PHY 4202
Semiconductor Physics and Technology

Joy Mitra

October 19, 2023

Contents

List of Figures

List of Tables

0.1 Preface

Semiconductors are a class of materials with electrical conductivity between that of conductors and insulators and have profoundly shaped the landscape of modern technology. Their discovery and development have opened doors to an array of electronic devices that are now an integral part of our daily lives.

In the mid-20th century, scientists began to explore the potential of semiconductors for electronic applications. Notably, the invention of the transistor in 1947 by John Bardeen, Walter Brattain, and William Shockley at the Bell Laboratories, USA revolutionised the field of electronics. Transistors, made from semiconductors like Si and Ge, replaced bulky and less efficient vacuum tubes, leading to the miniaturisation of electronic circuits and the birth of the modern digital age. Further advancements in semiconductor technology gave rise to integrated circuits (ICs) and microprocessors, exponentially increasing computational power and other complex electronic systems. Semiconductors are now at the core of diverse technologies e.g. computers, smartphones, medical devices, renewable energy systems, communication networks, and more.

This evolution and innovations in semiconductor materials and technology, including discovery of newer material systems like conducting polymers, and new fabrication techniques continue to drive modern research in quantum computing, nanotechnology, energy research and others. Over the years significant contributions in Semiconductor Physics and Technology and related areas have been recognised by the Nobel Prize as shown in the table below.

Contribution	**Recipients**	**Year**
Thermionic Phenomenon	O W Richardson	1928
Invention of Transistor	W B Shockley, J Bardeen, W H Brattain	1956
Tunneling in semiconductors and superconductors	L Esaki, I Giaever	1973
Electron Microscope	Ernst Ruska	1986
Scanning Tunneling Microscope	G Binnig, H Rohrer	1986
invention of the integrated circuit	J S. Kilby	2000
semiconductor heterostructures for high-speed & opto-electronics	Z I Alferov, H Kroemer	2000
Conducting Polymers#	A J Heeger, A G MacDiarmid, H Shirakawa	2000
CCD sensor	Willard S. Boyle George E. Smith	2009
Blue light-emitting diode	I Akasaki, H Amano, S Nakamura	2014
Quantum Dots#	M G Bawendi, L E Brus, A I Ekimov	2023

Table 1: Nobel Prizes awarded in areas related to Semiconductor Physics and Technology. # denotes awards in Chemistry and all others were awarded in Physics

IMPORTANT: These Notes are not intended as replacement of standard textbooks. Consult the Text and Reference Books for exposure to standard development and also experimental data.

0.2 Syllabus

- Review of Bulk semiconductors: crystals, compound semiconductors, band- structure, density of states, doping and carrier concentration, Fermi statistics. [4]
- Electrical Transport in Bulk Semiconductors: Drude model, Boltzmann transport; solutions in electric and magnetic field; moments of transport equation, continuity equation, diffusion, drift, thermal gradient etc. [6]
- Semiconductor Junctions: Schottky and heterojunctions, role of interfaces, band bending concept, self-consistent band bending equations (Poisson - Schrodinger etc). Band bending near surfaces and interfaces. Forward and reverse biased diodes. Applications: thermometry, tunnel diodes etc. [7]

- Optical Properties of metals and semiconductors: Optical interactions in metals and semiconductors, reflection, refraction, optical absorption, free carrier absorption, refraction, Kramers Kronig relation; classical and quantum mechanical description of optical absorption, excitons; spontaneous and stimulated emission, Einstein coefficients; Photoluminescence and Electroluminescence. [7]
- Quantum Heterostructures & Reduced dimensional systems: 3D, 2D, 1D electron gas and quantum dot systems; engineering heterostructures and superlattices; optical properties of reduced dimensional systems; Quantum confined Stark effect. [6]
- Screening in 3D and 2D electron systems: Lattice polarisation; screened Coulomb potential, remote doping and mobility. [3]
- Photovoltaic Devices: photoconductors, photodiodes, Light Emitting Diodes, Laser Diodes; Quantum cascade lasers etc. [3]

0.3 Books

1. Semiconductor Devices: Physics and Technology, 3rd Edition, SM Sze and M Lee, Wiley India
2. John Singleton, Band Theory and Electronic Properties of Solids, Oxford Master Series in Condensed Matter Physics, Oxford University Press 2001.
3. Neil W. Ashcroft & N. David Mermin, Solid State Physics, Sanders College Publishing 1976.
4. Seeger, K., Semiconductor Physics, Springer-Verlag, 1990.
5. Fox, M., Optical Properties of Solids, Oxford Master Series in Condensed Matter Physics, Oxford University Press 2001.
6. J. H. Davies, Physics of Low-Dimesnional Semiconductors, Cambridge, 1997.
7. R. F. Pierret, Semiconductor Device Fundamentals, Pearson India, 2006

Acknowledgements

Prof K Dasgupta (IIT Bombay)

Joy Mitra
j.mitra@iisertvm.ac.in
http://jmitra.wix.com/joygroup

1

Basics of Semiconductors

1.1 Carrier concentration and doping

At T = 0 in a pure semiconductor, the conduction band (C_B) is empty and the valence band (V_B) is full. A completely full or a completely empty band cannot conduct current. We will see soon that under these circumstances the Fermi energy (E_f) lies in the bandgap (E_g) between the valence and conduction band. The density of states at the E_f is zero. The semiconductor is an insulator at this point. In reality there are no qualitative distinction between semiconductors and insulators, but arises from the difference in E_g between the two. E_g of an insulator is large, e.g. Silicon oxide has $E_g \sim$ 9eV, Diamond has $E_g \sim$ 5eV and so on. In contrast E_g of typical semiconductors is in the range of nearly zero to $\leq$ 4 eV. At very high temperatures, if an insulator hasn't already melted, it will act as a semiconductor. Carriers in a semiconductor's bands come from two sources:
1. Thermally excited electrons in conduction band and the corresponding vacancies left behind in the valence band.
2. Some suitable foreign atoms called dopants which can put some electrons in CB or capture some electrons from VB. Sometimes crystal defects can also play the role of foreign atoms.

Consider a group V atom like Phosphorous replacing an atom of group IV Silicon in the lattice. P has one extra electron compared to Si. We keep aside the question about how to get the P atom to replace the Si for the time being - but that is not a trivial question. A "dopant" will not work as a dopant if it does not sit in the right place. It is possible for a P atom to somehow go in as an "interstitial", that will not work. Also the same atom may act as an acceptor or a donor in some cases. For example if Si is incorporated in GaAs lattice, replacing a Ga atom, it will act as a donor. If it replaces an As atom it will act as an acceptor. You can figure out the reason. P can act as a donor in the Si lattice if the binding energy of that extra electron becomes very low. We discuss a very simplified model - usually called the "hydrogenic impurity model". Assume that the outermost electron in P behaves as if it is tied to a hypothetical nucleus - that is the P^+ ion core. The binding energy and Bohr radius of an H atom (1s state) is

$$E = -\frac{me^4}{8\epsilon_0^2 h^2} \tag{1.1}$$

$$a_B = 4\pi\epsilon\frac{\hbar^2}{me^2} \tag{1.2}$$

Now we make two crucial claims. Inside the "medium" the free electron mass would be modified such that $m \to m_{eff}$ and $\epsilon_0 \to \epsilon_0\epsilon_r$. Typically $\epsilon_r \sim 10 - 15$ for most semiconductor lattices and $m_{eff} \sim 0.1m$. That means the binding energy would reduce by a factor of 1000 and the Bohr radius would increase by a factor of about $\sim$ 100. So instead of E = 13.6 eV the binding energy will be a few 1-10 meV, the Bohr radius will increase from 0.5 Å to 50 Å. This means that the electron will be exploring something of the order of lattice units. This in retrospect justifies the use of the lattice dielectric which is a quantity meaningful only if averaged over sum volume of the lattice. Also the fact that the electron gets spread over a large area, means that replacing the free electron mass with the band effective mass can be justified. If the binding energy drops to a few meV, it is clear that at room temperature ($k_{BT} = 25meV$) these can be almost fully ionised. The order of magnitude of these numbers ensure that semiconductors can be useful at room temperature.

Table 1.1: Vital Statistics of Si, Ge, GaAs and ZnO

Properties	Si	Ge	GaAs	ZnO
Atomic Density ($\times 10^{22} cc^{-1}$)	5.0	4.4	4.4	
Crystal Structure	Diamond	Diamond	Zincblende	Wurtzite
Density (gm/cc)	2.33	5.33	5.32	5.606
Dielectric constant	11.9	16	13.1	8.3
Electron affinity (eV)	4.1	4.0	4.1	4.29
Band gap at 300K (eV)	1.12	0.66	1.42	3.3
Excitonic binding energy (meV)	14	2.7	4.2	60
electron m_{eff} ($\times m_e$)	0.98	1.64	0.067	0.27
hole m_{eff} ($\times m_e$)	0.16, 0.49	0.044, 0.28	0.082, 0.45	
Effective density of states in C_B $N_C(cc^{-1})$ at 300K	2.8×10^{19}	1.04×10^{19}	4.7×10^{17}	4.8×10^{18}
Effective density of states in V_B $N_V(cc^{-1})$ at 300K	1.04×10^{19}	6.0×10^{18}	7.0×10^{18}	4.8×10^{18}
Intrinsic carrier concentration $n(cc^{-1})$ at 300K	1.5×10^{10}	2.4×10^{13}	1.8×10^{6}	-
Intrinsic electron mobility) $(cm^2V^{-1}s^{-1})$ at 300K	1350	3900	8500	200
Intrinsic hole mobility $(cm^2V^{-1}s^{-1})$ at 300K	480	1900	400	50
Electron diffusion coefficient (cm^2s^{-1}) at 300K	35	100	220	
n-type Donors (E_D meV)	P(45) As(12)	P(50) As(12.7)	Se(5.9) Te(5.8)	native defects
p-type Donors (E_A meV)	B(45) Al(60)	B(10.4) Al(10.2)	Be(28) Zn(30.7)	unknown

1.2 Fermi Level in an intrinsic (undoped) semiconductor

If the material is undoped, then all the electrons in the conduction band (CB) must have been thermally excited from the valence band (VB). This fact is sufficient to tell us where (E_f) should be. The electron and hole densities are given by,

$$n = \int_{E_C}^{\infty} D(E) f(E) dE \tag{1.3}$$

$$p = \int_{-\infty}^{E_V} D(E)(1 - f(E)) dE \tag{1.4}$$

Assuming parabolic CB and VB the dispersion relations can be shown to be given by,

$$E_e(k) = E_{\mathrm{C}} + \frac{\hbar^2 k^2}{2m_e} \tag{1.5}$$

$$E_h(k) \;=\; E_V - \frac{\hbar^2 k^2}{2m_h} \tag{1.6}$$

where E_C, E_V denote the bottom and the top of CB and VB respectively. The density of states in 3D, including spin degeneracy is then given by:

$$D(E) = \frac{1}{2\pi^2}(\frac{2m_e}{\hbar^2})^{3/2}(E - E_{\mathrm{C}})^{1/2} for E > E_{\mathrm{C}} \tag{1.7}$$

$$D(E) = \frac{1}{2\pi^2}(\frac{2m_h}{\hbar^2})^{3/2}(E_V - E)^{1/2} for E < E_V \tag{1.8}$$

To evaluate equations 1.3 & 1.4 we proceed as :

$$n = \frac{1}{2\pi^2}(\frac{2m_e}{\hbar^2})^{3/2}\int_{E_C}^{\infty} \mathrm{d}E(E-E_\mathrm{C})^{1/2}\frac{1}{e^{\beta(E-E_F)}+1} \tag{1.9}$$

$$= \frac{1}{2\pi^2}(\frac{2m_e}{\hbar^2})^{3/2}\frac{1}{\beta^{3/2}}\int_0^{\infty} \mathrm{d}u\frac{u^{1/2}}{e^u e^{-\beta(E_F-E_C)}+1} \text{where } u=\beta(E-E_\mathrm{C}) \tag{1.10}$$

$$= 2\,(\frac{2\pi m_e k_B T}{h^2})^{3/2}(\frac{2}{\sqrt{\pi}}\int_0^{\infty} \mathrm{d}u\frac{u^{1/2}}{e^u e^{\beta(E_C-E_F)}+1} \tag{1.11}$$

Now we identify the integral within the brackets as a Fermi-Dirac integral, dened as :

$$F_j(z) = \frac{1}{\Gamma(j+1)}\int_0^{\infty} \mathrm{d}x\frac{x^j}{e^z e^x+1} \tag{1.12}$$

Further the "effective density of states" in the conduction band is defined as

$$N_\mathrm{C} = 2(\frac{2\pi m_e k_B T}{h^2})^{3/2} \tag{1.13}$$

Note however that the dimension of N_C is not the same as $D(E)$. The electron and hole densities are then given by,

$$n = N_\mathrm{C}F_{1/2}(\frac{E_\mathrm{C}-E_F}{k_B T}) \tag{1.14}$$

$$p = N_V F_{1/2}(\frac{E_F-E_V}{k_B T}) \tag{1.15}$$

The Fermi-Dirac integrals appear often in physics. They are tabulated as "special functions". We can show that if E_F is reasonably *below* E_C, such that $\frac{E_F-E_C}{k_B T} < -4$ the integral is very closely approximated by $e^{\frac{E_F-E_C}{k_B T}}$. This is called the non-degenerate regime where the electron and hole densities are given by

$$n = N_C e^{\beta(E_F-E_C)} \tag{1.16}$$

$$p = N_V e^{\beta(E_V-E_F)} \tag{1.17}$$

For charge neutrality we must have $n_i = p_i$ for undoped (intrinsic) semiconductors only. Multiplying eqns 1.16 & 1.17 $n_i^2 = n_i p_i = N_C N_V e^{-\beta(E_C-E_V)}$

$$n_i = \sqrt{N_C N_V}\; e^{-\beta E_g/2} \tag{1.18}$$

Clearly the intrinsic carrier density falls rapidly with increasing band gap. We can compare the result with the data in table 1.1. Note however that the product of the carrier densities is independent of the location of the Fermi level even when $n = p$. This is a very important fact and allows us to write

$$np = n_i^2 \tag{1.19}$$

even when the source of the charges are dopants. In such cases (we will see in the next section) the Fermi level moves away from its intrinsic position. The electron and hole densities can them become vastly unequal - but they do so in such a way that the np product still remains the same. We now solve eqns 1.16 & 1.17 for E_f and get the intrinsic Fermi level (E_{fi})

$$E_{fi} = \frac{E_C+E_V}{2} + \frac{3}{4}k_B T ln\frac{m_h}{m_e} \tag{1.20}$$

1.3 Fermi level in a doped semiconductor

We now come to the more practical situation, where there are dopants and ask: where is the Fermi level? If there are dopants then n and p are no longer equal. In fact the number of carriers supplied by ionised dopants can be several orders larger than the intrinsic carrier densities. The fundamental point is that all the atoms of the host lattice and the dopants were initially neutral. But inside the semiconductor there are now four sources of charge :

1. Negatively charged electrons in the conduction band (n)
2. Unoccupied (positively charged) ionised donor atoms (N_d^+)
3. Negatively charged ionised acceptor atoms (N_a^-)
4. Unoccupied states (holes) in the valence band (p)

The sum total of all these four must continue to be zero. To start with the valence band was full and the conduction band was empty (intrinsic semiconductor at T = 0K), then we put in neutral donor atoms (capable of giving out an electron) and neutral acceptor atoms (capable of capturing an electron). So the sum total must remain zero. Thus if we can write down the carrier concentrations in the conduction and valence band and calculate the fraction of dopants which are ionised (as a function of E_f) then we can have an equation where E_f is the only unknown. This is how one determines the location of E_f

$$n + N_A^- = p + N_D^+ \tag{1.21}$$

We know how to calculate n and p as a function of E_f. What is the probability that a donor will ionise? Resort to statistical physics to calculate the fraction N_D^+/N_D. The donor site (*e.g* Phosphorous in Silicon) can exist in 4 states

1. lost its electron (charge = +1 , energy = 0)
2. occupied by a spin up electron (↑, charge = 0, energy = ED)
3. occupied by a spin down electron (↓, charge = 0, energy = ED)
4. occupied by one spin up and one spin down electron (↑↓, charge = -1 , energy = 2ED +U)

where "U" is the large repulsive energy cost of putting two electrons on the same site, making the state very improbable.) The dopant densities are not very large compared to the density of atoms of the host lattice. It is rarely more than 1 in 10^3 to 10^4. So we can treat each dopant atom in isolation and the electron can be localised on the atom[1]. Each dopant can exchange electrons with the "sea" of CB electrons. It is in equilibrium with a larger system (the Si lattice) and can exchange particles with it - thus its temperature and chemical potential must be the same as that of the larger system.

The grand cannonical partition function for a single donor (with μ, the chemical potential set as E_f) is given by,

$$\begin{aligned} Z_G &= \sum_E e^{-\beta(E-\mu N)} \\ &= e^{-\beta(0-0)} + 2e^{-\beta(E_D-E_f)} + e^{-\beta(2E_D+U-2E_f)} \\ &\simeq 1 + 2e^{-\beta(E_D-E_f)} \end{aligned} \tag{1.22}$$

The mean occupancy (probability that the dopant is not ionised) is then,

$$\begin{aligned} 1 - \frac{N_D^+}{N_D} &= 1 \times P(\uparrow) + 1 \times P(\downarrow) + 2 \times P(\uparrow\downarrow) \\ &= \frac{2e^{-\beta(E_D-E_f)}}{Z_G} \\ &= \frac{2e^{-\beta(E_D-E_f)}}{1 + 2e^{-\beta(E_D-E_f)}} \\ &= \frac{1}{\frac{1}{2}e^{-\beta(E_D-E_f)} + 1} \end{aligned} \tag{1.23}$$

The fraction of ionised donors is

$$\frac{N_D^+}{N_D} = \frac{1}{1 + 2e^{-\beta(E_D-E_f)}} \tag{1.24}$$

[1]However, for high dopant densities then the dopant states will no longer be localised but form extended band of finite width. This is known as "Mott transition".

It is worth noting that the above equation indicates that the probability of ionisation of a donor increases with increasing $(E_D - E_f)$. For, $E_D = E_f$ the probability of ionisation is $= \frac{1}{2}$.

A similar expression for the fraction of ionised (negatively charged) acceptors is

$$\frac{N_A^-}{N_A} = \frac{1}{1 + 4e^{-\beta(E_f - E_A)}} \tag{1.25}$$

The factor 4 is a result of the fact that the electron sitting on the acceptor could have come from four possible places - spin up/down from heavy hole band, spin up/down from light hole band. The split off band does not come into the picture because it is too far down.

One type of dopant only
If we neglect the valence band and the acceptors (which can be justified if only donors are present), combining eqns we get

$$N_C e^{-\beta(E_C - E_f)} = \frac{N_D}{1 + 2e^{-\beta(E_D - E_f)}} \tag{1.26}$$

E_f is the only unknown in eqn and can be solved (numerically if required).

