N_1\mu - \varepsilon_1)/T}
\]

Fermi-Dirac Distribution Function:

We are interested in “fermions”— particles (technically those with half-integer spin) that obey the Pauli-exclusion principle (e.g., electrons). A particular quantum-mechanical state (e.g., a -state in a band in a semiconductor, or an orbital in an atom) can only be occupied by one fermion of a given species; by species here we mean a particle of a particular spin. We can have both a spin-up electron and a spin-down electron in the same “state”, the spin-up and spin-down electrons being different “species.” we might want to avoid confusion by pretending we had a slight magnetic field on, which would make spin-up states and spin-down states have slightly separate energy, so that there are more obviously different states.

Let us now imagine that we have a single quantum mechanical state that can either be full (\( N = 1 \)) or empty (\( N = 0 \)). If the state is full (\( N = 1 \)), it has an energy \( \varepsilon \), otherwise (\( N = 0 \)) it has energy zero. This is the situation in the simplest model of a semiconductor, where the state in question is a spin-up (or spin-down) k-state in a band. The Gibbs sum for this state is

\[
Z = \exp((0 \times \mu - 0 \times \varepsilon)/T) + \exp((1 \times \mu - 1 \times \varepsilon)/T) = 1 + e^{(\mu - \varepsilon)/\varepsilon}
\]

so we have for the probability that the quantum mechanical state in question is occupied

\[
P(1, \varepsilon) \equiv f(\varepsilon) = \frac{e^{(\mu - \varepsilon)/\varepsilon}}{1 + e^{(\mu - \varepsilon)/\varepsilon}}
\]

where we have introduced the more usual notation \( f(\varepsilon) \), i.e.,

\[
f(\varepsilon) = \frac{1}{1 + e^{(\mu - \varepsilon)/\varepsilon}}
\]

This is called the Fermi-Dirac distribution function, and gives the occupation probability of an electron level (of given spin) of energy \( \varepsilon \) for temperature \( T \) and chemical potential \( \mu \).

In a more common notation for semiconductor problems, this expression is usually written as

\[
f(\varepsilon) = \frac{1}{1 + e^{(\mu - \varepsilon)/kT}}
\]

where the term \( kT \) is known as the “Fermi energy”, and we have reverted to the conventional notation for temperature. The “Fermi energy” in this relation is rigorously identical to the chemical potential. We will use this definition of Fermi energy.

Note: The use of the term “Fermi energy” in this expression is somewhat unfortunate, since Fermi energy is also used in solid state physics to describe the level up to which electrons occupy all states in a metal at zero temperature. This gives the impression that the Fermi level is a constant for a given density of electrons. Though the chemical potential is the level up to which all electron states are occupied at zero temperature, as temperature is increased, even for a constant average number of particles, the chemical potential is dependent on temperature for anything other than a constant density of states, and hence the impression that the Fermi level is a constant is misleading and confusing here. The concept that we are actually dealing with a chemical potential, which may vary with temperature or other parameters, is always correct.
Fermi-Dirac distribution:

Electrons in thermal equilibrium have a Fermi-Dirac distribution, probability \( f_F(E, T) \) of electron in state of energy \( E \) at temperature \( T \) is

\[
    f_F(E, T) = \frac{1}{1 + \exp\left(\frac{E - \mu}{k_B T}\right)}
\]

\( \mu \) is the chemical potential, \( k_B \) is Boltzmann's constant, note \( k_B \approx 25 \text{ meV} \) at room temperature.

Chemical potential (Fermi energy) corresponds to energy for which \( f = \frac{1}{2} \). At zero temperature Fermi-Dirac distribution is step function, with all states up to energy \( E_F \) (Fermi energy) totally full.

Free electron gas at zero temperature:

The free electron gas model represents the starting point for describing the properties of metals. As in many solid state physics problems, the first step is to figure out what the ground state is in the system.

Consider a single electron moving in a box of volume \( V = L^3 \). The orbital of the free particle wavefunction

\[
    \psi(x, y, z) = A \sin \frac{n_x \pi x}{L} \sin \frac{n_y \pi y}{L} \sin \frac{n_z \pi z}{L}
\]

The energy values are

\[
    E_n = \frac{\hbar^2}{2M} \left( \frac{n_x^2 + n_y^2 + n_z^2}{L^2} \right)
\]

For an \( N \)-fermion system, the ground state should correspond to the scenario where each orbital is filled with two electrons (one spin up, one spin down) up the Fermi energy.

Heat capacity of degenerate electron gas:

We derive a quantitative expression for the heat capacity of a degenerate Fermi gas of electrons in three dimensions. This calculation gives excellent agreement with the experimental results for the heat capacity.
Introduction to Solid State Physics

Chapter 10 – Fermi-Dirac Distribution

capacity of metals. The increase in the total energy of a system of N electrons when heated from 0 to \( t \) is denoted by

\[
\Delta U = \int_0^\infty \! dx \left( f(x) \! e^x \! D(x) \right) = \int_0^{e_F} \! dx e^x D(x)
\]

Here \( f(x) \) is the Fermi-Dirac function, and \( D(x) \) is the density of states. Notice that:

\[
\left( \int_0^{e_F} + \int_{e_F}^\infty \right) \! dx e^x f(x) D(x) = \int_0^{e_F} \! dx e^x D(x)
\]

hence:

\[
\Delta U = \int_0^{e_F} \! dx (e - e_F) D(x) \left( f(x) + \int_0^{e_F} \! dx (e - e_F)(1 - f(x)) D(x) \right)
\]

The first integral on the right-hand side gives the energy needed to take electron from \( e_F \) to the orbitals of energy \( e > e_F \), and the second integral gives the energy needed to bring the electrons to \( e_F \) from orbitals below \( e_F \).

