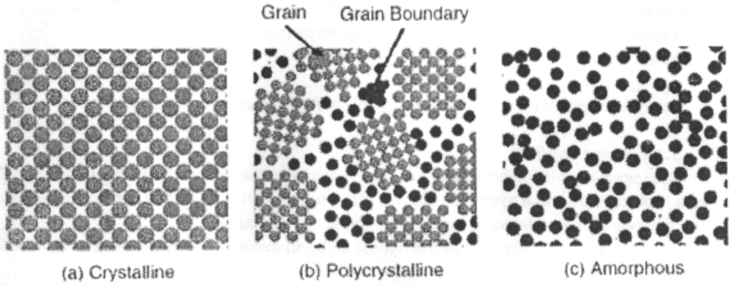
- Materials: crystalline, poly-crystalline, amorphous
- Bonds in Solids
- Crystal Structure
- Basis
- Lattice: Bravais Lattice vs. Non Bravais Lattice
- Unit Cell, Primitive Unit Cell, Wigner-Seitz Unit Cell
- Crystal symmetry
- Miller Indices

Why we call it solid-state?

States of the matter:

1. Solid: density ~ 10^{22} /cm^3

1. a: Crystal: long range order (lattice + basis) {Example: Epitaxial silicon and diamond}
1. b: Polycrystal: short range order (μm~10μm) {Example: Most metals (Al, Cu) Ploy-Si}
1. c: Amorphous: no order {Example: Glasses like SiO_2}



(a) Crystalline (b) Polycrystalline (c) Amorphous

Figure 1: Schematic representation of crystalline, polycrystalline, and amorphous material structures [Ref: Micro Electro Mechanical System Design - James J. Allen]

2. Liquids: no order, takes the shape of the container, still weak bounds; density ~ 10^{19} /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

---

٢٥٢٣٤

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 1: Abbreviated periodic chart of the elements

1 IA 1A							18 VIIIA 8A
1 H 1.008	2 IIA 2A	13 IIIA 3A	14 IVA 4A	15 VA 5A	16 VIA 6A	17 VIIA 7A	2 He 4.003
3 Li 6.941	4 Be 9.012	5 B 10.81	6 C 12.01	7 N 14.01	8 O 16.00	9 F 19.00	10 Ne 20.18
11 Na 22.99	12 Mg 24.31	13 Al 26.98	14 Si 28.09	15 P 30.97	16 S 32.07	17 Cl 35.45	18 Ar 39.95
19 K 39.10	30 Zn 65.39	31 Ga 69.72	32 Ge 72.59	33 As 74.92	34 Se 78.96	35 Br 79.90	36 Kr 83.80
37 Rb 85.47	48 Cd 112.4	49 In 114.8	50 Sn 118.7	51 Sb 121.8	52 Te 127.6	53 I 126.9	54 Xe 131.3
55 Cs 132.9	80 Hg 200.5	81 Tl 204.4	82 Pb 207.2	83 Bi 209.0	84 Po (210)	85 At (210)	86 Rn (222)

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)
   - II-VI, e.g., zinc selenide (ZnSe), cadmium telluride (CdTe)
   - IV-IV, e.g., silicon carbide (SiC)
   - IV-VI, e.g., lead sulphide (PbS)
3. Alloys,

e.g., ternary (three component)

- aluminum gallium arsenide ($Al_{1-x}Ga_xAs$)
  x is fraction of gallium in group III elements (still equal numbers overall of group III and group V elements)
- mercury cadmium telluride ($Hg_{1-x}Cd_xTe$)

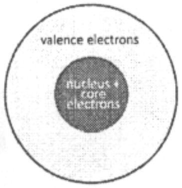
e.g., quaternary (four component)

- indium gallium aluminum arsenide ($In_{1-x-y}Ga_xAl_yAs$)
- indium gallium arsenide phosphide ($In_{1-x}Ga_xAs_{1-y}P_y$)

Bonds in solids: [Ref: U of Exeter]

Atoms vibrate with small amplitudes about fixed equilibrium positions. We assume that atoms are fixed, unless phonons are considered.

Atoms look like outer valence electrons orbiting around the core. Core consists of nucleus plus inner core electrons



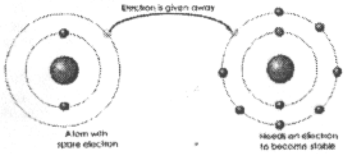
1. Ionic bonding:

Complete transfer of electrons from one atom (usually a metal) to another (non metal ion) (compounds only, not elemental solids). Bond comes from electrostatic attraction between ions.

$$Na + Cl \rightarrow Na^+ + Cl^- \rightarrow NaCl$$

All ionic compounds have a degree of covalent bonding. The larger the difference in electronegativity between two atoms, the more ionic the bond is.

- Bond is strong (high melting point, large elastic modulus)
- !? Not directional (high density, high coordination number)
- Compounds only
- Good insulators (except near melting point)
- Transparent up to UV (strong bonds → electrons need a lot of energy to become free)



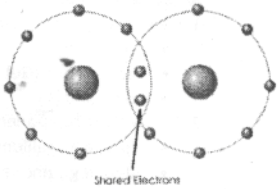
Mathematical form: Energy ~ 1/r, Example: Sodium Chloride

2. Covalent Bonding:

Equal sharing of electrons between atoms → both atoms have full shells (Example: Diamond, Silicon)

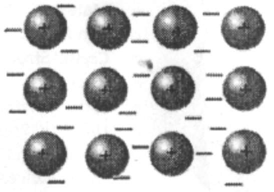
Note continuum of behavior, ionic → covalent (e.g. III-V compounds GaAs, InSb, are partially covalent and partially ionic.)

- Bond is strong (high melting point, large elastic modulus)
- Directional (from orientation of QM orbitals) → low density
- Saturable (limited number of bonds per atom) ↑
- Good insulators



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: $\Delta p \Delta x \geq h/2$ as for electrons m is small so the zero point energy $\Delta E = \Delta p^2/2m$ is very large; the electrons would shake themselves free and are therefore delocalized)



- Bonds are non directional (high coordination number, high density, malleable and ductile)
- Variable strength
- Free electrons → 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 (→ low melting point, large expansion coefficient)
- Non directional so high coordination number BUT
- Long bond lengths (→ 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

Bond	Energy (GPa)	Example of Bond
Covalent	1,000	Diamond
Ionic	30 - 100	Salt and Ceramics
Metallic	30 - 150	Metals
Hydrogen	8	Ice
Van der Vaals	2	Polythene

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.

Shortest possible a_i gives us *primitive vectors*

The volume cell enclosed by the primitive vectors is called the *primitive unit cell* $V_0 = \vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)$

- برای حالت دوبعدی سطح مورد نظر است
به عبارت دیگر می‌توانیم بگوییم که

$S = \hat{a}_1 \times \hat{a}_2$

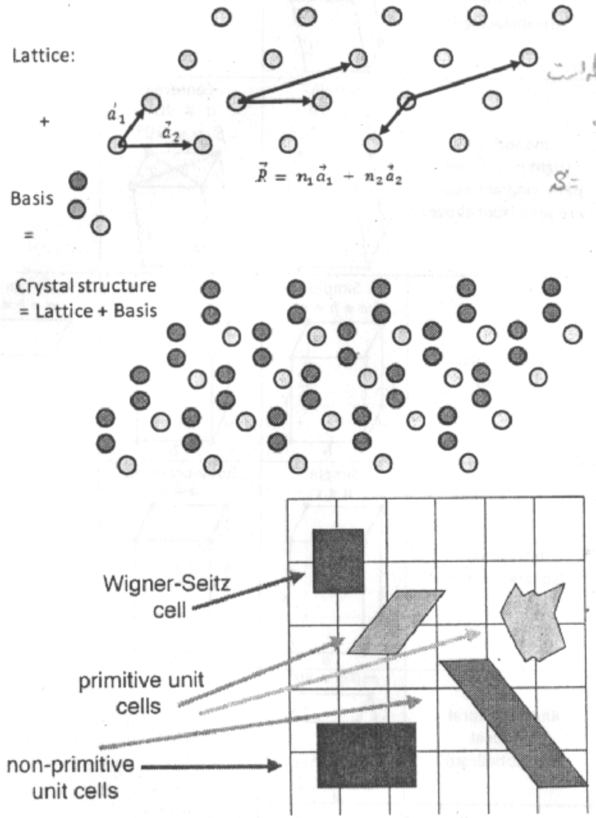


Figure 2: Lattice, basis, crystal Structures, and unit cells

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 l-amino acids, hence, reflection or inversion symmetries are not allowed.)

N-fold rotation symmetry, C_n: Rotation by an angle $2\pi/n$ 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 asymmetric object) (rotation will be about a certain point for 2D or axis for 3D)

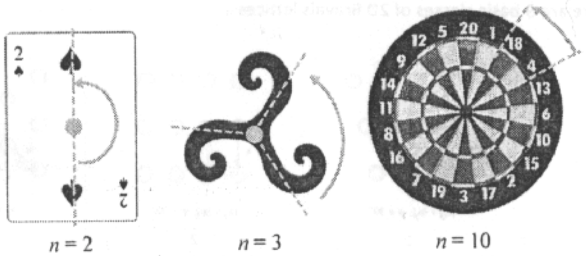


Figure 3 : Different orders of rotational symmetry

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), σ: 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:

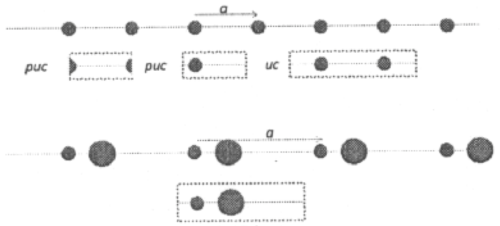


Figure 4: 1D lattices

Two-Dimensional Lattices:

There are 5 basic classes of 2D Bravais lattices

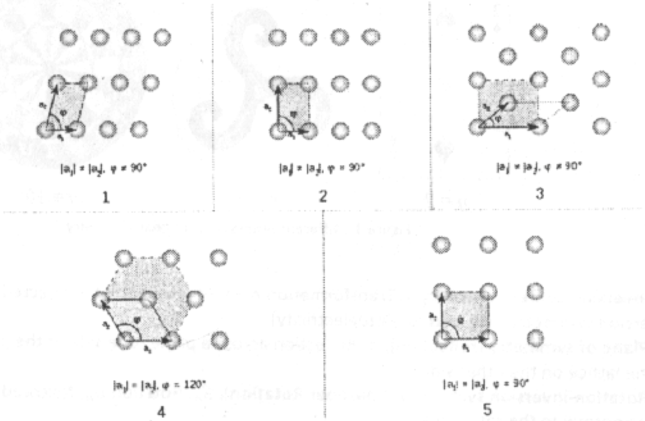
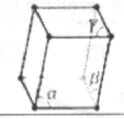
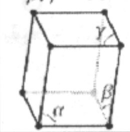
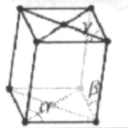
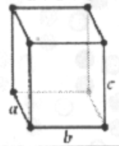
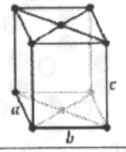
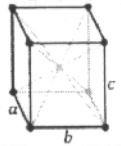
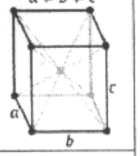
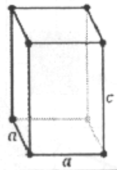
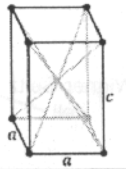
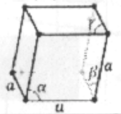
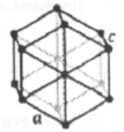
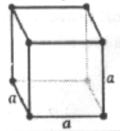
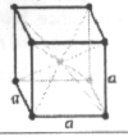
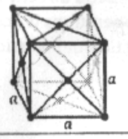


Figure 5 : 2D Bravais lattices: 1 oblique, 2 rectangular, 3 centered rectangular, 4 hexagonal, and 5 square [Ref: Wikipedia]

Three-Dimensional Lattices:

14 Bravais lattices

Table 3 : 14 Bravais lattice types into 7 lattice systems [Ref: Wikipedia]

The 7 lattice systems	The 14 Bravais Lattices			
Triclinic (parallelepiped)	$\alpha, \beta, \gamma \neq 90°$ 			
monoclinic (right prism with parallelogram base; here seen from above)	Simple $\alpha \neq 90°$ $\beta, \gamma = 90°$ 	Centered $\alpha \neq 90°$ $\beta, \gamma = 90°$ 		
orthorhombic (cuboid)	Simple $a \neq b \neq c$ 	base-centered $a \neq b \neq c$ 	body-centered $a \neq b \neq c$ 	face-centered $a \neq b \neq c$ 
tetragonal (square cuboid)	Simple $a \neq c$ 	body-centered $a \neq c$ 		
Rhombohedral (trigonal trapezohedron)	$\alpha = \beta = \gamma \neq 90°$ 			
hexagonal (centered regular hexagon)				
Cubic (isometric; cube)	Simple 	body-centered 	face-centered 	

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 Frankenheim (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 surroundings. (Why we do 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 4 : Crystal structure and lattice constants (at 300°k) [Ref: P, p.12]

Semiconductor	Crystal Structure	Lattice Constants (°A)
Si	Diamond	5.43095
Ge	Diamond	5.64613
GaAs	Zincblende	5.6536
CdS	Zincblende	5.8320
	Wurtzite	a=4.16, c=6.756
PbS	Rock-Salt	5.9362

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.

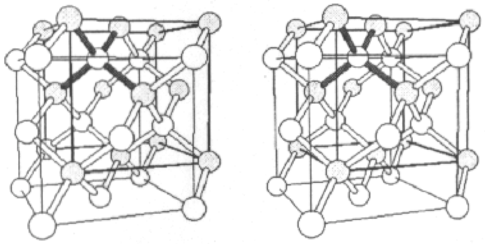


Figure 6: Diamond (Zincblende) lattice unit cell

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 from.

Wurtzite:

Named after the mineral wurtzite, 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 pyroelectricity.

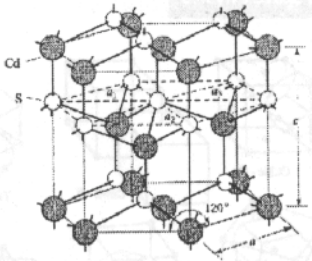


Figure 7 : Wurtzite structure

Rock salt:

Formed in cubic cells with two different atoms alternating atomic positions. Found in IV-VI lead-based semiconductors, such as PbS.

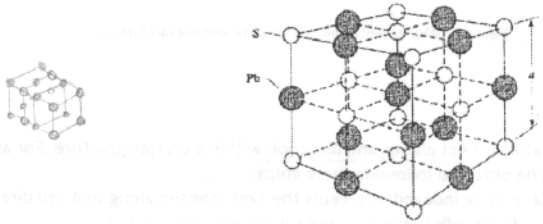


Figure 8 : Rock salt structure

The Wigner-Seitz unit cell:

The Wigner-Seitz primitive unit cell is defined as the locus of points in space that are closer to that lattice point than to any of the other lattice points

1. Choose a lattice point, and draw lattice vectors from it to its neighbors
2. Draw the perpendicular bisectors of these vectors
3. Take a smallest region formed by these bisectors about the point

(The Wigner–Seitz cell in itself is not of paramount importance in the direct space, it is extremely important in the reciprocal space. The Wigner–Seitz cell in the reciprocal space is called the Brillouin zone, which contains the information about whether a material will be a conductor, semiconductor or an insulator)

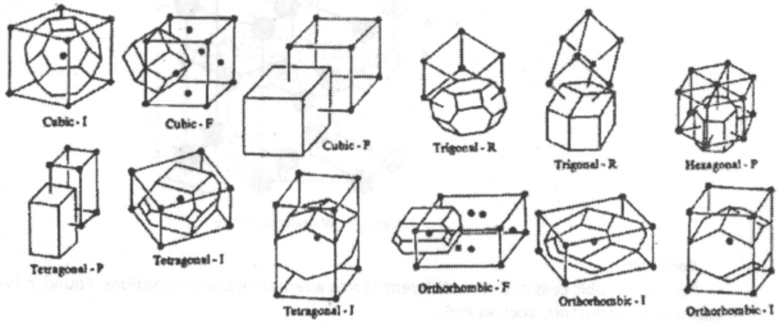


Figure 9 : Wigner-Seitz unit cell for different Bravais lattices

Miller indices:

That is a method to label distinct planes and direction within a crystal structure. For any given plane, Miller indices easily 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, 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., (1$\bar{1}$1).

Groups of equivalent planes, ((100), (010), (001), ($\bar{1}$00), (0$\bar{1}$0), and (00$\bar{1}$) 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 notated in curly brackets, i.e., {100} for the above set of equivalent planes.

که وجوه یک مکعب هستن
چون با دوران به هم میر

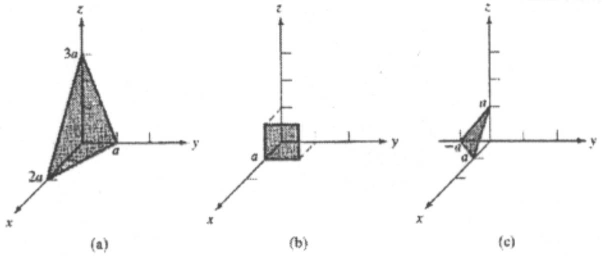


Figure 10 : Miller indices (a) The (362) plane. (b) The (100) plane. (c) The (1$\bar{1}$1) plane.

Similar procedure can be used to define Miller indices for directions.

1. Set up a vector of arbitrary length in the direction of interest (must be a crystal direction, i.e., connecting two crystal points)
2. Decompose the vector into its basis vector components in the a, b, and c directions
3. Convert the resulting numbers to the lowest possible set of integers by multiplying by an appropriate number

Directions are notated using square brackets, e.g., [1$\bar{1}$1]

For cubic crystals, directions perpendicular to particular crystal planes can be indexed using the same index as the plane. Sets of equivalent directions are specified by triangular brackets, e.g., ⟨100⟩

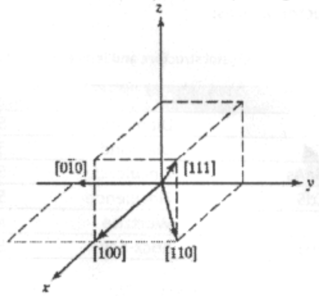


Figure 11: Sample direction vectors and Miller indices

- Wave-particle duality
- Time-independent Schrödinger Equation
- Probability densities
- Operators
- Properties of Eigenfunctions and Eigenvalues
- Time-Dependent Schrödinger Equation
- Free Particle, Particle in a Box, Quantum Wells, Tunneling

Why do we need quantum mechanics?

Not all of the electrons contribute in conductance.

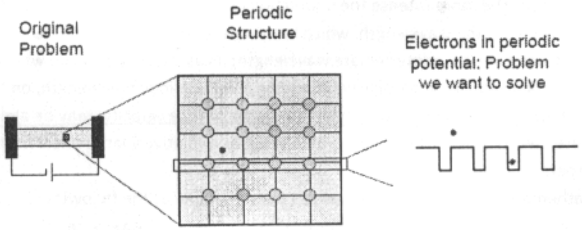


Figure 1: electrons in periodic potential

Carrier number = Number of states x filling factor

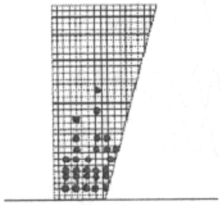
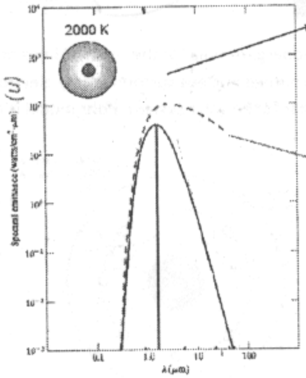


Figure 2

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 in 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.