1.3.1 Thermal ionisation (Saha equation) of the dopant system

It is instructive to calculate the fraction of ionised dopants in another way. We can think of the problem as a thermal ionization of bound states - in a way that is very similar to the method of calculating the ratio of ionised to unionised atoms (of a certain species) in a hot plasma. We want to find the "chemical equilibrium point" of the reaction:

$$dopant + ionisationenergy \leftrightarrow ioniseddopant + electron \tag{1.27}$$

A certain fraction of atoms will exist in the dissociated state and a certain fraction will remain in the undissociated state. The fraction which minimises the free energy of the entire system (at a certain temperature) will be the equilibrium point. Taking this approach we can calculate the ratio N+ D=ND by minimising the free energy of the entire system of free electrons and the dopants. First we write the free energy such that the free electron density n is the only variable.

$$F_{system} = F_{electrons} + F_{dopants} \tag{1.28}$$

Now $F_{electron}$ may be calculated as

$$F_{electrons} = k_B T ln \frac{z^n}{n!} \tag{1.29}$$

where, the single electron partition function $z = \sum e^{-\beta E} = 2V/h^3 \int d^3\mathbf{p} e^{\frac{-\beta p^2}{2m}} = 2V(2\pi m k_B T/h^2)^{3/2}$. Now since $N_D - n$ dopant sites are occupied we have for the internal energy (U) and entropy (S)

$$U = -\Delta(N_D - n) \tag{1.30}$$

$$S = kln(2^{N_D - n} \frac{N_D!}{n!(N_D - n)!}) \tag{1.31}$$

$$F_{dopants} = U - TS \tag{1.32}$$

1.4 General Method of Solving for E_f

Consider a situation where a semiconductor is doped with ND donors and NA acceptors. We want the general solution for the location of EF and all the carrier densities, ionisation probabilities. Since the semiconductor is overall neutral we have using the charge neutrality condition

$$n + N_A^- = p + N_D^+ \tag{1.33}$$

$$N_C F_{1/2}(\frac{E_C - E_f}{k_B T}) + \frac{N_A}{1 + g_A e^{\beta(E_A - E_f)}} = N_V F_{1/2}(\frac{E_f - E_V}{k_B T}) + \frac{N_D}{1 + g_D e^{\beta(E_f - E_D)}} \tag{1.34}$$

- here $g_A = 4$ and $g_D = 2$ are the acceptor and donor degeneracies. E_A and E_D are the acceptor and donor levels.
- Since E_f is the only unknown here, we can plot the LHS and RHS by treating E_f as an inde pendent variable. The position where they intersect must be the solution. Figure illustrates the situation.
- Notice that E_f is temperature dependent.
- Once E_f is determined all the quantities can be determined. In general this cannot be done analytically.

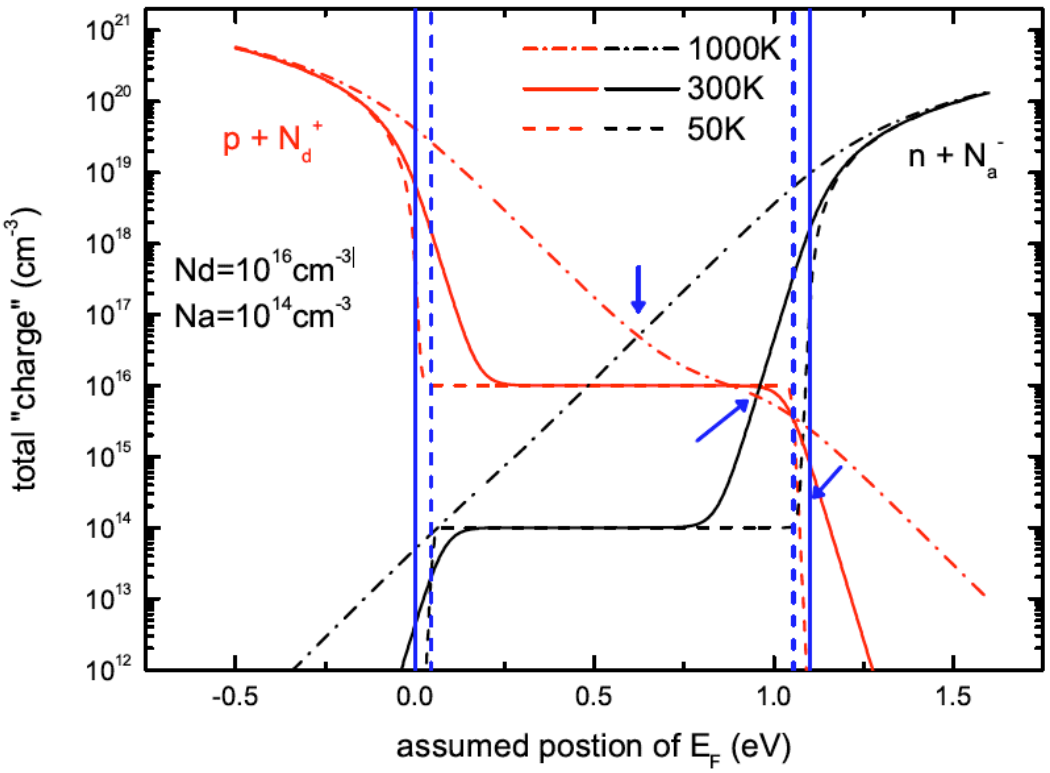


Figure 1.1: Plot of the LHS and RHS of Equation 1.34 as a function of E_f

1.5 Degenerate vs Nondegenerate Semiconductors

In the discussion on the effect of doping and the ensuing calculations, we have implicitly assumed that the concentration of dopant atoms is much smaller in comparison to that of the parent semiconductor atoms. These impurity atoms are sparsely distributed with large separation from each other, effectively preventing any interactions or overlap of electronic wave functions. We've also assumed that these impurities introduce discrete, non-interacting and localised donor energy levels within the n-type semiconductor (non-interacting acceptor states within the p-type semiconductor). Such semiconductors are commonly referred to as non-degenerate semiconductors.

However, as the impurity concentration increases, the separation between these impurity atoms diminishes, eventually reaching a point where donor electron wavefunctions overlap each other. Thereafter, the single electron, discrete donor energy level begins to split into a band of energy levels. With a further increase in donor concentration, this band of donor states widens and may even overlap with the lower edge of the conduction band. These states which are added to the conduction or valence bands are called bandtail states. Evidently, as a consequence of bandtailing, the bandgap narrows and is observed in the absorption spectrum of heavily doped semiconductors. Bandgap narrowing has important consequences in the operation of laser diodes. Here the E_F lies within the expanded conduction band and the material behaves like a metal, thus the system undergoes an insulator to

metal transition and the carrier freeze out temperature of the semiconductor is lowered. Typically this happens when $N_D \sim N_C$ ($N_A \sim N_V$). If n_e N_C the Fermi energy assumes a position within the conduction band. Such a semiconductor is termed a degenerate n-type semiconductor and has $N_D > 10^{20}/cc$. As a rule of thumb, for nondegenerate semiconductors the E_D lies around $k\mathrm{T}/3$ below E_C and the E_F lies $3k\mathrm{T}$ below the E_C.

Carrier mobility (μ), discussed in Chapter 2 is the speed at which electrons and holes can move through the material, which in turn affects the device performance. Higher carrier mobility leads to faster device switching speeds and higher current densities and lower carrier mobility may lead to slower switching speeds and higher power consumption. In heavily doped semiconductors, μ is limited by a number of factors, including scattering from dopants, phonon scattering, and impurity scattering, all of which can lead to degradation of device performance.

1.6 Beyond the Bulk

From the last section we have understood the method of calculating the position of E_f as a function of temperature and doping in bulk semiconductors. Assuming a semiconductor is n-type doped and ignoring holes in the V_B one can obtain an expression for charge density inside the semiconductor as follows;

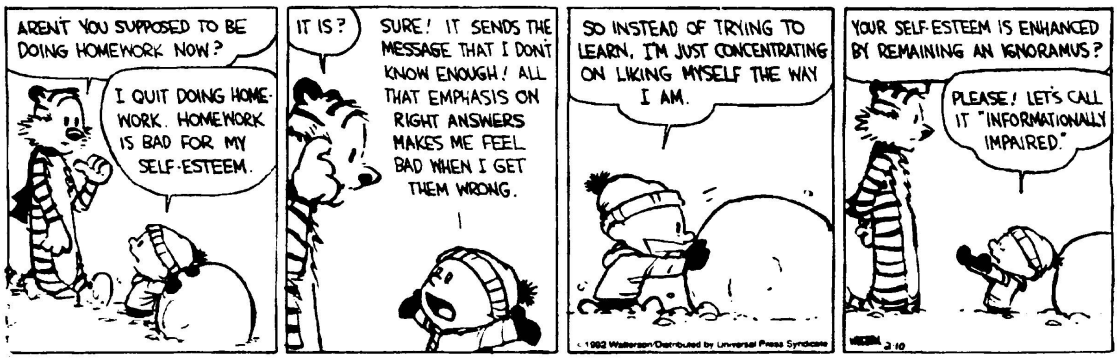
$$\rho = -e(n - N_D^+) = -eN_C e^{\beta(E_f - E_C)} - N_D \frac{1}{1 + 2e^{\beta(E_D - E_f)}} \tag{1.35}$$

Further, this charge density is related to the electrostatic potential (V) by the Poisson equation;

$$\nabla^2 V = -\frac{\rho(\mathbf{r})}{\epsilon_r \epsilon_0} \tag{1.36}$$

where, the scalar potential is essentially the bottom of the C_B. Within the bulk of a semiconductor the potential and the charge density are in general homogeneous. However the situation vastly changes at semiconductor surfaces or interface, especially at junctions.

1.7 PROBLEMS



1. Calculate the density of states in 2D and 1D systems.

2. Consider an electron in a 1D lattice of lattice constant a. The energy levels are given by $\epsilon(k) = -2T\cos(ka)$. A small dc electric field $\vec{E}$ is imposed parallel to the lattice. Plot the energy dispersion $(E - k)$ diagram in absence of the field and write the equation of motion of electrons. Obtain an expression for the time evolution of k and describe qualitatively the motion of the electron in k-space and in real space in both the absence and presence of scattering. What is meant by small $\vec{E}$ and what can happen in a real (multiband) crystal when $\vec{E}$ is no longer small?

3. Show that the deviation of the electron density (n) from intrinsic density (n_i) and the deviation of E_f from the intrinsic Fermi level (E_{fi}) are related as

$$n = n_i e^{\beta(E_f - E_{fi})} \tag{1.37}$$

4. The effective masses of electrons and holes are different in a semiconductor. Discuss an experiment that can determine the effective mass of electrons and holes in a semiconductor.

5. Show that for a semiconductor with only donor dopants, the carrier density can be obtained by solving the following equation: (which is in turn obtained by using eqn)

$$n^2 + nN_C\frac{e^{-\beta\Delta}}{2} - N_D N_C\frac{e^{-\beta\Delta}}{2} = 0 \tag{1.38}$$

where $\Delta = E_C - E_D$. Also show that the Fermi level may be obtained by solving

$$x^2 + x\frac{e^{-\beta\Delta}}{2} - \frac{N_D}{N_C}\frac{e^{-\beta\Delta}}{2} = 0 \tag{1.39}$$

where $x = e^{\beta(E_f - E_C)}$. If you put $N_D = 0$ in either of the two equations you would get an unphysical answer. Why is this so?

6. Consider an intrinsic semiconductor whose electronic density of states $N(E)$ is plotted in the figure 1.2. Show that the Fermi level is given by the expression $E_F = (E_C + E_V)/2$, where E_C and E_V are the conduction and valence band edges. Estimate the density of conduction band electrons at room temperature. n_0 is the number density of electron states .

7. In a system with N_D donors and N_D acceptors, N_D^+ donors and N_A^- acceptors are ionised. Each donor (acceptor) level has a degeneracy of g_D (g_A). There are n electrons in C_B and p holes in V_B. (In general $g_D = 2$, but g_A may be different from 2.). Then

$$N_D^+ = \frac{N_D}{(g_D n/N_C)e^{\beta(E_C - E_D)} + 1} \tag{1.40}$$

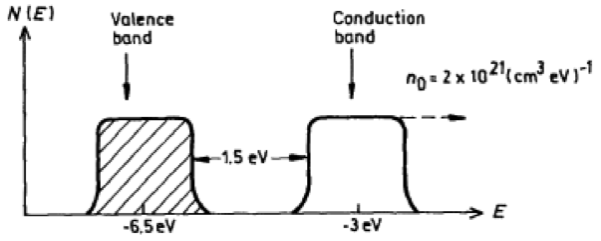


Figure 1.2: Density of States

And the corresponding result for the acceptors:

$$N_A^- = \frac{N_A}{(g_A p/N_V)e^{\beta(E_A - E_V)} + 1} \tag{1.41}$$

Here N_C and N_V are the conduction and valence band effective density of states. Notice that E_f does not appear in these relations. A semiconductor may be doped with both (acceptors and donors) types of dopants. In a situation where there are a large number of donors and a few acceptors (i.e $N_D \gg N_A$), how would the eqn in the previous problem be modified?

8. Determine the temperature at which 90 % of the Boron acceptor atoms in p-type Si are ionized. (given $N_A \sim 10^{16}$/cc).

9. Minimise $F_{system} = F_{electrons} + F_{dopants}$ w.r.t n, using Stirling's approximation for factorials as needed and show that you get exactly the same result as eqn. This is essentially a variant of the "Saha ionisation" equation, applied to a situation where the atoms and ions are not mobile, but only the electrons are.

2

Electrical Transport

Before we delve into electrical transport i.e. motion of electrons in matter, lets revisit the system in which the abundance of electrons is the most obvious, metals.

2.1 What are metals?

- More than $\frac{2}{3}$ of elements in the periodic table are metals.
- Metals occupy the left hand side of the periodic table. Atoms with noble gas core + loosely bound outer electrons.
- Metals form crystal structures with large number of nearest neighbours
- Metals have large interatomic distances but small ionic radii. Thus leaving large volumes for the conduction electrons to swim around.

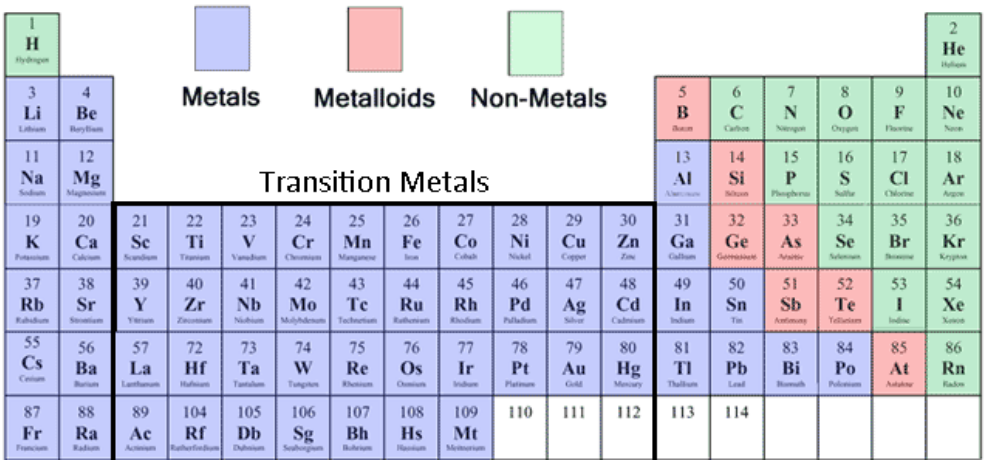


Figure 2.1: Periodic table of elements showing metals, mettaloids and non-metals

2.2 The Drude Model

The Drude model predates the quantum theory and was developed at the turn of the 20th century by Paul Drude, a few years after J.J. Thompson discovered the electron in 1897.

Assumptions

1. Electron "ideal gas" permeates the fixed crystal lattice composed of core ions.
2. Collisions between e's and ions are instantaneous, uncorrelated events.

3. All other interactions (i.e. long range potentials of other ions or $e-e$ interactions) are neglected except externally applied fields.

4. Electrons travel in straight lines between scattering events.

5. The average time between subsequent collisions of such an electron is τ. Probability of an electron having a collision in a time interval dt is dt/τ. τ does not depend on the electron position or momentum.

6. Collisions "thermalise" electrons and restore equilibrium with lattice.

Density of conduction electrons.

In Sodium (Na) the density of conduction electrons, n is:

$$n = N_A \frac{Z_c \rho_m}{A} = 6.02 \times 10^{23} atoms/mol \frac{1e/atom \cdot 1 \times 10^6 g/m^3}{29 g/mol} = 2 \times 10^{28}/m^3 \tag{2.1}$$

where N_A is Avogadro's number, ρ_m is the density of the metal, A is the atomic number of the element and the numbers are for Na.

2.2.1 Equations of motion

The first thing you need is to figure out how an electron's momentum, on average, will evolve over time. To do this we'll just find the average equation of motion for an electron. To find this let's start with the momentum of an electron at time t, $\vec{p}(t)$, and find it at time $t+dt$. If the electrons had a collision it would on average have no momentum ($\vec{p}_c(t+dt) = 0$) at time $t+dt$ and by the third assumption above this has the probability, $P_c = dt/\tau$. This means that the probability of no collision is $P_{nc} = (1 - dt/\tau)$ this is because $P_c + P_{nc} = 1$. If there were no collision the electrons would have evolved normally and the electrons momentum becomes, $\vec{p}_{nc}(t+dt) = \vec{p}(t) + \vec{F}(t)dt$. Thus the average value of the electrons momentum after time dt is,

$$\vec{p}(t+dt) = P_c \cdot \vec{p}_c(t+dt) + P_{nc} \cdot \vec{p}_{nc}(t+dt) = (1 - \frac{dt}{\tau})[\vec{p}(t) + \vec{F}(t)dt] \tag{2.2}$$

The above can be unscrambled to give,

$$\frac{d}{dt}\vec{p}(t) = \frac{\vec{p}(t+dt) - \vec{p}(t)}{dt} = -\frac{\vec{p}(t)}{\tau} + \vec{F}(t) \tag{2.3}$$

And you have the equation of motion (EoM) averaged over electrons.

Of note there are a few regimes and solutions to consider.