The heat capacity of the electron gas is found by differentiating \( \Delta U \) with respect to \( t \). The only temperature-dependent term in equation above is \( f(x) \). Therefore, we can group terms to obtain:

\[
C_{el} = \frac{d\Delta U}{dt} = \int_0^{e_F} \! dx \left( e - e_F \right) D(x) \frac{df}{dt}
\]

In a typical metal, the concentration of electron is such that \( e_F/k_B = 50000 \) K. Thus, the typical temperature of interest in metals is \( t/e_F < 0.01 \). Since \( f \) is like a step function, the derivative \( df/dt \) is large on at energies near \( e_F \). It is a good approximation to evaluate the density of orbitals \( D(x) \) at energy \( e_F \) and take it outside of the integral:

\[
C_{el} \approx D(e_F) \int_0^{e_F} \! dx (e - e_F) \frac{df}{dt}
\]

One could typically ignore the temperature dependence of the chemical potential in the Fermi-Dirac distribution and replace \( m \) by the constant \( e_F \). We have then:

\[
\frac{df}{dt} = \frac{e - e_F}{t^2} \cdot \frac{e^{(e-e_F)/t}}{1 + e^{(e-e_F)/t}}
\]

We set

\[
x \equiv \left( (e - e_F)/t \right)
\]

\[
C_{el} = tD(e_F) \int_0^{e_F} \! dx e^x \frac{e^x}{(1 + e^x)^2} = \int_0^{e_F} \! dx x^2 \frac{e^x}{(1 + e^x)^2}
\]

Therefore, in the region \( t \ll t_F \), we have

\[
C_{el} = \frac{1}{3} t^2 D(e_F)
\]

The key results are the linear dependence of the heat capacity on temperature, and the dependence of the heat capacity on the density of states at the Fermi energy. Both of these results are confirmed by experimental observations. The strong dependence of properties on the density of states at Fermi surface is in fact, a rather general result.

Electron distribution:

Electrons in thermal equilibrium have a Fermi-Dirac distribution, probability \( f_e(E, T) \) of electron in state of energy \( E \) at temperature \( T \) is

\[
f_e(E, T) = \frac{1}{1 + \exp \left( \frac{E - \mu}{k_B T} \right)}
\]

Hole distribution:

Holes correspond to the absence of an electron, hence the probability \( f_h \) of finding a hole in a given state is the probability that there is not an electron in that state, hence \( f_h = 1 - f_e \), and so

\[
f_h(E, T) = \frac{1}{1 + \exp \left( \frac{E - \mu}{k_B T} \right)}
\]

Hence the holes also have a Fermi-Dirac distribution, can look at holes this way, in which hole energies are increasingly negative as they acquire kinetic energy, or can look at hole energies "upside down", changing \( -(E - \mu) \) to \( (E - \mu) \).

Pure semiconductors, donors and acceptors:

Let \( n_d \) denote the concentration of conduction electrons, and \( n_a \) denote the concentration of holes. In a pure semiconductor the two will be equal: \( n_d = n_a \).

Most semiconductors as used in devices have been intentionally doped with impurities that may become thermally ionized in the semiconductor at room temperature. Impurities that give an electron to the crystal (and become positively charged in the process) are called donors. Impurities that accept the electron from the valence band (and become negatively charged in the process) are called acceptors.

Let \( n_d^0 \) be the concentration of positively charged donors and \( n_a^0 \) the concentration of negatively charged acceptors. The electrical neutrality condition becomes:

\[
n_d - n_a = \Delta n = n_d^0 - n_a^0
\]

This specifies the difference between electron and hole concentration.

Group V elements [e.g., P] are usually donors in silicon.

Group III elements [e.g., Al] are usually acceptors in silicon.

In III-V semiconductors, the donors are often Group VI materials [e.g., tellurium, tin], and acceptors often Group II materials [e.g., zinc, beryllium]. Often Group IV materials are used. Their behavior is less obvious. One way of rationalizing their behavior is to note that Group IV atoms might occupy either Group III or Group V sites in the lattice. In the first case they might be expected to behave more like donors, in the second case more like acceptors. For example, silicon is a commonly used n-type dopant in GaAs, carbon is sometimes used as a p-type dopant in GaAs, and germanium is often used as an n-type dopant in forming contacts to GaAs (e.g., by depositing Ge-doped gold metal and "alloying" it into the n-type materials being contacted). Under some circumstances, silicon can become a p-type dopant in GaAs.

n-typed and p-typed semiconductors:

A semiconductor in which the added number of donors dominates is called an "n-type" semiconductor, and "majority carrier" conduction takes place through electrons in the conduction band.

A semiconductor in which the added number of acceptors dominates is called a "p-type" semiconductor, and "majority carrier" conduction takes place through holes in the valence band.
Introduction to Solid State Physics  Chapter 10—Fermi-Dirac Distribution

Semiconductor statistical mechanics - classical limit:
For energies far above the chemical potential, the Fermi-Dirac distribution behaves like classical
Maxwell-Boltzmann distribution, \( f_{MB}(E,T) \), i.e., for electrons,

\[
f_{MB}(E,T) = A e^{-E/k_B T}
\]

where \( A = \exp(\mu/k_B T) \). This is to be expected, since at these high energies occupation probabilities
are small. Hence little influence of the Pauli exclusion restriction that no more than one electron may
occupy a state.