Rayleigh-Jeans Formula (ترمودینامیک کلاسیک)

$u(\lambda,T) \propto \underline{k_B T / \lambda^4} \rightarrow$ ⊕ برای λ های بزرگ

$\log(u) = -4\log(\lambda) + \log(T)$ طبق شکل صحیح است.

Wein's Formula

$$u \propto \frac{e^{-\beta/\lambda T}}{\lambda^5}$$

Plank's fitting formula

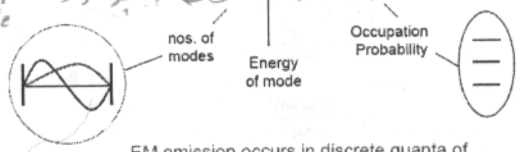
$$u(\lambda,T) \propto \frac{1}{\lambda^5}\left[\frac{1}{e^{\beta/\lambda T}-1}\right]$$

Figure 3: Blackbody radiation

$$u(f,T) = u(\lambda,T)\frac{d\lambda}{df} \sim \frac{1}{\lambda^5}\left[\frac{1}{e^{\beta/\lambda T}-1}\right]\frac{d\lambda}{df} \qquad \lambda = \frac{c}{f}$$

$$\sim f^2 \times hf \times \left(\frac{1}{e^{hf/kT}-1}\right)$$

چرا f^2 را nos of mode گرفته است.



EM emission occurs in discrete quanta of

$E = hf$ n=1,2, ……. N

Figure 4: Planck's interpretation of the formula

2. Photoelectric Effect: In 1905 Albert Einstein took Plank'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.

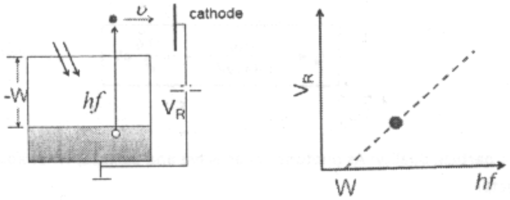


Figure 5: Photoelectric effect

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.

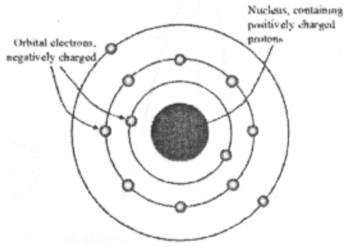


Figure 6: Bohr model

He assumed that the angular momentum is quantized.

$$\boxed{\begin{gathered} L_n = m_0 v r_n = n\hbar \\ \upsilon = n\hbar / m_0 r_n \end{gathered}} \qquad n = 1, 2, 3, \ldots$$

$$\frac{m_0 v^2}{r_n} = \frac{q^2}{4\pi\varepsilon_0 r_n^2}$$

$$r_n = \frac{4\pi\varepsilon_0 (n\hbar)^2}{m_0 q^2}$$

$$r_n = \frac{4\pi\varepsilon_0 (n\hbar)^2}{m_0 q^2}$$

$$\text{K.E.} = \frac{1}{2} m_0 v^2 = \frac{1}{2}(q^2/4\pi\varepsilon_0 r_n)$$

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

$$E_n = \text{K.E.} + \text{P.E.} = -\frac{1}{2}(q^2/4\pi\varepsilon_0 r_n)$$

$$\boxed{E_n = -\frac{m_0 q^4}{2(4\pi\varepsilon_0 n\hbar)^2} = -\frac{13.6}{n^2}\,\text{eV}} \qquad E_{m,n} = const \times ($$

* ؟ آیا دسته‌ی طیف هروقت En ، متناسب با $\frac{1}{n^2}$
و هروقت [illegible] متناسب با n^2 است

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

$$E = mc^2 = \sqrt{m_0^2 c^4 + p^2 c^2}$$
$$photons(m_0 = 0) \rightarrow h\nu = pc$$

$$p = \frac{h\nu}{c} = \frac{h}{\lambda} \rightarrow \lambda = \frac{h}{p}$$
$$h \cong 6.626 \times 10^{-34} \text{ Joule.sec}$$

This is called De Broglie's wavelength.

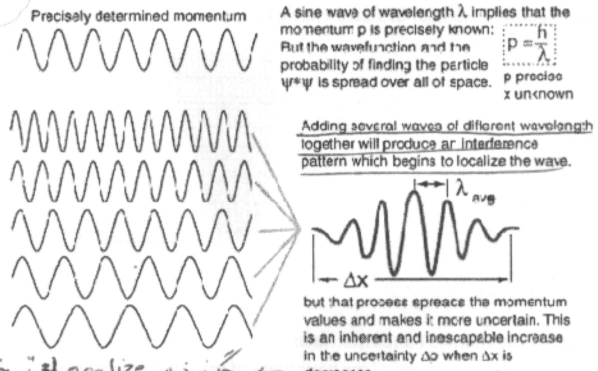
Uncertainty principle:

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 (بالا و پایین رفتن) 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.

← به عبارتی ذره localized خواهد بود.

Mathematically we describe the uncertainty principle as the following

$$\Delta x \Delta p \geq \frac{1}{2}\hbar \qquad \text{or} \qquad \Delta E \Delta t \geq \frac{1}{2}\hbar$$



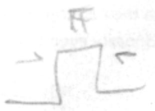
A sine wave of wavelength λ implies that the momentum p is precisely known: But the wavefunction and the probability of finding the particle ψ*ψ is spread over all of space.

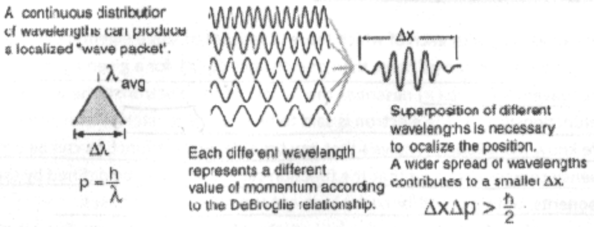
Adding several waves of different wavelength together will produce an interference pattern which begins to localize the wave.

but that process spreads the momentum values and makes it more uncertain. This is an inherent and inescapable increase in the uncertainty Δp when Δx is decreases.

$$\Delta x \Delta p > \frac{\hbar}{2}$$

← چگونه ذره localize تر خواهد گردید طبق شکل؟





One of the implications of this principle is to calculate the confinement energy.

limited

Schrödinger Equation:

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

$$-\frac{\hbar^2}{2m_0}\nabla^2\Psi = \frac{p^2}{2m_0}\Psi$$

But $p^2/2m_0$ is the kinetic energy of electron and in general

Total energy (E) = Kinetic energy ($K.E.$) +Potential energy (V)

Hence we could rewrite the above equation as:

$$-\frac{\hbar^2}{2m_0}\nabla^2\Psi = (E - V(r))\Psi$$

Or

$$\left(-\frac{\hbar^2}{2m_0}\nabla^2 + V(r)\right)\Psi = E\Psi$$

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

$$E = \sqrt{m_0^2c^4 + p^2c^2} \approx m_0c^2\left[1 + p^2c^2/2m_0^2c^4 + \ldots\right]$$

$$\begin{aligned} E - m_0c^2 &= V + (p^2/2m_0) \\ \searrow \quad & \qquad\qquad \downarrow \\ hf = \hbar\omega &= V + (\hbar^2k^2/2m_0) \\ \hbar\omega &= (\hbar^2k^2/2m_0) + V \end{aligned}$$

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

$$d\Psi/dt = -i\omega\Psi \quad \text{and} \quad d^2\Psi/dx^2 = -k^2\Psi$$

$$i\hbar\frac{d\Psi}{dt} = \left(-\frac{\hbar^2}{2m_0}\frac{d^2\Psi}{dx^2}\right) + V\Psi$$

Assume

$$-\frac{\hbar^2}{2m_0}\frac{d^2\Psi}{dx^2} + U(x)\Psi = i\hbar\frac{d\Psi}{dt} \qquad \Psi(x,t) = \psi(x)e^{-iEt/\hbar}$$

$$-e^{\frac{-iEt}{\hbar}}\frac{\hbar^2}{2m_0}\frac{d^2\psi(x)}{dx^2} + e^{\frac{-iEt}{\hbar}}U(x)\psi(x) = i\hbar\frac{-iE}{\hbar}\psi(x)e^{\frac{-iEt}{\hbar}}$$

$$-\frac{\hbar^2}{2m_0}\frac{d^2\psi}{dx^2} + U(x)\psi = E\psi$$

$$\frac{d^2\psi}{dx^2} + \frac{2m_0}{\hbar^2}(E - U)\psi = 0$$

Probability densities:

In practice $P(r)$, the probability of finding an electron at specific point r, is proportional to $|\Psi(r)|^2$, where $\Psi(r)$ is the solution of the Schrodinger's equation. Without changing this equation one can normalize wave function such that

$$\int |\Psi(r)|^2 \, d^3r = 1$$

Operators:

In quantum mechanics, a system is always presumed to be in some "state". The state may be described by some set of numbers $c_1, c_2, \ldots$. We could write all of these numbers out as a vector $[c_1, c_2, \ldots]^T$. For example, we are saying that the state of the electron may be described by the set of numbers that is the set of values of the wavefunction at each point in space, $[\psi(x_1), \psi(x_2), \ldots]^T$, where $x_1, x_2, \ldots$ are all the points in the space. (The vector for our wavefunction would have an infinite number of elements, one for each point in space, though we will ignore the mathematical complications of this infinity, and think of the wavefunction as a finite but very large vector.) There are other numbers required actually to completely specify the state of an electron, including its "spin" (+1 (up) or −1 (down)), but we will omit this here for simplicity.)

A convenient way of thinking of an operator (and one that defines it as a "linear" operator) is to think of the operator as a matrix M that multiplies the vector. In this way, the result depends possibly on all of the values in the vector that describe the state

$$M\begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix}$$

When the vector is an eigenvector of the matrix (or the state is an eigenstate of the operator), we are saying that we have the remarkable situation that this process results in a vector of exactly the same form as when we started, multiplied only by a constant that is the "eigenvalue"

$$M\begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix} = a\begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix}$$

The Schrödinger equation is just such an eigenvalue equation, though up to now we have only though of it with $\hat{H}$ as a differential operator. We could also think of it as a matrix, in which case we would find it as some kind of "stripe" diagonal قطری 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 $\hat{H} \equiv -\frac{\hbar^2}{2m_0}\nabla^2 + V(r)$ 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. Schrodinger equation now can be written as $\hat{H}\,\Psi = E\,\Psi$. ($\hat{H}$ is an operator but E is simply a number, the energy of electron) This equation can be read as:

When $\hat{H}$ operates on Ψ, the result is the total energy E.

Momentum Operator:

We can see for the specific case of a plane wave in the z direction $\psi \propto exp(ikz)$ that

$$\frac{\hbar}{i}\frac{\partial}{\partial z}\psi \propto \hbar k\psi = p\psi$$

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

$$\hat{p} \equiv \frac{\hbar}{i}\nabla$$

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

$$\frac{1}{2m_0}\hat{p}\cdot\hat{p} = -\frac{\overset{\text{۲}}{\hbar}}{2m_0}\nabla^2$$

The (vector) momentum is

$$\vec{p} = \hbar\vec{k}$$

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

Eigenstates:

Starting from Schrödinger equation

$$\left(-\frac{\hbar^2}{2m_0}\nabla^2 + V(r)\right)\Psi \Rightarrow E\,\Psi$$

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 (∇^2) 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

X در انتگرال فقط یکی باید * باشد

"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 eigenfunctions$\psi_n(r)$, the electron is said to be in an "eigenstate", with corresponding total energyE_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 measureable quantity form a complete set. The eigenfunctions of such an operator are all "orthogonal" to one another.

$$X \quad \int \psi_n^*(r)\psi_m^*(r)\,d^3r = 0, \quad n \neq m$$

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

$$\psi(r) = \sum_n a_n\,\psi_n(r)$$

where a_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 measureable 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" (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

$$E_{photon} = \hbar\omega$$

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

$$E = \frac{p^2}{2m} + V$$

X $e^{-iE_m t/\hbar}$

The energy E is related to the angular frequency ω, 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:

↓
guess

$$i\hbar\frac{\partial \Psi}{\partial t} = \hat{H}\Psi$$

or

$$\left(-\frac{\hbar^2}{2m_0}\nabla^2 + V(r)\right)\Psi = i\hbar\frac{\partial \Psi}{\partial t}$$

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

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

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

$$\left(-\frac{\hbar^2}{2m_0}\nabla^2 + V(r)\right)\psi(r) = E\psi(r)$$

(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 $t = 0$

$$\boxed{\Psi(r,0) = \sum_m c_m\,\Phi_m(r)}$$

Where Φ_m is an eigenstate at the energy E_m, i.e.

$$\boxed{\hat{H}\Phi_m = E_m\Phi_m}$$

At an arbitrary time t, the state evolves to

$$\boxed{\Psi(r,t) = \sum_m c_m\,e^{iE_m t/\hbar}\Phi_m(r)} \quad !? \quad e^{-iE_m t/\hbar}$$

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

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

Solution will be in form of

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

X

چرا $e^{-iEt/\hbar}$, $e^{+iEt/\hbar}$
داریم! آیا زمان در هر دو جهت است.

Adding the time dependence

$$\Psi(x,t) = \underset{\substack{\downarrow \\ \text{wave in}+x\text{ direction}}}{A_+ e^{i\left(kx-\frac{Et}{\hbar}\right)}} + \underset{\substack{\downarrow \\ \text{wave in}-x\text{ direction}}}{A_- e^{-i\left(kx-\frac{Et}{\hbar}\right)}}$$

For a free particle:

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

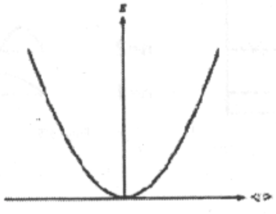


Figure 7

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

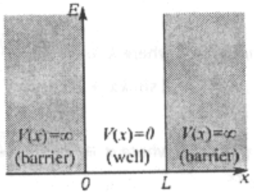


Figure 8

Schrodinger 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. Cosider $V(x)$ as

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

As $V = \infty$ outside the box there is no possibility of finding particle outside the box, hence ψ must be zero at the walls ot 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

$$\psi_n(x) = A_n \sin\frac{n\pi x}{L}$$

Where A_n is a normalization constant. Eigenenergies will be as

$$E_n = \frac{\hbar^2}{2m}\left(\frac{n\pi}{L}\right)^2 \qquad n = 1,2,\cdots$$

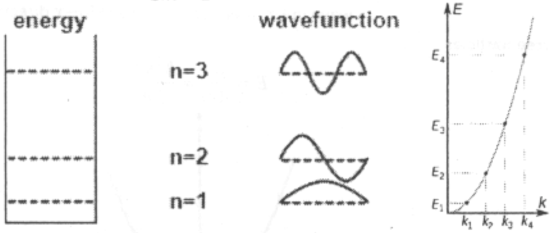


Figure 9

Finite potential well:

Again for simpilicity we consider 1-D finite potential well, the time-indepent schrodinger 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 $E > U_0$: $\frac{d^2\psi(x)}{dx^2} + k^2\psi(x) = 0$ where $k \equiv \frac{\sqrt{2m(E-U_0)}}{\hbar}$

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

If $E < U_0$: $\frac{d^2\psi(x)}{dx^2} - \alpha^2\psi(x) = 0$ where $\alpha \equiv \frac{\sqrt{2m(U_0-E)}}{\hbar}$

so $\psi(x) = De^{-\alpha x} + Ee^{\alpha x}$

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

Step 1. solve wave equation $\frac{d^2\psi(x)}{dx^2} + k^2\psi(x) = 0 \rightarrow 2N$ unkonwns.

Step 2. Wavefunction at infinity: $\psi(x=-\infty) = \psi(x=+\infty) = 0 \rightarrow 2N-2$ unkonwns

Step 3. Continuity of wavefunction: for remaining $N-1$ borders between N regions:

$\psi|_{x_B^-} = \psi|_{x_B^+}$ and $\left.\frac{d\psi}{dx}\right|_{x_B^-} = \left.\frac{d\psi}{dx}\right|_{x_B^+}$ → set $2N-2$ equations. (DONE!)

Step 4. find coefficients by setting ***det***(coefficient matrix) $= 0$

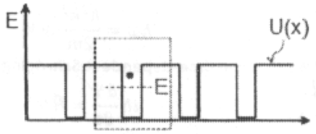
Step 5. Normalization $\int_{-\infty}^{+\infty}|\psi(x)|^2dx = 1$ دوطرف region ها، $x=+\infty$, $x=-\infty$ هستند.

Once ψ is known, anything else can be computed by its operator as;

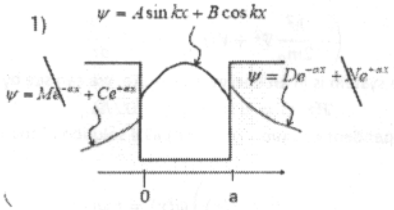
$$p = \int_{-\infty}^{+\infty} \psi^* \left[\frac{\hbar}{i}\frac{d}{dx}\right]\psi\, dx$$

$$E = \int_{-\infty}^{+\infty} \psi^* \left[-\frac{\hbar}{i}\frac{d}{dt}\right]\psi\, dx$$

Step 1&2:



1)



2) Boundary conditions

$\psi(x=-\infty) = 0$

$\psi(x=+\infty) = 0$

Step 3:

$$C = B$$
$$\alpha C = -kA$$
$$A\sin ka + B\cos ka = De^{-\alpha a}$$
$$kA\cos ka - kB\cancel{A}\sin ka = -\alpha De^{-\alpha a}$$

In matrix form:

$$\begin{pmatrix} 0 & 1 & -1 & 0 \\ k & 0 & \alpha & 0 \\ \sin ka & \cos ka & 0 & -e^{-\alpha a} \\ \cos ka & -\sin ka & 0 & \alpha e^{-\alpha a}/k \end{pmatrix}\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

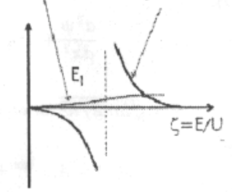
Step 4:

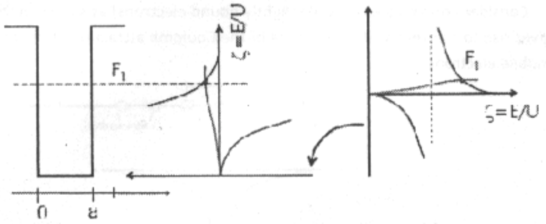
$$det\begin{pmatrix} 1 & 1 & -1 & 0 \\ k & 0 & \alpha & 0 \\ \sin ka & \cos ka & 0 & -e^{-\alpha a} \\ \cos ka & -\sin ka & 0 & \alpha e^{-\alpha a}/k \end{pmatrix} = 0$$

$$\tan\left(\alpha a\sqrt{\xi}\right) = \frac{2\sqrt{\xi(1-\xi)}}{2\xi-1} \quad \text{where } \xi \equiv \frac{E}{U_0}, \alpha \equiv \sqrt{\frac{2mU_0}{\hbar^2}}$$

Graphic solution

$$\tan(\alpha_0 a\sqrt{\xi}) = \frac{2\sqrt{\xi(1-\xi)}}{2\xi-1}$$





Step 5:

$$\begin{pmatrix} 1 & 1 & -1 & 0 \\ k & 0 & \alpha & 0 \\ \sin ka & \cos ka & 0 & -e^{-\alpha a} \\ \cos ka & -\sin ka & 0 & \alpha e^{-\alpha a}/k \end{pmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & \alpha & 0 \\ \cos ka & 0 & -e^{-\alpha a} \end{pmatrix} \begin{bmatrix} B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ -kA \\ -A\sin ka \end{bmatrix}$$

$$\begin{bmatrix} B \\ C \\ D \end{bmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & \alpha & 0 \\ \cos ka & 0 & -e^{-\alpha a} \end{pmatrix}^{-1} \begin{bmatrix} 0 \\ -kA \\ -A\sin ka \end{bmatrix}$$

B, C, and A is written in terms of A , now A can be found as:

$$\int_{-\infty}^{+\infty} |\psi|^2 dx \quad = 1 \quad \int_{-\infty}^{0} C^2 e^{2\alpha x} dx + \int_{0}^{a} (A\sin kx + B\cos kx)^2 dx + \int_{a}^{+\infty} D^2 e^{-2\alpha x} dx$$

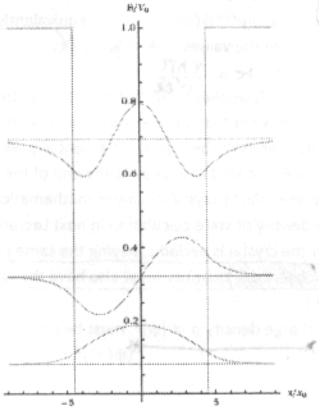


Figure 10

and → an

- Crystal potential energy structure
- Bloch theorem
- Near free-electron model
- Kronig-Penney model

What is Electronic Band Structure?