- If $\vec{F}(t) = 0$ the solution to this homogeneous equation is $\vec{p}(t) = \vec{p}(0)e^{-t/\tau}$ which is why τ is called the relaxation time. If you impart momentum to the electrons on average they will relax back to no momentum exponentially with a time constant τ

- With a constant $\vec{F}$ e.g. a dc electric field you can show that the solution to the momentum $\vec{p}(t) = \vec{p}(0)e^{-t/\tau} + \vec{F}\tau$

- After a long time, $t \gg \tau$, the exponential term becomes negligible leaving $\vec{p}(t) = \vec{F}\tau$

2.2.2 Ohm's law and dc Conductivity

$$V = IR \tag{2.4}$$

V is the Voltage applied across the metal, I is the resulting current, and R is a proportionality constant. Recasting the equation in geometry independent form, we get

$$\vec{j} = \sigma \vec{E} \tag{2.5}$$

where σ is the conductivity of the metal and $\vec{j}$ is the current density, i.e. $\vec{j} = 1/nev$ where n is the electron density, e the electron charge, and v is the average drift velocity of the electrons. In an applied electric field $\vec{E}$ the EoM for $t \gg \tau$, gives us $\vec{p}(t) = e\vec{E}\tau$ and $\vec{v}(t) = \frac{e\vec{E}\tau}{m}$.

$$\implies \vec{j} = \frac{ne^2\tau}{m}\vec{E} \tag{2.6}$$

which is Ohm's law with $\sigma_o = ne^2\tau/m$ (the dc conductivity). For a metal like Na with a resistivity, $\rho_{Na} = 1/\sigma_o = 50$ n$\Omega\cdot$ the relaxation time is $\sim 10^{-14}$ s. The dc conductivity is also expressed as $\sigma_o = ne\mu$ where μ is the carrier mobility $= e\tau/m$

2.2.3 ac Conductivity

For an oscillating E-field ($\vec{E}(t) = Re\{\vec{E}(\omega)e^{i\omega t}\}$) Ohm's law takes the form

$$\vec{j}(\omega) = \frac{ne^2\tau}{m}\frac{1}{i\omega\tau + 1}\vec{E}(\omega) \tag{2.7}$$

where the ω dependent conductivity is given by

$$\sigma(\omega) = \frac{\sigma_o}{i\omega\tau + 1} \tag{2.8}$$

If the electric field originates from an incident EM wave e.g. incident light, the response of the system will then be given by Maxwell's equation of the form,

$$\vec{\nabla}\cdot\vec{E} = 0 \qquad \vec{\nabla}\cdot\vec{B} = 0$$
$$\vec{\nabla}\times\vec{E} = -\frac{d\vec{B}}{dt} \quad \vec{\nabla}\times\vec{B} = \mu_0\vec{j} + \mu_0\epsilon_0\frac{d\vec{E}}{dt}$$

Inserting the expression for $\vec{j}(\omega)$ and $\vec{E}(t)$, the last equation then reads as

$$\vec{\nabla}\times\vec{B} = (\mu_0\sigma(\omega) + i\omega\mu_0\epsilon_0)\vec{E}$$

Curl of the third equation then leads us to,

$$\vec{\nabla}\times\vec{\nabla}\times\vec{E} = -\vec{\nabla}^2\vec{E} = \frac{d\vec{\nabla}\times\vec{B}}{dt} = i\omega(\mu_0\sigma(\omega) + i\omega\mu_0\epsilon_0)\vec{E}$$

After some rearranging this can be written as

$$-\vec{\nabla}^2\vec{E} = \frac{\omega^2}{c^2}(\frac{i\sigma(\omega)}{\omega\epsilon_0} - 1)\vec{E} \tag{2.9}$$

Now notice that if you take the limit where $\omega\tau \gg 1$, $\sigma(\omega) \to \sigma_0/i\omega\tau$ leaving us with

$$-\vec{\nabla}^2\vec{E} = \frac{\omega^2}{c^2}(\frac{ne^2}{m\epsilon_0\omega^2} - 1)\vec{E} \tag{2.10}$$

Assuming that the electric field has a plane wave spatial structure (i.e. $\vec{E}(t) = Re\{\vec{E}(\omega)e^{i\omega t}e^{i\vec{k}\cdot\vec{r}}\}$)[1] then the above equation simplifies to the dispersion relation,

$$k^2c^2 = (\omega^2 - \omega_p^2)$$

with $\omega_p = ne^2/m\epsilon_0$ being the plasma frequency.

This dispersion relationship tells us that k is imaginary for $\omega < \omega_p$ which is the result for an exponentially decaying electric field. This means that the EM wave doesn't propagate into a metal for frequencies $\omega < \omega_p$ but are instead reflected. On the other hand for $\omega > \omega_p$ it yields a real k which is the form for a travelling wave. This implies that for frequencies higher than the plasma frequency metals become transparent. For metals the plasma frequency is in the ultra violet, e.g. for Na $\omega_p = 10^{15}\text{s}^{-1}$ which corresponds to a wavelength of λ 200nm.

[1] Direct magnetic field effects are neglected since they are much smaller than the electric field forces.

2.2.4 Thermal conductivity

The electrons in a metal aren't just good conductors of electricity, they are also good at conducting heat. So we're asking about the amount of heat that travels along a sample of metal as heat is applied. There is a simple phenomenological model that is analogous to Ohm's law, it is called Fourier's law and it looks like:

$$\vec{j}^q = -\kappa\vec{\nabla}T$$

The minus sign is because heat flows to lower temperature, and $\vec{j}^q$ is the heat current density $\Delta Q/\text{Time}\cdot A$, heat per time per area in units of [Watts/m^2] and κ is the thermal conductivity which has units of [Watts/mK]. So what does the Drude model predict?

A couple of assumptions are important here. First we need to remember the last assumption of the Drude model. This assumption means that after a collision an electron carries the thermal energy of the local environment. Also we need to assume that T varies a little over l (the mean free path electrons travel before scattering).

To calculate the heat delivered to a point we need to simplify this to a 1-D problem and then we need to calculate the heat it gets from the left and subtract the heat it gets from the right. Let's write the thermal energy an electron at temperature T has as $\mathcal{E}(T)$. Then at a point x the average electron coming from the left brings with it an energy $\mathcal{E}(T[x-v\tau])$ the average electron from the right delivers $\mathcal{E}(T[x+v\tau])$. So for electrons of density n and average velocity v (remembering that half will travel towards the point x and half away) this leaves us with

$$\vec{j}^q = \frac{1}{2}nv\{\mathcal{E}(T[x-v\tau] - \mathcal{E}(T[x+v\tau]\}$$

$$\vec{j}^q = \frac{1}{2}nv(\frac{d\mathcal{E}}{dT})(-\frac{dT}{dx})2v\tau$$

$$\vec{j}^q = nv^2\tau(\frac{d\mathcal{E}}{dT})(-\frac{dT}{dx})$$

Now note a few things $n\frac{d\mathcal{E}}{dT} = \frac{N}{V}\frac{d\mathcal{E}}{dT} = \frac{1}{V}\frac{dE}{dT} = c_V$ or the specific heat per volume. Also when we generalize to 3-D we note that $< v_x^2 >=< v_y^2 >=< v_z^2 >= \frac{1}{3} < v >^2$. Putting all this together we have

$$\vec{j}^q = \frac{< v >^2 \tau}{3}c_V(-\vec{\nabla}T)$$

So the prefactor up there is equal to the thermal conductivity.

There is another way to look at the specific heat. It's called the Wiedemann-Franz ratio. Which is given by

$$\frac{\kappa}{\sigma T} = \frac{m < v >^2 \tau c_V}{3e^2}$$

Now in a classical gas using the equipartition theorem the total thermal energy per volume is given by $E = n\frac{3}{2}k_BT$ so the specific heat per volume (the derivative of the energy with respect to energy) is $c_V = n\frac{3}{2}k_B$. And also note that the total thermal energy is the kinetic energy is given by $\frac{1}{2}m < v >^2 = n\frac{3}{2}k_BT$ which we can use to get rid of the $m < v >^2$ in the W-F ratio. All together this gives

$$\frac{\kappa}{\sigma T} = \frac{3}{2}(\frac{k_B}{e})^2 = 1.11 \times 10^{-8}(W\Omega/K^2)$$

This number only has fundamental constants in it and it is 1/2 the real value you get experimentally. But is really close for such a simple model.

Having calculated the equilibrium local charge density in doped systems we now address the issue of electrical transport in such devices. The Boltzmann Transport equation (BTE) allows us to take a step out of equilibrium thermodynamics. In general the concept of equilibrium implies that there is no net particle flow from one point of a system to another. An equivalent statement is that the electrochemical potential is same throughout the system. Yet, in reality, all electrical devices have currents flowing through otherwise it wouldn't be interesting at all. Importantly, some effects of current flow are irreversible -like Joule heating. We will see how the presence of electro-magnetic fields, electrochemical potential gradients and thermal gradients drive current. We will do so by calculating a distribution function in a situation driven *slightly* away from equilibrium. Our main target is derive analytic expressions for current in the presence of external fields.

2.3 The Boltzman Transport Equation

Let's consider the 6-dimensional phase space of two canonically conjugate co-ordinates **r** and **p** pertaining to the position and momentum of the *free* quasi particles in a solid. The probability of dN molecules which all have **r** and **p** within $d^3\mathbf{r}d^3\mathbf{p}$ is a function f which gives this probability per unit phase-space volume, or probability per unit length cubed per unit momentum cubed, at an instant of time t. This is a probability density function: f(r, p, t), defined so that,

$$\mathrm{d}N = f(\mathbf{r}, \mathbf{p}, t)\, \mathrm{d}^3\mathbf{r}\, \mathrm{d}^3\mathbf{p} \tag{2.11}$$

What happens to the points in the "volume element" after some time has elapsed? Unless the electrons are scattered they change their co-ordinates according to the following rule:

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \frac{p}{m}\delta t \tag{2.12}$$

$$\mathbf{p}(t + \delta t) = \mathbf{p}(t) + \mathbf{F}\delta t \tag{2.13}$$

Thus at time $t + \delta t$ we can write;

$$f(\mathbf{r}(t) + \frac{p}{m}\delta t, \mathbf{p}(t) + \mathbf{F}\delta t)d^3\mathbf{r}'d^3\mathbf{p}' = f(r(t), p(t), t)d^3\mathbf{r}d^3\mathbf{p} \tag{2.14}$$

The volume element around the point distorts, but preserves its volume. Something that was a square at time t, may become a parallelogram at $t + \delta t$. This would happen if the external force conservative i.e. is derivable from a potential. Now the time derivative of the distribution function may be written as

$$\frac{df}{dt} = \frac{\mathbf{p}}{m}.\nabla_r f + \mathbf{F}.\nabla_p f + \frac{\partial f}{\partial t} = 0 \tag{2.15}$$

Now we take another step, to convert this classical equation into a semiclassical one. We change momentum to wavevector. This indeed means that we are using the concept of phase space (simultaneously defined momentum and position) in quantum mechanical scenario. An analysis of how far this can give correct results, is non-trivial. For carriers in a 'band-solid' the semiclassical dynamics of a wavepacket are described by the equations

$$\frac{d\mathbf{r}}{dt} = \mathbf{v}_n(k) = \frac{1}{\hbar}\nabla_k \mathcal{E}_n(k) \tag{2.16}$$

$$\frac{dk}{dt} = \frac{q}{\hbar}E(r, t) - \frac{q}{\hbar c}v_n(k) \times B(r, t) \tag{2.17}$$

where, n is the band index and $\mathcal{E}_n(k)$ is the dispersion relation for the band n. The wavevector is k ($\hbar k$ is the 'crystal momentum'), and $\mathcal{E}_n(k)$ is periodic under $k \to k + G$, where G is any reciprocal lattice vector. With these inputs the *semi-classical* version of df/dt is given as;

$$\frac{df}{dt} = \frac{d\mathbf{r}}{dt}.\nabla_r f + \frac{d\mathbf{k}}{dt}.\nabla_k f + \frac{\partial f}{\partial t} \tag{2.18}$$

In the absence of collisions, the distribution function must satisfy the continuity equation in the phase space,

$$\nabla_{rk}.(\mathbf{u}f) + \frac{\partial f}{\partial t} = 0$$

where **u** are the ($\dot{\mathbf{r}}$) and ($\dot{\mathbf{k}}$), and ∇_{rk} are the six component gradient in the phase space namely ($\frac{\partial}{\partial x}$,..., $\frac{\partial}{\partial k_x}$,....). Further since the phase space flow is *incompressible* since $\mathcal{E}_n(k)$ is independent of **r**, in the absence of collision the continuity equation takes the form

$$\mathbf{u}.\nabla_{rk}(f) + \frac{\partial f}{\partial t} = 0 \tag{2.19}$$

.

2.3.1 Collisions - Relaxation Time Approximation

In the presence of collisions the equality $df/dt = 0$ fails to hold and the amount by which it to hold, must be attributed to collisions. Collisions, as in the Drude model holds the key to restoring equilibrium of the system once it is disturbed. Say $f^0(r, k, t)$ is the equilibrium distribution function and when it deviates from equilibrium, a "restoring effect" arises in the system, that tries to push the distribution back towards equilibrium. This modifies the continuity equation to include the *restoring effect* of *collisions*

$$\frac{df}{dt} = \frac{d\mathbf{r}}{dt}.\nabla_r f + \frac{d\mathbf{k}}{dt}.\nabla_k f + \frac{\partial f}{\partial t} = \frac{df}{dt}|_{collision} \tag{2.20}$$

The above is the semiclassical form of the Boltzmann Transport Equation. Its crucial departure from the basic phase space continuity equation is the inclusion of the collision term, which in general may be a function of **r**, **k**, t. The RHS of the above equation may be modelled via varrious schemes. One straightforward sche is the relaxation time approximation, where;

$$\frac{df}{dt}|_{collision} = -\frac{f - f^0}{\tau} \tag{2.21}$$

$f^0(r, k)$ is a static distribution function which describes a local equilibrium at **r**. The quantity $\tau(k)$ is the relaxation time, which is allowed to be energy-dependent and may be determined in terms of various scattering mechanisms systems. The best justification of the relaxation time approximation is that it works in many cases! With these inputs, the *semiclassical* version of the BTE is given as,

$$\frac{d\mathbf{r}}{dt}.\nabla_r f + \frac{d\mathbf{k}}{dt}.\nabla_k f + \frac{\partial f}{\partial t} = -\frac{f - f^0}{\tau} \tag{2.22}$$

In moving forward we assume;

1. External fields are sufficiently weak
2. Interband transitions are neglected - we drop the band index n. They neglect, for example, Zener tunneling processes in which an electron may change its band index as it traverses the Brillouin zone.
3. Neglect spin-orbit interactions
4. f has no explicit time dependence, $\frac{\partial f}{\partial t} = 0$
5. $\nabla_r f = 0$, which in general means that there is no density gradient of particles across the system. This assumption is correct if we are dealing with a piece of copper wire at constant temperature, but not necessarily true for systems with a thermal gradient.

We will assume that the charge of each particle is "q". For the most common case of electrons in the conduction band we would need to put q = -|e| to get the correct sign of the terms.

2.3.2 Electric field only

With these assumptions, the BTE in the presence of only an electric field reduces to

$$\frac{q}{\hbar}E.\nabla_k f = -\frac{f - f^0}{\tau} \tag{2.23}$$

Assuming that the change to the distribution function, under the action of the E field is small, we make the first order approximation by taking the derivative around the equilibrium value, f^0,

$$f(k) = f^0(k) + \delta f = f^0(k) - \frac{q\tau}{\hbar}E.\nabla_k f^0 \tag{2.24}$$

$$= f^0(k - \frac{q\tau}{\hbar}E) \tag{2.25}$$

This means that the equilibrium distribution function has retained its functional form but just got shifted by a certain amount. Think of how the graph of a function $f(x)$ would be related to $f(x - a)$.

In the figure we have drawn it for a Fermi distribution in 2 dimensions. Note that if the relaxation mechanism is strong then τ would be small. On the other hand if the particle suffers very little scattering then τ would be large and the displacement of the Fermi circle (or sphere) would also be large. Our target is to calculate the current produced by this state:

$$\begin{aligned} \mathrm{j} &= q\sum_{\mathrm{k}} \mathrm{v}\delta f \\ &= \frac{2q}{(2\pi)^3}\int \mathrm{d^3kv}\delta f \end{aligned} \tag{2.26}$$

Consider the equilibrium Fermi distribution, which is a space-independent and time-independent solution to the Boltzmann equation.

$$f^0(\mathrm{k}) = \frac{1}{e^{\beta(\mathcal{E}(k)-\mu)}+1} \tag{2.27}$$

In general however, since collisions act locally in space, they act on short time scales to establish a local equilibrium that may be described by a distribution function,

$$f^0((\mathrm{r},\mathrm{k},\mathrm{t})) = \frac{1}{e^{\beta(r,t)(\mathcal{E}(k)-\mu(r,t))}+1} \tag{2.28}$$

This is, however, not a solution to the full Boltzmann equation due to the 'streaming terms' $r.\partial_r$ and $k.\partial_k$.

Here we will neglect any time and spatial dependence of μ ($=\mathcal{E}_f$) and β and calculate δf.

$$\begin{aligned} \nabla_k f^0 &= -(\frac{1}{e^{\beta(\mathcal{E}-\mathcal{E}_f)}+1})^2 e^{\beta(\mathcal{E}-\mathcal{E}_f)}\nabla_k\beta(\mathcal{E}-\mathcal{E}_f) \\ &= -\beta f^0(1-f^0)\nabla_k\mathcal{E} \\ &= -\beta f^0(1-f^0)\hbar\mathrm{v_g} \end{aligned} \tag{2.29}$$

Note that $\mathcal{E}_f$ is not a function of k. The above equation can also be written as,

$$\nabla_k f^0 = \frac{\partial f^0}{\partial\mathcal{E}}\hbar\mathrm{v_g} \tag{2.30}$$

Thus we get δf as

$$\delta f = q\tau\beta f^0(1-f^0)\mathrm{E.v_g} \tag{2.31}$$

Notice that the change occurs only near the Fermi surface. This is the generic reason phenomena like electrical or heat conduction are often referred to as a "Fermi surface property". Now we calculate the current as dened in eqn A.10

$$\begin{aligned} \mathrm{j} = &\frac{q}{4\pi^3}\int \mathrm{d^3k v_g}\Big(q\tau\beta f^0(1-f^0)\mathrm{E.v_g}\Big) \\ &= nq\Big(\frac{q}{4\pi^3 n}\int \mathrm{d^3k}\tau\mathrm{v_g}\otimes\mathrm{v_g}\Big(-\frac{\partial f^0}{\partial\mathcal{E}}\Big)\Big).\mathrm{E} \end{aligned} \tag{2.32}$$

Notice that the part within the large brackets is determined by equilibrium properties of the system only. The outer product ($\otimes$) of two vectors is an object with two indices and can be written out like a matrix. For example

$$\mathrm{C} = \mathrm{A}\otimes\mathrm{B} \implies C_{ij} = A_iB_j \tag{2.33}$$

We will call the quantity inside the bracket as mobility. But it is often not necessary to evaluate this is full generality. We assume that the dispersion relation is spherically symmetric and evaluate the expression for low temperature. Low temperature implies that the Fermi distribution has a sharp drop near E_f and behaves like a step function at that point. The derivative of a step function is a (Dirac) delta function which would pick out the contribution of the integrand around its peak. So we can write

$$\lim_{T\to 0} -\frac{\partial f^0}{\partial\mathcal{E}} = \delta(\mathcal{E}-\mathcal{E}_f) \tag{2.34}$$

Let's go through the steps for evaluating the mobility integral:

$$\overleftrightarrow{\mu} = \frac{q}{4\pi^3 n}\int \mathrm{d^3k}\tau \mathrm{v_g}\otimes \mathrm{v_g}(-\frac{\partial f^0}{\partial \mathcal{E}}) \tag{2.35}$$

$$= \frac{q}{n}\int \mathrm{d}\mathcal{E}D(\mathcal{E})\tau \mathrm{v_g}\otimes \mathrm{v_g}(-\frac{\partial f^0}{\partial \mathcal{E}}) \tag{2.36}$$

$$= \frac{q}{n}\int \mathrm{d}\mathcal{E}D(\mathcal{E})\tau \mathrm{v_g}\otimes \mathrm{v_g}\delta(\mathcal{E}-\mathcal{E}_f) as \mathrm{T}\to 0 \tag{2.37}$$

Now since $\mathrm{v_g} = \hbar \mathrm{k}/m$, we can write:

$$\mu_{ij} = \frac{q}{n}\int \mathrm{d}\mathcal{E}D(\mathcal{E})\tau\Big(\frac{\hbar}{m}\Big)^2 k_i k_j \delta(\mathcal{E}-\mathcal{E}_f) \tag{2.38}$$

This form works in all dimensions, provided the density n is interpreted correctly. Now μ_{ij} will average to zero if $i \neq j$, due to symmetry. If we fix k_i, we can find corresponding pairs of points at k_j and $-k_j$, which will add up to zero. So we need to calculate only the diagonal terms. Since there is nothing to distinguish the x, y or z directions, all the diagonal components must be equal. This allows us to write:

$$\mu_{ii} = \frac{q}{n}\int \mathrm{d}\mathcal{E}D(\mathcal{E})\tau(\frac{\hbar}{m})^2\frac{k_x^2+k_y^2+k_z^2}{3}\delta(\mathcal{E}-\mathcal{E}_f)$$

$$= \frac{q}{3n}\int \mathrm{d}\mathcal{E}D(\mathcal{E})\tau\frac{2\mathcal{E}}{m}\delta(\mathcal{E}-\mathcal{E}_f) \tag{2.39}$$

Using the expression for density of states in 3D, eqn A.23 reduces to:

$$\mu_{ii} = \frac{q}{3\pi^2 mn}\int \mathrm{d}\mathcal{E}k^3\tau\delta(\mathcal{E}-\mathcal{E}_f)$$

$$= \frac{q\tau}{m} \qquad \text{since} \qquad k_F^3 = 3\pi^2 n \tag{2.40}$$

The above treatment assumes that there is no explicit t dependence of $f(\mathbf{k})$, and the solution is obtained with $\frac{\partial f}{\partial t} = 0$ i.e. a steady state is reached under the acion of the $\mathbf{E}$ field. Now say the electric field was acting on the system and at time $t = 0$ it is turned off. Now the $f(\mathbf{k})$ will relax back to $f^o(\mathbf{k})$ and the corresponding equation is given by;

$$\frac{\partial f}{\partial t} = -\frac{f-f^0}{\tau} \tag{2.41}$$

Note that $f(\mathbf{k})$ now has an explicit t dependence and given by the solution of the equation;

$$\frac{\partial}{\partial t}(f^o+\delta f) = -\frac{f-f^0}{\tau} \tag{2.42}$$

Since $\frac{\partial f^o}{\partial t} = 0$ the solution of the above equation is given as $\delta f(t) = \delta f(0)\exp(-t/\tau)$ where $\delta f(0) = -\frac{q\tau}{\hbar}E.\nabla_k f^0 = q\tau\beta f^0(1-f^0)\mathrm{E.v_g}$. Thus the steady state distribution relaxes back to the equilibrium Fermi distribution with a time constant τ.