If \( f_e \ll 1 \) for all conduction states, and \( f_v \ll 1 \) for all valence states, the semiconductor is said to be
in a nondegenerate regime, i.e. classical regime. This corresponds to the case when the chemical
potential is deep inside the gap, far from both the conduction band edge and the valence band edge.
The total number of conduction electrons in the form:

\[
N_e = \sum_{E_B} e^{-E_e/k_B T} = \int_{E_B} e^{-E_e/k_B T} dE
\]

Or

\[
N_e = N_e e^{-(\mu-\phi)/k_B T}
\]

Here

\[
N_e = \int_{E_B}^{E_F} dV D(E) e^{-E/k_B T}
\]

Whereas we have discussed before

\[
D(E) = \frac{m^*_e \sqrt{2m^*_e(E-e_v)}}{n^2 h^3}
\]

And \( m^*_e \) is the density of state effective mass for the conduction band. Evaluating the integral we have

\[
N_e = 2 \frac{(m^*_e)^{3/2}}{(2\pi h^2)^3} V
\]

Defining the quantum concentration \( n_e \) for the conduction electrons as:

\[
n_e = \frac{N_e}{V} = 2 \frac{(m^*_e)^{3/2}}{(2\pi h^2)^3}
\]

The conduction electron concentration \( n_e = N_e/V \) becomes

\[
n_e = n_e e^{-(\mu-\phi)/k_B T}
\]

Similarly, for the concentration of holes in the valence band, we have

\[
n_h = n_e e^{-(\mu-\phi)/k_B T}
\]

where

\[
n_e = 2 \frac{(m^*_h)^{3/2}}{(2\pi h^2)^3}
\]

Law of Mass Action:
If we are dealing with semiconductors in the Maxwell-Boltzmann limit, i.e., any of the criteria

- large effective masses
- or low doping densities
- or very high temperatures

but not III-V semiconductors with large numbers of electrons, we have, \( n_e n_h = n_i^2 \) where

\[
n_i = (n_e n_h)^{1/2} e^{(\mu-\phi)/2k_B T}
\]

This equation is called the Law of Mass Action (the term comes from chemistry)
\( n_i \) is known as the "intrinsic concentration," with no doping (i.e., an "intrinsic" semiconductor),
the number of electrons, \( n_e \), and the number of holes, \( n_h \), must be equal by charge neutrality, since
\( n_e n_h = n_i^2 \), the number of each must be \( n_i \).
Intrinsic concentrations become large in narrow gap semiconductors or at high temperature. E.g.,
for silicon at OC, \( n_i = 1.04 \times 10^{16} \text{ cm}^{-3} \)
for silicon at 50C, \( n_i = 7.06 \times 10^{16} \text{ cm}^{-3} \)
for GaAs at OC, \( n_i = 1.02 \times 10^{16} \text{ cm}^{-3} \)
for GaAs at 50C, \( n_i = 2.18 \times 10^{15} \text{ cm}^{-3} \)

Fermi level in intrinsic semiconductors:
In an intrinsic semiconductor, the number of electrons is equal to the number of holes. Thus,

\[
n_i = n_e e^{(\mu-\phi)/k_B T} = (n_e n_h)^{1/2} e^{(\mu-\phi)/2k_B T}
\]

By rearranging, we have

\[
\left( \frac{n_i}{n_e} \right)^{1/2} = \exp \left[ \frac{E_g - \mu}{k_B T} \right]
\]

or, after some algebra,

\[
\mu = E_g + \frac{k_B T}{2} \log \left( \frac{n_i}{n_e} \right) = E_g + \frac{k_B T}{2} \frac{3k_B T}{4} \log \left( \frac{m^*_e}{m^*_h} \right)
\]

so that the Fermi level in an intrinsic semiconductor is always very close to the middle of the
bandgap, since \( k_B T \ll E_g \) for most semiconductors.

Simple, fully ionized, non-degenerate case:
Consider one simple, limiting case. Suppose we have \( n_d \) donors per unit volume, and \( n_a \) acceptors
per unit volume. We assume that each acceptor atom takes away an electron and each donor atom adds
an electron (we may have to be at high temperature to assure that this is the case); this is called the
approximation of fully ionized impurities. Hence the net excess number of electrons is

\[
\Delta n = n_e - n_h = n_d - n_a
\]

Assume also that all the carrier distributions can be approximated as non-degenerate. Hence we can use
the law of mass action. Also, assume \( \Delta n > 0 \), i.e., n-doped semiconductor.
Since, from the law of mass action,

\[
n_h = n_t^2/n_e
\]

Or

\[
n_e^2 = n_t^2 \Delta n = n_i^2
\]

Therefore, we have

\[
n_e = \frac{1}{2} \left( (\Delta n)^2 + 4n_i^2 \right)^{1/2} + \Delta n
\]

And
Chapter 10 – Fermi-Dirac Distribution

Fermi Level in Extrinsic Semiconductor:

From our previous expressions, in the non-degenerate limit

\[ n_x = n_x e^{-\Delta E/kT} \]

Or

\[ n_x = n_x e^{(\Delta E - \mu)/kT} \]

where we have chosen the energy origin at the top of the valence band for convenience. Hence we have

\[ \mu = E_g - kT \log(n_v/n_x) \]

or, equivalently

\[ \mu = kT \log(n_v/n_x) \]

Semiconductor statistical mechanics - degenerate case:

When one of the carrier concentrations is increased and approaches the quantum concentration, we may no longer use the classical distribution for that carrier. Instead, the calculation of the carrier concentration now follows the treatment of the Fermi gas. (except the chemical potential is now related to both the electron, and the hole concentration). The number of electrons or holes, is written as an integral over the density of states times the distribution function

\[ N = \int dE D(x)f(x) \]

Use the density of state effective mass for electrons, we have

\[ n_x = \frac{1}{2\pi^2} \left( \frac{2m^*}{\hbar^2} \right)^{3/2} \int_{E_g}^{\infty} \frac{(e-E_e)^{1/2}}{1 + e^{(E-E_e)/kT}} dE \]

Let \( x \equiv (e-E_e)/kT \) and \( \eta \equiv (\mu-E_e)/kT \) We have

\[ n_x = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{x^{1/2}}{1 + e^{(x-\eta)}} dx \]

where \( f(\eta) \) is known as the Fermi-Dirac integral.