The ranges of energy that an electron is *forbidden* or *allowed* to have, is called electronic band structure. The material's electronic and optical properties depend on its band structure. We saw that the electrons in the simple Bohr's model may only have certain energies. Considering Pauli Exclusion Principle and electron's spin at each energy level, two electrons can exist. By adding another atom, each electron is under influence of its own nucleus and the nucleus of the next atom, hence each energy level will split into two energy levels close to one another to accommodate (جا دادن) 4 electrons. Similarly, set of N atoms will make N different energy levels for $2N$ electrons, which are so close that can be considered as a band of allowed energy. Between this band and the next allowed energy band there is a forbidden gap that no electron can exist with that energy. These band energies are plotted for electrons with different kinetic energies, in this chapter we are looking after the allowed energies for electrons for different certain momentum.

?! تا چه تعدادی اتم به هم باند شوند به این شکل میگردد آیا محدودیتی نیست.

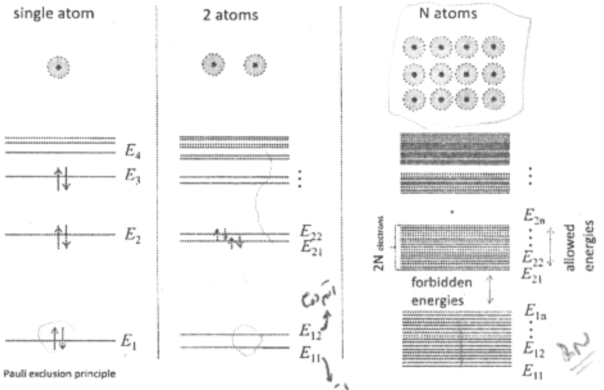


Figure 1:

1-D System:

Consider and ideal lattice with no defects or imperfections. For simplicity, we examine a 1-D case assume that N atoms in the crystal with a lattice constant a. Think of all the electrons (except the one that we are interested in) and all the nuclei in the crystal as giving some fixed potential $U(x)$ in the x direction. That single electron that we are interested in moves against this essentially fixed background. This assumption is known as *Hartree-Fock* approximation.

⊛ فرض Hartree fock : الکترون آزاد در این ساختار که به شدت به هم وابسته است حرکت میکند و آزادی دارد. البته بقیه الکترونها fix هستند.

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.

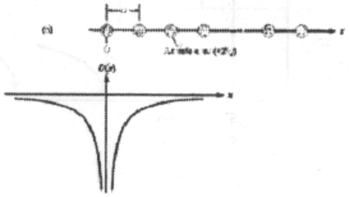


Figure 2

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:

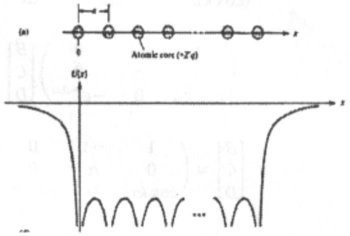


Figure 3

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 $\psi(x+a) = e^{ika}\psi(x)$ or, equivalently, $\psi(x) = e^{ikx}u(x)$, $u(x) = u(x+a)$

where k can take on the values: $k = \cancel{2\pi}n/\cancel{2\pi}n$ while $n = 0, \pm1, \pm2, \cdots, \pm N/2$

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

Bloch theorem strictly applies to an "infinite" crystal. In studying an infinite crystal, however, a common trick is to consider first a finite crystal, with N atoms, and impose a periodic boundary condition at the both ends of the crystal. So essentially we imagine connecting the right-most unit cell to the left-most unit cell in a kind of a loop. At the end of the calculation, we will then assume N to be very large. In this way, the infinite crystal becomes mathematically more tractable. You will encounter this trick again in the density of state calculation in next Lectures.

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 observable quantity 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

$$|\psi(x)|^2 = |\psi(x+a)|^2$$

which means

چرا $\rho \propto |\psi|^2$

charge density

$$\psi(x) = C\,\psi(x+a)$$

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 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.,

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

Or

$$C = exp\left(2i\pi(\frac{s}{N}+m)\right) \quad ; \quad s = 0,1,2,\cdots,N-1 \text{ and } m \text{ any integer}$$

Replacing C into equation for ψ:

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

Where

$$k = \frac{2\pi s}{Na} \quad ; \quad s = 0,1,2,\cdots,N-1$$

Or

$$k = \frac{2\pi n}{Na} \quad ; \quad n = 0,\pm1,\pm2,\cdots,\pm N/2$$

ی سریها .

Which still gives essentially N states, nut 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

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

!؟ k فقط در رنج $-\frac{\pi}{a}$ تا $+\frac{\pi}{a}$ خواهد بود، چگونه است که در سایر قسمتها نیز به تناوب تکرار می‌گردد؟

Where $\vec{a}$ is any crystal lattice vector, and considering the three crystal basis vector directions, 1, 2, and 3, with lattice constants a_1, a_2, and a_3, and numbers of atoms N_1, N_2, and N_3,

$$k_1 = \frac{2\pi n}{N_1 a} \quad ; \quad n = 0,\pm1,\pm2,\cdots,\pm\frac{N_1}{2}$$

and similarly for the other two components of k in the other two crystal basis vector directions.

Bragg scattering in one-dimension: (Kittle از روی کتاب)

Bragg scattering occurs when wave interacts with a periodic structure whose periodicity is comparable to the wavelength. In one-dimension, assuming that we have a periodic array of scatters, each scatters is characterized by a small reflectivity r.

The total reflectivity is then

$$R = \sum_m re^{i2kma}$$

π/a ۲π/a

which diverges when the Bragg condition $ka = n\pi$ 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 other. Thus,

$$\begin{cases} K = \frac{۲\pi}{\lambda} \\ k = \frac{\pi}{a} \end{cases} \Rightarrow \lambda = ۲a$$

the incident wave will be completely reflected back if the crystal is large enough. The eigenstate of the system when Bragg condition is satisfied should therefore be standing waves.

چرا ؟!

Nearly-Free Electron Model:

Suppose we have a wave with wavevector $k = \pm\pi/a$, where a is the lattice constant in the direction of the wave or $\lambda = 2a$ will satisfy the "Bragg reflection" condition for waves along this axis → will get "standing waves."

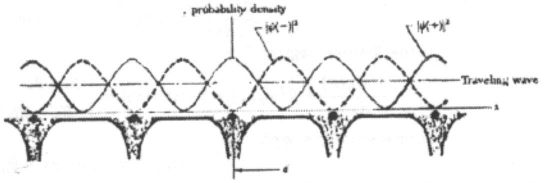
If the potential energy is very "weak", the standing waves will be very close to sinusoidal waves (i.e., the electron is "nearly free")

$$\psi(+) = e^{i\pi x/a} + e^{-i\pi x/a} = 2\cos\frac{\pi x}{a}$$
$$\psi(-) = e^{i\pi x/a} - e^{-i\pi x/a} = 2i\sin\frac{\pi x}{a}$$

=> . چرا از کجا آمده است؟

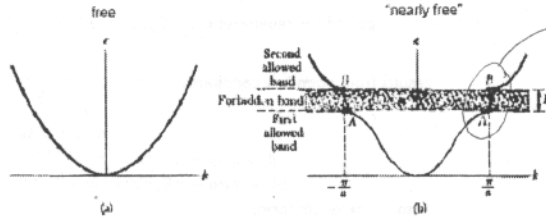
With corresponding charg densities

$$\rho(+) = |\psi(+)|^2 \propto \cos^2\frac{\pi x}{a}$$
$$\rho(-) = |\psi(-)|^2 \propto \sin^2\frac{\pi x}{a}$$



the two probability densities have their maxima at different points in the unit cell. $\rho(+)$ has maxima at the atomic cores, where the potential energy is low and $\rho(-)$ has maxima between the atomic cores, where the potential energy is high. there is an "energy gap" between these two possibilities

More detailed calculation in this model yields the energy – momentum relationship below



در این نواحی الکترون نوعی تغییر در تراز را حس می‌کند گویی که نوعی فشار به آن وارد می‌گردد.

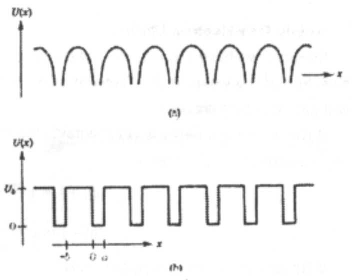
emergence of "energy gap"
region of energy where there are no propagating (or standing) waves possible

=> Kittle از روی

Kronig-Penney Model:

take non-trivial (though not necessarily very realistic) potential that we can solve completely, rectangular "Kronig-Penney" potential as:

k — $k + \frac{2\pi}{a}$



Previously we solved shrodinger eduation in the "well" region: $0 < x < a$

$$\frac{d^2\psi_a(x)}{dx^2} + \alpha^2\psi_a(x) = 0 \quad \text{where } \alpha = \sqrt{2mE/\hbar^2}$$

In the "barrier" region: $0 > x > -b$

$$\frac{d^2\psi_b(x)}{dx^2} + \beta^2\psi_b(x) = 0 \quad \text{where } \beta = \begin{cases} i\beta_- \ ; \beta_- = \sqrt{2m(U_0 - E)/\hbar^2} & (0 < E < U_0) \\ \beta_+ \ \ ; \beta_{\mp} = \sqrt{2m(E - U_0)/\hbar^2} & (E > U_0) \end{cases}$$

Formal solution in well:

$$\psi_a(x) = A_a \sin\alpha x + B_a \cos\alpha x$$

Formal solution in barrier:

$$\psi_b(x) = A_b \sin\beta x + B_b \cos\beta x$$

Boundary conditions:

continuity requirment: $\psi_a(0) = \psi_b(0) \quad ; \quad \left.\frac{d\psi_a}{dx}\right|_0 = \left.\frac{d\psi_b}{dx}\right|_0$

periodicity requirment: $\psi_a(a) = e^{ik(a+b)}\psi_b(-b) \quad ; \quad \left.\frac{d\psi_a}{dx}\right|_0 = e^{ik(a+b)}\left.\frac{d\psi_b}{dx}\right|_{-b}$

results from boundary conditions

$$B_a = B_b$$
$$\alpha A_a = \beta A_b$$
$$A_a \sin\alpha a + B_a \cos\alpha a = e^{ik(a+b)}[-A_b \sin\beta b + B_b \cos\beta b]$$
$$\alpha A_a \cos\alpha a - \alpha B_a \sin\alpha a = e^{ik(a+b)}[\beta A_b \cos\beta b + \beta B_b \sin\beta b]$$

+k از رابطه $\psi(x+a) = e^{ika}\psi(x)$
بروز می نماید.

In mmatrix vector form:

$$\begin{pmatrix} \sin\alpha a + (\alpha/\beta)e^{ik(a+b)}\sin\beta b & \cos\alpha a - e^{\widehat{ik(a+b)}}\cos\beta b \\ \alpha\cos\alpha a - \alpha e^{ik(a+b)}\cos\beta b & -\alpha\sin\alpha a - \beta e^{ik(a+b)}\sin\beta b \end{pmatrix} \begin{bmatrix} A_a \\ B_a \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Taking the determinant of the matrix and setting it equal to zero as usual for linear equantion, gives the "characteristic equation" that determines the conditions under which the system has solutions:

$$\boxed{-\frac{\alpha^2 + \beta^2}{2\alpha\beta}\sin\alpha a \cdot \sin\beta b + \cos\alpha a \cdot \cos\beta b = \cos k(a+b)}$$

Finding the mathematical solutions to this equation will determine the allowed energies and the wavefunctions associated with them

---

X sinh → sin

A more convenient form

$$\beta = \begin{cases} i\beta_- \ ; & (0 < E < U_0) \\ \beta_+ \ ; & (E > U_0) \end{cases}$$

Defining

$$\alpha_0 \equiv \sqrt{2mU_0/\hbar^2} \quad ; \quad \xi \equiv E/U_0$$

Such that

$$\alpha = \alpha_0\sqrt{\xi} \quad ; \quad \beta_- = \alpha_0\sqrt{1-\xi} \quad ; \quad \beta_+ = \alpha_0\sqrt{\xi - 1}$$

gives

for $0 < E < U_0$:

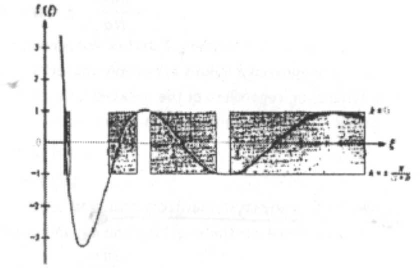
$$\frac{1 - 2\xi}{2\sqrt{\xi(1-\xi)}}\sin\alpha_0 a\sqrt{\xi} \cdot \sinh\alpha_0 b\sqrt{1-\xi} + \cos\alpha_0 a\sqrt{\xi} \cdot \cos\alpha_0 b\sqrt{1-\xi} = \cos k(a+b)$$

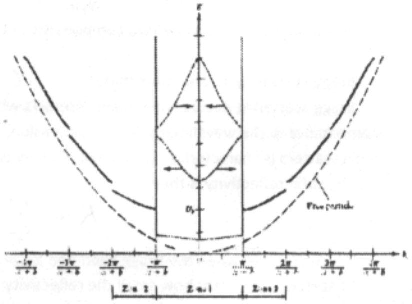
for $E > U_0$:

$$\frac{1 - 2\xi}{2\sqrt{\xi(\xi-1)}}\sin\alpha_0 a\sqrt{\xi} \cdot \underline{\cancel{\sinh}}\,\alpha_0 b\sqrt{\xi - 1} + \cos\alpha_0 a\sqrt{\xi} \cdot \cos\alpha_0 b\sqrt{\xi - 1} = \cos k(a+b)$$

Left hand side of these euqations depend only on E and righ hand sides only on k, hence determins the allowed values of E corresponing to a given k .

$f(\xi)$ is the left hand side of equations above for specific choice of $\alpha_0 a = \alpha_0 b = \pi$. Values of $|f(\xi)| > 1$ correspond to no wave solution – band gap.





- Brillouin zone and Zone folding
- Particle motion, Group Velocity, Effective mass
- Carrier and Electron and Hole Current

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

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

and

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

We should rewrite the Schrodinger equation as:

$$\frac{-\hbar^2}{2m}\left[\frac{d^2}{dx^2} + 2ik\frac{d}{dx} - k^2\right]u + Vu = Eu \; (*)$$

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, $u_{n,k}(x)$, and (generally) different eigenvalues for the energy, $E_{n,k}(x)$, (where n 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 graphed. In fact, there is a major simplification possible here that also reconciles these two pictures.

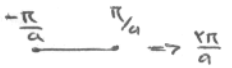
Suppose we know a solution of the above Equation for some particular value of $k = k_0$

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

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



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

Where

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

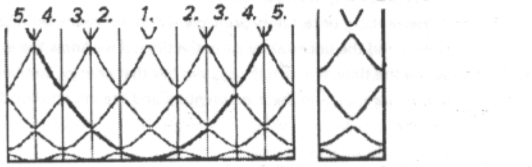
Note that exp $(i\,2\pi jx/a)$ is also periodic with the lattice periodicity. Hence, $v(x)$ is also periodic with the lattice periodicity (being the product of two functions with the lattice periodicity).

Hence we only need to know the solutions for a range of k of a total of 2π/a, and we automatically know the solutions for any other k. Therefore, to solve for the eigen solutions for the case of $k = k_0 + 2\pi j/a$, we can substitute the $\psi(x) = v(x)e^{ik_0x}$ into the Schrödinger equation, which gives us exactly the same result as Eq. (*) above, except with u replaced by v. Since the equations are identical, the set of solutions must be identical. Hence, if $u_{q,k_0}(x)$ is a solution for the case of $k = k_0$, with associated eigenenergy $E_{q,k}$, then it is also a solution for the case of $k = k_0 + 2\pi j/a$.

A specific range of size $2\pi/a$ is known as a "Brillouin zone", conventionally chosen so that the first

Hence, we have proved that the band structure is repetitive in k-space, repeating with period $2\pi/a$. Therefore, we only need to know the solutions for a range of k of a total of $2\pi/a$, and we automatically know the solutions for any other k .

Brillouin zone is symmetrically disposed about $k = 0$, i.e. from $-\pi/a$ to π/a .



Note that all the properties of the band structure are completely specified by specifying all the bands in one Brillouin zone, usually called the "folded zone" representation. There are exactly N, equally-spaced, allowed values of k in a given band in a given Brillouin zone. Note the states in this "first" Brillouin zone account for all the possible states in the crystal. There are N states in each band, one for each atom, and one band corresponding to each atomic state in the atom. The states in the other Brillouin zones are the same states as in the first Brillouin zone – they are not distinct states.