Xtra: Temperature dependence of mobility and conductivity

The mobility is a temperature dependent quantity-the T dependence of conductivity for example arises from changes in mobility as well as carrier density of a system. Usually scattering calculations give us the scattering rate ($\tau(\mathcal{E})$) or the collision cross section as a function of $\mathcal{E}$. How do we use this information to calculate $\mu(T)$. Let's consider the diagonal element (say μ_{xx}) from equation A.15 which relates j and E

$$\mu_{ij} = \frac{q\int_0^\infty d\mathcal{E}D(\mathcal{E})\tau(\mathcal{E})v_i v_j(-\frac{\partial f}{\partial \mathcal{E}})}{n}$$

$$\mu_{xx} = \frac{q\int_0^\infty d\mathcal{E}D(\mathcal{E})\tau(\mathcal{E})v_x^2(-\frac{\partial f}{\partial \mathcal{E}})}{\int_0^\infty d\mathcal{E}D(\mathcal{E})f(\mathcal{E})}$$

Now if we are working in d dimensions, then in general we have

$$D(E) \propto \mathcal{E}^{d/2-1}$$
$$E \propto m v_x^2 \frac{d}{2} \tag{2.43}$$

Using these two results and a partial integration of the denominator we get:

$$\mu_{xx} = \frac{2q}{md} \frac{\int_0^\infty d\mathcal{E}\mathcal{E}^{d/2}\tau(\mathcal{E})(-\frac{\partial f}{\partial E})}{\int_0^1 d\mathcal{E}\frac{\mathcal{E}^{d/2}}{d/2}(-\frac{\partial f}{\partial \mathcal{E}})}$$
$$= \frac{q}{m} \frac{\int_0^\infty d\mathcal{E}\mathcal{E}^{d/2}\tau(\mathcal{E})\frac{\partial f}{\partial \mathcal{E}}}{\int_0^\infty d\mathcal{E}\mathcal{E}^{d/2}\frac{\partial f}{\partial \mathcal{E}}}$$

Since $\mu = \frac{q\tau}{m}$, we usually write,

$$\langle \tau(T) \rangle = \frac{\int_0^\infty d\mathcal{E}\mathcal{E}^{d/2}\tau(\mathcal{E})\frac{\partial f}{\partial \mathcal{E}}}{\int_0^\infty d\mathcal{E}\mathcal{E}^{d/2}\frac{\partial f}{\partial \mathcal{E}}} \tag{2.44}$$

$\tau(\mathcal{E})$ is often available from scattering calculations and the integral gives the energy range over which we need to average it. The presence of the term $\frac{\partial f}{\partial \mathcal{E}}$ ensures that the important part is centred at Fermi energy, the spread of the region increases with increasing temperature.

Xtra: Conservation of the phase space volume

We will apply the BTE to a situation where the "forces" will have some velocity dependence, like the Lorentz force. So let's prove that the "volume" will still be conserved. Part of the proof is left as an exercise. We will work with two variables only for simplicity. Consider the points $(x,\ p)$ and a small area element $\delta x \delta p$ around it as before. What happens to the corner points after time δt ? Both $\dot{x}$ and $\dot{p}$ can be functions of x and p, but we do not write all the functional dependances explicitly. See the following table:

point	time $= t$	time $= t + \delta t$
$1 \to 1'$	$(p + \dot{p}\delta t x + \dot{x}\delta t)$	
$2 \to 2'$	$(x + \delta x p)$	$(x + \delta x + (\dot{x} + \frac{\partial \dot{x}}{\partial x}\delta x)\delta t p + (\dot{p} + \frac{\partial \dot{p}}{\partial x}\delta x)\delta t)$
$4 \to 4'$	$(p + \delta p x)$	$(p + \delta p + (\dot{p} + \frac{\partial \dot{p}}{\partial p}\delta p)\delta t x + (\dot{x} + \frac{\partial \dot{x}}{\partial p}\delta p)\delta t)$

2.3.3 Electric and Magnetic Field

Consider the case where an electric and magnetic field. The deviation of the distribution function $f(\mathrm{k})$ from its equilibrium distribution ($f^0(\mathrm{k})$) should now read:

$$f(\mathrm{k}) = f^0(\mathrm{k}) - \frac{q\tau}{\hbar}(\mathrm{E} + \mathrm{v} \times \mathrm{B}).\nabla_k f^0 \tag{2.45}$$

Since the force term is $\mathbf{k}$ dependent, the solution for $f(\mathbf{k})$ is not readily obtained by inspection, as was done in the case for $\mathbf{E}$ field only. However, we now try a solution of the same form, with an unknown vector $\mathbf{Z}$. Our target is to write $\mathbf{Z}$ as a function of $\mathbf{E}$ and $\mathbf{B}$, but free of $\mathbf{k}$ and $\mathbf{v}_g$. Thus we want $\mathbf{Z}$, such that

$$f(\mathrm{k}) = f^0(\mathrm{k} - \frac{q\tau}{\hbar}\mathrm{Z}) \tag{2.46}$$

Hence,

$$\delta f = f(\mathbf{k}) - f^o(\mathbf{k}) = -\frac{q\tau}{\hbar}\mathrm{Z}.\nabla_k f^0 \tag{2.47}$$

The BTE now becomes,

$$\frac{q}{\hbar}(\mathrm{v} \times \mathrm{B}).(\nabla_k(f^0 + \delta f)) + \frac{q}{\hbar}\mathrm{E}.\nabla_k(f^0 + \delta f) = -\frac{\delta f}{\tau} \tag{2.48}$$

We already know that $\nabla_k f^0$ points along $\mathrm{v_g}$ and hence the first term gives zero. Now comparing the above 2 equations give us,

$$\frac{q}{\hbar}(\mathrm{v}\times\mathrm{B}).\nabla_k\delta f + \frac{q}{\hbar}\mathrm{E}.\nabla_k f^0 = \frac{q}{\hbar}\mathrm{Z}.\nabla_k f^0 \tag{2.49}$$

Now we need to calculate $\nabla_k\delta f$.

$$\begin{aligned}\nabla_k\delta f &= \nabla_k\frac{q\tau}{\hbar}\mathrm{Z}.\nabla_k f^0\\ &= \frac{q\tau}{\hbar}\nabla_k(-\beta f^0(1-f^0)\mathrm{Z}.\hbar\mathrm{v_g})\\ &= -\beta q\tau((1-f^0)\ (\mathrm{Z}.\mathrm{v_g})\nabla_k f^0 + f^0(\mathrm{Z}.\mathrm{v_g})\nabla_k(1-f^0) + f^0(1-f^0)\nabla_k(\mathrm{Z}.\mathrm{v_g}))\end{aligned} \tag{2.50}$$

Once again the first two terms in the RHS of A.33 will give zero when dotted with $\mathrm{v}\times\mathrm{B}$ as they are $\propto \mathrm{v_g}$. The only term left is

$$\nabla_k Z.\mathrm{v_g} = \nabla_k Z.\frac{\hbar\mathrm{k}-q\mathrm{A}}{m} = \frac{\hbar}{m}\mathrm{Z} \tag{2.51}$$

In eqn A.34, A denotes the vector potential of the magnetic field, $\mathrm{v_g}$ is related to the canonical momentum in presence of a magnetic field in the usual way. Combining eqns A.33 and A.34 we can write:

$$(\mathrm{v}\times\mathrm{B}).\nabla_k\delta f = -\beta q\tau f^0(1-f^0)(\mathrm{v}\times\mathrm{B}).\frac{\hbar}{m}\mathrm{Z} \tag{2.52}$$

So eqn A.32 now simplies to:

$$\begin{aligned}&-\frac{\hbar}{m}\beta q\tau f^0(1-f^0)(\mathrm{v}\times\mathrm{B}).\mathrm{Z} + (\mathrm{E}-\mathrm{Z}).\nabla_k f^0 = 0\\ &\therefore -\frac{\hbar}{m}\beta q\tau f^0(1-f^0)(\mathrm{v}\times\mathrm{B}).\mathrm{Z} + (\mathrm{E}-\mathrm{Z})\beta f^0(1-f^0)\hbar\mathrm{v_g} = 0\\ &\therefore \frac{q\tau}{m}(\mathrm{v_g}\times\mathrm{B}).\mathrm{Z} + (\mathrm{E}-\mathrm{Z}).\mathrm{v_g} = 0\\ &\therefore \frac{q\tau}{m}(\mathrm{B}\times\mathrm{Z}).\mathrm{v_g} + (\mathrm{E}-\mathrm{Z}).\mathrm{v_g} = 0\\ &\mathrm{E} = \mathrm{Z} - \frac{q\tau}{m}\mathrm{B}\times\mathrm{Z}\end{aligned} \tag{2.53}$$

We call Z as the Hall vector. When both E and B fields are present, this quantity in some way, "replaces" the electric field in the transport equation. But we still need to express Z explicitly in terms of E and B, with $\mu = q\tau/m$. The proof is left as homework.

$$\mathrm{Z} = \frac{\mathrm{E}+\mu\mathrm{B}\times\mathrm{E}+\mu^2(\mathrm{B}.\mathrm{E})\mathrm{B}}{1+\mu^2B^2} \tag{2.54}$$

Continuing the similarity between **Z** and **E** further, we write the expression for current in presence of a magnetic field by replacing E by Z,

$$\mathbf{j} = \sigma_0\mathbf{Z} = \frac{nq^2\tau}{m}\mathbf{Z} = nq\mu\mathbf{Z} \tag{2.55}$$

A very general expression with arbitrary E and B can be written, but is not very useful. Rather, we consider a situation where the magnetic field points along $\hat{z}$, and the electric field is in the xy plane. So we have :

$$\begin{aligned}\mathrm{E} &= E_x\hat{\mathrm{x}} + E_y\hat{\mathrm{y}}\\ \mathrm{B} &= B_0\hat{z}\end{aligned} \tag{2.56}$$

and hence:

$$\begin{aligned}Z_x &= \frac{E_x - \mu B_0 E_y}{1+\mu^2B_0^2}\\ Z_y &= \frac{E_y + \mu B_0 E_x}{1+\mu^2B_0^2}\end{aligned}$$

Figure 2.2: Two device geometries commonly used in experiments with 2-dimensional systems

Eqn A.38 then can be written out in 2×2 matrix form as :

$$\begin{pmatrix} j_x \\ j_y \end{pmatrix} = \frac{\sigma_0}{1+\mu^2 B_0^2} \begin{pmatrix} 1 & -\mu B_0 \\ \mu B_0 & 1 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \tag{2.57}$$

Which can be inverted to give the resistivity matrix:

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} \rho_0 & \frac{B_0}{nq} \\ -\frac{B_0}{nq} & \rho_0 \end{pmatrix} \begin{pmatrix} j_x \\ j_y \end{pmatrix} \tag{2.58}$$

where we have written ρ_0 for $1/\sigma_0$.

How do we relate the above to experimental situations? Consider a rectangular block in the xy plane, with the current injecting contacts placed as shown. Sufficiently away from the contacts, the current component j_y must vanish, because there are no current sourcing/withdrawing contacts on the long sides. This allows us to interpret the ratio E_x/j_x as the longitudinal voltage drop and E_y/j_x as the Hall (transverse) voltage. The off-diagonal terms are linear in B and offers the most common way of measuring the electron density in a 2-dimensional system.

It is important to understand that resistance or conductance can no longer be specied by a single number in presence of a magnetic field. They must be understood in a matrix sense. In fact by inverting the resistivity matrix you can easily show that in a magnetic field both σ_{xx} and ρ_{xx} can be simultaneously zero, which appears counter-intuitive at first glance-but there is no contradiction in it.

2.4 Moments of the transport equation: Continuity & Drift-diffusion

Taking the moments of a differential equation means multiplying both sides of the equation with some function and integrating over all states/space. How does that help? The integration "removes" some variable and results in a simpler looking equation. Of course the "simpler" equation is no longer as detailed or informative as the original one-but sometimes we may need focus on a broad feature while removing some details. The BTE refers to the distribution function which is not always possible (or necessary) to know. We show cases where focussing on quantities averaged over the distribution $f(\mathrm{r},\ \mathrm{k})$ is immensely useful.

2.4.1 Integrating the BTE over k space

Integrate/sum over all k states.

$$\frac{\partial f}{\partial t} + \frac{q}{\hbar}(\mathrm{E}+\mathrm{v}\times\mathrm{B})\nabla_{\mathrm{k}} f + \mathrm{v}\nabla_{\mathrm{r}} f = \frac{df}{dt}\Big|_{collision} \tag{2.59}$$

The first term in LHS gives

$$\int \frac{d^3\mathrm{k}}{(2\pi)^3}\frac{\partial f}{\partial t} = \frac{\partial n(\mathrm{r})}{\partial t} \tag{2.60}$$

where $n(\mathrm{r})$ is the conventional particle density at r. The second term in LHS can be written as,[2]

$$(\mathrm{E}+\mathrm{v}\times\mathrm{B}).\nabla_{\mathrm{k}} f = \nabla_{\mathrm{k}}.f(\mathrm{E}+\mathrm{v}\times\mathrm{B}) - f\nabla_{\mathrm{k}}.(\mathrm{E}+\mathrm{v}\times\mathrm{B}) \tag{2.61}$$

The integral of $\nabla_{\mathrm{k}}.f(\mathrm{E}+\mathrm{v}\times\mathrm{B})$ can be converted to a surface integral. With the fields independent of k, the Fermi distribution $\sim e^{-k^2}$[3], and the surface growing as k^2 this integral will vanish. Integral of $f\nabla_{\mathrm{k}}.(\mathrm{E}+\mathrm{v}\times\mathrm{B}$ also goes to zero, as seen below.

$$\begin{aligned} \nabla_{\mathrm{k}}.(\mathrm{E}+\mathrm{v}\times\mathrm{B}) &= \frac{\partial E_i}{\partial k_i} + \epsilon_{ijk}\frac{\partial}{\partial k_i}\frac{\partial \mathcal{E}}{\partial k_j}B_k \\ &= 0 + \epsilon_{ijk}\frac{\partial^2 \mathcal{E}}{\partial k_i k_j}B_k \\ &= 0 + 0 \end{aligned} \tag{2.62}$$

[2] use vector identity. $\mathrm{A}.\nabla f = \nabla.f\mathrm{A} - f\nabla.\mathrm{A}$. f is a scalar and A is a vector.

[3] FD $\rightarrow$ MB distribution for large k

Now, integral of the third term in LHS of equation 2.55 is

$$\begin{aligned}\int \frac{d^3\mathrm{k}}{(2\pi)^3}\mathrm{v}.\nabla_\mathrm{r} f &= \nabla_\mathrm{r}. \int \frac{d^3\mathrm{k}}{(2\pi)^3}\mathrm{v} f \\ &= \nabla_\mathrm{r}.n(\mathrm{r})\langle \mathrm{v}\rangle \end{aligned} \tag{2.63}$$

The RHS of equation 2.55 must give zero when integrated over all *k-space* because the particles which are scattered out of a certain volume must be appearing in some other volume. Thus the integral of the BTE over all k-space yields,

$$\frac{\partial n(\mathrm{r})}{\partial t} + \nabla_\mathrm{r}.n(\mathrm{r})\langle v\rangle = 0 \tag{2.64}$$

Which is nothing but the expected continuity equation.

2.4.2 Drift-diffusion equation

Multiply both sides of the BTE by velocity (or momentum) and integrating over all states. Lets consider a 1D BTE with an electric field only. The 3D case with both electric and magnetic fields will follow.

$$\frac{\partial f}{\partial t} + \frac{q}{\hbar}E\frac{\partial f}{\partial k} + v\frac{\partial f}{\partial x} = -\frac{f-f^0}{\tau} \tag{2.65}$$

multiply by v and integrate over all k. The first term in the LHS gives:

$$\int \frac{dk}{2\pi} v\frac{\partial f}{\partial t} = \int \frac{dk}{2\pi}\frac{\partial}{\partial t} fv = \frac{\partial}{\partial t} n\langle v\rangle \tag{2.66}$$

With $v = \frac{\hbar k}{m}$, the second term in LHS gives:

$$\int \frac{dk}{2\pi} v\frac{\partial f}{\partial k} = \int \frac{dk}{2\pi}(\frac{\partial}{\partial k} fv - f\frac{\partial v}{\partial k}) = fv|_{-\infty}^{\infty} - \frac{\hbar}{m}n \tag{2.67}$$

The third term in LHS gives:

$$\begin{aligned}\int \frac{dk}{2\pi} v^2\frac{\partial f}{\partial x} &= \frac{\partial}{\partial x}\int \frac{dk}{2\pi} v^2 f = \frac{\partial}{\partial x} n(x)\langle v^2\rangle \\ &= \frac{\partial}{\partial x} n(x)\Big\langle \frac{2E}{m}\Big\rangle = \frac{kT}{m}\frac{\partial n(x)}{\partial x}\end{aligned} \tag{2.68}$$

Note the use of thermal average kinetic energy from the classical MB distribution in the last step. This will ultimately lead to a relation between mobility and diffusion constant.
The RHS term:

$$\begin{aligned}-\int \frac{dk}{2\pi} v\frac{f-f^0}{\tau} &= \frac{1}{\tau}\int \frac{dk}{2\pi} fv \\ &= \frac{1}{\tau} n\langle v\rangle\end{aligned} \tag{2.69}$$

Reassembling the four integrals together and multiplying with τ and q allover we get:

$$q\tau\frac{\partial}{\partial t} n\langle v\rangle + qn\langle v\rangle + \underbrace{\frac{q^2\tau}{\hbar}E\Big(-\frac{\hbar}{m}n\Big)}_{\text{drift term}(\mu=\frac{q\tau}{m})} + \underbrace{q\tau\frac{kT}{m}\frac{\partial n(x)}{\partial x}}_{\text{diffusion term}(D=\frac{\tau kT}{m})} = 0 \tag{2.70}$$

$$q\tau\frac{\partial}{\partial t} n\langle v\rangle + qn\langle v\rangle + \underbrace{nq\mu E}_{\text{drift current}} + \underbrace{qD\frac{\partial n(x)}{\partial x}}_{\text{diffusion current}} = 0 \tag{2.71}$$

where D is the diffusion constant and the ratio $\frac{D}{\mu} = \frac{kT}{q}$, is called the Einstein relation. This is correct for a classical distribution and applicable to degenerate semiconductors only. Notice $\hbar$ has disappeared, another indication that the result is essentially classical. The relation between drift and diffusion components would be different if full Fermi-Dirac distribution used. However at room temperatures in most devices this holds well for motion of electrons/holes in a band. The values of D are thus readily obtained from the mobility values.