Fermi integral \( F(\eta) = \int_0^\infty \frac{x^{1/2}}{1 + e^{(x-\eta)}} dx \)

Two useful analytical limits to the Fermi integral:

- low-temperature and/or high-density
- high-temperature and/or low density

Low-temperature and/or high-density "degenerate" limit:

As temperature goes towards zero, Fermi function tends towards being 1 up to the chemical potential (Fermi energy), and zero above, leading to, formally,

\[ F(\eta) \approx \frac{2}{\eta^{3/2}} \]

limit of low temperature is also essentially the limit at high densities, where \( k_BT \) becomes small compared to the chemical potential, useful for chemical potentials above about \( 4k_BT \), and becomes relatively accurate above about \( 6k_BT \).
Knowing that \( \int_0^\infty x^{1/2} e^{-x} \, dx = \sqrt{\pi} / 2 \), so

\[
F(\eta) = \frac{\sqrt{\pi}}{2} e^{-\eta}
\]

is useful for chemical potentials about \( k_B T \) below the bandgap energy, and accurate from about 2\( k_B T \) below. Therefore,

\[
\frac{n_{\text{avg}}}{n_e} = e^{(\mu - E_g)/kT}
\]

![Figure 4: exact value - solid line, high-temp./low-density approx. - dashed line](image)

Fermi integral approximations:
- Transport in semiconductors: usually take Maxwell-Boltzmann approx:
  - In germanium and silicon: good approx. because electron and hole masses are large
  - In direct bandgap semiconductors used for optoelectronic devices: conduction band effective mass is relatively small, so often n-doped semiconductors have degenerate distributions. In laser diodes degenerate distribution essential to device operation. In most diodes, n-doped contact regions typically doped \( \sim 10^{18} \, \text{cm}^{-3} \) or higher, putting electron Fermi energy in the conduction band. So Maxwell-Boltzmann approx. not strictly valid for optoelectronic diodes.

In the range between the two extreme physical approximations of the non-degenerate (Maxwell-Boltzmann) case and the degenerate case, it is sometimes useful to have approximations to the Fermi integral and its inverse.

One approximation to the Fermi integral is the Aymerich-Humert approximation (X. Aymerich-Humert, F. Serra-Mestres, and J. Millan, Solid State Electronics 24, 581 (1981)), which is

\[
F(x) = \frac{\sqrt{\pi}}{2} e^{-x} + \xi(x)
\]

Where

\[
\xi(x) = \frac{3}{\sqrt{\pi}} \left[ x + 2.13 + \left( x - 2.13 \right)^2 + 9.6 \right]^{-3/2}
\]

This approximation is accurate to about 0.5% in the transitional range \(-5 < x < 10\) (and asymptotes relatively well outside those regions). It is also useful to have approximations that enable us to calculate the Fermi energy (chemical potential) from the carrier density, which is the inverse of the Fermi integral. One such approximation is the Joyce-Dixon approximation. For the conduction band, with the energy of the edge of the conduction band as \( E_g \), defining a "reduced" chemical potential as

\[
\eta = (\mu - E_g) / e
\]

and a "reduced" carrier concentration as \( r = N / N_c \), where \( N_c \) is the effective density of states at the conduction band edge previously defined, the Joyce-Dixon approximation (W.B. Joyce and R.W. Dixon, Appl. Phys. Lett. 31, 354 (1977)) is

\[
\eta \equiv \log r + \frac{1}{\sqrt{8 \pi}} r - \left( \frac{3}{16} \right) \sqrt{3} \pi r^2 + \ldots
\]

i.e.,

\[
\eta \equiv \log r + 3.53553 \times 10^{-1} r - 4.95009 \times 10^{-3} r^2 + 4.78386 \times 10^{-4} r^3 - 4.24563 \times 10^{-5} r^4
\]

The Joyce-Dixon approximation is valid for all negative values of \( \eta \) since it asymptotes to the low density limit, but is not valid for positive values greater than about 5.

Another approximation, which asymptotes to both the low-density (non-degenerate) and high-density (degenerate) limits and hence is valid for all \( \eta \) is an approximation due to Nilsson (N. G. Nilsson, Phys. Sta. Solidi (a) 19, 775 (1973)), which is accurate within about 1%

\[
\eta \equiv \frac{\log r}{1 - r} + \left( \frac{3\sqrt{3}}{\sqrt{8 \pi}} \right)^{1/2} \sqrt{r} + \frac{6\sqrt{r}}{3(4 + \sqrt{r})^{3/2}}
\]

This approximation limits quite obviously to the low density (log \( r \)) and high density \((3\sqrt{3}/\sqrt{8 \pi})^{1/2}\) limits, and adds "joining function" terms chosen to give an empirically good fit.

Example for calculating the Fermi level
- Dopant ionization consideration
- PN junction
- Concept of quasi Fermi level

As a starting point we have

\[ n_e = \frac{1}{2\pi^2} \left( \frac{2m^*_e}{\hbar^2} \right)^{3/2} \int_{E_F}^{\infty} \left( \frac{\epsilon - E_F}{1 + e^{(\epsilon-E_F)/T}} \right) d\epsilon \]

\[ N = 2.83 \times 10^{19} \left( \frac{m^* \cdot T}{m_0 \cdot 300} \right)^{3/2} F(E_F/k_BT) \ cm^{-3} \]

where we are expressing \( E_F \) relative to the bottom of the conduction band. For the case of a degenerate distribution, from last section we had

\[ F(x) \approx \frac{x}{3} \]

so then we have

\[ N = \frac{2}{3} \times 2.83 \times 10^{19} \left( \frac{m^* \cdot T}{m_0 \cdot 300} \right)^{3/2} \left( \frac{E_F}{k_BT} \right)^{3/4} \]

or

\[ E_F = \frac{3N/2}{2.83 \times 10^{19}} \left( \frac{m^* \cdot T}{m_0 \cdot 300} \right)^{2/3} \left( \frac{1.89 \times 10^{19}}{N} \right)^{2/3} \]

For the case of a nondegenerate distribution, we have

\[ E_F = \ln \left( \frac{N}{2.51 \times 10^{16}} - \frac{3}{2} \right) \left( \frac{m^* \cdot T}{m_0 \cdot 300} \right) \]

Example

Consider \( 10^{19} \) cm\(^3\) electrons in InP at room temperature. Find the Fermi level and state whether the distribution is degenerate or nondegenerate.

**Solution**

This concentration is large and the electron effective mass is small, so we guess the electron distribution is degenerate and try to check that.