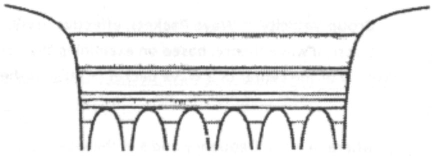
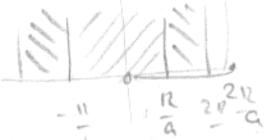


Figure 1: Energy Bands – in "Real Space"



X $v_g = \frac{d\omega}{dk}$ ۶ X $f(x,t) = \text{۲} \sin\left(\frac{\omega t - kx}{\text{۲}}\right) \cos\left(\frac{\delta\omega t - \delta k x}{\text{۲}}\right)$

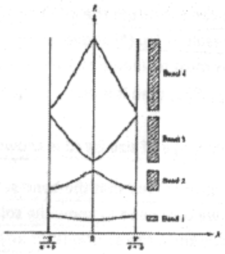


Figure 2 : Energy Bands – in "k-Space", for Kronig-Penney example

Particle Motion and Effective Mass:

Heisenberg uncertainty principle states, for example, $\Delta E \Delta t \ge \hbar$, for the uncertainty ΔE in the energy and the uncertainty Δt in the time; if we know the energy accurately, we cannot accurately define the time at which the system has that energy. Similarly $\Delta p_x \Delta x \ge \hbar$, for the uncertainty in momentum, Δp_x, in the x momentum and the uncertainty Δx 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"

$$v_g = \frac{d\omega}{dx} \quad X$$

where ω 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

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

Since

$$\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$$

then

$$f(t,x) = 2\sin(\omega t - kx)\cos(\omega t - kx) \quad X \qquad X$$

which can be viewed as an underlying wave $2\sin(\omega t - kx)$ modulated by an envelope $\cos(\omega t - kx)$.

X $v_g = \frac{d\omega}{dk}$

The underlying wave moves at the "phase velocity"

$$v_p = \frac{\omega}{k}$$

but the envelope moves at a "group velocity"

$$v_g = \frac{\delta\omega}{\delta x} \quad X$$

or, in the limit of very small $\delta\omega$ and δk,

$$v_g = \frac{d\omega}{dx} \quad X$$

For the quantum mechanical case, since a wave of a given energy E has a time dependence $\propto \exp(iEt/\hbar)$, $\omega \equiv E/\hbar$, and so

$$v_g = \frac{1}{\hbar}\frac{dE}{dk}$$

is the group velocity of a quantum mechanical electron wavepacket.

Suppose now we apply an "external" force F to the moving wavepacket (or equivalently to the moving "particle"), the work done in applying the force through a distance dx is

$$dE = F dx$$

The distance dx is equal to the group velocity, v_g, times the time, dt , for which the force is applied, so we have

$$dE = F dx = F v_g dt$$

Or making use of $v_g = dE/\hbar\, dk$

$$F = \frac{d(\hbar k)}{dt}$$

(Note, incidentally, that, since "effective" momentum is essentially $\hbar k$, we are now saying that the force is equal to the rate of change of momentum, in agreement with the classical situation)

To find out how fast we are accelerating the wavepacket, we differentiate the group velocity with respect to time, to obtain

$$\frac{dv_g}{dt} = \frac{1}{\hbar}\frac{d}{dt}\left(\frac{dE}{dk}\right) = \frac{1}{\hbar}\frac{d^2E}{dk^2}\frac{dk}{dt} = \frac{1}{\hbar^2}\frac{d^2E}{dk^2}\frac{d(\hbar k)}{dt}$$

Or equivalently using the expression for F from

$$\frac{dv_g}{dt} = \frac{1}{\hbar^2}\frac{d^2E}{dk^2}F$$

which we can rewrite as

$$F = m^* \frac{dv_g}{dt}$$

where the quantity

$$m^* = \left(\frac{1}{\hbar^2}\frac{d^2E}{dk^2}\right)^{-1}$$

برای الکترون


$m^* > ۰$


$m^* < ۰$

خمیدگی

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 be where we are most likely to find electrons participating in transport (i.e., transport of charge, or current). There

$$E - E_{edge} \cong (\text{constant})\left(k - k_{edge}\right)^2$$

so

!?

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

چرا ثابت می‌شود؟!

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 in the linear "crystal." There are N states 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.

Obviously, an empty band can carry no current – there are no electrons. 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

Notice that

$$I = \frac{q}{L}\sum_i v_i n_i = \frac{q}{\hbar L}\sum_i \left(\frac{dE}{dk_i}\right) n_i$$

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 bands (if 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.)

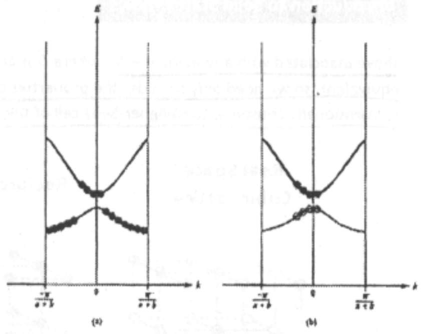


Figure 3

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 $\vec{a}_1, \vec{a}_2$, and $\vec{a}_3$, the wavefunction should satisfy the Bloch theorem:

$$\psi(\vec{r}) = e^{i\vec{k}.\vec{r}}u(\vec{r})$$

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

in other words

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

$$E(\vec{k} + \vec{G}) = E(\vec{k})$$

- در واقع برای شبکه Bravais است که با بردار خواص کل شبکه را تکرار می‌کنیم.

where $e^{i\vec{G}.\vec{a}} = 1$ for any lattice vector a. The set $\{\vec{G}\}$ is the reciprocal lattice for the crystal.

The reciprocal lattice of a Bravais lattice is the set of all vectors $\vec{K}$ such that

$$e^{i\vec{K}.\vec{R}} = 1$$

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

$$\vec{b}_1 = 2\pi\frac{\vec{a}_2 \times \vec{a}_3}{\vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)}$$

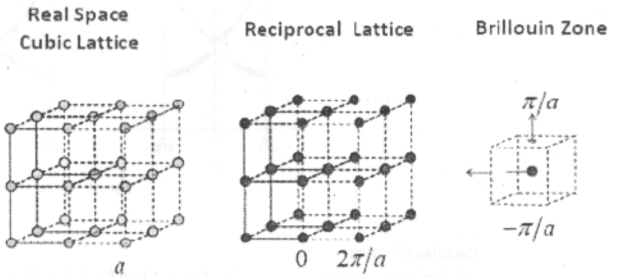
$$\vec{b}_2 = 2\pi\frac{\vec{a}_3 \times \vec{a}_1}{\vec{a}_2 \cdot (\vec{a}_3 \times \vec{a}_1)}$$

$$\vec{b}_1 = 2\pi\frac{\vec{a}_1 \times \vec{a}_2}{\vec{a}_3 \cdot (\vec{a}_1 \times \vec{a}_2)}$$

Or

$$\boxed{\left[\vec{b}_1\,\vec{b}_2\vec{b}_3\right]^T = 2\pi[\vec{a}_1\,\vec{a}_2\vec{a}_3]^{-1}}$$

There are corresponding reciprocal lattice vectors $\vec{G} = m_1\vec{b}_1 + m_2\vec{b}_2 + m_3\vec{b}_3$ 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 those associated with any point $\vec{k}+\vec{G}$, where $\vec{G}$ is any reciprocal lattice vector. All Brillouin zones are equivalent, so we need only consider the properties of the first Brillouin zone. The first Brillouin zone is conventionally chosen as the Wigner-Seitz cell of the reciprocal lattice.



Follow W-S algorithm, but now for reciprocal lattice

Figure 4: Brillouin zone in cubic lattice

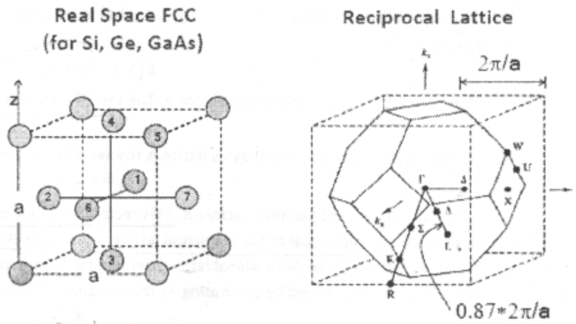
از Fcc شروع

per[illegible] ⇓

$a' = \frac{۲\pi}{a}$ ، $b' = \frac{۲\pi}{b}$

$a' = \frac{۲\pi}{\frac{\sqrt{۲}}{۲} p}$ ، $b' = \frac{۲\pi}{p}$ p = lattic Constant

$a' = \sqrt{۲} \times \frac{۲\pi}{p}$ ، $b' = \frac{۲\pi}{p} \Longrightarrow$



Note unlike cubic lattice, zone edge is not at π/a

Figure 5:

⟹ * شبکه Fcc در kspace bcc خواهد بود.

Each point (hkl) in the reciprocal lattice corresponds to a set of lattice planes (hkl) in the real space lattice. The direction of the reciprocal lattice vector corresponds to the normal to the real space planes. The magnitude of the reciprocal lattice vector is given in reciprocal length and is equal to the reciprocal of the interplanar spacing of the real space planes.

!؟ ⟵

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.

$$F(\vec{k}) = \int_{-\infty}^{\infty} f(r) e^{-i\vec{k}\cdot\vec{r}} d^3r$$

!؟

مگر unique vector ای که بیان ساخته
با تکراری کنیم G نیست

A reciprocal lattice is a periodic set of points in this space, and contains the $\vec{k}$ 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:

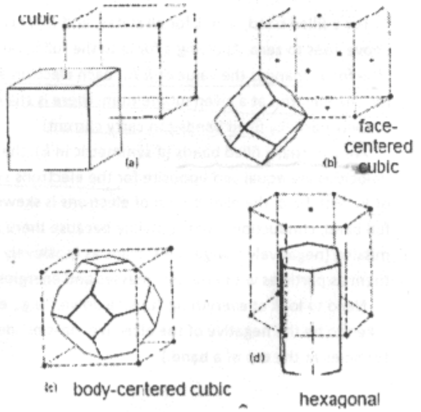
A common "standardized" choice of primitive unit cell. The Wigner-Seitz cell is the region round about a lattice point that is closer to that point than to any other point.

The Witner-Seitz unit cell can be constructed by

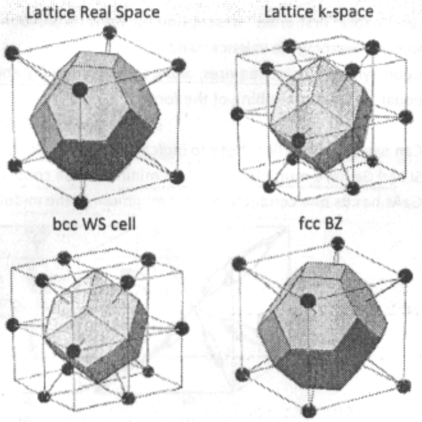
(i) drawing lines between the point and (in principle) all other points in the lattice (in practice only the points reasonably close to the one of interest need be considered),

(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

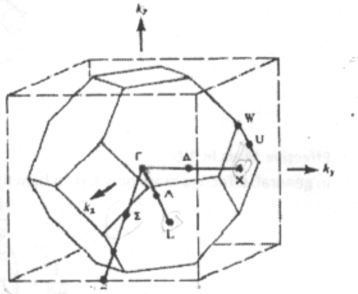


Figure 6

Notations:

Γ - identifies zone center (k=0)

X - denotes zone end along a <100> direction

L - denotes zone end along a <111> direction

Note that the size of Brillouin zone is different in the different directions.

Usually we plot band structures along only a few, specific directions, because of the symmetry of semiconductor crystals. Most interesting properties usually found along "high-symmetry" directions (e.g., band maxima and minima.) Because diamond structures have a center of symmetry, only need to plot for +ve (or −ve) k values in any particular direction

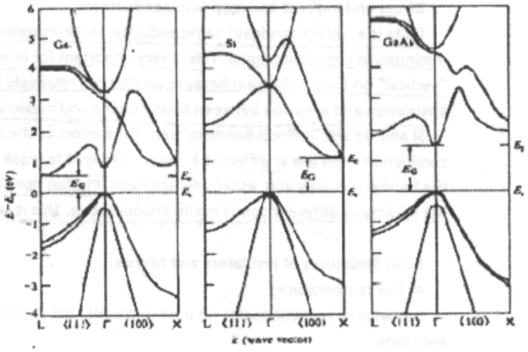


Figure 7

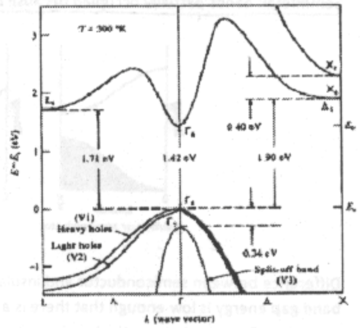


Figure 8: GaAs band structure

Features:

Valence bands: essentially full of electrons ; maximum at Γ ; three valence bands (heavy hole, light hole and split off hole)

؟! همه در Γ ماکس دارند

Conduction bands: essentially empty of electrons ;

Ge: minimum at L, six equivalent minima; ۴تا نگفته ؟!

Si: minimum at 0.8(2π/a) along the ΓX direction, eight equivalent minima → ؟ آیا در راستای Σ مین ندارد.

GaAs: minimum at Γ, one minimum

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

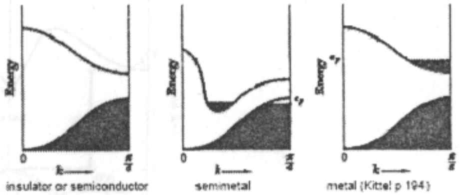


Figure 9

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, $E - E_c$, just above the energy, E_c, of the minimum or minima. In general:

$$E - E_c \cong Ak_1^2 + Bk_2^2 + Ck_3^2$$

because the curvature of the band structure is, in general, different in the three different crystal directions

In simple cubic symmetry, at zone center, must have the same behavior in all three directions, so might expect something like

$$E - E_c \cong A(k_1^2 + k_2^2 + k_3^2)$$

reasonable "first order" description for GaAs conduction band at zone center (though can have significant "warping" in valence bands)

چرا دو جهت دیگر کدام است که
symmetry ندارد

? Cubic symmetry also requires, along the direction ΓX and ΓL, that two of the coefficients A, B, and C be equal, so have something of the form

$$E - E_c \cong Ak_1^2 + B(k_2^2 + k_3^2)$$

Can use equations like these to indicate "constant energy surfaces"

Si and Ge have multiple equivalent minima in the conduction bands (8 for Ge, 6 for Si)

GaAs has its one conduction band minimum in the middle of the Brillouin zone

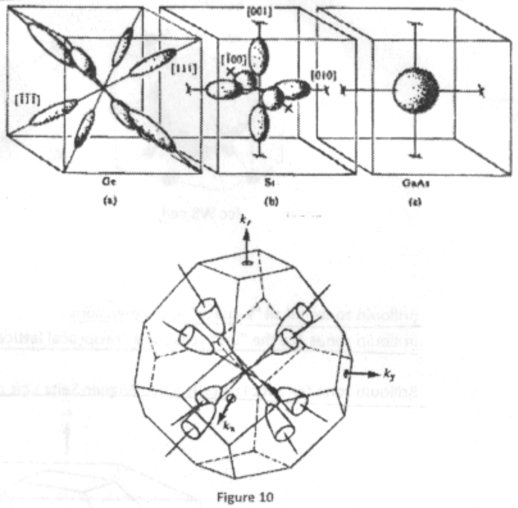


Figure 10

Effective Mass in 3-D:

In general, effective mass now has to be written as a tensor

$$\frac{d\vec{v}}{dt} = \frac{1}{m^*} \cdot \vec{F}$$

$$\frac{1}{m^*} = \begin{pmatrix} m_{xx}^{-1} & m_{xy}^{-1} & m_{xz}^{-1} \\ m_{yx}^{-1} & m_{yy}^{-1} & m_{yz}^{-1} \\ m_{zx}^{-1} & m_{zy}^{-1} & m_{zz}^{-1} \end{pmatrix}$$

where, in general

$$\boxed{\frac{1}{m_{ij}} = \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k_i \partial k_j}}$$

Note in general that acceleration and force are not colinear. the particle does not necessarily go in the direction you push it!

هم خط ، برای استاد

$$\frac{d\vec{v}}{dt} = m_{xx}^{-1}F_x\hat{a}_x + m_{yx}^{-1}F_x\hat{a}_y + m_{zx}^{-1}F_x\hat{a}_z$$

For conduction bands at zone center in diamond and zinc-blende semiconductors, behavior of band is approximately isotropic and "parabolic"

$$E - E_c = A(k_1^2 + k_2^2 + k_3^2)$$

So masses are formally

$$m_{xx}^{-1} = m_{yy}^{-1} = m_{zz}^{-1} = \frac{2A}{\hbar^2} \qquad \frac{dE}{dk_1} = ۲Ak_1 \rightarrow \boxed{\frac{d^۲E}{dk_1^۲} = ۲A}$$

غیرقطری

and all "off-diagonal" terms in the effective mass tensor are (approximately) zero. hence can define single scalar electron effective mass, m_e^*

$$E - E_c = \frac{\hbar^2}{2m_e^*}(k_1^2 + k_2^2 + k_3^2)$$

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

$$(\ast)\ \frac{d\vec{v}}{dt} = \frac{\vec{F}}{m^*}$$

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

$$E - E_c = Ak_1^2 + B(k_2^2 + k_3^2) \longrightarrow$$

چرا نزدیک نقطه مینیمم اینگونه میشود؟

where k_1, k_2 , and k_3 are along the principal axis. (For Si, k_1 is along the <100> directions, for Ge, k_1 is along the <111> directions).

so we can define

$$m_{11}^{-1} = \frac{2A}{\hbar^2}$$

$$m_{22}^{-1} = m_{33}^{-1} = \frac{2B}{\hbar^2}$$

leading to

$$E - E_c = \frac{\hbar^2}{2m_l^*}k_1^2 + \frac{\hbar^2}{2m_t^*}(k_2^2 + k_3^2) \qquad \ldots \qquad \text{Ge, Si}$$

بیضی گون

$$\frac{m_l^*}{m_t^*} = \left(\frac{\text{Length of the ellipsoid along the axis of revolution}}{\text{Maximum width of the ellipsoid perpendicular to the axis-of revolution}}\right)^2$$

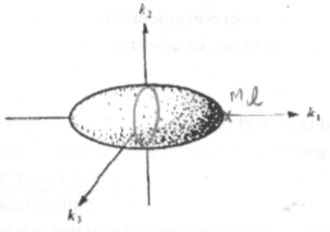


Figure 11

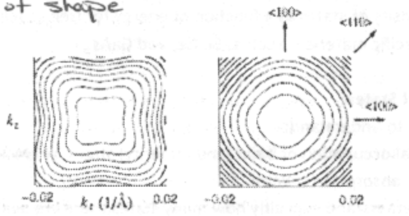
Valence Bands in Diamond and Zinc-Blende Crystals:

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

$$E_v - E = Ak^2 \pm \left[B^2k^4 + C^2(k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2)\right]^{1/2}$$

Note this has "cross terms" between the different directions (e.g, $k_x^2k_y^2$) and shows valence band "warping"

warp: turn out of shape



Heavy Hole

Light Hole

Figure 12

- Density of states in k-space (periodic boundary condition)
- Density of states in k-space (complete reflecting boundary condition)
- Density of states as a function of energy for free space
- Specific materials, such as Si, Ge, and GaAs

Density of States:

Is needed to understand:

1. Thermal occupation of band (and impurity) states by electrons and/or holes
2. Optical absorption spectra

✓ ⊛ 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

✓ ⊛ Density of states (in energy): number of possible electron states per unit energy per unit volume

To do so we start by calculating how many states there are per unit k.