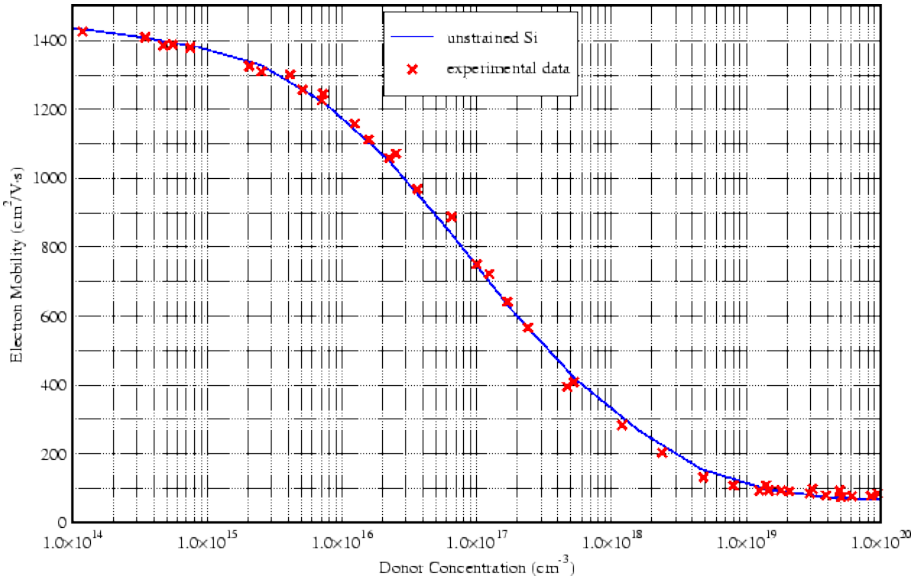


Figure 2.3: Variation in mobility of Si with dopant density.

Drift diffusion in 3D

Removing the simplifying assumptions and take the moment of the BTE after multiplying with v. We need to work with

$$\underbrace{\int \frac{d^3k}{(2\pi)^3} v \frac{\partial f}{\partial t}}_{\text{LHS1}} + \underbrace{\frac{q}{\hbar}\int \frac{d^3k}{(2\pi)^3} v(E + v \times B).\nabla_k f}_{\text{LHS2}} + \underbrace{\int \frac{d^3k}{(2\pi)^3} vv.\nabla_r f}_{\text{LHS3}} = -\int \frac{d^3k}{(2\pi)^3} v \frac{f - f^0}{\tau} \tag{2.72}$$

The complete calculation can proceed along the following lines. Prove the following results. The relation between velocity and the wavevector is $m\mathrm{v} = \hbar\mathrm{k} - q\mathrm{A}$

1. LHS 1 : This gives

$$\frac{\partial}{\partial t} n(\mathrm{r})\langle \mathrm{v} \rangle$$

2. LHS 2 : Notice the occurrence of the *averaged velocity* in the Lorentz term. The calculation is somewhat non-trivial. Do it carefully! You should get

$$-\frac{q}{m} n(\mathrm{r})(\mathrm{E} + \langle \mathrm{v} \rangle \times \mathrm{B})$$

3. LHS 3: The diffusion term requires averaging over the distribution. You should get

$$\nabla_\mathrm{r}.n(\mathrm{r})\langle v_i v_j \rangle$$

For Maxwell-Boltzmann distribution $\langle v_i v_j \rangle = \dfrac{kT}{m}\delta_{ij}$

4. RHS : This gives the current term

$$-n(\mathrm{r})\frac{\langle \mathrm{v} \rangle}{\tau}$$

Adding all the results will give the drift diffusion relation.

2.5 Thermal and Electrochemical Gradients

The preceding sections may give you the impression that current (particle ow) must be associated with response to electric or magnetic field. This is not true as was seen in the context of the diffusion current term. There are also very striking instances where there is a strong electric field but no charge

flow. Also at the end of this section we will be able to answer the question - what does a voltmeter measure? We have more or less got accustomed to the idea that connecting a voltmeter between two points would measure the line integral of the electric field ("potential difference") between the points. While this is indeed true in many circumstances there are situations where it is not. For example if you look near the surface of a metal or a semiconductor or a *pn* junction you will find very strong electric field at equilibrium. But a voltmeter connected between the surface of a metal and somewhere inside would give zero. So would a voltmeter connected across a p-n junction in equilibrium. The purpose of this section is to show that particle flow is ultimately related to the gradients of the "electrochemical potential" (Fermi level) and thermal gradients in the system. It is an important conceptual point for treating electrons moving in conduction band of a semiconductor that may have spatial variation due to changes in composition or effects of accumulated charges or external gates.

So let us consider a general case where the equilibrium distribution function changes in response to external stimuli. To obtain a solution to the BTE we write;

$$\frac{d\mathbf{r}}{dt}.\nabla_r f + \frac{d\mathbf{k}}{dt}.\nabla_k f + \frac{\partial f}{\partial t} = -\frac{f - f^0}{\tau}$$

with

$$f(r,k,t) = f^0(r,k,t) + \delta f(r,k,t)$$

Assuming that the collisions acting locally on a short time scale establish local equilibrium(s) we can write a local form of the distribution function as,

$$f^0(\mathrm{r},\mathrm{k},\mathrm{t}) = \frac{1}{e^{\beta(r,t)(\mathcal{E}(r,k)-\mu(r,t))}+1}^{4} \tag{2.73}$$

Assume $\mu = \mu(r)$, $T = T(r)$, $\mathcal{E} = \mathcal{E}(r,k)$ and neglect any explicit t dependence we can compute δf^0 as,

$$df^0 = kT\frac{\partial f^0}{\partial \mathcal{E}} d(\frac{\mathcal{E}-\mu}{kT}) \tag{2.74}$$

$$= kT\frac{\partial f^0}{\partial \mathcal{E}}\{\frac{d\mathcal{E}}{kT} - \frac{d\mu}{kT} - \frac{\mathcal{E}-\mu}{kT^2}dT\} \tag{2.75}$$

$$= \underbrace{\frac{\partial f^0}{\partial \mathcal{E}}\nabla_k \mathcal{E}}_{\nabla_k f^0}.dk \underbrace{-\frac{\partial f^0}{\partial \mathcal{E}}\{-\nabla_r \mathcal{E} + \nabla_r \mu + \frac{\mathcal{E}-\mu}{T}\nabla_r T\}}_{\nabla_r f^0}.dr \tag{2.76}$$

which along with the BTE yields δf^0 as;

$$\delta f^0 = -\tau(v_g.\nabla_r f^0 + \frac{q(E+v\times B)}{\hbar}.\nabla_k f^0) \tag{2.77}$$

which then yields the current density as;

$$j = -\frac{2q}{(2\pi)^3}\int d^3k v_g(\tau(v_g.\nabla_r f^0 + \frac{q(E+v\times B)}{\hbar}\nabla_k f^0) \tag{2.78}$$

$$= \frac{2q}{(2\pi)^3}\int d^3k \frac{\partial f^0}{\partial \mathcal{E}}\tau v_g\{v_g.(-\nabla_r \mathcal{E} + \nabla_r \mu + \frac{\mathcal{E}-\mu}{T}\nabla_r T) - \frac{q(E+v\times B)}{\hbar}.\nabla_k \mathcal{E}\} \tag{2.79}$$

$$= \frac{2q}{(2\pi)^3}\int d^3k \frac{\partial f^0}{\partial \mathcal{E}}\tau v_g\{v_g.(-\nabla_r \mathcal{E} + \nabla_r \mu + \frac{\mathcal{E}-\mu}{T}\nabla_r T) - \frac{q(E+v\times B)}{\hbar}.\hbar v_g\} \tag{2.80}$$

$$= \frac{2q}{(2\pi)^3}\int d^3k \frac{\partial f^0}{\partial \mathcal{E}}\tau v_g\{v_g.(-\nabla_r \mathcal{E} + \nabla_r \mu + \frac{\mathcal{E}-\mu}{T}\nabla_r T) + qv_g.\nabla_r \phi - q(v\times B).v_g\} \tag{2.81}$$

$$\tag{2.82}$$

[4] this function is not a solution of the BTE in equilibrium

2.6 PROBLEMS

1. Starting with the expression for ac conductivity (Eqn 2.8), given by the Drude model derive expressions for the real and imaginary components of the dielectric constant ($\varepsilon(\omega) = \varepsilon'(\omega) + \varepsilon''(\omega)$) of a free electron system. Plot the components identifying the plasma frequency (ω_p) of the system. Discuss significance of the plasma frequency.

2. The free electron density in Copper is $n = 8.5 \times 10^{28} m^{-3}$ and near room temperature the relaxation time of most metals is of the order of 10-15$\times 10^{-14}$ sec. From this data estimate the fractional shift of the distribution on the scale of the Fermi wavevector (k_F) for an electric field of 10V/m,(i.e. calculate $\Delta k/k_F$).

3. Certain combinations of the Fermi function, occur very frequently in expressions that involve scattering or transitions. It is useful to be familiar with the combination $f^0(1 - f^0)$. Plot $f^0(1 - f^0)$ as a function of energy. How does the area under the curve of $f^0(1 - f^0)$ vary with temperature?

4. If $\mathbf{E}$=$\mathbf{Z}$–$\mathbf{A}\times$ $\mathbf{Z}$, then show that,
$\mathbf{Z} = \frac{\mathbf{E}+\mathbf{A}\times\mathbf{E}+(\mathbf{A}.\mathbf{E})\mathbf{A}}{1+A^2}$
Hint : Explore $\mathbf{A} \times \mathbf{E}$ and $\mathbf{A}.\mathbf{E}$.

5. Consider the "Corbino-disk" geometry shown in figure. Current flows between the inner (central) contact and the outer (circumferential) contact. Show by symmetry arguments that one of the components of the electric field (E_y in figure) must be zero. Can you roughly sketch the current flow paths from the center to the circumference? Ref: Journal of Applied Physics 31, 2176 (1960); https://doi.org/10.1063/1.1735520

6. Invert the matrix in the equation ??. Call this the conductivity matrix whose elements are σ_{ij}. What will be the value of ρ_{xx} if $\sigma_{xx} = 0$, when no magnetic field is present? How would your answer be modied when a finite strong magnetic field is present?

3

Junctions

3.1 Metal-semiconductor junctions

From chapter 1 we know the following about doped semiconductors;

1. How to calculate the charge density, if we know the location of E_f. In a n-type semiconductor, ignoring holes and acceptors, keeping the number of terms to a minimum, we have

$$n(x) - N_D^+(x) = N_C e^{\beta(E_f(x)-E_C(x))} - N_D \frac{1}{1+2e^{-\beta(E_D(x)-E_f(x))}} \tag{3.1}$$

$$\rho(x) = -|e|\{n(x) - N_D^+(x)\} \tag{3.2}$$

2. The charge density is related to the electrostatic potential (V) as

$$\nabla^2 V(x) = -\frac{\rho(x)}{\epsilon_r \epsilon_0} \tag{3.3}$$

3. The scalar potential is essentially the bottom of the conduction band.
4. In equilibrium E_f is constant, recall that current flow requires a gradient in the electrochemical potential or Fermi level.

3.1.1 A Metal-Semiconductor Junction at equilibrium

Now let's see how we can put this in practice – a (somewhat idealised) metal in contact with a semiconductor, forming a M-S junction. The metal work function (φ_m) is the energy an electron sitting at the E_f of the metal needs to escape from inside the metal to outside (vacuum level) and is typically about 4 - 5 eV. It depends on which crystal face is considered and how clean the surface is. Such effects are ignored here, hence the discussion is a bit idealised here. φ_s is the work function of the semiconductor in question and the difference $\varphi_C = |\varphi_m| - |\varphi_s|$ is the contact potential which we assume to be positive.

The two objects are brought in contact, so that they can exchange electrons and establishing equilibrium requires electron transfer from the semiconductor to the metal, which is energetically favourable. However, in a semiconductor there are no electrons at E_f, but they transfer from the bottom of the CB to the metal establishing thermal equilibrium across the entire system with E_F same on both sides. As the electrons transfer charge separation gives rise to an electrostatic potential - and it is reasonable to expect that the bands would start bending in a way that would result in a barrier, preventing further flow and establish equilibrium. At this point the metal and the semiconductor's E_f must be identical across the M-S junction. Applying these set of conditions produces the band diagram shown in Fig. 3.1. The energy difference between the vacuum level and the bottom of the CB, sufficiently deep inside the bulk is defined as the electron affinity (χ_s) of the semiconductor. Thus once the M-S junction is formed and the CB edge bends the work done in transferring an electron from the CB edge into the metal would be $\phi_m - \chi_s$, which is the work done in overcoming a barrier of height $\phi_B = \phi_m - \chi_s$ and defined as the barrier height of the junction. Since deep in the bulk of the semiconductor (where the surface should have no effect) the bottom of the CB is no longer bent

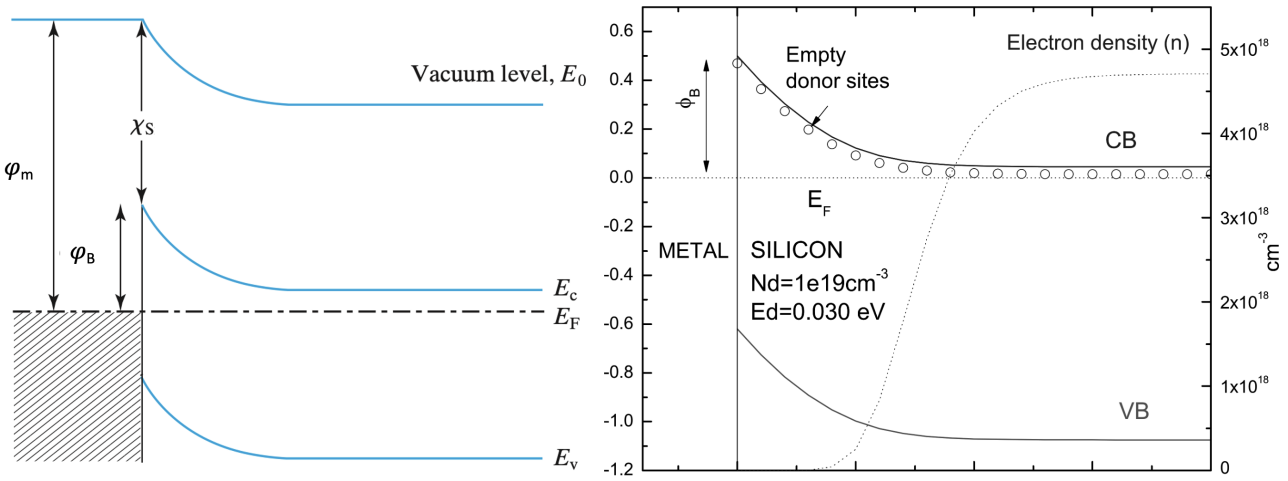


Figure 3.1: (left) Band bending near the surface of a n-type semiconductor to metal contact ($|\varphi_s| < |\varphi_m|$) (right) shows a calculated band bending for n-Si Au contact. Dotted line shows variation in local electron density. Greg Snider (Notre Dam University)

and the E_f must continue to be separated by $\phi_s - \chi$, this dictates that the total bend in the CB of the semiconductor is the built-in potential ϕ_{bi}.

$$\phi_{bi} = \phi_B - (E_C - E_f) = \phi_B - kT \ln \frac{N_C}{N_D} \tag{3.4}$$

As the figure (right) shows some charge has moved from the semiconductor to the metal. This charge came from the dopants sitting considerably above E_f, closer to the junction. Equation 1.24 indicates that the probability of dopant ionisation $\propto (1 + 2e^{-\beta(E_D - E_f)})^{-1}$, i.e. as dopant energy E_D is pushed above E_f then it must be ionised, because the electron cannot reside at a site sitting much above E_f. The bands in the metal didn't have to bend a lot to accommodate this extra charge, because the density of states of a metal near E_f is very large.

To (numerically) solve eqns 3.1, 3.2 and 3.3 we can proceed as follows. 1. Since E_f is constant i.e. independent of x, other energies can be measured relative to E_f, by setting $E_f = 0$.
2. The gradient of the scalar potential is the same as the gradient of the CB edge, $E_C(x)$
3. We make a guess for $E_C(x)$ and use this to calculate the expected charge density by using eqns 3.1 and 3.2.
4. This calculated charge density should provides a new initial value for the potential via Poisson's equation (eqn 3.3).
5. We use this potential and go back to step 3.
6. The iterative process can continue till the change in two successive iterations becomes very small (our convergence criteria)
7. Note: These differential equations are defined and solved only in the semiconductor and needs proper boundary conditions. In the calculation of Fig. 3.1, we set the slope $dE_C/dx = 0$ deep inside the material i.e. as $x \to \infty$ and $E_C = \varphi_B$ at $x = 0$ (the junction). Choosing the correct boundary condition depends on the physical situation being addressed.

The above procedure can also be executed assuming an initial variation of the expected charge density inside the semiconductor.

Either way having solved the above equation iteratively we can calculate some important parameters for the junction.
1. The Depletion width w_d.

$$w_d = \sqrt{\frac{2\epsilon_0\epsilon_r(\phi_B - V)}{qN_d}} \tag{3.5}$$

2. The junction Capacitance C_j (per unit area).

$$C_J = |\frac{dQ}{dV}| = \sqrt{\frac{2q\epsilon_0\epsilon_r N_d}{2(\phi_B - V)}} = \frac{\epsilon_0\epsilon_r}{w_d} \tag{3.6}$$

3. The Electric field at the junction E_j.

$$E_j = \frac{qN_d w_d}{\epsilon_0 \epsilon_r} \tag{3.7}$$

4. The φ_B is dependent on the metal semiconductor pair and the table lists the values for n and p type

Metal	Mg	Ti	Cr	W	Mo	Pd	Au	Pt
φ_m(eV)	3.7	4.3	4.5	4.6	4.6	5.1	5.1	5.7
φ_{Bn}(eV)	0.4	0.5	0.61	0.67	0.68	0.77	0.8	0.9
φ_{Bp}(eV)	-	0.61	0.50	-	0.42	-	0.3	-

Table 3.1: Schottky barrier heights for electrons on n-type Si (φ_{Bn}) and for holes on p-type Si (φ_{Bp})

Si with different metals. The sum of φ_{Bn} and φ_{Bp} is approximately equal to E_g of the semiconductor (E_g^{Si} $\simeq$1.12 eV). Note that for a given semiconductor φ_{Bn} increases with increasing φ_m, which is commensurate with the relation $\varphi_B = \varphi_m - \chi_s$.