We use \( m^* = 0.079 \) and use \( T = 300 \) K for simplicity. From Eq., we have 2/3

\[ E_F = \frac{m^*_e \cdot 300}{1.89 \times 10^{19}} \left( \frac{N}{1.89 \times 10^{19}} \right)^{2/3} \]

\[ \left( \frac{E_F}{k_BT} \right)^{3/4} = 0.079 \]

This number is \( \approx 4 \), i.e., the calculated Fermi level based on this guess is many \( k_BT \) above the bottom of the band, so the degenerate assumption is valid. Completing the calculation, at 300 K in electron-volts.

\[ f_{\text{F}}(E,T) = \frac{1}{1 + \exp \left( \frac{E - \mu}{k_BT} \right)} \]

\( \mu \) is the chemical potential, \( k_B \) is Boltzmann's constant. Note \( k_B \approx 25 \) meV at room temperature. Chemical potential (Fermi energy) corresponds to energy for which \( f = \frac{1}{2} \). At zero temperature Fermi-Dirac distribution is step function, with all states up to energy \( E_F \) (Fermi energy) totally full.

**Free electron gas at zero temperature:**

The free electron gas model represents the starting point for describing the properties of metals. As in many solid-state physics problems, the first step is to figure out what the ground state is in the system. Consider a single electron moving in a box of volume \( V = L^3 \). The orbital of the free particle wavefunction

\[ \psi(x,y,z) = A \sin \left( \frac{n_x \pi x}{L} \right) \sin \left( \frac{n_y \pi y}{L} \right) \sin \left( \frac{n_z \pi z}{L} \right) \]

The energy values are

\[ E_n = \frac{h^2}{2M} \left( \frac{n_x^2 + n_y^2 + n_z^2}{L^2} \right) \]

For an \( N \)-fermion system, the ground state should correspond to the scenario where each orbital is filled with two electrons (one spin up, one spin down) up the Fermi energy

\[ E = \frac{h^2}{2M} \left( \frac{\eta r^2}{L^2} \right) \]

Here \( \eta_r \) is the radius of sphere (in the space of the integers \( n_x, n_y, n_z \)) that separates filled and empty orbitals. For the system to hold \( N \)-electrons the orbitals must be filled up to \( \eta_r \) determined by

\[ N = 2 \times 1 \times \frac{4\pi}{3} \eta_r^3 \]

I.e.

\[ \eta_r = \left( \frac{3N/\pi}{4} \right)^{1/3} \]
Here the factor 2 arises because an electron has two possible spin orientations. The factor 1/6 arises because only triplets $n_x, n_y, n_z$ in the positive octant of the sphere in $n$ space are to be counted.

Thus, we may calculate the Fermi energy as:

$$\varepsilon_F = \frac{h^2}{2m} \left( \frac{3^n}{V} \right)^{1/3}$$

This relates the Fermi energy to the electron concentration. The total energy of the system in the ground state is:

$$U_0 = 2 \sum_{n \neq 0} \varepsilon_n = 2 \times \frac{1}{8} \pi^2 \frac{m}{h^2} \int_0^{\varepsilon_F} \text{d}n \varepsilon_n^2$$

which, after taking into account $n_x$, gives

$$U_0 = \frac{3}{5} N \varepsilon_F$$

Free electron gas at finite temperature:

For the case of finite temperature, for an electron gas that are kept at a constant temperature $\tau$ and a constant chemical potential $\mu$, the occupation number for each orbital is:

$$f(\varepsilon) = \frac{1}{1 + \exp \left( \frac{\varepsilon - \mu}{\tau} \right)}$$

Therefore, the thermal averages for the independent particle problems have the form:

$$(X) = \sum_{n} f(\varepsilon_n, \tau, \mu) X_n$$

where $n$ denotes the quantum orbital; $X_n$ is the value of the quantity $X$ in the orbital $n$; and $f(\varepsilon_n, \tau, \mu)$ is the thermal average occupancy, called the distribution function, of the orbital $n$. We express $(X)$ as an integral over the orbital energy $\varepsilon$, Then above Eq becomes:

$$(X) = \int \text{d}\varepsilon D(\varepsilon) f(\varepsilon, \tau, \mu) X(\varepsilon)$$

where the sum over orbitals has been transformed to an integral. $D(\varepsilon)$ is the density of state that we have discussed extensively before. The density of state for free electron (including spins) is:

$$D(\varepsilon) = \frac{V}{2\pi^2} \left( \frac{2m}{\hbar^2} \right)^{3/2} e^{\varepsilon/\tau}$$

Heat capacity of degenerate electron gas:

We derive a quantitative expression for the heat capacity of a degenerate Fermi gas of electrons in three dimensions. This calculation gives excellent agreement with the experimental results for the heat capacity of metals. The increase in the total energy of a system of $N$ electrons when heated from 0 to $\tau$ is denoted by

$$\Delta U = \int_0^\varepsilon \text{d}\varepsilon D(\varepsilon) f(\varepsilon) - \int_0^{\varepsilon_F} \text{d}\varepsilon D(\varepsilon)$$

Here $f(\varepsilon)$ is the Fermi-Dirac function, and $D(\varepsilon)$ is the density of states. Notice that:

$$\int_0^{\varepsilon_F} \text{d}\varepsilon f(\varepsilon) D(\varepsilon) = \int_0^{\varepsilon_F} \text{d}\varepsilon f(\varepsilon) D(\varepsilon)$$

hence:

$$\Delta U = \int_0^{\varepsilon_F} \text{d}\varepsilon (\varepsilon - \varepsilon_F) D(\varepsilon) f(\varepsilon) + \int_0^{\varepsilon_F} \text{d}\varepsilon (\varepsilon - \varepsilon_F) (1 - f(\varepsilon)) D(\varepsilon)$$

The first integral on the right-hand side of gives the energy needed to take electron from $\varepsilon_F$ to the orbitals of energy $\varepsilon > \varepsilon_F$, and the second integral gives the energy needed to bring the electrons to $\varepsilon_F$ from orbitals below $\varepsilon_F$.