✓ ⊛ 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

$$\frac{2\pi}{a}, \frac{2\pi}{b}, \frac{2\pi}{c}$$

⊛ a, b و c طول قطعه device است.

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)

$a = L = N a_{unitcell}$

Cyclic boundary conditions: correctly "counts" all the states if we consider states only in one Brillouin zone. In a total range of k_x of $2\pi/a_{unitcell}$, for example, we will have N_x states, where N_x is the number of unit cells in the x direction, since $N_x a_{unitcell} = a$

فرق a و $a_{unitcell}$ ؟ ✓

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

چیه؟

چگونه؟! ✓

$$N_k = \frac{abc}{(2\pi)^3}$$

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)

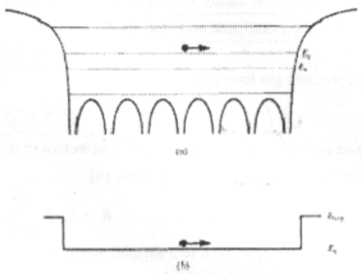
$-\frac{\pi}{a_{unitcell}}$ —— $\frac{\pi}{a_{unitcell}}$ => N_x حالت داریم

از طرفی $N_x \cdot a_{unitcell} = a$ =>

$\frac{2\pi}{a_{unitcell}}$	N_x حالت
واحد طول (در فضای k)	N حالت

=> $N = \frac{N_x \cdot a_{unitcell}}{2\pi} = \frac{a}{2\pi}$

تعداد حالتها در واحد طول را نشان می‌دهد $N = \frac{a}{2\pi}$ ⊛



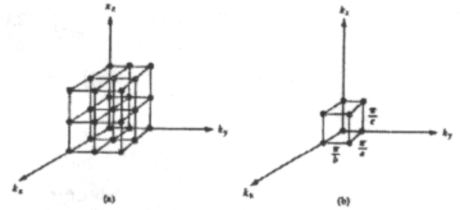
Requiring the waves to be ~ zero at the walls of the box (regardless of the specific form of the unit cell wavefunction) requires that the "envelope" be standing waves, so

$$\psi(x,y,z) = u(x,y,z)\sin k_x x \sin k_y y \sin k_z z$$

where

$$k_x = \frac{n_x\pi}{a}\ ;\ k_y = \frac{n_y\pi}{a}\ ;\ k_z = \frac{n_z\pi}{a} \quad (n_x, n_y, n_z = \pm1, \pm2, \pm3, \cdots)$$

This would give



$$\left(\frac{\text{solutions}}{\text{unit volume of k-space}}\right) = \frac{abc}{\pi^3}$$

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

$$\left(\frac{\text{allowed energy states}}{\text{unit volume of k-space}}\right) = \frac{abc}{8\pi^3}$$

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

$$\left(\frac{\text{allowed electron energy states}}{\text{unit volume of k-space}}\right) = \frac{abc}{4\pi^3}$$

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

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

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

$$(\text{energy states with } k \text{ between } k \text{ and } k+dk) = 4\pi k^2 dk\,(\frac{abc}{4\pi^3})$$

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

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

so the number of states within this energy range is

$$(\text{energy states with } E \text{ between } E \text{ and } E+dE) = (abc)\frac{m\sqrt{2mE}}{\pi^2\hbar^3}dE$$

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

$$g(E) = \frac{(\text{energy states with } E \text{ between } E \text{ and } E+dE)}{abc.\,dE}$$

$$g(E) = \frac{m\sqrt{2mE}}{\pi^2\hbar^3}$$

Density of States in Energy in Conduction and Valence Bands:

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

$$g_c(E) = \frac{m_n^*\sqrt{2m_n^*(E-E_c)}}{\pi^2\hbar^3} \qquad E \geq E_c$$

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

$$g_v(E) = \frac{m_p^*\sqrt{2m_p^*(E_v-E)}}{\pi^2\hbar^3} \qquad E \leq E_v$$

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 $m_n^* = m_p^*$ (this mass is very light ($\sim 0.07 m_0$) which corresponds to steeply curved band, low overall density of states)

Density of States in Si and Ge Conduction Band:

The N_{el} minima in conduction band, each approximately "ellipsoidal"

$$E - E_c = \frac{\hbar^2}{2m_l^*}k_1^2 + \frac{\hbar^2}{2m_t^*}(k_2^2+k_3^2)$$

Can be rewrite in form of:

$$\frac{k_1^2}{\alpha^2} + \frac{k_2^2+k_3^2}{\beta^2} = 1$$

where

$$\alpha \equiv \frac{\sqrt{2m_l^*(E-E_c)}}{\hbar} \quad ; \quad \beta \equiv \frac{\sqrt{2m_t^*(E-E_c)}}{\hbar}$$

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

$$N_{el}\left(\frac{4}{3}\pi\alpha\beta^2\right) = \frac{4}{3}\pi k_{eff}^3$$

where k_{eff} is the radius of sphere of the same volume.

A single density of state equivalent mass m_n^* can be defined as:

$$(m_n^*)^{3/2} = N_{el}(m_l^* m_t^{*2})^{1/2}$$

Giving:

$$m_n^* = 6^{2/3}(m_l^* m_t^{*2})^{1/3} \quad \cdots \text{Si}$$

$$m_n^* = 4^{2/3}(m_l^* m_t^{*2})^{1/3} \quad \cdots \text{Ge}$$

- **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

$$g_v(E) = \frac{m_p^*\sqrt{2m_p^*(E_v-E)}}{\pi^2\hbar^3} = \frac{m_{hh}^*\sqrt{2m_{hh}^*(E_v-E)}}{\pi^2\hbar^3} + \frac{m_{lh}^*\sqrt{2m_{lh}^*(E_v-E)}}{\pi^2\hbar^3}$$

or

$$m_p^* = [(m_{hh}^*)^{3/2} + (m_{lh}^*)^{3/2}]^{2/3}$$

- 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 "indirect" 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 ; "optical" and acoustic)
- *dispersion relation* between frequency, ω, and wavevector $\vec{k}$ 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 of frequency 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)

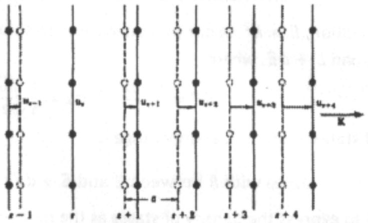


Figure 1: Longitudinal vibrations

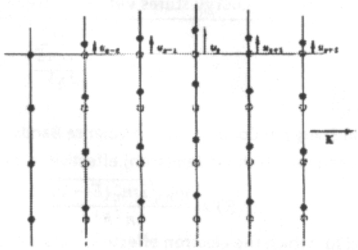


Figure 2: Transverse Vibrations

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_s pushing a given plane is the balance of the forces from the two adjacent planes of atoms:

$$F_s = C(u_{s+1} - u_s) + C(u_{s-1} - u_s)$$

where C is an effective *spring constant*. Hence the equation of motion of the plane of atoms of mass M is:

$$M\frac{d^2u_s}{dt^2} = C(u_{s+1} + u_{s-1} - 2u_s)$$

We make an intelligent guess that there are solutions to this equation that are monochromatic oscillations of some frequency ω , i.e., with some time dependent part of the form $exp(\pm i\omega t)$. Hence we have

$$-M\omega^2 u_s = C(u_{s+1} + u_{s-1} - 2u_s)$$

We now make the further intelligent guess that these solutions are waves of the form $u_s = e^{i(Kz \pm \omega t)}$. With this assumption, we conclude that

$$u_{s\pm1} = ue^{isKa}e^{\pm iKa}$$

Replacing the above Eq

$$-M\omega^2 ue^{isKa} = Cu\{e^{i(s+1)Ka} + e^{i(s-1)Ka} - 2e^{isKa}\}$$

$$X(s) = \int x(t)\, e^{-i\omega t}\, dt$$

$$X'(s) = \int x(t+a)\, e^{-i\omega t}\, dt = X(s)\, e^{i\omega a}$$

Or

$$M\omega^2 = -C\{e^{iKa} + e^{-iKa} - 2\} = 2C(1 - \cos Ka)$$

$$\omega = \sqrt{\frac{4C}{M}}\left|\sin\frac{Ka}{2}\right|$$

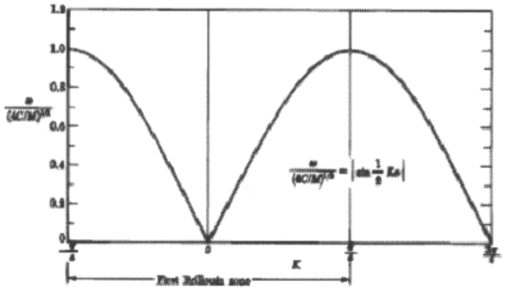


Figure 3: The "dispersion relation", between ω and K is clearly periodic

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 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

$$v_g = d\omega/dK$$

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

$$v_g = \sqrt{\frac{Ca^2}{M}}\cos\tfrac{1}{2}Ka$$

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

$$\omega = a\sqrt{\frac{C}{M}}K$$

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 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 .

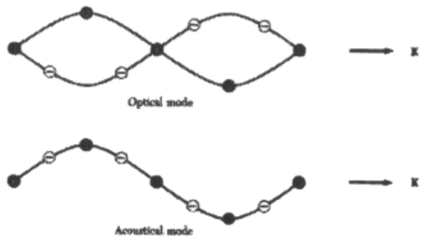


Figure 4

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

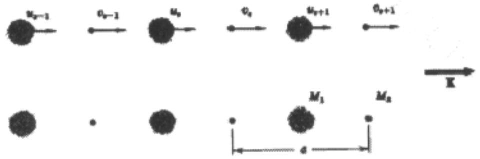


Figure 5

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:

$$M_1\frac{d^2u_s}{dt^2} = C(v_{s+1} + v_{s-1} - 2u_s)$$
$$M_2\frac{d^2v_s}{dt^2} = C(u_{s+1} + u_{s-1} - 2v_s)$$

Proposing wave equation as:

$$u_s = ue^{isKa}e^{-i\omega t} \quad ; \quad v_s = ve^{isKa}e^{-i\omega t}$$

So

$$-M_1\omega^2 u = Cv\left(1+e^{-iKa}\right) - 2Cu$$
$$-M_2\omega^2 v = Cu\left(1+e^{+iKa}\right) - 2Cv$$

In matrix form

$$\begin{pmatrix} 2C - M_1\omega^2 & -C\left(1+e^{-iKa}\right) \\ -C\left(1+e^{+iKa}\right) & 2C - M_2\omega^2 \end{pmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

To have a solution the determinant should be zero

$$\begin{vmatrix} 2C - M_1\omega^2 & -C\left(1+e^{-iKa}\right) \\ -C\left(1+e^{+iKa}\right) & 2C - M_2\omega^2 \end{vmatrix} = 0$$

which gives

$$M_1M_2\omega^4 - 2C(M_1+M_2)\omega^2 + 2C(1-\cos Ka) = 0$$

quadratic – gives two solutions ω for each K:

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

$$\omega^2 \cong \frac{C/2}{M_1+M_2}K^2a^2 \qquad \text{near } K = 0$$
$$\omega^2 \cong \frac{2C}{M_1} \qquad \text{at } K = \pm\pi/a$$

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)

$$\omega^2 \cong 2C\left(\frac{1}{M_1}+\frac{1}{M_2}\right) \qquad \text{near } K = 0$$
$$\omega^2 \cong \frac{2C}{M_2} \qquad \text{at } K = \pm\pi/a$$

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

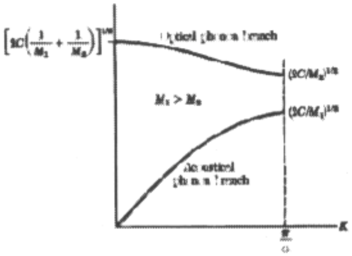


Figure 6

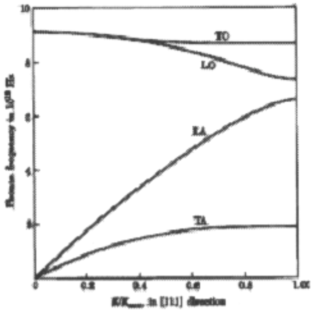


Figure 7: Phonon Dispersion Relations in Germanium. TA – transverse acoustic , LA – longitudinal acoustic, TO – transverse optical, LO – longitudinal optical, One LA, one LO, two TA and two TO

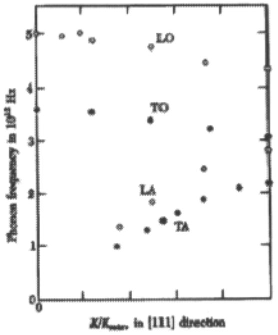


Figure 8: Phonon Dispersion Relations in KBr

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)

$$E = \frac{p^2}{2m} + \tfrac{1}{2}Cx^2 \qquad (*)$$

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:

$$\left[-\frac{\hbar^2}{2M}\frac{\partial^2}{\partial x^2} + Cx^2\right]\varphi(x) = E\varphi(x)$$

We can now solve this equation for the allowed values of E, and the associated wavefunctions $\varphi(x)$. If we were simply dealing with a mass M on a spring, for example, $|\varphi(x)|^2$ 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

$$E_n = \left(n + \tfrac{1}{2}\right)\hbar\omega \quad (**)$$

Where $n = 0,1,2,\cdots$ 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 ~10s of meV. Phonons can also be viewed as having a "crystal" momentum $\hbar K$ if they have a wavevector K.

- Joint density of states
- Absorption in direct band gap semiconductors
- Absorption in indirect band gap semiconductors

Semiconductors (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 semiconductors as far as optical properties are concerned are:

1. Direct gap semiconductors
    - strong optical absorption and emission above the bandgap energy
    - very abrupt onset of optical absorption
2. Indirect gap semiconductors
    - relatively weaker optical absorption and emission above the bandgap energy
    - relatively smooth onset of optical absorption

Direct Gap Optical Absorption:

Optical wavelength λ is very long compared to the separation between atoms a (Typical wavelength of interest ~ 1 µm; Lattice constant of GaAs ~ 0.5 nm). Therefore optical wavevector $k_{opt} = 2\pi/\lambda$ 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 $\hbar k$.

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 a transition, absorbing a photon in the process, i.e., we would have a transition rate

$$W = A\,\delta\left(E_c(\vec{k}) - E_v(\vec{k}) - \hbar\omega\right)$$

Here A is a constant representing the strength of the absorption corresponding to this transition, $E_c(\vec{k})$ is the energy in the conduction band and $E_v(\vec{k})$ the energy in the valence band corresponding to the wavevector $\vec{k}$, 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:

$$W_{TOT} = A\sum_{\vec{k}} \delta\left(E_c(\vec{k}) - E_v(\vec{k}) - \hbar\omega\right)$$

This formula tells us how the transition rate, and therefore the absorption coefficients for the incident light, varies with the frequency ω 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)

$$\sum_{\vec{k}} \approx \int_{\vec{k}} g(\vec{k})d^3\vec{k}$$

where $g(\vec{k})$ is the density of states in k-space. Change variables in the integral to energy $E_J = E_c(\vec{k}) - E_v(\vec{k})$. Assuming parabolic bands, can define E_J as

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

m_h is valence band effective mass, and reduced effective mass, μ_{eff}, is defined as

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

hence, define "joint density of states," $g_J(E_J)$, and write

$$g_J(E_J)dE_J = g(\vec{k})d^3\vec{k}$$

leading, for $E_J \geq E_g$, to

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

hence

$$W_{TOT} = A\sum_{\vec{k}} \delta\left(E_c(\vec{k}) - E_v(\vec{k}) - \hbar\omega\right) = A\int_{E_J \geq E_g} \frac{1}{2\pi^2}\left(\frac{2\mu_{eff}}{\hbar^2}\right)^{3/2}\left(E_J - E_g\right)^{1/2}\delta(E_J - \hbar\omega)dE_J$$

i.e (for $\hbar\omega \geq E_g$)

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

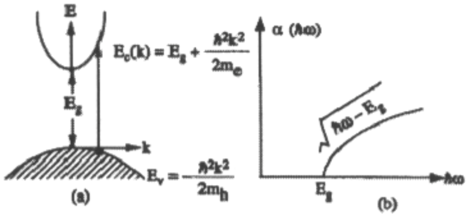


Figure 1

Optical absorption length:

Since α 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

$$\frac{dn_p}{dz} = -n_p\alpha$$

$$n_p(z) = n_p(0)e^{-\alpha z}$$

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

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

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

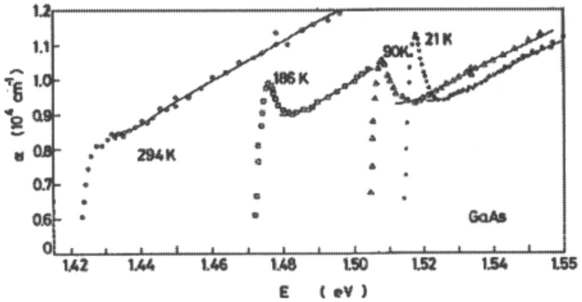


Figure 2: Optical absorption of GaAs at various temperatures

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

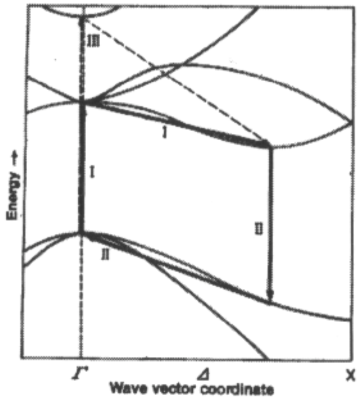


Figure 3

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

$$\alpha \propto B_{abs}\left(\hbar\omega + E_{phonon} - E_g\right)^2 + B_{em}\left(\hbar\omega - E_{phonon} - E_g\right)^2$$

where E_{phonon} is the phonon energy, and the factors B_{abs} and B_{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)

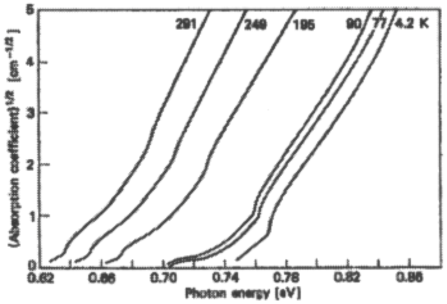


Figure 4: Absorption spectrum of Ge (plotted so that quadratic increase of absorption with energy gives a straight line)

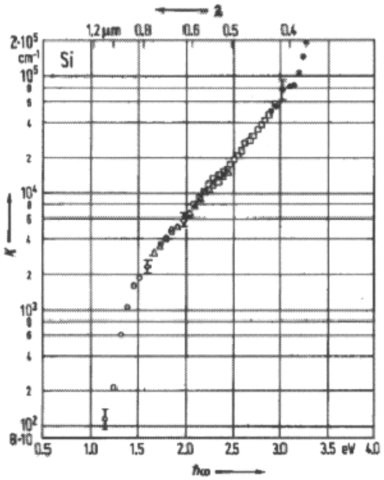


Figure 5: Absorption spectrum of Si

Indirect absorption spectra do not have the abruptly rising character of the direct absorption spectrum (generally significantly weaker than direct absorption.)