3.2 $p-n$ Junctions

The above process can be applied to $p-n$ junctions as well to calculate band bending, the junction field, depletion width and junction capacitance, as shown in Fig. 3.2. We take a closer look at the

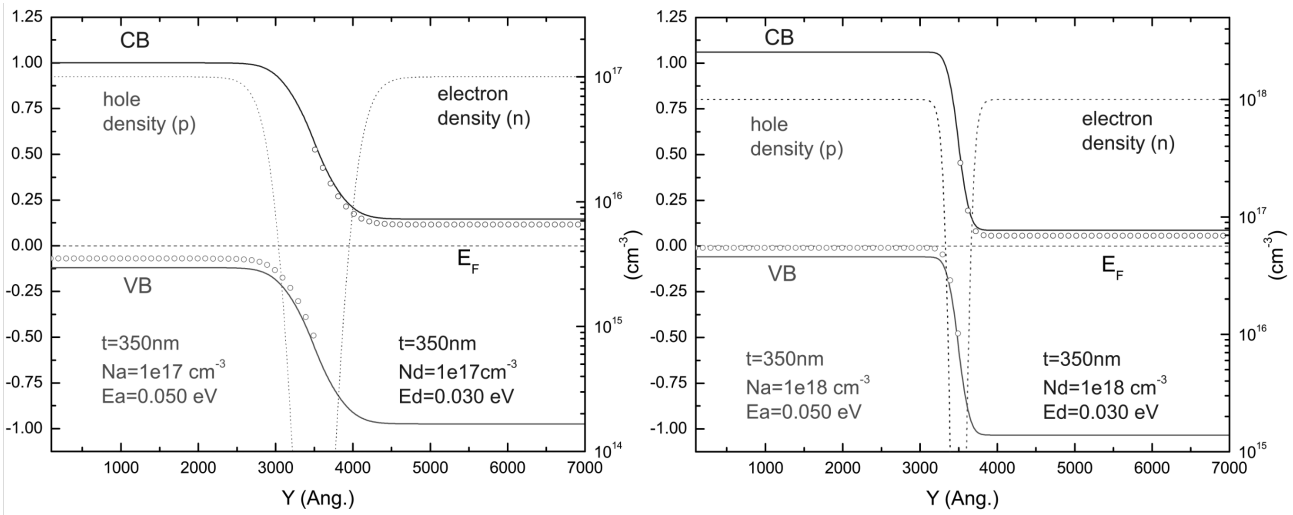


Figure 3.2: Band bending at pn junction. The doping density N_D and N_A are lower in the left than right figure. The junction becomes sharper and the E_F moves closer to the dopant levels with increased doping. Dotted line shows variation in local electron density. Greg Snider (Notre Dame University)

band diagram of a PN junction as shown in Fig. 3.2. The figure shows that E_c and E_v are not flat and they are vertically displaced, across the junction indicating the presence of a voltage differential. The energy (voltage) differential between the band edges across the junction is called the built-in potential, ϕ_{bi}. A built-in potential is present at the interface of ANY two dissimilar materials. It is the ϕ_{bi} that limits the initial diffusion current across a dissimilar junction. The energy difference of the CB edge

from the equilibrium E_F in the n and p side may be calculated as follows;

$$n-\text{side}: n(\simeq N_D) = N_C e^{-(E_C^n - E_F)/kT} \implies E_C^n - E_F = kT \ln \frac{N_C}{N_D} \tag{3.8}$$

$$p-\text{side}: n(\simeq \frac{n_i^2}{N_A}) = N_C e^{-(E_C^p - E_F)/kT} \implies E_C^p - E_F = kT \ln \frac{N_C N_A}{n_i^2} \tag{3.9}$$

$$\phi_{bi} = E_C^p - E_C^n = kT(\ln \frac{N_C N_A}{n_i^2} - \ln \frac{N_C}{N_D}) \tag{3.10}$$

$$\phi_{bi} = kT \ln \frac{N_D N_A}{n_i^2} \tag{3.11}$$

ϕ_{bi} increases with N_A and N_D and is typically ~ 0.9 V for a silicon pn junction. Comparison of the two cases shown in Fig. 3.2 shows that with increasing doping (i) depletion width decreases, (ii) band bending increases and (iii) the Fermi Energy moves closer to the dopant levels.

3.3 How realistic are these calculations?

We remarked at the beginning of this section that there are some idealisations. The work function of a metal in reality depends on which crystal face we are using, how clean it is etc. This means that if we deposit a thin film of a metal (say gold on silicon) on a semiconductor, we can't really take the values for a crystal of gold and clean silicon and predict what the barrier will be. Also the density of states near the surface of a semiconductor is modied by the presence of surface states-which ultimately mean that the Schottky barrier needs to determined experimentally. However the band diagram of the barrier that we drew and the principles for solving the band-bending are sufficiently generic.

3.4 "Schottky" – "Ohmic"

In the previous section we considered the work function of the metal to be larger $|\varphi_s| < |\varphi_m|$ and the semiconductor to be n-doped. As a consequence some electrons flowed from the semiconductor to the metal. What if $|\varphi_s| > |\varphi_m|$? To establish equilibrium electrons now flow into the semiconductor and the E_D and the CB edge now bends downwards towards E_f at the junction. Probability of ionisation of dopants decrease as $E_D \to E_f$ and remain unionised if $E_D \leq E_f$. The CB accommodates the incoming electrons from the metal and if the CB edge dips below E_F it creates an accumulation of free electrons at the interface as shown in Fig. 3.3 [1]. The junction does not have any Schottky barrier at all.

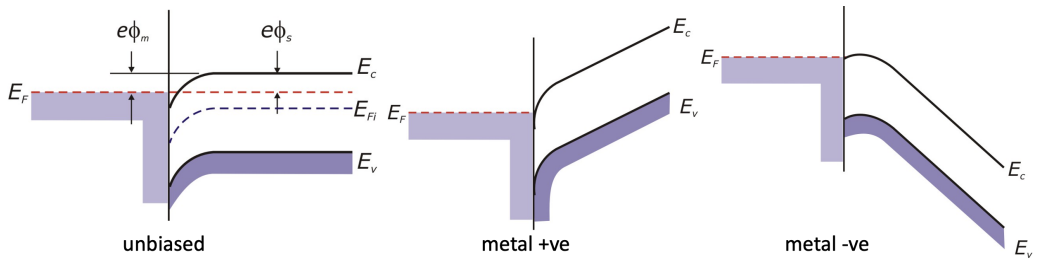


Figure 3.3: Band bending for metal-semiconductor Ohmic contacts when $|\varphi_s| > |\varphi_m|$

Real ohmic contacts

Ohmic contacts are low resistance junctions with linear IV characteristics (non-rectifying) for sourcing/draining current to semiconductors. In reality ohmic contacts on a semiconductor are made by depositing an alloy that often contains one noble metal (Gold) and another element that can act as a dopant. For example an alloy of Gold-Germanium is commonly used to make ohmic contacts to

[1] https://warwick.ac.uk/fac/sci/physics/current/postgraduate/regs/mpagswarwick/ex5/devices/hetrojunction/

n-type Gallium Arsenide. After depositing the alloy the sample is generally annealed (heated to a high temperature) very rapidly so that the Germanium diffuses into the surface, heavily dopes the region around it allowing the Gold to make a contact with no barrier. Enhancing local doping also decreases the depletion width, which improves tunnelling probability through a barrier thus decreasing junction resistance. The microscopic mechanisms of Ohmic contact formation are non trivial and significant research is devoted to developing high quality contacts. Gold-Beryllium alloy can be used to contact p-type Gallium Arsenide. Gold-Antimony alloy can be used to contact n-type Silicon.

3.5 Situations with varying E_f: what more is needed?

We noted earlier that the current is related to the gradient of the electrochemical potential (in a 1-dimensional case) as

$$j = -n(x)\mu \frac{\mathrm{d}}{\mathrm{d}x} E_f(x) \tag{3.12}$$

So we now have three variables to deal with-other than the charge density and the profile of the CB edge. We also need to calculate the profile of $E_f(x)$. We expect mobility (μ) to be a function of n. In general the product $n\mu$ would increase with increasing carrier density, it doesn't necessarily imply that μ will be larger at higher densities. However at least in a 1-dimensional situation it is easy to see that wherever $n\mu$ is large, $\mathrm{d}\mu/\mathrm{d}x$ must be small. This reminds us of what to expect if we apply a voltage across a string of resistances (in series). The largest voltage drop must occur across the largest resistance, because the current through each of them is constant. When the current flow is very small we can approximate the situation by saying that all the drop in E_f must be across the most resistive region (like a barrier) if we can identify one. This is however an approximation to get around the fact that the variation of μ with n is in general a non-trivial and a system dependent problem. In Fig. 3.1 we plotted the band diagram of a metal-semiconductor junction with no voltage applied (E_f = constant). No current flows at equilibrium and the cross-over rates from both sides balance each other.

Forward Bias

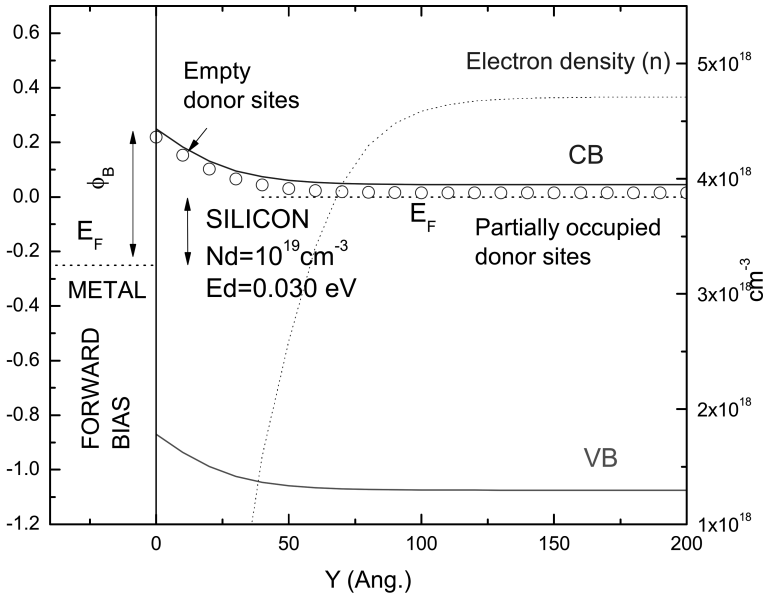


Figure 3.4: Approximate band bending near a forward biased metal-semiconductor junction. Dotted line shows variation in local electron density. Greg Snider (Notre Dame University)

Now if the electron energies of the semiconductor is raised by connecting a negative potential (V) to the semiconductor the relative band position across the junction are modified which is shown in Fig. 3.4. The metal being at a higher potential net current flow from metal to semiconductor and electrons in the reverse under FB. The E_F are not equilibrium values but denote the difference in

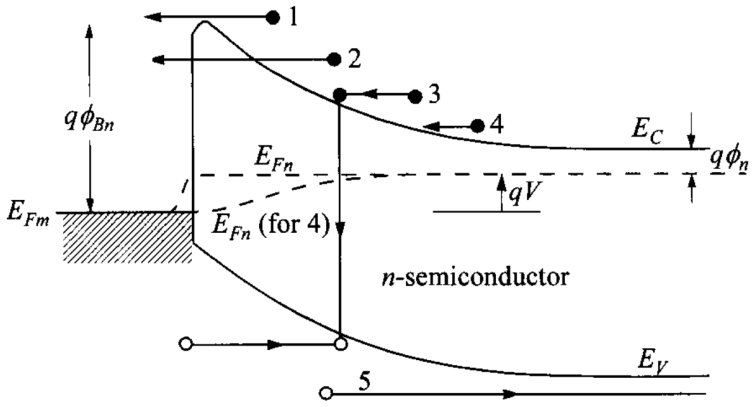


Figure 3.5: Basic transport processes under forward bias (1)Thermionic emission (2)Tunneling (3) Recombination. (4) Diffusion of electrons (5) Diffusion of holes (Sze)

chemical potential of free electrons in either system away from the junction. Note the E_F on either side differ by the applied potential V.

- Fig. 3.5 shows the five basic transport processes under forward bias. (1)Thermionic emission (2)Tunneling (3) Recombination. (4) Diffusion of electrons (5) Diffusion of holes
- Electrons that try to cross over from the metal to the semiconductor encounter a barrier ϕ_B and is has low probability of crossover. A minority hole current from metal to semiconductor via recombination and diffusion contributes to the overall forward current.
- Electrons crossing from the semiconductor to the metal now see a lower barrier than the unbiased case and constitute a majority of the forward current via the thermionic emission and tunnel current.
- Tunnelling probability through a barrier increases exponentially as the height of the barrier is lowered. $J_{s\to m} \propto e^{-(\varphi_B - V)/kT}$
- Taking the difference of the left going and the right going currents to get the total current which is the well known diode equation: $J = J_0(e^{eV/kT} - 1)$, where J_0, the reverse saturation current density, determined by the height of the Schottky barrier given by, $I_0 = A_0 T^2 e^{-\phi_B/kT}$, where $A_0 = \frac{4\pi m k^2 q_e}{h^3} = 1.20173 \times 10^6\,\mathrm{A\,m^{-2}\,K^{-2}}$ is the Richardson constant.
- Under FB: band bending and depletion width decreases, junction capacitance increases and the junction field decreases.

Reverse Bias

Under reverse bias (RB) electron energies on the semiconductor side are lowered by connecting it to a positive potential (or applying a negative potential to the metal), as shown in Fig. 3.6. Again the drop in the electrochemical potential must happen predominantly over the depletion region and now current flows from the semiconductor to metal and net electrons movement happens in reverse.

- Electrons which try to cross over from the metal to the semiconductor still see almost the same barrier. The current that can pass through is $J_{m\to s} \simeq J_0 = AT^2 e^{-\varphi_B/kT}$
- But electrons that try to cross from the semiconductor to the metal now see a higher barrier ($\sim (\phi_B + V)$).
- Tunnelling probability through a barrier drops exponentially with the height of the barrier. $J_{s\to m} = AT^2 e^{-(\phi_B + V)/kT}$, where V is the voltage bias on the semiconductor w.r.t. the metal. In this case $V > 0$. Remember that positive voltage bias lowers electron energies.

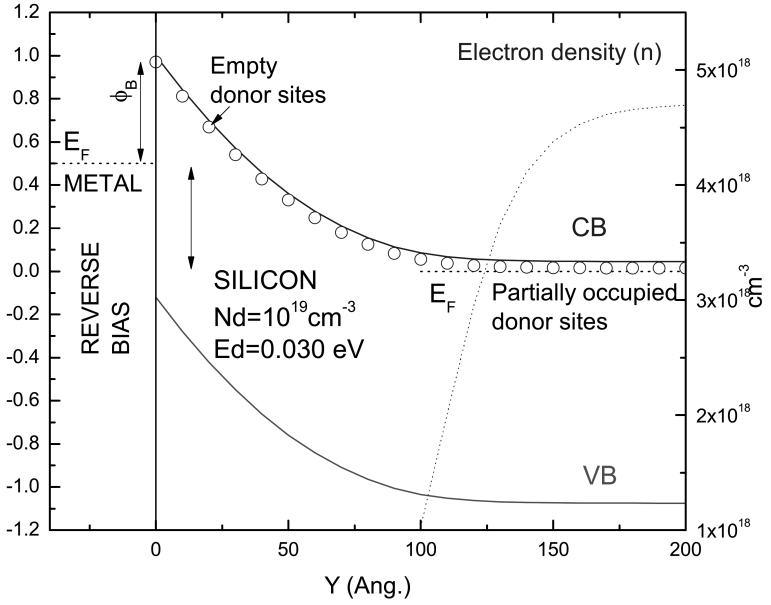


Figure 3.6: Approximate band bending near a reverse biased metal-semiconductor junction. The calculation has been done using a program written by Greg Snider (Notre Dame University)

- Thus under RB tunnelling + diffusion are two primary contributors to the net current flows which is $\simeq J_0$.
- Further, under RB: depletion width widens, junction capacitance decreases and the junction field ($\vec{E}_J$) increases. The increase is useful for separating $e-h$ pairs thus increasing their lifetime.

Refer to the Chapters 2 and 3 of the book The Physics of Semiconductor Devices by S M Sze for a more detailed discussion on the various transport processes across pn and metal-semiconductor junctions.

- **Schottky Diode**: Transport across a diode is mathematically modelled in terms of the current density as;

$$J(V,\phi_B,T) = J_0(e^{eV/\eta kT} - 1) \tag{3.13}$$

where η is the ideality factor, J_0, the reverse saturation current density, determined by the height of the Schottky barrier given by, $J_0 = A_0T^2e^{-\phi_B/kT}$, where $A_0 = \frac{4\pi mk^2q_e}{h^3} = 1.20173 \times 10^6\,\mathrm{A\,m^{-2}\,K^{-2}}$ is the Richardson constant. Both J_0 and η are functions of T and dopant density as shown in the Fig. 3.7. For $eV >> k_BT$ the current density may be written as $J = J_0e^{eV/\eta kT}$

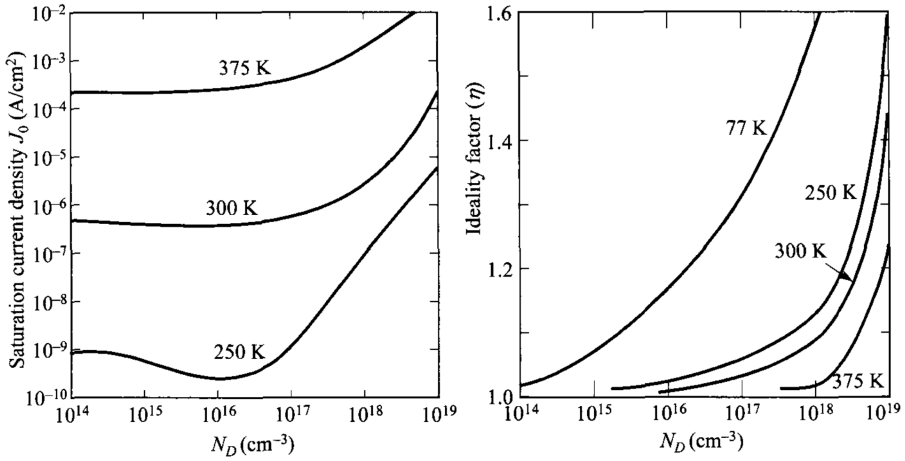


Figure 3.7: J_0 versus N_D and η versus N_D at different temperatures for an Au-Si junction. (Sze)

which is used to linearise the equation as $\ln J = \ln J_0 + eV/kT$. The intercept of a linear fit to

the $\ln I$ vs. V plot in the high voltage regime gives $\ln I_0$ is used to determine the barrier height at a particular T.