The heat capacity of the electron gas is found on differentiating $\Delta U$ with respect to $\tau$. The only temperature-dependent term in equation above is $f(\varepsilon)$. Therefore, we can group terms to obtain:

$$C_{el} = \frac{\Delta U}{\tau} = \int_0^{\varepsilon_F} \text{d}\varepsilon (\varepsilon - \varepsilon_F) f(\varepsilon)$$

In a typical metal, the concentration of electron is such that $\varepsilon_F/k_B = 50000$ K. Thus, the typical temperature of interest in metals is $\tau \varepsilon_F < 0.01$. Since $f(\varepsilon)$ is like a step function, the derivative $df/d\tau$ is large on at energies near $\varepsilon_F$. It $\tau$ is a good approximation to evaluate the density of orbitals $D(\varepsilon)$ at energy $\varepsilon_F$ and take it outside of the integral:

$$C_{el} = D(\varepsilon_F) \int_0^{\varepsilon_F} \text{d}\varepsilon (\varepsilon - \varepsilon_F) \frac{df}{d\tau}$$

One could typically ignore the temperature dependence of the chemical potential in the Fermi-Dirac distribution and replace $m$ by the constant $\varepsilon_F$. We have then:

$$\frac{df}{d\tau} = \frac{\varepsilon_F}{\tau^2} \left[ 1 + e^{(\varepsilon_F - \varepsilon)/\tau} \right]^{-2}$$

We set:

$$\frac{df}{d\tau} = \int \text{d}x D(\varepsilon_F) \int_0^{\varepsilon_F} \text{d}x \frac{\varepsilon^2}{(1 + e^{(\varepsilon_F - \varepsilon)/\tau})^2} \int_0^{\varepsilon_F} \text{d}x \frac{e^x}{(1 + e^x)^2}$$

Therefore, in the region $\tau \ll \tau_F$, we have:

$$C_{el} = \frac{1}{2} \pi^2 D(\varepsilon_F) \tau$$

The key results are the linear dependence of the heat capacity on temperature, and the dependence of the heat capacity on the density of states at the Fermi energy. Both of these results are confirmed by experimental observations. The strong dependence of properties of metals on the density of states at Fermi surface is in fact, a rather general result.

Electron distribution:

Electrons in thermal equilibrium have a Fermi-Dirac distribution, probability $f(E, T)$ of electron in state of energy $E$ at temperature $T$ is:

$$f(E, T) = \frac{1}{1 + \exp \left( \frac{E - \mu}{k_B T} \right)}$$

Hole distribution:

Holes correspond to the absence of an electron, hence the probability $f_h$ of finding a hole in a given state is the probability that there is not an electron in that state. Hence $f_h = 1 - f$, and so
Chapter 10 —Fermi level

Introduction to Solid State Physics

\[ f_{e}(E, T) = \frac{\exp \left( \frac{E - \mu}{k_{B}T} \right)}{1 + \exp \left( \frac{E - \mu}{k_{B}T} \right)} = \frac{1}{1 + \exp \left( \frac{-E + \mu}{k_{B}T} \right)} \]

Hence the holes also have a Fermi-Dirac distribution. Can look at holes this way, in which hole energies are increasingly negative as they acquire kinetic energy, or can look at hole energies “upside down”, changing \(-E + \mu\) to \(E - \mu\).

Pure semiconductors, donors and acceptors:

Let \(n_{d}\) denote the concentration of conduction electrons, and \(n_{a}\) denote the concentration of holes.

In a pure semiconductor the two will be equal: \(n_{d} = n_{a}\).

Most semiconductors as used in devices have been intentionally doped with impurities that may become thermally ionized in the semiconductor at room temperature. Impurities that give an electron to the crystal (and become positively charged in the process) are called donors. Impurities that accept the electron from the valence band (and hence become negatively charged in the process) are called acceptors.

Let \(n_{d}^{+}\) be the concentration of positively charged donors and \(n_{a}^{-}\) the concentration of negatively charged acceptors. The electrical neutrality condition becomes

\[ n_{d}^{+} - n_{a}^{-} = \Delta n = n_{d}^{+} - n_{a}^{-} \]

This specifies the difference between electron and hole concentrations.

Group V elements (e.g., P) are usually donors in silicon.

Group III elements (e.g., Al) are usually acceptors in silicon.

In III-V semiconductors, the donors are often Group IV materials (e.g., tellurium, tin), and acceptors are often Group II materials (e.g., zinc, beryllium). Often Group IV materials are used. Their behavior is less obvious. One way of rationalizing their behavior is to note that Group IV atoms might occupy either Group III or Group V sites in the lattice. In the first case they might be expected to behave more like donors, in the second case more like acceptors. For example, silicon is a commonly used n-type dopant in GaAs, and carbon is sometimes used as a p-type dopant in GaAs, and germanium is often used as an n-type dopant in forming contacts to GaAs (e.g., by depositing Ge-doped metal and “allaying” it into the n-type materials being contacted). Under some circumstances, silicon can become a p-type dopant in GaAs.

n-type and p-type semiconductors:

A semiconductor in which the added number of donors dominates is called an “n-type” semiconductor, and “majority carrier” conduction takes place through electrons in the conduction band.

A semiconductor in which the added number of acceptors dominates is called a “p-type” semiconductor, and “majority carrier” conduction takes place through holes in the valence band.

Semiconductor statistical mechanics - classical limit:

For energies far above the chemical potential, the Fermi-Dirac distribution behaves like classical Maxwell-Boltzmann distribution, \(f_{MB} \pm\), i.e., for electrons,

\[ f_{MB}(E, T) = A e^{-E/k_{B}T} \]

where \(A = \exp(\mu/k_{B}T)\). This is to be expected, since at these high energies occupation probabilities are small. Hence little influence of the Pauli exclusion restriction that no more than one electron may occupy a state.