- 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

$$\uparrow_1 \uparrow_2 \downarrow_3 \downarrow_4 \downarrow_5 \uparrow_6 \downarrow_7 \uparrow_8 \uparrow_9 \uparrow_{10}$$

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

$$(\uparrow_1+\downarrow_1)(\uparrow_2+\downarrow_2)(\uparrow_3+\downarrow_3)\cdots(\uparrow_N+\downarrow_N) \qquad (*)$$

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

$$(\uparrow_1+\downarrow_1)(\uparrow_2+\downarrow_2) = \uparrow_1\uparrow_2+\uparrow_1\downarrow_2+\downarrow_1\uparrow_2+\downarrow_1\downarrow_2$$

The above statement reads in words as: 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

$$M = N, N-2, N-4, \cdots, -N+2, -N$$

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.g.,

there is only one state with $M = N$, i.e., $\uparrow\uparrow\uparrow\uparrow\cdots\uparrow\uparrow\uparrow\uparrow$

but there are N states with $M = N-2$, i.e., $\downarrow\uparrow\uparrow\uparrow\cdots\uparrow\uparrow\uparrow\uparrow$, $\uparrow\downarrow\uparrow\uparrow\cdots\uparrow\uparrow\uparrow\uparrow$, $\uparrow\uparrow\downarrow\uparrow\cdots\uparrow\uparrow\uparrow\uparrow$, etc

Enumeration of States: For simplicity in subsequent algebra, we introduce notation. $N_\uparrow$ as number of magnets up, and $N_\downarrow$ as number of magnets down. Presume total number of magnets N is even for simplicity. Define "spin excess", s, as

$$N_\uparrow - N_\downarrow = 2s$$

so

$$N_\uparrow = \frac{N}{2} + s \quad ; \quad N_\downarrow = \frac{N}{2} - s$$

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 correct 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:

$$(\uparrow + \downarrow)^N$$

Like for $N = 2$ we have

$$(\uparrow + \downarrow)^N = \uparrow\uparrow + 2\uparrow\downarrow + \downarrow\downarrow$$

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

$$(x+y)^N = \sum_{s=-N/2}^{N/2} \frac{N!}{\left(\frac{N}{2}+s\right)!\left(\frac{N}{2}-s\right)!} x^{\frac{N}{2}+s} y^{\frac{N}{2}-s}$$

Or we can rewrite $(\uparrow + \downarrow)^N$ as:

$$(\uparrow + \downarrow)^N = \sum_{s=-N/2}^{N/2} \frac{N!}{\left(\frac{N}{2}+s\right)!\left(\frac{N}{2}-s\right)!} \uparrow^{\frac{N}{2}+s} \downarrow^{\frac{N}{2}-s}$$

The coefficient

$$g(N,s) = \frac{N!}{\left(\frac{N}{2}+s\right)!\left(\frac{N}{2}-s\right)!} = \frac{N!}{N_\uparrow!\, N_\downarrow!}$$

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

s	g(10,s)
-5	1
-4	10
-3	45
-2	120
-1	210
0	252
1	210
2	120
3	45
4	10
5	1

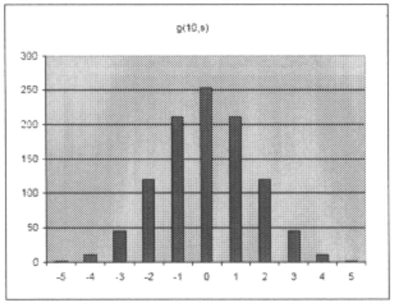


Figure 1: Multiplicity function for 10 magnets

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 factorial 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

$$\log N! \cong \tfrac{1}{2}\log 2\pi + \left(N + \tfrac{1}{2}\right)\log N - N$$

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

$$g(N,s) \cong g(N,0)e^{-2s^2/N}$$

where

$$g(N,0) = (2/\pi N)^{1/2} 2^N$$

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 $\sim 10^{-11}$ 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

states can have "degeneracy," means several quantum states with identical energy. example of what is known as "multiplicity" in thermal physics, number of different states with the same energy. For example: 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

$$\langle f \rangle = \sum_s f(s)P(s)$$

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 $2N$ 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

$$\langle s^2\rangle = \frac{(2/\pi N)^{1/2}2^N}{2^N}\int dss^2\, e^{-2s^2/N} = (2/\pi N)^{\frac{1}{2}}(N/2)^{\frac{3}{2}}\int_{-\infty}^{+\infty} dxx^2e^{-x^2}$$

$$\langle s^2\rangle = \frac{N}{4} \quad ;\ \langle(2s)^2\rangle = N$$

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^{20} particles, the fluctuation F is $\sim 10^{-10}$!!!!

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 the g accessible states is

$$P(q) = 1/g$$

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

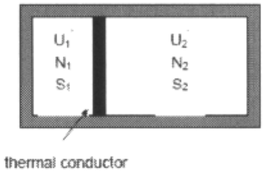


Figure 2: U: Energy, N: Number of particles, S: System

The question is: What determines how much energy flows from one system to the other? (not simply which system has the most energy). Most probably division of energy is that for which the combined system has the maximum number of accessible states. Let's try to understand the characteristics of the condition that corresponds to the maximum number of accessible states for the combined system.

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_1, which we will call $\hat{s}_1$, that corresponds to the configuration with the highest multiplicity, i.e., the most probable configuration for which $g_1(N_1, \hat{s}_1)\cdot g_2(N_2, s - \hat{s}_1)$ is the largest. We know that, for a large system, the configurations for which s_1 is close to $\hat{s}_1$ 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 $\hat{s}_1$, hence the properties of the system tend in practice to be the properties of those configurations with s_1 is close to $\hat{s}_1$.

We expect to find the system in the most probable configuration (or those with very similars_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 equilibration*. 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

$$U(s) = U_1(s_1) + U_2(s_2) = -2mB(s_1 + s_2) = -2mBs$$

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

$$g_1(N_1, s_1) \cdot g_2(N_2, s_2) = g_1(0)g_2(0)\exp\left(-\frac{2s_1^2}{N_1} - \frac{2s_2^2}{N_2}\right)$$

Or in terms of $s = s_1 + s_2$

$$g_1(N_1, s_1) \cdot g_2(N_2, s - s_1) = g_1(0)g_2(0)\exp\left(-\frac{2s_1^2}{N_1} - \frac{2(s - s_1)^2}{N_2}\right)$$

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

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

Differentiating to find the maximum, we have at a maximum

$$\frac{\partial\big(g_1(N_1, s_1)g_2(N_2, s - s_1)\big)}{\partial s} = -\frac{4s_1}{N_1} + \frac{4(s - s_1)}{N_2}$$

(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

$$\frac{s_1}{N_1} = \frac{s - s_1}{N_2} = \frac{s_2}{N_2}$$

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 $\hat{s}_1$ and $\hat{s}_2$ denoting the values of s_1 and s_2 at the maximum multiplicity (i.e., in the most probably configuration), we can rewrite

$$\frac{\hat{s}_1}{N_1} = \frac{\hat{s}_2}{N_2} = \frac{s}{N}$$

Hence at our most probably configuration

$$(g_1 g_2)_{\max} = g_1(\hat{s}_1)g_2(s - \hat{s}_1) = g_1(0)g_2(0)e^{-2s^2/N}$$

Now let us investigate how sharp this maximum is. Consider a small deviation δ from the maximum, so that $s_1 = \hat{s}_1 + \delta$; $s_2 = \hat{s}_2 - \delta$

hence,

$$g_1(N_1, s_1)g_2(N_2, s_2) = (g_1 g_2)_{\max}\exp\left(-\frac{4\hat{s}_1\delta}{N_1} - \frac{2\delta^2}{N_1} + \frac{4\hat{s}_2\delta}{N_2} - \frac{2\delta^2}{N_2}\right)$$

$$g_1(N_1, \hat{s}_1 + \delta)g_2(N_2, \hat{s}_2 \overset{-}{\oplus} \delta) = (g_1 g_2)_{\max}\exp\left(-\frac{2\delta^2}{N_1} - \frac{2\delta^2}{N_2}\right)$$

If either of N_1 or N_2 is large, this is a very sharp distribution

e.g., for $N_1 = N_2 = 10^{22}$ and $\delta = 10^{12}$, i.e., $\delta/N_1 = 10^{-10}$, then $2\delta^2/N_1 = 200$, and the product $g_1 g_2$ is reduced to $e^{-400} \approx 10^{-174}$ 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

$$g(N, U) = \sum_{U_1} g_1(N_1, U_1)g_2(N_2, U - U_1)$$

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

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

Remember that the notation $(\partial g_1/\partial U_1)_{N_1}$ 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 infinitesimally 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

$$dg = \left(\frac{\partial g_1}{\partial U_1}\right)_{N_1} g_2 dU_1 - g_1 \left(\frac{\partial g_2}{\partial U_2}\right)_{N_2} dU_1$$

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 $\sigma(N,U) \equiv \log g(N,U)$. Then we have, for the configuration with the largest multiplicity (the most probably configuration)

$$\left(\frac{\partial \sigma_1}{\partial U_1}\right)_{N_1} = \left(\frac{\partial \sigma_2}{\partial U_2}\right)_{N_2}$$

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 are 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.,

$$\frac{1}{T} = k_B \left(\frac{\partial \sigma}{\partial U}\right)_N$$

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 can be more convenient to work with "fundamental temperature", τ, which we can define as

$$\frac{1}{\tau} = \left(\frac{\partial \sigma}{\partial U}\right)_N$$

or $\tau = k_B T$. Note that the fundamental temperature has dimensions of energy, which is the real unit of temperature – other units (Kelvin, Celcius, Fahrenheit, Rankine) are technically redundant even if practically convenient (300K is ~ 4.14 x 10^{-21}Joules)

In classical thermodynamics

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)_N$$

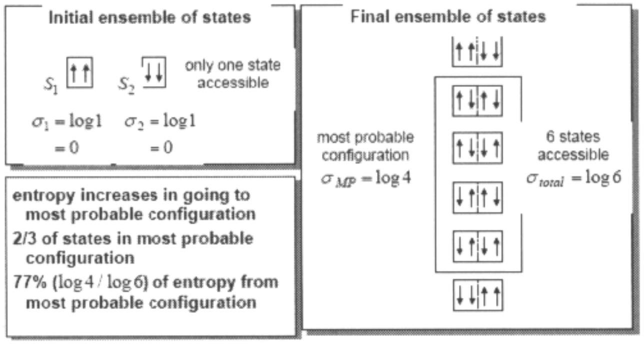
where the (thermodynamic) entropy S corresponds with our statistical "fundamental" entropy through

$$S = k_B \sigma$$

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 increases in going to most probable configuration

2/3 of states in most probable configuration

77% ($\log 4 / \log 6$) of entropy from most probable configuration

Entropy and heat flow:

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

$$\Delta\sigma = \left(\frac{\partial \sigma_1}{\partial U_1}\right)_{N_1}(-\Delta U) + \left(\frac{\partial \sigma_2}{\partial U_2}\right)_{N_2}(\Delta U) = \Delta U\left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)$$

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 0.1J of energy has been transferred? (Presume large objects (large thermal mass) so that there has been negligible change in temperature during the transfer.)

$$\Delta S_1 = \frac{-0.1\text{ J}}{350\text{ K}} = -2.86 \times 10^{-4}\text{ JK}^{-1}$$
$$\Delta S_2 = \frac{0.1\text{ J}}{290\text{ K}} = 3.45 \times 10^{-4}\text{ JK}^{-1}$$
$$\Delta S = \Delta S_1 + \Delta S_2 = 0.59 \times 10^{-4}\text{ JK}^{-1}$$

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

$$\Delta\sigma = \frac{0.59 \times 10^{-4}}{k_B} = 0.43 \times 10^{19}$$

i.e., the number of states accessible to the system has increased by a factor $e^{0.43 \times 10^{19}}$

Law of increase of entropy:

Proof that: total entropy always increases when two systems are brought into thermal contact. If total energy $U = U_1 + U_2$ is a constant, the total multiplicity (i.e., the total number of accessible states) after the systems are in thermal contact, is

$$g(U) = \sum\nolimits_{U_1} g_1(U_1) g_2(U - U_1)$$

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 $g_1(U_{1o})g_2(U-U_{1o})$ of the total system before the two parts were brought into thermal contact (where U_{1o} is the initial energy of system 1). but the sum also contains various other terms, all of which are positive. Hence

$$g(U) > g_1(U_{1o})g_2(U-U_{1o})$$

Therefore the total entropy has increased.

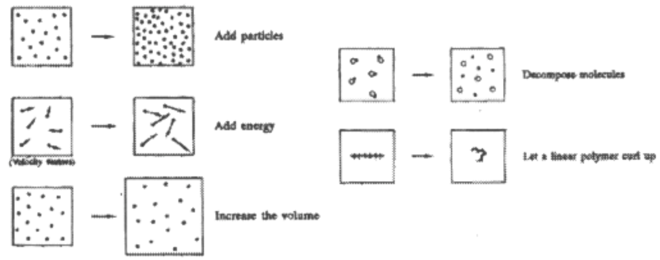


Figure 3: Ways of increasing entropy

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 alphabetic 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

$$H(X) = -\sum_i e^{-\sigma}\log e^{-\sigma} = \sigma e^{-\sigma}\sum_i 1 = \sigma$$

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":

$$1\text{nit} = \log_2 e \text{ bits} \cong 1.44 \text{ bits}$$

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 σ, I need to give σ nits $= (\log_2 e)\,\sigma$ 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

$$\left(\frac{\partial \log g_1}{\partial U_1}\right)_{N_1} = \left(\frac{\partial \log g_2}{\partial U_2}\right)_{N_2} \text{ and } \left(\frac{\partial \log g_2}{\partial U_2}\right)_{N_2} = \left(\frac{\partial \log g_3}{\partial U_3}\right)_{N_3}$$

Then

$$\left(\frac{\partial \log g_1}{\partial U_1}\right)_{N_1} = \left(\frac{\partial \log g_3}{\partial U_3}\right)_{N_3}$$

Or if $\tau_1 = \tau_2$ and $\tau_2 = \tau_3$ then $\tau_1 = \tau_3$.

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:

Concept of reservoir: reservoir R is a very large system, so large that bringing it into equilibrium with another smaller system S of interest makes no substantial difference to its properties (such as temperature) (though there may be transfer of, e.g., energy and entropy into and out of the reservoir). The total system $R+S$ is a closed system, total energy $U_0 = U_R + U_S$ is constant. e.g., if system S has energy ε , then the reservoir has energy $U_0 - \varepsilon$.

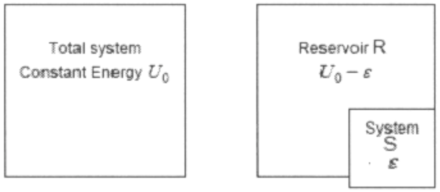


Figure 4

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 τ for the total system. Consider two possible states of the system S, state 1 and state 2, with associated energies ε_1 and ε_2 . The probability that the system S is in state 1 is proportional to the multiplicity of the reservoir R when it has energy $U_0 - \varepsilon$. 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

$$\frac{P_1}{P_2} = \frac{\text{Multiplicity of } R \text{ at energy } U_0 - \varepsilon_1}{\text{Multiplicity of } R \text{ at energy } U_0 - \varepsilon_2}$$

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 ε_1 or energy ε_2, we are calculating probabilities of the system being in a specific state of energy ε_1 or energy ε_2.

Now we know that

$$\text{Multiplicity of } R \text{ at energy } U_0 - \varepsilon_1 = e^{\sigma_R(U_0 - \varepsilon_1)}$$

and so we have

$$\frac{P_1}{P_2} = \frac{e^{\sigma_R(U_0-\varepsilon_1)}}{e^{\sigma_R(U_0-\varepsilon_2)}}$$

By expanding $\sigma_R(U_0 - \varepsilon)$ about the point U_0

$$\sigma_R(U_0 - \varepsilon) \cong \sigma_R(U_0) - \varepsilon\left(\frac{\partial \sigma_R}{\partial U_2}\right)_{U=U_0} = \sigma_R(U_0) - \frac{\varepsilon}{\tau}$$

$$\frac{P_1}{P_2} = e^{-(\varepsilon_1 - \varepsilon_2)/\tau}$$

A term of the form $e^{-\varepsilon/\tau}$ is called a Boltzmann factor. It tells us the relative occupation probability of two quantum states separated by energy ε 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 Boltzmann factor as an "unnormalized" occupation probability. We can normalize the probabilities by dividing by the sum of all such "unnormalized" probabilities. That sum is called the "partition function"

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

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 ε_q is

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

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

$$\langle\varepsilon\rangle = \sum_q \varepsilon_q P(\varepsilon_q) = \frac{1}{Z}\sum_q \varepsilon_q e^{-\varepsilon_q/\tau} = \tau^2 \frac{\partial \log Z}{\partial \tau}$$

It is a useful expression for solving many thermal physics 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 τ, we want to find the energy of the system as a function of τ. The partition function for the two state system of the particle is:

$$Z = e^{-0/\tau} + e^{-\varepsilon/\tau} = 1 + e^{-\varepsilon/\tau}$$

The average energy is:

$$U = \langle\varepsilon\rangle = \frac{1}{Z}\varepsilon e^{-\varepsilon/\tau} = \varepsilon\frac{e^{-\varepsilon/\tau}}{1 + e^{-\varepsilon/\tau}}$$

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* $F = U - \tau\sigma$

It is at a minimum in equilibrium at constant temperature τ and volume V .

Proof that Helmholtz energy is minimized at constant τ, 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

$$dF = dU_S - \tau\, d\sigma_s$$

where we have used the constancy of the temperature to eliminate the term $\sigma_s d\tau$ from the differential. But

$$\frac{1}{\tau} = \left(\frac{\partial \sigma_S}{\partial U_S}\right)_V$$

so

$$dF_S = 0$$

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 $N_\uparrow$ magnets up and $N_\downarrow$ down, with $N = N_\uparrow + N_\downarrow$, and spin excess $2s = N_\uparrow - N_\downarrow$, and magnetic moment (per magnet) m . We could evaluate the entropy of this system to be (using Stirling's approximation, simplified somewhat further for very large N)

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

The energy of this system in a magnetic field B is $U = -2smB$. so the Helmholtz free energy is

$$F = U - \tau\sigma = -2smB + \left(\frac{N}{2}+s\right)\tau\log\left(\frac{1}{2}+\frac{s}{N}\right)+\left(\frac{N}{2}-s\right)\tau\log\left(\frac{1}{2}-\frac{s}{N}\right)$$

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 τ and N constant, giving

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

Hence the thermal equilibrium value of $2s$, which we denote by $\langle 2s\rangle$, is the value that satisfies the above equation , i.e.,

$$\frac{N+\langle 2s\rangle}{N-\langle 2s\rangle} = e^{2mB/\tau}$$

or, after some rearrangement

$$\langle 2s\rangle = N\frac{e^{2mB/\tau}-1}{e^{2mB/\tau}+1} = N\tanh\frac{mB}{\tau}$$

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

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.