- *pn* **Junction Diode**: Transport across a *pn* junction diode is mathematically modelled in terms of the current density J, given as;

$$J(T, E_g, V) = J_0(e^{eV/kT} - 1) \tag{3.14}$$

where J_0 is the reverse saturation current density given as $J_0 = qn_i^2(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A})$, where $n_i = \sqrt{N_C N_V}\ e^{-E_g/2kT}$ is the intrinsic carrier density, L_n and L_p are the e and h diffusion lengths, D_n and D_p are the diffusion coefficients. The expression for J_0 may further be simplified to yield $J_0 \propto T^{(3+\gamma)}e^{-E_g/kT}$ such that the IV characteristics is given by the Shockley equation;

$$I(V, E_g, T) = \underbrace{AT^{(3+\gamma)}e^{-E_g/kT}}_{I_0}(e^{eV/kT} - 1) \tag{3.15}$$

This idealised Shockley equation primarily gives qualitative agreement for most semiconductors at low current densities. Note that the dominant T dependence of the diode equation is ascribed to the exponential term rather than the $T^{(3+\gamma)}$ term. Thus the reverse current is determined by $\exp(-E_g/kT)$ and the forward current dominated by $\exp(-(E_g - V)/kT)$.

3.6 Across the Barrier

The calculation of how many electrons can make it through the barrier at a metal semiconductor interface is very similar to the way we calculate how many electrons a hot filament can emit. The electrons in a metal can be effectively pictured as being inside a box with the outside world (vacuum level) at a height ϕ_m above the Fermi energy of the metal and U above the bottom of the conduction band of the metal. All energies are measured from the bottom of the conduction band, which is usual for the free electrons. E_F is Fermi level and hence $\phi_m = U - E_F$. This is the work function barrier which keeps the free electrons from jumping out of the "box".

- The electrons are in random motion and some of them will hit the boundary. Do they have enough energy to come out? The potential barrier has to be finite for tunnelling to happen. To be able to come out of the metal, the x component of the electron's velocity should satisfy $\frac{mv_x^2}{2} > U$
- How many such electrons will hit an area a (normal to the boundary) in time $\triangle t$? This is commonly calculated in kinetic theory of gasses. The total charge coming out is;

$$Q = e\frac{2m^3}{h^3}\int_{\sqrt{2U/m}}^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z (av_x \triangle t) f(E) \tag{3.16}$$

The factor $\frac{2m^3}{h^3}$ comes from the density of states ($k = mv/\hbar$). Here $f(E)$ is the Fermi distribution, but for high temperatures we can approximate it with the Boltzmann distribution and hence $f(E) \approx e^{-\beta(2E_{kin} - E_F)}$, where $E_{kin} = 1/2(mv_x^2 + mv_y^2 + mv_z^2)$. The integral can be computed analytically to give,

$$J = \frac{Q}{a\triangle t} = \frac{4\pi em}{h^3}(kT)^2 e^{-\phi_m/kT} \tag{3.17}$$

The same reasoning is valid for a metal semiconductor interface though the work function will be different, much less than the 4.5 eV or so, typical of a tungsten filament.

A similar calculation can be done for tunnelling electrons instead of thermionic emission. Fig. 3.8 shows the various currents under FB and RB condition and variation with dopant density and temperature.

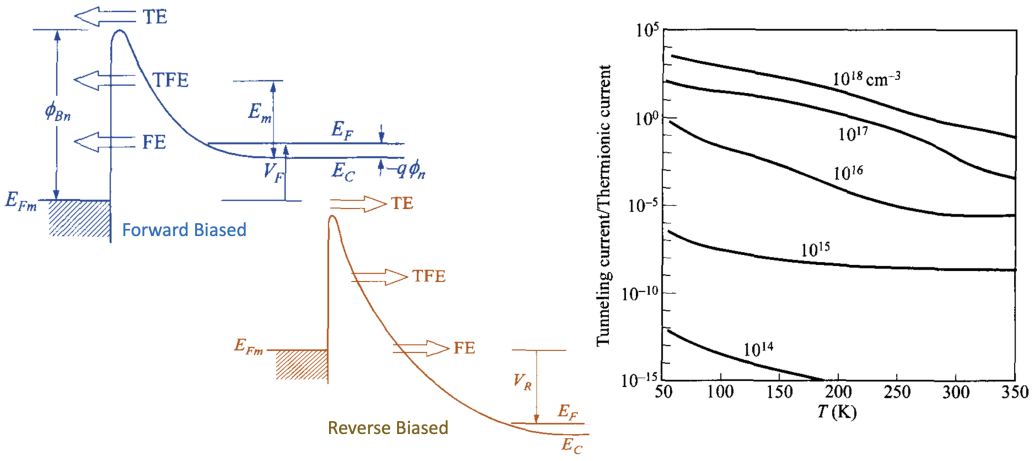


Figure 3.8: (left) energy-band diagrams showing various currents (TE: thermionic emission, TFE: thermionic field-emission, FE: field-emission) in a Schottky diode under forward and reverse bias. (right) Ratio of tunnelling to thermionic current across a Au-Si diode. Tunnelling dominates at higher doping and lower T (Sze)

3.7 Applications

3.7.1 Schottky Field Emission

In electron guns, the thermionic electron emitter is biased negative relative to its surroundings, which creates an electric field of magnitude **E** at the emitter surface. Without the field, the surface barrier seen by an escaping Fermi-level electron has height ϕ_m. The **E** lowers the surface barrier by an amount $(\Delta\phi)_E = \sqrt{e^3E/4\pi\epsilon_0}$, and increases the emission current. This is known as the Schottky effect (named for Walter H. Schottky) or field enhanced thermionic emission. It can be modelled by a simple modification of the Richardson equation

$$J = \frac{Q}{a\triangle t} = \frac{4\pi em}{h^3}(kT)^2 e^{-(\phi_m - \Delta\phi_E)/kT} \tag{3.18}$$

Electron emission that takes place in the field-and-temperature-regime where this modified equation applies is often called Schottky emission. This equation is relatively accurate for electric field strengths lower than about $10^8 V/m$. For electric field strengths higher than $10^8 V/m$, so-called Fowler-Nordheim (FN) tunnelling begins to contribute significant emission current. In this regime, the combined effects of field-enhanced thermionic and field emission can be modelled by the Murphy-Good equation for thermo-field (T-F) emission. At even higher fields, FN tunnelling becomes the dominant electron emission mechanism, and the emitter operates in the so-called cold field electron emission regime.

3.7.2 Tunnel Diode

Tunnel diodes, also known as Esaki diodes, are a type of pn junction diode that exhibit **negative differential resistance** (NDR) in their voltage-current characteristic curve, Fig. 3.9(left). The tunnel diode was invented by Leo Esaki in 1957 for which he received the Nobel Prize in Physics in 1973. Tunnel diodes are unique devices that show a decrease in current with increase in voltage in a limited voltage regime. Note that $dI/dV < 0$ but $I/V > 0$ in the specified voltage regime. Tunnel diodes are extensively used in high-frequency circuits, oscillators, and microwave applications, benefiting from their fast response times and frequency stability. In tunnel diodes, both the p and n sides are degenerately doped such that their respective E_F lie just within the band edge of the valence and conduction bands. The equilibrium band structure of the diode, creating a narrow energy barrier within the bandgap is shown in Fig. 3.9 (right).

- Above the E_F there are no filled states (electrons) on either side of the junction, and below the Fermi level there are no empty states (holes) available on either side of the junction. Hence, the net tunneling current at zero applied voltage is zero.

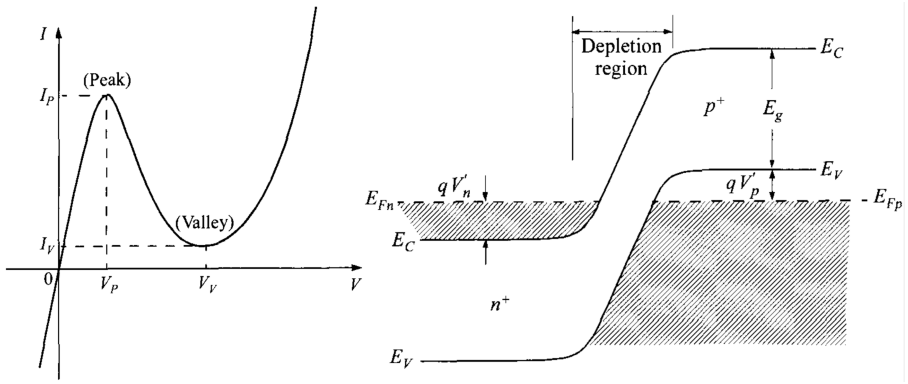


Figure 3.9: (left) Static current-voltage characteristics of a tunnel diode I_P and V_P, are the peak current and peak voltage. I_V and V_V are the valley current and valley voltage. (right) Energy-band diagram of tunnel diode in thermal equilibrium. (Sze)

- Under FB e may tunnel from the filled states of the CB of n-side to the unfilled states of the VB of the p-side allowing a large forward current with increasing bias (Fig. 3.10a-b) upto the point (I_P, V_P).
- For $V \leq V_P$ e tunnel across the band gap of the material and not a physical barrier.
- At I_P, V_P the CB edge of n-side is energetically aligned with the VB edge of the p side.

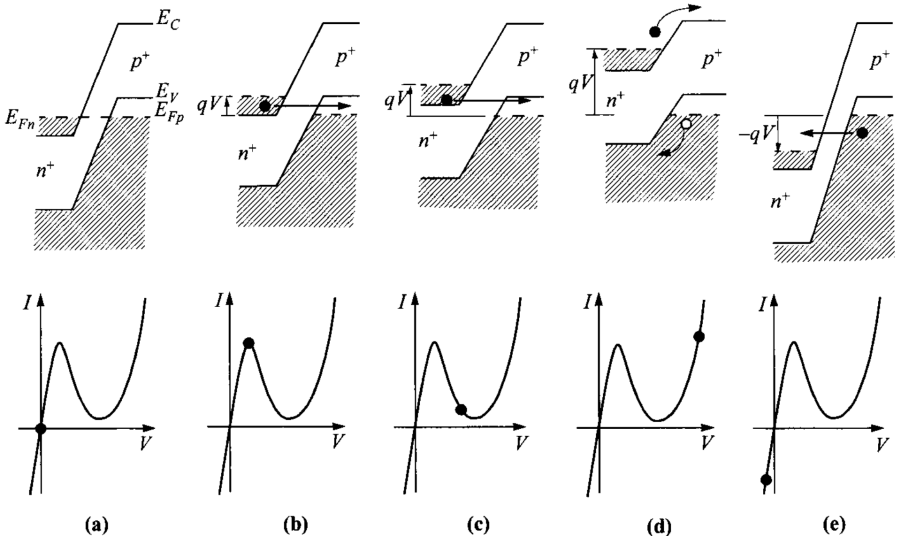


Figure 3.10: Energy-band diagrams of tunnel diode at (a) thermal equilibrium, zero bias; (b) forward bias V such that peak current is obtained; (c) forward bias approaching valley cur- rent; (d) forward bias with diffusion current and no tunnelling current; and (e)reverse bias with increasing tunnelling current. (Sze)

- Negative Differential Resistance (NDR): Increasing voltage above V_P no longer increases current but decreases it. Since the e at the CB edge of the n-side cannot tunnel to the p-side since at that energy there are no vacant states within the p-side. Thus the current begins to decrease with increasing voltage, resulting in a region of negative differential resistance ($dI/dV < 0$) in the voltage-current characteristic curve (Fig. 3.10c) for $V_P < V < V_V$ in the FB regime.
- With further increase in$V > V_V$ intraband diffusion current dominate transport and current monotonically increases with voltage (Fig. 3.10d).
- Fig. 3.10e shows electron tunneling from the VB to the CB under reverse bias. Here, the tunneling current may increase indefinitely with bias.

Fig. 3.11 shows the FB IV characteristics of tunnel diodes made of different semiconductors. Note that the NDR region is typically restricted to around 500 mV. Interestingly, the IVs originate from

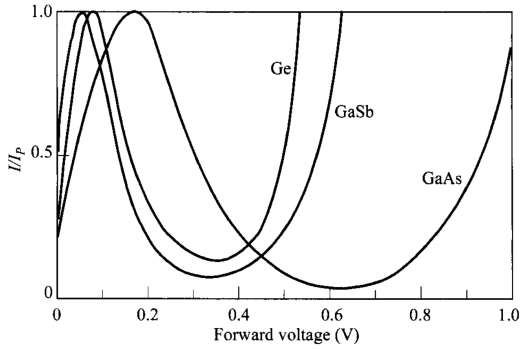


Figure 3.11: Typical IV characteristics of Ge, GaSb, and GaAs tunnel diodes at 300 K(Sze)

the quantum tunnelling process which dominates transport for $V < V_V$ as opposed to drift-diffusion in generic diodes. Since tunnelling is an extremely fast process tunnel diodes are used as active elements in switching circuits and oscillators. Their low capacitance allows them to function at microwave frequencies, far above the range of ordinary diodes and transistors. However, low output power, tunnel diodes, limited to a few hundred milliwatts limit their applications. The resonant-tunneling diode (RTD) has achieved some of the highest frequencies of any solid-state oscillator.

3.7.3 Diode Thermometry

Diode thermometry is based on T dependence of the FB V drop in a pn diode at a constant current, $I_{dc}(\sim 10\mu A)$. Fig. 3.12(left) shows IV characteristics at various T for a Si diode[2]. At higher T it takes a smaller V for the diode to conduct at a given I, due to a larger n_i in eqn. 3.15. Since the

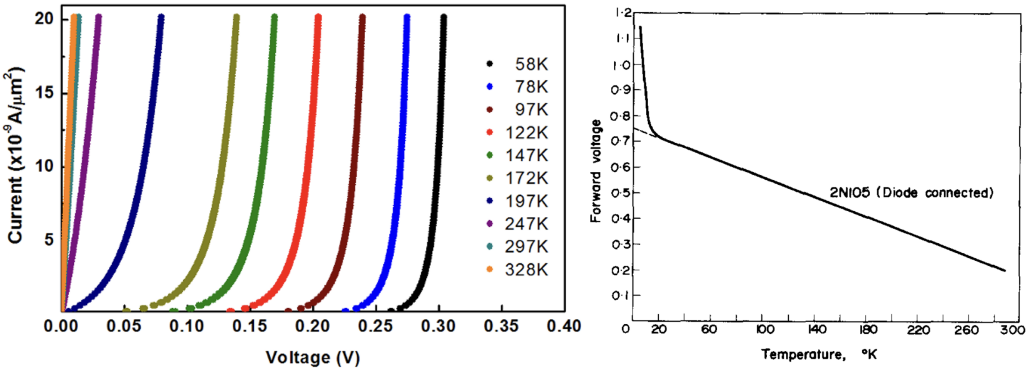


Figure 3.12: (left)IV characteristics of a Si diode at various T (right) 2NI05 Ge diode V vs T at constant current.

resulting diode $V >> kT$ between 0.1 - 6 V, eqn. 3.15 may be re-written as $(V - \phi_B) \simeq kT \ln(I_{dc}/A)$. A constant $I_{dc} \implies V \propto T$. Fig. 3.12(right) shows the VT for a Ge diode[3]. Typically $dV/dT \sim$ -2 mV/K from 300 - 20 K. Fig. 3.13 shows the characteristics of a Si diode DT-670 from LakeShore Cryotronics[4]. See this page for simulating the V vs. T graph for diodes; `http://lampx.tugraz.at/~hadley/psd/L6/VT_I.php`.

[2] An alternative methodology in Schottky diode physics. J. of Appl. Phys. 117, 244501, 2015; doi: 10.1063/1.4922974

[3] Solid-State Electronics, 15, 473, 1972

[4] https://www.lakeshore.com/products/categories/overview/temperature-products/cryogenic-temperature-sensors/dt-670-silicon-diodes

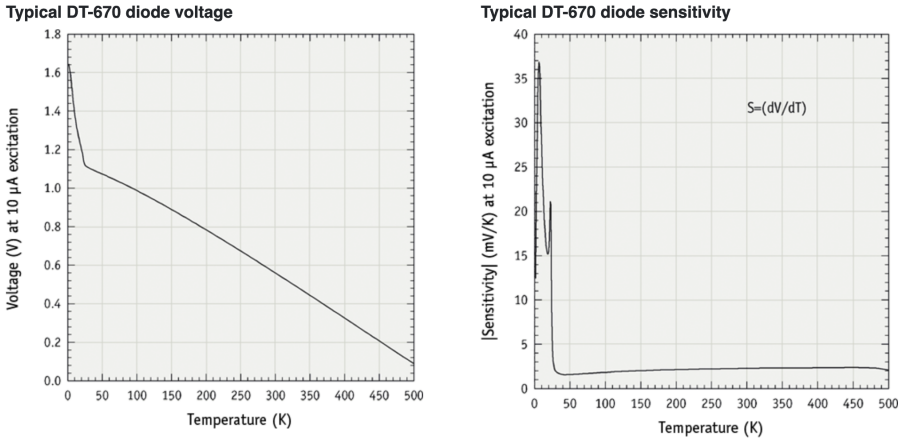


Figure 3.13: Si diode DT-670 (left)VT characteristics, (right) Sensitivity dV/dT vs. T.

Problems

1. Band bending at the p-n junction. The total drop in the prole of the bands shown in Fig. 2.2 can be calculated in two different ways. First let us see the method given in most text books. The flow of charge through the junction can be thought to have a drift (forced by the electric field) and a diffusion (forced by density gradient) component-at equilibrium, when the electrochemical potential in constant, these two components must add upto zero. So we get for zero electron current

$$J_{drift} + J_{diffusion} = 0 \tag{3.19}$$

$$-ne\mu\frac{dV}{dx} - De\frac{dn}{dx} = 0 \tag{3.20}$$

We have used the standard relation between current, diffusion constant and density gradient. (A full justification of this set of equations require the Boltzmann transport formulation.) Then solve the differential equation using $D/\mu = kT/e$ and the assumption that all dopants are ionised. So that the electron density on the n-side is $n = N_D$ and on the p-side it is $n = n_i^2/N_A$. You should get the result for the total change in electrostatic potential as one moves from one side of the junction to the other. The electron bands are higher on the p-side.

$$\triangle V = \frac{kT}{e}\ln\frac{N_A N_D}{n_i^2} \tag{3.21}$$

Now think of the same in another way. Let us not mention diffusion constant at all, but use the fact that the electrochemical potential (E_f) is constant. Here the free energy of the electrons can be written, including the electrostatic potential as

$$F = -k_B T \ln\frac{Z^n}{n!} + neV \tag{3.22}$$

where the electron density $n(x)$ is a function of position. And

$$Z = 2\Omega(\frac{2\pi m k_B T}{h^2})^{3/2} \tag{3.23}$$

is the partition function of a single free electron moving in the conduction band. Ω is the volume which should drop out of the calculation. Since we assume full ionisation we can neglect the entropy contribution coming from possible number of ways to distribute the bound electrons among the dopants. Differentiating this w.r.t. n to get the electrochemical potential, first write

$$E_f(x) = \frac{dF}{dn(x)} \tag{3.24}$$

And then show that setting $E_f(x) =$ constant leads to exactly the same condition as before. Convince yourself that in both cases the approximations that we made are actually identical. They are both consequences of Boltzmann statistics applied to the free electron gas in the conduction band.