If \(\mu < \epsilon_{F}\) for all conduction states, and \(\mu > \epsilon_{F}\) for all valence states, the semiconductor is said to be in a nondegenerate regime. (i.e classical regime). This corresponds to the case when the chemical potential is deep inside the gap, far from both the conduction band edge and the valence band edge. The total number of conduction electrons in the form:

\[ N_{e} = \sum_{\epsilon_{F}} e^{-(\epsilon_{F} - \mu)/k_{B}T} \]

Or

\[ N_{e} = \int_{-\infty}^{\infty} \frac{d\epsilon}{e^{(\epsilon - \mu)/k_{B}T} + 1} \]

Here

\[ N_{e} = \int_{-\infty}^{\epsilon_{F}} e^{-(\epsilon - \epsilon_{F})/k_{B}T} \]

Where as we have discussed before

\[ D(\epsilon) = \frac{m_{e}^{*} \sqrt{2m_{e}^{*}(\epsilon - \epsilon_{F})}}{\pi h^{3}} \]

And \(m_{e}^{*}\) is the density of state effective mass for the conduction band. Evaluating the integral we have

\[ N_{e} = \int \frac{m_{e}^{*} \sqrt{2m_{e}^{*}(\epsilon - \epsilon_{F})}}{\pi h^{3}} \]

Defining the quantum concentration \(n_{e}\) for the conduction electrons as:

\[ n_{e} = \frac{N_{e}}{V} = 2 \left( \frac{m_{e}^{*}}{2\pi h^{2}} \right)^{3/2} \]

The conduction electron concentration \(n_{e} = N_{e}/V\) becomes

\[ n_{e} = n_{e} e^{-(\epsilon - \epsilon_{F})/k_{B}T} \]

Similarly, for the concentration of holes in the valence band, we have

\[ n_{h} = n_{h} e^{-(\epsilon - \epsilon_{F})/k_{B}T} \]

where

\[ n_{h} = 2 \left( \frac{m_{e}^{*}}{2\pi h^{2}} \right)^{3/2} \]

Law of Mass Action:

If we are dealing with semiconductors in the Maxwell-Boltzmann limit, i.e., any of the criteria

- large effective masses
- or low doping densities
- or very high temperatures

but not III-V semiconductors with large numbers of electrons, we have, \(n_{e}n_{h} = n_{F}^{2}\) where

\[ n_{F} = \left( n_{F} n_{F} \right)^{1/2} \]

This equation is called the Law of Mass Action (the term comes from chemistry).
Introduction to Solid State Physics

Chapter 10 – Fermi level

\[ n_i = \text{known as the "intrinsic concentration," with no doping (i.e., an "intrinsic" semiconductor), the number of electrons, } n_e, \text{ and the number of holes, } n_h, \text{ must be equal by charge neutrality. since } n_e n_h = n_i^2, \text{ the number of each must be } n_i. \]

Intrinsic concentrations become large in narrow gap semiconductors or at high temperature. E.g.,

- for silicon at 0°C, \( n_i = 1.04 \times 10^{10} \text{ cm}^{-3} \)
- for silicon at 50°C, \( n_i = 2.06 \times 10^{10} \text{ cm}^{-3} \)
- for GaAs at 0°C, \( n_i = 1.02 \times 10^{15} \text{ cm}^{-3} \)
- for GaAs at 50°C, \( n_i = 2.18 \times 10^{15} \text{ cm}^{-3} \)

\( n_i \) in intrinsic semiconductors:

In an intrinsic semiconductor, the number of electrons is equal to the number of holes. Thus,

\[ n_i = n_e = n_h = (n_i n_p)^{1/2} = n_i^{1/2} \exp \left( \frac{E_g - \mu}{k_B T} \right) \]

By rearranging, we have

\( \left( \frac{n_i}{n_e} \right)^{1/2} = \exp \left( \frac{E_g - \mu}{k_B T} \right) \)

or, after some algebra,

\[ \mu = \frac{E_g}{2} + k_B T \log \left( \frac{n_i}{n_e} \right) = \frac{E_g}{2} + \frac{k_B T}{4} \log \left( \frac{n_i^2}{n_e^2} \right) \]

so that the Fermi level in an intrinsic semiconductor is always very close to the middle of the bandgap, since \( k_B T \ll E_g \) for most semiconductors.

Simple, fully ionized, non-degenerate case:

Consider one simple, limiting case. Suppose we have \( n_d \) donors per unit volume, and \( n_a \) acceptors per unit volume. We assume that each acceptor atom takes away an electron and each donor atom adds an electron (we may have to be at high temperature to assure that this is the case); this is called the approximation of fully ionized impurities. Hence the net excess number of electrons is

\[ \Delta n = n_d - n_a = n_e - n_h \]

Assume also that all the carrier distributions can be approximated as non-degenerate. Hence we can use the law of mass action. Also, assume \( \Delta n > 0 \), i.e., n-doped semiconductor.

Since, from the law of mass action,

\[ n_e = n_i^2 / n \]

Or

\[ n_e^2 - n_e \Delta n = n_i^2 \]

Therefore, we have

\[ n_e = \frac{3}{2} \left( \Delta n^2 + 4 n_i^2 \right)^{1/2} + \Delta n \]

And

\[ n_h = \frac{3}{2} \left( \Delta n^2 + 4 n_i^2 \right)^{1/2} - \Delta n \]

Introduction to Solid State Physics

Chapter 10 – Fermi level

For most practical situations, the intrinsic concentration, \( n_i \), is a very small number relative to any doping concentration or resulting carrier concentration. Presuming for the moment that \( n_i \) is indeed negligibly small (we may check this a posteriori), we have

\[ n_e \approx n_i \]

\[ n_h \approx n_i \Delta n \]

In other words, as long as the difference in doping concentrations is much larger than the intrinsic concentration, the resulting net number of electrons in the conduction band (or, if negative, the net number of holes in the valence band) is simply the difference in the doping concentrations, \( \Delta n \). A semiconductor where \( \Delta n \gg n_i \) is called an "extrinsic" semiconductor.