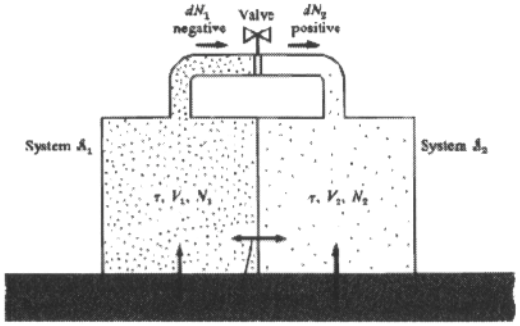


Figure 5: systems in thermal equilibrium with each other and with a reservoir, and in diffusive contact if valve is opened, and in diffusive contact if the valve is opened

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

$$F = F_1 + F_2 = U_1 + U_2 - \tau(\sigma_1 - \sigma_2)$$

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

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

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

Hence

$$dF = \left[\left(\frac{\partial F_1}{\partial N_1}\right)_\tau - \left(\frac{\partial F_2}{\partial N_2}\right)_\tau\right]dN_1 = 0 \quad (*)$$

so that at equilibrium

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

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*** μ, i.e.,

$$\mu(\tau, V, N) \equiv \left(\frac{\partial F}{\partial N}\right)_{\tau,V}$$

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_1 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_1 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,

$$\mu_j \equiv \left(\frac{\partial F}{\partial N}\right)_{\tau,V,N_1,N_2,\dots}$$

where in the differentiation the numbers of all particles are held constant except for the species j.

- Chemical potential and entropy
- Chemical potential and potential energy change
- Gibbs sum and Gibbs factor

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

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

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

$$\frac{1}{\tau} = \left(\frac{\partial\sigma}{\partial \underset{U}{\cancel{N}}}\right)_{N,V}$$

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

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

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 dS, dU, and dN such that the temperature change $d\tau$ is zero, and we denote this restricted combination by $(\delta\sigma)_\tau$, $(\delta U)_\tau$, and $(\delta N)_\tau$. Then $d\tau = 0$ when

$$(\delta\sigma)_\tau = \left(\frac{\partial\sigma}{\partial U}\right)_N (\delta U)_\tau + \left(\frac{\partial\sigma}{\partial N}\right)_U (\delta N)_\tau$$

Dividing by $(\delta N)_\tau$

$$\frac{(\delta\sigma)_\tau}{(\delta N)_\tau} = \left(\frac{\partial\sigma}{\partial U}\right)_N \frac{(\delta U)_\tau}{(\delta N)_\tau} + \left(\frac{\partial\sigma}{\partial N}\right)_U$$

The ratio $(\delta\sigma)_\tau/(\delta N)_\tau$ is $(\partial\sigma/\partial N)_\tau$, and $(\delta U)_\tau/(\delta N)_\tau$ is $(\partial U/\partial N)_\tau$, all at constant volume. Hence, using $\tau^{-1} = (\partial\sigma/\partial U)_{V,N}$, we have

$$\left(\frac{\partial\sigma}{\partial N}\right)_{\tau,V} = \frac{1}{\tau}\left(\frac{\partial U}{\partial N}\right)_{\tau,V} + \left(\frac{\partial\sigma}{\partial N}\right)_{U,V}$$

i.e.,

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

But we know from our original definition of chemical potential that

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

Hence we have

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

Chemical potential and potential energy changes:

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 ΔV 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_o, so we had a chemical potential

$$\mu_o = \left(\frac{\partial F_o}{\partial N}\right)_{\tau,V}$$

After this change as in ΔF, we have a new chemical potential

$$\mu = \left(\frac{\partial F}{\partial N}\right)_{\tau,V} = \left(\frac{\partial F_o}{\partial N}\right)_{\tau,V} + \left(\frac{\partial \Delta F}{\partial N}\right)_{\tau,V} = \mu_o + q\Delta V$$

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

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, μ_{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 μ

2. "external chemical potential", μ_{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, $\mu_{tot} = \mu_{int} + \mu_{ext}$

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 $Al_xGa_{1-x}As$ 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.

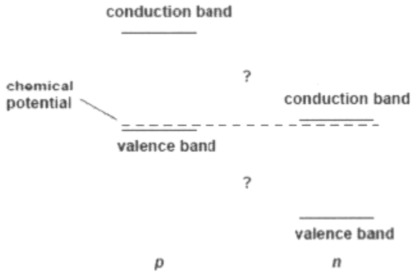


Figure 1: Chemical potentials and dissimilar joined materials

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_0 identical particles, and energy U_0. When the system S has N particles, the reservoir has $N - N_0$ particles, When the system S has energy ε_S, the reservoir has energy $U_0 - \varepsilon_S$.

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 ε_S and N particles.

State 1 is a state in which system has energy ε_1 and N_1 particles.

State 2 is a state in which system S has energy ε_2 and N_2 particles.

The ratio of the probabilities of S being in state 1 or state 2 is

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

Now we know that, by definition, the multiplicity g is just the exponential of the entropy, σ, i.e.,

$$g = e^{\sigma}$$

so

$$\frac{P(N_1,\varepsilon_1)}{P(N_2,\varepsilon_2)} = \frac{e^{\sigma(N_0-N_1,U_0-\varepsilon_1)}}{e^{\sigma(N_0-N_2,U_0-\varepsilon_2)}} = e^{\Delta\sigma}$$

where

$$\Delta\sigma = \sigma(N_0 - N_1, U_0 - \varepsilon_1) - \sigma(N_0 - N_2, U_0 - \varepsilon_2)$$

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 ε, to obtain

$$\sigma(N_0 - N, U_0 - \varepsilon) = \sigma(N_0, U_0) - N\left(\frac{\partial\sigma}{\partial N_0}\right)_{U_0} - \varepsilon\left(\frac{\partial\sigma}{\partial N_0}\right)_{N_0} + \cdots$$

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

$$\Delta\sigma \cong -(N_1 - N_2)\left(\frac{\partial\sigma}{\partial N_0}\right)_{U_0} - (\varepsilon_1 - \varepsilon_2)\left(\frac{\partial\sigma}{\partial N_0}\right)_{N_0}$$

and so, using the relations

$$\frac{1}{\tau} \equiv \left(\frac{\partial \sigma}{\partial U_0}\right)_{N_0} \quad and \quad -\frac{\mu}{\tau} \equiv \left(\frac{\partial \sigma}{\partial N_0}\right)_{U_0}$$

we have

$$\Delta\sigma = \frac{(N_1 - N_2)\mu - (\varepsilon_1 - \varepsilon_2)}{\tau}$$

Hence

$$\frac{P(N_1, \varepsilon_1)}{P(N_2, \varepsilon_2)} = \frac{e^{(N_1\mu - \epsilon_1)/\tau}}{e^{(N_2\mu - \epsilon_2)/\tau}}$$

A factor of the form exp $((N\mu - \varepsilon)/\tau)$ is called a Gibbs factor, and represents an unnormalized probability that the system S can be in a state of energy ε 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.

$$Z(\mu, \tau) = \sum_{N=0}^{\infty} \sum_{S(N)} \exp\left(\frac{N\mu - \epsilon_{S(N)}}{\tau}\right) = \sum_{ASN} \exp\left(\frac{N\mu - \epsilon_{S(N)}}{\tau}\right)$$

where ASN 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 ε_1 is

$$P(N_1, \varepsilon_1) = \frac{1}{Z} e^{(N_1\mu - \epsilon_1)/\tau}$$

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 ε, otherwise ($N = 0$) it has energy 0. 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

$$Z = \exp((0 \times \mu - 0 \times \varepsilon)/\tau) + \exp((1 \times \mu - 1 \times \varepsilon)/\tau) = 1 + e^{(\mu - \varepsilon)/\tau}$$

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

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

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

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

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

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

$$f(\varepsilon) = \frac{1}{1 + e^{(\varepsilon - E_F)/k_B T}}$$

where the term E_F 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
- Fermi Gas in three dimensions: ground state and specific heat
- Classical carrier distribution in semiconductors
- Law of mass action
- Extrinsic semiconductors
- Fermi levels in extrinsic semiconductors
- Degenerate semiconductors

Fermi-dirac 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

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

μ is the chemical potential, k_B is Boltzmann's constant. note $k_B \sim 25$ meV at room temperature. Chemical potential (Fermi energy) corresponds to energy for which f = ½. At zero temperature Fermi-Dirac distribution is step function, with all states up to energy E_F (Fermi energy) totally full.

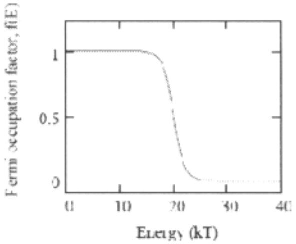


Figure 1:

Free electron gas at zero temperature:

The free eletron gas model represents the starting point for describing the properties of metals. As in many solid-state physics problem, 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

$$\varepsilon_n = \frac{\hbar^2}{2M}\left(\frac{\pi}{L}\right)^2\left(n_x^2+n_y^2+n_z^2\right)$$

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

$$\varepsilon_f = \frac{\hbar^2}{2M}\left(\frac{\pi n_f}{L}\right)^2$$

Here n_f 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 n_f determined by

$$N = 2\times\frac{1}{8}\times\frac{4\pi}{3}n_F^3$$

i.e.

$$n_F = (3N/\pi)^{1/3}$$

Here the factor 2 arises because an electron has two possible spin orientations. The factor 1/8 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:

$$\varepsilon_f = \frac{\hbar^2}{2M}\left(3\pi\frac{N}{V}\right)^{2/3}$$

This relates the Fermi energy to the electron concentration. The total energy of the system in the ground state is:

$$U_0 = 2\sum_{n\le n_f}\varepsilon_f = 2\times\frac{1}{8}\times 4\pi\int_0^{n_f} dn n^2\varepsilon_n$$

which, after taking into account n_F, 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 τ and a constant chemical potential μ, the occupation number for each orbital is:

$$f(\varepsilon) = 2\times\frac{1}{1+\exp\left(\frac{\varepsilon_n-\mu}{\tau}\right)}$$

Therefore, the thermal averages for the independent particle problems have the form:

$$\langle X\rangle = \sum_n f(\varepsilon_n,\tau,\mu)X_n$$

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 often express $\langle X\rangle$ as an integral over the orbital energy ε, Then above Eq becomes:

$$\langle X\rangle = \int d\varepsilon D(\varepsilon)f(\varepsilon,\tau,\mu)X(\varepsilon)$$

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:

$$D(\varepsilon) = \frac{V}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\varepsilon^{1/2}$$

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 τ is denoted by

$$\Delta U = \int_0^\infty d\varepsilon D(\varepsilon) f(\varepsilon) - \int_0^{\varepsilon_F} d\varepsilon \varepsilon D(\varepsilon)$$

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

$$\left(\int_0^{\varepsilon_F} + \int_{\varepsilon_F}^\infty\right) d\varepsilon \varepsilon_F f(\varepsilon) D(\varepsilon) = \int_0^{\varepsilon_F} d\varepsilon \varepsilon_F D(\varepsilon)$$

hence:

$$\Delta U = \int_{\varepsilon_F}^\infty d\varepsilon(\varepsilon - \varepsilon_F) D(\varepsilon) f(\varepsilon) + \int_0^{\varepsilon_F} d\varepsilon(\varepsilon - \varepsilon_F)(1 - f(\varepsilon)) D(\varepsilon)$$

The first integral on the right-hand side of gives the energy needed to take electron from ε_F to the orbitals of energy $\varepsilon > \varepsilon_F$, and the second integral gives the energy needed to bring the electrons to ε_F from orbitals below ε_F.

The heat capacity of the electron gas is found on differentiating ΔU with respect to τ. The only temperature-dependent term in equation above is $f(\varepsilon)$. Therefore, we can group terms to obtain:

$$C_{el} = \frac{dU}{d\tau} = \int_0^\infty d\varepsilon(\varepsilon - \varepsilon_F) \frac{df}{d\tau} D(\varepsilon)$$

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

$$C_{el} \cong D(\varepsilon_F) \int_0^\infty d\varepsilon(\varepsilon - \varepsilon_F) \frac{df}{d\tau}$$

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

$$\frac{df}{d\tau} = \frac{\varepsilon - \varepsilon_F}{\tau^2} \cdot \frac{e^{(\varepsilon - \varepsilon_F)/\tau}}{[1 + e^{(\varepsilon - \varepsilon_F)/\tau}]^2}$$

We set

$$x \equiv (\varepsilon - \varepsilon_F)/\tau$$

$$C_{el} = \tau D(\varepsilon_F) \int_{-\varepsilon_F/\tau}^\infty dx x^2 \frac{e^x}{(1 + e^x)^2} = \tau D(\varepsilon_F) \int_{-\infty}^\infty dx x^2 \frac{e^x}{(1 + e^x)^2}$$

Therefore, in the region $\tau \ll \tau_F$, we have

$$C_{el} = \tfrac{1}{3}\pi^2 D(\varepsilon_F)\tau$$

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 dependency 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 $f_e(E,T)$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

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

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_e denote the concentration of conduction electrons, and n_h denote the concentration of holes. In a pure semiconductor the two will be equal: $n_e = n_h$

Most semiconductors as used in devices have been intentionally doped with impurities that may become thermally ionied 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^+ be the concentration of positively charged donors and n_a^- the concentration of negatively charged acceptors. The electrical neutrality condition becomes

$$n_e - n_h = \Delta n = n_d^+ - n_a^-$$

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.

Semiconductor statistical mechanics - classical limit:

For energies far above the chemical potential, the Fermi-Dirac distribution behaves like classical Maxwell-Boltzmann distribution, f_{M-B}, i.e., for electrons,

$$f_{M-B}(E,T) = Ae^{-E/k_BT}$$

where $A = \exp(\mu/k_BT)$. 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 state, and $f_h \ll 1$ 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.

$$N_e = \sum_{CB} e^{-(\varepsilon-\mu)/\tau} = e^{-(\varepsilon_c-\mu)/\tau} \sum_{CB} e^{-(\varepsilon-\varepsilon_c)/\tau}$$

Or

$$N_e = N_c e^{-(\varepsilon_c-\mu)/\tau}$$

Here

$$N_c = \int_{\varepsilon_c}^{\infty} d\varepsilon \, VD(\varepsilon)\, e^{-(\varepsilon-\varepsilon_c)/\tau}$$

Where as we have discussed before

$$D(\varepsilon) = \frac{m_e^* \sqrt{2m_e^*(\varepsilon - \varepsilon_c)}}{\pi^2 \hbar^3}$$

And m_e^* is the density of state effective mass for the conductor band. Evaluating the integral we have

$$N_c = 2\left(\frac{m_e^* \tau}{2\pi\hbar^2}\right)^{3/2} V$$

Defining the quantum concentration n_c for the conduction electrons as:

$$n_c = \frac{N_c}{V} = 2\left(\frac{m_e^* \tau}{2\pi\hbar^2}\right)^{3/2}$$

The conduction electron concentration $n_e = N_e/V$ becomes

$$n_e = n_c e^{-(\varepsilon_c-\mu)/\tau}$$

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

$$n_h = n_v e^{-(\mu-\varepsilon_v)/\tau}$$

where

$$n_c = 2\left(\frac{m_h^* \tau}{2\pi\hbar^2}\right)^{3/2}$$

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, $n_e n_h = n_i^2$ where

$$n_i = (n_c n_v)^{1/2} e^{-E_g/2k_BT}$$

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 0C, n_i=1.04x10^9 cm^{-3}

for silicon at 50C, n_i=7.06 x10^{10} cm^{-3}

for GaAs at 0C, n_i=1.02 x105^5 cm^{-3}

for GaAs at 50C, n_i=2.18 x10^7 cm^{-3}

Fermi level in intrinsic semiconductors:

In an intrinsic semiconductor, the number of electrons is equal to the number of holes. Thus,

$$n_i \cong n_c e^{(\mu-\varepsilon)/k_BT} = (n_c n_v)^{1/2} e^{-E_g/2k_BT}$$

By rearranging ,we have

$$\left(\frac{n_c}{n_v}\right)^{1/2} = \exp\left[\frac{\frac{\mathrm{E_g}}{2} - \mu}{k_BT}\right]$$

or, after some algebra,

$$\mu = \frac{\mathrm{E_g}}{2} + \frac{k_BT}{2}\log\left(\frac{n_v}{n_c}\right) = \frac{\mathrm{E_g}}{2} + \frac{3k_BT}{4}\log\left(\frac{m_h^*}{m_e^*}\right)$$

so that the Fermi level in an intrinsic semiconductor is always very close to the middle of the bandgap, since $k_BT \ll E_g$ 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, assumes $\Delta n > 0$, i.e, n-doped semiconductor

Since, from the law of mass action,

$$n_h = n_i^2/n_e$$

Or

$$n_e^2 - n_e\Delta n = n_i^2$$

Therefore, we have

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

And

$$n_h = \tfrac{1}{2}\left(\left[\Delta n^2 + 4n_i^2\right]^{1/2} - \Delta n\right)$$

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 it is indeed negligibly small (we may check this a posteriori), we have

$$n_e \cong \Delta n$$

$$n_h \cong n_i^2/\Delta n$$

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, Δn. 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

$$n_h \cong n_v e^{-\mu/k_B T}$$

Or

$$n_e \cong n_c e^{(E_g-\mu)/k_B T}$$

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

$$\mu = E_g - k_B T \log(n_c/n_e)$$

or, equivalently

$$\mu = k_B T \log(nv_c/n_h)$$

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

$$N = \int d\varepsilon D(\varepsilon) f(\varepsilon)$$

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

$$n_e = \frac{1}{2\pi^2}\left(\frac{2m_e^*}{\hbar^2}\right)^{3/2} \int_{\varepsilon_c}^{\infty} \frac{(\varepsilon-\varepsilon_c)^{1/2}}{1+e^{(\varepsilon-\mu)/\tau}} d\varepsilon$$

Let $x \equiv (\varepsilon - \varepsilon_c)/\tau$ and $\eta \equiv (\mu - \varepsilon_c)/\tau$ We have

$$\frac{n_e}{n_c} = I(\eta) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{x^{1/2}}{1+e^{(x-\eta)}} dx$$

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

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

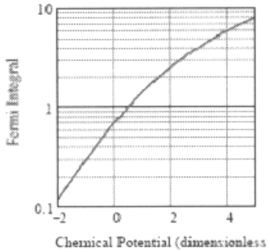


Figure 2

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,

$$F(\eta) \cong \frac{2}{3}\eta^{3/2}$$

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 $6k_B T$.