Note this analysis is not necessarily valid at low temperatures where the dopants would not be fully ionized, nor is it valid at high concentrations where the distributions were degenerate.

Fermi Level in Extrinsic Semiconductor:

From our previous expressions, in the non-degenerate limit

\[ n_e \approx n_i \exp \left( \frac{E_g - \mu}{k_B T} \right) \]

Or

\[ n_e \approx n_i \exp \left( \frac{E_g - \mu}{k_B T} \right) \]

where we have chosen the energy origin at the top of the valence band for convenience. Hence we have

\[ \mu = E_g - k_B T \log \left( n_e / n_i \right) \]

or, equivalently

\[ \mu = E_g - k_B T \log \left( n_i / n_e \right) \]

Semiconductor statistical mechanics - degenerate case:

When one of the carrier concentrations is increased and approaches the quantum concentration, we may no longer use the classical distribution for that carrier. Instead, the calculation of the carrier concentration now follows the treatment of the Fermi gas. (except the chemical potential is now related to both the electron and the hole concentration). The number of electrons or holes, is written as an integral over the density of states times the distribution function

\[ N = \int \frac{d\mathcal{E}}{e} f(\mathcal{E}) \]

Use the density of state effective mass for electrons, we have

\[ n_e = \frac{1}{2\pi^2} \left( \frac{2m_e^*}{\hbar^2} \right)^{1/2} \int_{\epsilon_c}^{\infty} \frac{(\epsilon - \epsilon_c)^{1/2}}{1 + e^{(\epsilon - \mu)/k_B T}} d\epsilon \]

Let \( x \equiv (\epsilon - \epsilon_c)/k_B T \) and \( y \equiv (\mu - \epsilon_c)/k_B T \) We have

\[ n_e = \frac{1}{k_B T} \int_{-\infty}^{\infty} x^{1/2} e^{-x} dx \]

where \( f(\eta) \) is known as the Fermi-Dirac integral.

Fermi integral \( f(\eta) = \int_{0}^{\infty} \frac{e^{\eta x}}{1 + e^{x}} dx \)
Two useful analytical limits to the Fermi integral:

- Low-temperature and/or high-density
- High-temperature and/or low-density

Low-temperature and/or high-density "degenerate" limit:

As temperature goes towards zero, Fermi function tends towards being 1 up to the chemical potential (Fermi energy), and zero above, leading to, formally,

$$ F(\eta) \approx \frac{1}{3} \eta^{3/2} $$

limit of low temperature is also essentially the limit at high densities, where $k_B T$ becomes small compared to the chemical potential, useful for chemical potentials above about $4k_B T$, and becomes relatively accurate above about $6k_B T$.

Fermi integral – approximations:

- Transport in semiconductors – usually take Maxwell-Boltzmann approx
  - In germanium and silicon: good approx. because electron and hole masses are large
  - In direct bandgap semiconductors used for optoelectronic devices: conduction band effective mass is relatively small, so often n-doped semiconductors have degenerate distributions. In laser diodes degenerate distribution essential to device operation. In most diodes, n-doped contact regions typically doped $\approx 10^{18} \text{ cm}^{-3}$ or higher, putting electron Fermi energy in the conduction band. So Maxwell-Boltzmann approx. not strictly valid for optoelectronic diodes

- In the range between two extreme physical approximations of the non-degenerate (Maxwell-Boltzmann) case and the degenerate case, it is sometimes useful to have approximations to the Fermi integral and its inverse.

One approximation to the Fermi integral is the Aymerich-Humet approximation (X. Aymerich-Humet, F. Serra-Mestres, and J. Millan, Solid State Electronics 24, 981 (1981)), which is

$$ F(\eta) = \frac{1}{2} \left( \frac{\eta}{\xi(\eta)} \right)^{1/2} $$

Where

$$ \xi(x) = \frac{3}{2} \left( x + 2.13 + (x - 2.13)^2 + 9.6 \right)^{1/2} $$

This approximation is accurate to about 0.5% in the transitional range $-5 < x < 10$ (and asymptotes relatively well outside those regions). It is also useful to have approximations that enable us to calculate the Fermi energy (chemical potential) from the carrier density, which is the inverse of the Fermi integral. One such approximation is the Joyce-Dixon approximation. For the conduction band, with the energy of the edge of the conduction band as $E_c$, defining a “reduced” chemical potential as
\[ \eta = \frac{(u - E_c)}{t} \text{ and a "reduced" carrier concentration as } r = N/N_e. \text{ (where } N_e \text{ is the effective density of states at the conduction band edge previously defined, the Joyce-Dixon approximation [W.B. Joyce and R. W. Dixon, Appl. Phys. Lett. 31, 354 (1977)] is} \]

\[ \eta \approx \log r + \frac{1}{\sqrt{16}} r - \left( \frac{3}{16} \sqrt{3} \right) r^2 + \ldots \]

i.e.,
\[ \eta \approx \log r + 3.53553 \times 10^{-1} r - 4.95009 \times 10^{-2} r^2 + 1.48386 \times 10^{-3} r^3 - 4.42563 \times 10^{-4} r^4 \]

The Joyce-Dixon approximation is valid for all negative values of \( \eta \) since it asymptotes to the low density limit, but is not valid for positive values greater than about 5.

Another approximation, which asymptotes to both the low-density (non-degenerate) and high-density (degenerate) limits (and hence is valid for all \( \eta \)) is an approximation due to Nilsson (N. G. Nilsson, Phys. Sta. Solidi (a) 19, K75 (1973)), which is accurate within about 1%:

\[ \eta \approx \frac{\log r}{1 - r} + \left( \frac{3\sqrt{\pi \tau}}{4} \right) \frac{2^2}{3} + \frac{3\sqrt{\pi \tau}}{3(4 + \sqrt{\pi \tau})^2} \]

This approximation limits quite obviously to the low density (\( \log r \)) and high density (\( 3\sqrt{\pi \tau} / 4 \)) limits, and adds "joining function" terms chosen to give an empirically good fit.