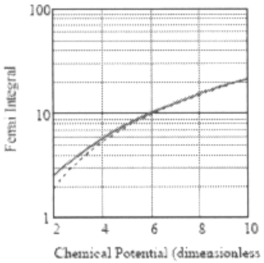


Figure 3: solid line - exact result, dashed line - low-temperature and/or high density approximation (energy units $k_B T$)

High-temperature and/or low-density "non-degenerate" limit:

Fermi-Dirac becomes like Maxwell-Boltzmann because μ (i.e. η) is very negative

$$\frac{1}{1+e^{(x-\eta)}} \cong e^{(\eta-x)}$$

So Fermi integral becomes

$$F(\eta) \cong e^{\eta} \int_0^{\infty} x^{1/2} e^{-x} dx$$

Knowing that $\int_0^\infty x^{1/2}e^{-x}\,dx = \sqrt{\pi}/2$ so

$$F(\eta) \cong \frac{\sqrt{\pi}}{2}e^{\eta}$$

Is useful for chemical potentials about $k_B T$ below the bandgap energy, and accurate from about $2k_B T$ below. Therefore,

$$\frac{n_e}{n_c} = e^{(\mu-\varepsilon_c)/\tau}$$

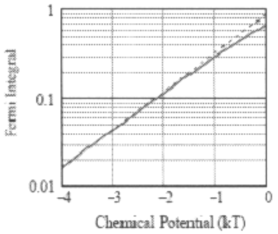


Figure 4: exact value - solid line, high-temp. /low-density approx. - dashed line

Fermi integral – approximations:

transport in semiconductors - usually take Maxwell-Boltzmann approx:

a. in germanium and silicon: good approx. because electron and hole masses are large
b. 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 ~ 10^{18} 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-Humet approximation (X. Aymerich-Humet, F. Serra-Mestres, and J. Millan, Solid State Electronics 24, 981 (1981)), which is

$$F(x) = \frac{\sqrt{\pi}}{2}\frac{1}{e^{-x} + \xi(x)}$$

Where

$$\xi(x) = 3\sqrt{\frac{\pi}{2}}\left[x + 2.13 + (|x - 2.13|^{2.4} + 9.6)^{\frac{5}{12}}\right]^{-3/2}$$

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_c, defining a "reduced" chemical potential as $\eta = (\mu - E_c)/\tau$ 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

$$\eta \cong \log r + \frac{1}{\sqrt{8}}r - \left(\frac{3}{16} - \frac{\sqrt{3}}{9}\right)r^2 + \cdots$$

i.e.,

$$\eta \cong \log r + 3.53553 \times 10^{-1}r - 4.95009 \times 10^{-3}r^2 + 1.48386 \times 10^{-4}r^3 - 4.42563 \times 10^{-6}r^4$$

The Joyce-Dixon approximation is valid for all negative values of η 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 η) is an approximation due to Nilsson (N. G. Nilsson, Phys. Sta. Solidi (a) 19, K75 (1973)), which is accurate within about 1%

$$\eta \cong \frac{\log r}{1 - r} + \left(\tfrac{3}{4}\sqrt{\pi}r\right)^{2/3} + \frac{8\sqrt{\pi}r}{3\left(4 + \sqrt{\pi}r\right)^2}$$

This approximation limits quite obviously to the low density ($\log r$) and high density $\left(3\sqrt{\pi}r/4\right)^{2/3}$ limits, and adds "joining function" terms chosen to give an empirically good fit.

Various approximations are discussed by J. S. Blakemore, "Approximations for Fermi-Dirac Integrals, Especially the Function $F_{1/2}(\eta)$ Used to Describe Electron Density in a Semiconductor," Solid-State Electronics, 25, pp1067-1076 (1982).

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

As a starting point we have

$$n_e = \frac{1}{2\pi^2}\left(\frac{2m_e^*}{\hbar^2}\right)^{3/2}\int_{\varepsilon_c}^{\infty}\frac{(\varepsilon-\varepsilon_c)^{1/2}}{1+e^{(\varepsilon-\mu)/\tau}}d\varepsilon$$

$$N = 2.83\times10^{19}\left(\frac{m^*}{m_0}\frac{T}{300}\right)^{3/2} F(E_F/k_BT)\ cm^{-3}$$

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

$$F(x)\cong\tfrac{2}{3}x^{3/2}$$

so then we have

$$N = \frac{2}{3}\times2.83\times10^{19}\left(\frac{m^*}{m_0}\frac{T}{300}\right)^{3/2}(E_F/k_BT)^{3/2}$$

Or

$$\frac{E_F}{k_BT} = \left(\frac{3N/2}{2.83\times10^{19}}\right)^{2/3}\frac{m_0}{m^*}\frac{300}{T} = \frac{m_0}{m^*}\frac{300}{T}\left(\frac{N}{1.89\times10^{19}}\right)^{2/3}$$

For the case of a nondegenerate distribution, we have

$$N = 2.51\times10^{19}\left(\frac{m^*}{m_0}\frac{T}{300}\right)^{3/2}\frac{\sqrt{\pi}}{2}e^{(E_F/k_BT)}$$

$$\frac{E_F}{k_BT} = \ln\left(\frac{N}{2.51\times10^{19}}\right)-\frac{3}{2}\ln\left(\frac{m^*}{m_0}\frac{T}{300}\right)$$

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

$$\frac{E_F}{k_BT} = \frac{m_0}{m^*}\frac{300}{T}\left(\frac{N}{1.89\times10^{19}}\right)^{2/3} = \frac{1}{0.079}\left(\frac{10^{19}}{1.89\times10^{19}}\right)^{2/3} = 8.28$$

This number is >>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.

$$f_e(E,T) = \frac{1}{1+\exp\left(\frac{E-\mu}{k_BT}\right)}$$

μ is the chemical potential, k_B is Boltzmann's constant. note $k_B\sim 25$ meV at room temperature. Chemical potential (Fermi energy) corresponds to energy for which f = ½. At zero temperature Fermi-Dirac distribution is step function, with all states up to energy E_F (Fermi energy) totally full.

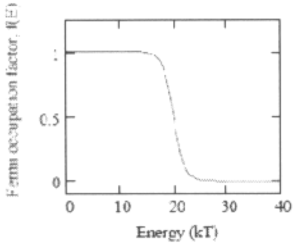


Figure 1:

Free electron gas at zero temperature:

The free eletron gas model represents the starting point for describing the properties of metals. As in many solid-state physics problem, 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

$$\varepsilon_n = \frac{\hbar^2}{2M}\left(\frac{\pi}{L}\right)^2(n_x^2+n_y^2+n_z^2)$$

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

$$\varepsilon_f = \frac{\hbar^2}{2M}\left(\frac{\pi n_f}{L}\right)^2$$

Here n_f 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 n_f determined by

$$N = 2\times\frac{1}{8}\times\frac{4\pi}{3}n_F^3$$

i.e.

$$n_F = (3N/\pi)^{1/3}$$

Here the factor 2 arises because an electron has two possible spin orientations. The factor 1/8 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:

$$\varepsilon_f = \frac{\hbar^2}{2M}\left(3\pi\frac{N}{V}\right)^{2/3}$$

This relates the Fermi energy to the electron concentration. The total energy of the system in the ground state is:

$$U_0 = 2\sum_{n\leq n_f} \varepsilon_f = 2\times\frac{1}{8}\times 4\pi\int_0^{n_f} dnn^2\varepsilon_n$$

which, after taking into account n_F, 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 τ and a constant chemical potential μ, the occupation number for each orbital is:

$$f(\varepsilon) = 2\times\frac{1}{1+\exp\left(\frac{\varepsilon_n-\mu}{\tau}\right)}$$

Therefore, the thermal averages for the independent particle problems have the form:

$$\langle X\rangle = \sum_n f(\varepsilon_n, \tau, \mu)X_n$$

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 often express $\langle X\rangle$ as an integral over the orbital energy ε, Then above Eq becomes:

$$\langle X\rangle = \int d\varepsilon D(\varepsilon)\, f(\varepsilon, \tau, \mu)X(\varepsilon)$$

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:

$$D(\varepsilon) = \frac{V}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\varepsilon^{1/2}$$

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 τ is denoted by

$$\Delta U = \int_0^\infty d\varepsilon D(\varepsilon)f(\varepsilon) - \int_0^{\varepsilon_F} d\varepsilon\varepsilon D(\varepsilon)$$

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

$$\left(\int_0^{\varepsilon_F} + \int_{\varepsilon_F}^\infty\right) d\varepsilon\varepsilon_F f(\varepsilon)D(\varepsilon) = \int_0^{\varepsilon_F} d\varepsilon\varepsilon_F D(\varepsilon)$$

hence:

$$\Delta U = \int_{\varepsilon_F}^\infty d\varepsilon(\varepsilon-\varepsilon_F)D(\varepsilon)f(\varepsilon) + \int_0^{\varepsilon_F} d\varepsilon(\varepsilon-\varepsilon_F)(1-f(\varepsilon))D(\varepsilon)$$

The first integral on the right-hand side of gives the energy needed to take electron from ε_F to the orbitals of energy $\varepsilon > \varepsilon_F$, and the second integral gives the energy needed to bring the electrons to ε_F from orbitals below ε_F.

The heat capacity of the electron gas is found on differentiating ΔU with respect to τ. The only temperature-dependent term in equation above is $f(\varepsilon)$. Therefore, we can group terms to obtain:

$$C_{el} = \frac{dU}{d\tau} = \int_0^\infty d\varepsilon(\varepsilon-\varepsilon_F)\frac{df}{d\tau}D(\varepsilon)$$

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 is like a step function, the derivative $df/d\tau$ is large on at energies near ε_F. It is a good approximation to evaluate the density of orbitals $D(\varepsilon)$ at energy ε_F and take it outside of the integral:

$$C_{el} \cong D(\varepsilon_F)\int_0^\infty d\varepsilon(\varepsilon-\varepsilon_F)\frac{df}{d\tau}$$

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

$$\frac{df}{d\tau} = \frac{\varepsilon-\varepsilon_F}{\tau^2}\cdot\frac{e^{(\varepsilon-\varepsilon_F)/\tau}}{[1+e^{(\varepsilon-\varepsilon_F)/\tau}]^2}$$

We set

$$x \equiv (\varepsilon-\varepsilon_F)/\tau$$

$$C_{el} = \tau D(\varepsilon_F)\int_{-\varepsilon_F/\tau}^\infty dxx^2\frac{e^x}{(1+e^x)^2} = \tau D(\varepsilon_F)\int_{-\infty}^\infty dxx^2\frac{e^x}{(1+e^x)^2}$$

Therefore, in the region $\tau \ll \tau_F$, we have

$$C_{el} = \tfrac{1}{3}\pi^2 D(\varepsilon_F)\tau$$

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 dependency 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 $f_e(E,T)$of electron in state of energy E at temperature T is

$$f_e(E,T) = \frac{1}{1+\exp\left(\frac{E-\mu}{k_BT}\right)}$$

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

$$f_e(E,T) = \frac{\exp\left(\frac{E-\mu}{k_BT}\right)}{1+\exp\left(\frac{E-\mu}{k_BT}\right)} = \frac{1}{1+\exp\left(\frac{-(E-\mu)}{k_BT}\right)}$$

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_e denote the concentration of conduction electrons, and n_h denote the concentration of holes. In a pure semiconductor the two will be equal: $n_e = n_h$

Most semiconductors as used in devices have been intentionally doped with impurities that may become thermally ionied 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^+ be the concentration of positively charged donors and n_a^- the concentration of negatively charged acceptors. The electrical neutrality condition becomes

$$n_e - n_h = \Delta n = n_d^+ - n_a^-$$

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.

Semiconductor statistical mechanics - classical limit:

For energies far above the chemical potential, the Fermi-Dirac distribution behaves like classical Maxwell-Boltzmann distribution, f_{M-B}, i.e., for electrons,

$$f_{M-B}(E,T) = Ae^{-E/k_BT}$$

where $A = \exp(\mu/k_BT)$. 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 state, and $f_h \ll 1$ 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.

$$N_e = \sum_{CB} e^{-(\varepsilon-\mu)/\tau} = e^{-(\varepsilon_c-\mu)/\tau} \sum_{CB} e^{-(\varepsilon-\varepsilon_c)/\tau}$$

Or

$$N_e = N_c e^{-(\varepsilon_c-\mu)/\tau}$$

Here

$$N_c = \int_{\varepsilon_c}^{\infty} d\varepsilon\, VD(\varepsilon)\, e^{-(\varepsilon-\varepsilon_c)/\tau}$$

Where as we have discussed before

$$D(\varepsilon) = \frac{m_e^*\sqrt{2m_e^*(\varepsilon-\varepsilon_c)}}{\pi^2\hbar^3}$$

And m_e^* is the density of state effective mass for the conductor band. Evaluating the integral we have

$$N_c = 2\left(\frac{m_e^*\tau}{2\pi\hbar^2}\right)^{3/2} V$$

Defining the quantum concentration n_c for the conduction electrons as:

$$n_c = \frac{N_c}{V} = 2\left(\frac{m_e^*\tau}{2\pi\hbar^2}\right)^{3/2}$$

The conduction electron concentration $n_e = N_e/V$ becomes

$$n_e = n_c e^{-(\varepsilon_c-\mu)/\tau}$$

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

$$n_h = n_v e^{-(\mu-\varepsilon_v)/\tau}$$

where

$$n_c = 2\left(\frac{m_h^*\tau}{2\pi\hbar^2}\right)^{3/2}$$

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, $n_e n_h = n_i^2$ where

$$n_i = (n_c n_v)^{1/2} e^{-E_g/2k_BT}$$

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 0C, n_i=1.04x10^9 cm^{-3}
for silicon at 50C, n_i=7.06 x10^{10} cm^{-3}
for GaAs at 0C, n_i=1.02 x105^5 cm^{-3}
for GaAs at 50C, n_i=2.18 x10^7 cm^{-3}

Fermi level in intrinsic semiconductors:

In an intrinsic semiconductor, the number of electrons is equal to the number of holes. Thus,

$$n_i \cong n_c e^{(\mu-\varepsilon)/k_BT} = (n_c n_v)^{1/2} e^{-E_g/2k_BT}$$

By rearranging ,we have

$$\left(\frac{n_c}{n_v}\right)^{1/2} = \exp\left[\frac{\frac{\mathrm{E_g}}{2} - \mu}{k_BT}\right]$$

or, after some algebra,

$$\mu = \frac{\mathrm{E_g}}{2} + \frac{k_BT}{2}\log\left(\frac{n_v}{n_c}\right) = \frac{\mathrm{E_g}}{2} + \frac{3k_BT}{4}\log\left(\frac{m_h^*}{m_e^*}\right)$$

so that the Fermi level in an intrinsic semiconductor is always very close to the middle of the bandgap, since $k_BT \ll E_g$ 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, assumes $\Delta n > 0$, i.e, n-doped semiconductor

Since, from the law of mass action,

$$n_h = n_i^2/n_e$$

Or

$$n_e^2 - n_e\Delta n = n_i^2$$

Therefore, we have

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

And

$$n_h = \tfrac{1}{2}\left(\left[\Delta n^2 + 4n_i^2\right]^{1/2} - \Delta n\right)$$

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 it is indeed negligibly small (we may check this a posteriori), we have

$$n_e \cong \Delta n$$
$$n_h \cong n_i^2/\Delta n$$

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, Δn. 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

$$n_h \cong n_v e^{-\mu/k_BT}$$

Or

$$n_e \cong n_c e^{(E_g-\mu)/k_BT}$$

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

$$\mu = E_g - k_BT\log(n_c/n_e)$$

or, equivalently

$$\mu = k_BT\log(nv_c/n_h)$$

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

$$N = \int d\varepsilon D(\varepsilon) f(\varepsilon)$$

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

$$n_e = \frac{1}{2\pi^2}\left(\frac{2m_e^*}{\hbar^2}\right)^{3/2}\int_{\varepsilon_c}^{\infty}\frac{(\varepsilon-\varepsilon_c)^{1/2}}{1+e^{(\varepsilon-\mu)/\tau}}d\varepsilon$$

Let $x \equiv (\varepsilon - \varepsilon_c)/\tau$ and $\eta \equiv (\mu - \varepsilon_c)/\tau$ We have

$$\frac{n_e}{n_c} = I(\eta) = \frac{2}{\sqrt{\pi}}\int_0^{\infty}\frac{x^{1/2}}{1+e^{(x-\eta)}}dx$$

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

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

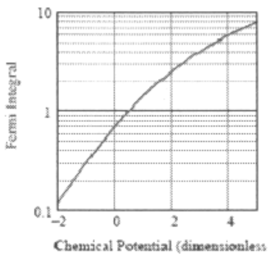


Figure 2

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,

$$F\,(\eta) \cong \frac{2}{3}\eta^{3/2}$$

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 $6k_BT$.

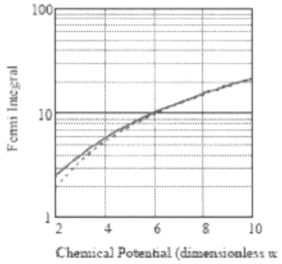


Figure 3: solid line - exact result, dashed line - low-temperature and/or high density approximation (energy units k_BT)

High-temperature and/or low-density "non-degenerate" limit:

Fermi-Dirac becomes like Maxwell-Boltzmann because μ (i.e. η) is very negative

$$\frac{1}{1+e^{(x-\eta)}} \cong e^{(\eta-x)}$$

So Fermi integral becomes

$$F(\eta) \cong e^{\eta}\int_0^{\infty} x^{1/2}e^{-x}\,dx$$

Knowing that $\int_0^{\infty} x^{1/2}e^{-x}\,dx = \sqrt{\pi}/2$ so

$$F(\eta) \cong \frac{\sqrt{\pi}}{2}e^{\eta}$$

Is useful for chemical potentials about k_BT below the bandgap energy, and accurate from about $2k_BT$ below. Therefore,

$$\frac{n_e}{n_c} = e^{(\mu-\varepsilon_c)/\tau}$$

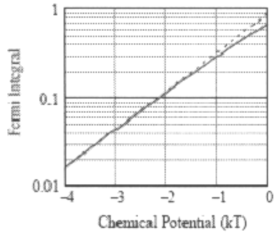


Figure 4: exact value - solid line, high-temp. /low-density approx. - dashed line

Fermi integral – approximations:

transport in semiconductors - usually take Maxwell-Boltzmann approx:

a. in germanium and silicon: good approx. because electron and hole masses are large
b. 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 ~ 10^{18} 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-Humet approximation (X. Aymerich-Humet, F. Serra-Mestres, and J. Millan, Solid State Electronics 24, 981 (1981)), which is

$$F(x) = \frac{\sqrt{\pi}}{2}\frac{1}{e^{-x}+\xi(x)}$$

Where

$$\xi(x) = 3\sqrt{\frac{\pi}{2}}\left[x + 2.13 + (|x-2.13|^{2.4} + 9.6)^{\frac{5}{12}}\right]^{-3/2}$$

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_c, defining a "reduced" chemical potential as

$\eta = (\mu - E_c)/\tau$ 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

$$\eta \cong \log r + \frac{1}{\sqrt{8}} r - \left(\frac{3}{16} - \frac{\sqrt{3}}{9}\right) r^2 + \cdots$$

i.e.,

$$\eta \cong \log r + 3.53553 \times 10^{-1} r - 4.95009 \times 10^{-3} r^2 + 1.48386 \times 10^{-4} r^3 - 4.42563 \times 10^{-6} r^4$$

The Joyce-Dixon approximation is valid for all negative values of η 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 η) is an approximation due to Nilsson (N. G. Nilsson, Phys. Sta. Solidi (a) 19, K75 (1973)), which is accurate within about 1%

$$\eta \cong \frac{\log r}{1 - r} + \left(\tfrac{3}{4}\sqrt{\pi} r\right)^{2/3} + \frac{8\sqrt{\pi} r}{3\left(4 + \sqrt{\pi} r\right)^2}$$

This approximation limits quite obviously to the low density ($\log r$) and high density $\left(3\sqrt{\pi} r/4\right)^{2/3}$ limits, and adds "joining function" terms chosen to give an empirically good fit.

Various approximations are discussed by J. S. Blakemore, "Approximations for Fermi-Dirac Integrals, Especially the Function $F_{1/2}(\eta)$ Used to Describe Electron Density in a Semiconductor," Solid-State Electronics, 25, pp1067-1076 (1982).