2. Try to draw the band diagram of the metal-semiconductor contact when $|\varphi_s| > |\varphi_m|$ and the semiconductor is p-type. Where will the semiconductor accommodate the electrons flowing in? Why doesn't the depletion zone extend to the metal as well?

3. Consider a gold-GaAs Schottky diode with a capacitance of 1 pF at -1 V. What is the doping density of the GaAs? Also calculate the depletion layer width at zero bias and the field at the surface of the semiconductor at -10 V bias voltage. The area of the diode is 10-5 cm^2.

4. Using the work functions listed in table to predict which metal-semiconductor junctions are expected to be ohmic contacts. Use the ideal interface model.

5. Design a Pt-Si diode with a capacitance of 1 pF and a maximum electric field less than 10^4 V/cm at -10 V bias. Provide a possible doping density and area. Make sure the diode has an area between 10^{-5} and 10^{-7} cm^2. Is it possible to satisfy all requirements if the doping density equals 10^{17} cm^{-3}?

6. A Pt-Si diode (area = 10^{-4} cm-2, Nd = 10^{17} cm-3) is part of an LC tuning circuit containing a 100 nH inductance. The applied voltage must be less than 5 V. What is the tuning range of the circuit? The resonant frequency $=\frac{1}{2\pi\sqrt{LC}}$, where L is the inductance and C is the diode capacitance.

7. Consider two Schottky diodes with built-in potential $V_{bi} = 0.6$ V. The diodes are connected in series and reversed biased. The diodes are identical except that the area of one is four times larger than that of the other one. Calculate the voltage at the middle node, V_{out}, as a function of the applied voltage, V_{in}. Assume there is no dc current going through either diode so that the charge at the middle node is independent of the applied voltage.

8. A metal-semiconductor junction consists of Pt and GaAs with $N_d = 10^{17}$ cm-3. The applied voltage equals -3 V. Calculate the electric field in the semiconductor at the metal-semiconductor interface. Use $V_{bi} = 0.8$ V.

9. An Al-Si Schottky diode has a breakdown voltage of 5 V. The silicon is p-type with a doping $N_a = 10V^{18}$ cm-3. Calculate the breakdown field and the depletion layer width. (FM = 4.0 V)

10. A metal-semiconductor junction, biased at an unknown voltage, has a doping density of $10V^{17}$ cm-3 and a capacitance of 1 pF. The semiconductor is p-type Ge, $V_{bi} = 0.5$ V and the diode area is $10V^{-4}$ cm^2. Calculate the depletion layer width and the applied voltage.

11. A metal-semiconductor junction, biased at an unknown voltage, has a maximum electric field of 10^5 V/cm and a capacitance of 1 pF. The semiconductor is n-type GaAs, the built-in potential of the junction is 0.7 V and the diode area is $10V^{-4}$ cm^2. Calculate the doping density and the applied voltage.

12. SiC contains three cigar-shaped conduction band minima. The constant energy surfaces of these are ellipsoids which are given by:

    $E = E_{min} + \hbar^2[\frac{k_x - k_{xo}}{m_t} + \frac{k_y - k_{yo}}{m_t} + \frac{k_z}{m_i}]$

    and $(k_{xo}, k_{yo}, 0) = k_0(1,0,0), k_0(\frac{-1}{2}, \frac{\sqrt{3}}{2}, 0), and k_0(\frac{-1}{2}, \frac{-\sqrt{3}}{2}, 0)$

    Derive an expression for the effective mass used in the Richardson constant for a flow of carriers in the z-direction. Derive an expression for the Richardson velocity in the z-direction.

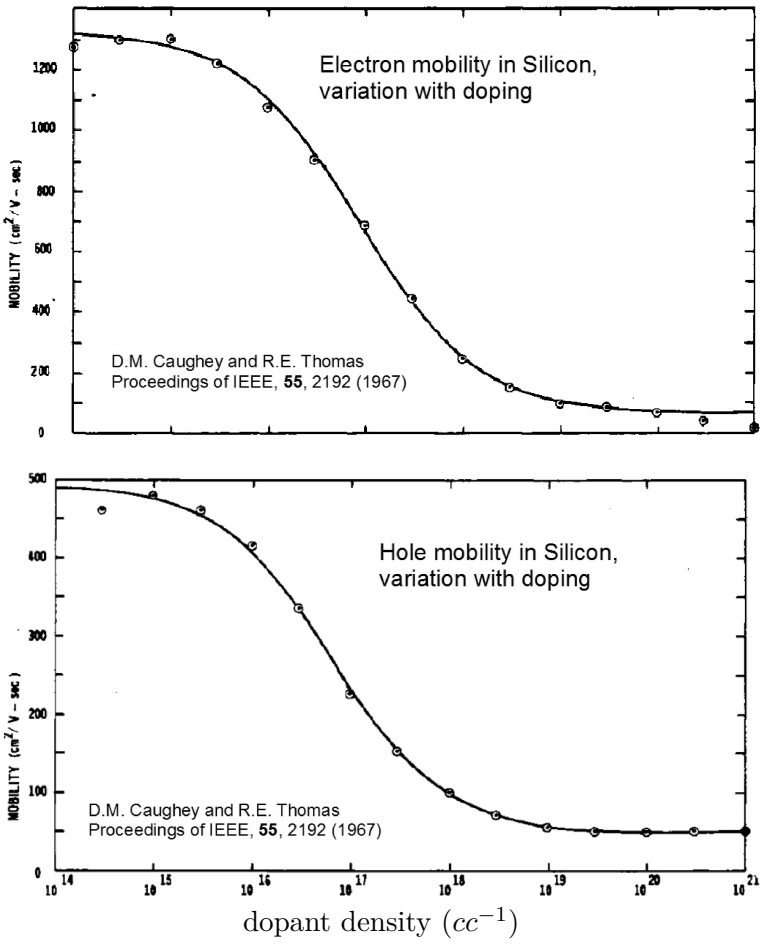


Figure 3.14: The variation of mobility with doping and carrier density can be empirically modelled from experimental data and used to solve the current equation numerically.

4

Optical Processes in Solid State Systems

When electromagnetic radiation, especially in the near UV to the IR (200 nm - 10^4 nm) is incident on a solid state system, its response is experimentally evidenced by recording a range of signals, corresponding to distinct light matter that may be classified into;

1. Reflection (occurs at the interface between two materials with different refractive indices)
2. Refraction
3. Absorption (if $h\nu$ of light matches characteristic energy gaps e.g. band gap or other excitations)
4. Scattering (i) Elastic e.g. Rayleigh (ii) Inelastic e.g. Raman
5. Luminescence (emission of light due to de-excitation of electrons from a higher energy state. Excitation may be via optical (photo-) or electron injection (electro-, cathodo-, tunnelling induced etc.))
6. Other non-linear optical processes

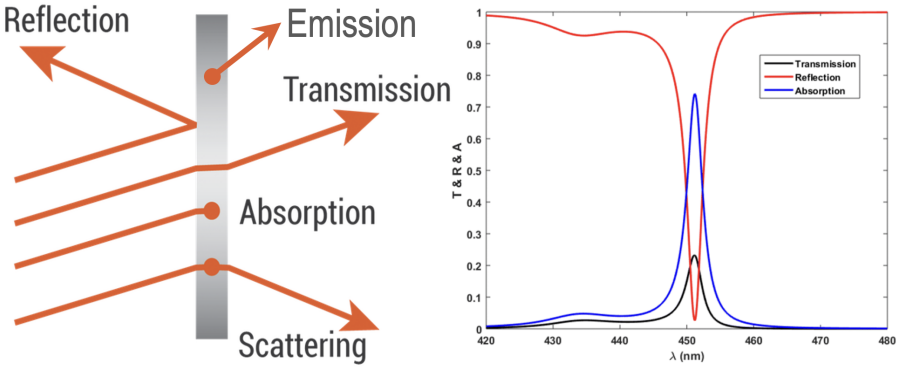


Figure 4.1: (left) Optical processes (right) spectral dependence of R, T, A coefficients

4.1 Quantifying Optical Coefficents

If light of intensity I_o (power per unit area) is incident on a slab of solid as shown in figure 4.1 and the reflected, transmitted light have intensities I_R and I_T then we can define the **reflectance** and **transmittance** as $R = I_R/I_o$ and $T = I_T/I_o$. In the absence of absorption and scattering $R+T$=1. If the media is absorbing then the **absorptance** (A) is quantified by its absorption coefficient α, defined as the fraction of the power absorbed in a unit length of the medium. If the beam is propagating in

the z direction, and the intensity at position z is $I(z)$, then the decrease of intensity in an incremental slice of thickness dz is given by:

$$dI = -\alpha dz \times I(z) \tag{4.1}$$

which gives the Beer's law;

$$I(z) = I_o e^{-\alpha z} \tag{4.2}$$

In general the absorption coefficient is a function of wavelength and thus may absorb specific wavelengths and transmit others. This is the operating principle of many optical filters that absorb all wavelength except a narrow band. Fig. 4.1(right) shows the spectra of a sample that has strong absorption and transmission around 450 nm and high reflection at other wavelengths. In the absence of other processes the sum of the coefficients $R + T + A = 1$ at all wavelengths.

The transmissivity T of an absorbing rectangular slab of thickness l is given by:

$$T = (1 - R_1)e^{-\alpha l}(1 - R_2) \tag{4.3}$$

where R_1 and R_2 are the reflectivities of the front and back surfaces respectively. If $R_1 = R_2 = R$ then the above simplifies to;

$$T = (1 - R)^2 e^{-\alpha l} \tag{4.4}$$

Note: Multiple reflections between front and back surfaces are neglected in the above derivation. See assignment problems on this topic.

In inhomogenous media, scattering may be mathematically modelled in an analogous way to absorption i.e. intensity decreases exponentially as it propagates into the medium according to:

$$I(z) = I_o e^{-N\sigma_s z} \tag{4.5}$$

where N is the number of scattering centres per unit volume, and σ_s is the scattering cross-section of the scattering centre. Scattering by scatterers much smaller than the wavelength of the incident light is known as Rayleigh scattering, where $\sigma_s \propto 1/\lambda^4$. This implies that inhomogeneous materials tend to scatter short wavelengths more strongly than longer wavelengths, giving the day sky its blue colour and red hue at sunset.

4.2 Optical Properties of Materials

Optical properties materials vary widely according to energetics of light - matter interactions in different systems. Materials can be loosely classified as follows;

- solids - insulators, semiconductors, metals - electronic bands
- molecular species - discrete hybridised electronic states
- atomic species - discrete electronic states

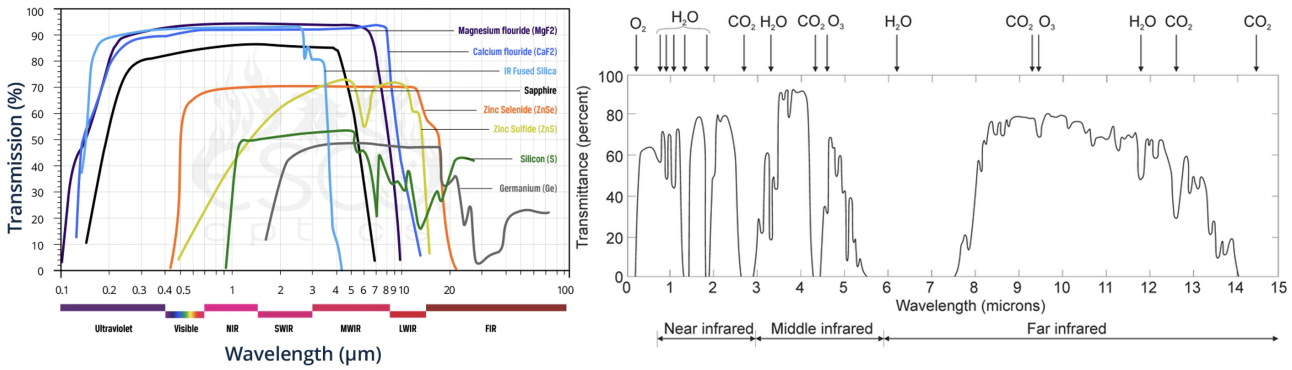


Figure 4.2: Transmission Spectra (left) various semiconductors and optical materials (right) earth's atmosphere containing various molecular species

Fig. 4.2 shows the transmittance spectra of a range of optical materials that are either semiconductors or insulators and that of the earth's atmosphere containing various molecular species. For example, the figure shows that the semiconductor ZnSe has high transmission (low absorption) between 0.45 - 15 μm and absorbs at shorter and longer λ. In contrast molecular absorption are spectrally narrow - denoting transition between discrete states. Similarly atomic absorption spectra also show narrow peaks at energies matching the energy difference between discrete states. Molecular absorption spectra not only evidence transition between electronic states but also absorption due to vibrational and rotational degrees of freedom and associated energy states.

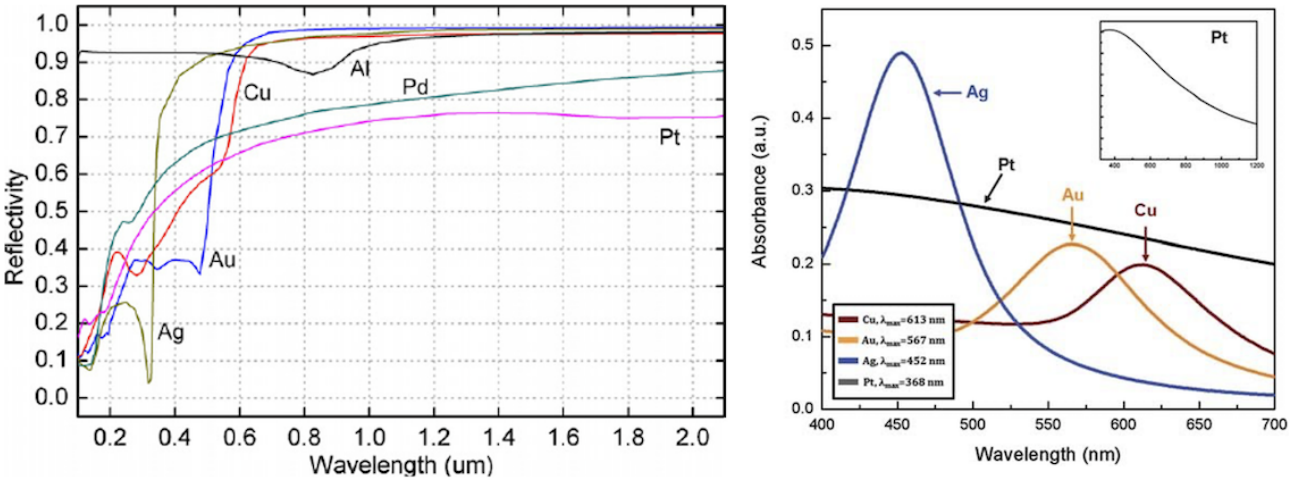


Figure 4.3: (left) Reflection Spectra of various metals (right) Absorption Spectra of various metal nanoparticles

Metals are distinguished from other solids by their shiny lustre and high reflectivity in the visible, which originates from interaction of light with the large free electron density of metals. Fig. 4.3(left) shows the reflectivity spectra of various metals evidencing $R \gtrsim 80\%$ in the VIS and IR. Note that R decreases sharply for light of frequencies above a specific frequency known as the plasma frequency ω_p, typically in the UV. Consequently, metals reflect IR and VIS wavelengths but transmit in the UV, a phenomenon known as the "ultraviolet transparency of metals." Additionally, certain metals display distinctive colours; copper - pinkish hue, gold - yellowish tint, silver - pale yellow etc.These colours stem from interband electronic transitions that supplement the reflection effects induced by free carriers. In contrast metal nanoparticles display distinct absorption peaks in the visible, as shown in Fig. 4.3(right). The absorption originates from excitation of localised surface plasmons - free electron oscillations in the nanoparticles.

For bulk systems, refractive index ($\tilde{n}$) is a parameter that quantifies interaction of light with matter.

Refractive Index

Absorption, reflection and refraction of a medium can be described by a single quantity called the complex refractive index $\tilde{n} = n + i\kappa$. The real part is the normal refractive index defined by the ratio c/v and the imaginary part κ, called the extinction coefficient is directly related to loss quantified by the absorption coefficient α of the medium. Consider an electromagnetic wave propagating in the z direction, given by $\vec{E} = E_o e^{i(kz-\omega t)}$. In a non-absorbing medium of refractive index n, the wavelength of the light is reduced by a factor n compared to the free space wavelength λ. k and ω satisfy the relation $k = n\omega/c$. This can be generalized to the case of an absorbing medium as $k = \tilde{n}\omega/c$. This then gives $\vec{E} = E_o e^{\kappa\omega z/c} e^{i(\omega n z/c-\omega t)}$, demonstrating that κ leads to an exponential decay of the field (wave) in the absorbing medium. Noting that the intensity of light is proportional to the square of the electric field we can show that;

$$\alpha = 4\pi\kappa/\lambda \tag{4.6}$$

Since the refractive index is related to the dielectric constant or relative permittivity ($\tilde{\epsilon}_r$) of a media, note that;

$$\begin{aligned} \tilde{n}^2 &= \tilde{\epsilon}_r = \epsilon_1 + \epsilon_2 \\ \epsilon_1 &= n^2 - \kappa^2 \\ \epsilon_2 &= 2n\kappa \end{aligned}$$

and

$$n = \frac{\sqrt{\epsilon_1 + \sqrt{(\epsilon_1^2 + \epsilon_2^2)}}}{\sqrt{2}} \tag{4.7}$$

$$\kappa = \frac{\sqrt{-\epsilon_1 + \sqrt{(\epsilon_1^2 + \epsilon_2^2)}}}{\sqrt{2}} \tag{4.8}$$

If $\kappa << n$, the above relations can be written as $n = \sqrt{\epsilon_1}$ and $\kappa = \epsilon_2/2\sqrt{\epsilon_1}$, indicating that the refractive index is basically determined by the real part of the dielectric constant, while the absorption is mainly determined by the imaginary part. This generalization is obviously not valid if the medium has a significant absorption coefficient.

4.3 Light - Matter Interactions: a Classical Model

The classical model of light propagation was developed at the end of the 19^{th}C following Maxwell's theory of electromagnetic waves and the introduction of the concept of the dipole oscillator. We assume that there are several different types of dipole oscillators within a medium, each with their own characteristic resonant frequency (ω_o). The different types of oscillators may be;

- Atomic Vibrations: bound electron oscillations - near IR, visible, UV (10^{14}-10^{15} Hz)
- Lattice Vibrations: ionic oscillations - infrared (10^{12}-10^{13} Hz)
- Free carrier (e or h) oscillators (characteristic resonant frequency = 0)

The Dipole Model: Lorentz Oscillator

Consider the interaction between a light wave and an atom with a single resonant frequency ω_o that binds the electron cloud to the nucleus.The electric field of the light wave induces forced oscillations of the atomic dipole through the driving forces exerted on the electrons. Since the nuclear mass is much larger than the mass of electrons ($m_N >> m_o$) we ignore the motion of the nucleus. The EOM of the electron is given by;

$$m_o\ddot{x} + m_o\gamma\dot{x} + m_o\omega_o^2 x = -q_e E \tag{4.9}$$

where $\omega_o \simeq \sqrt{K/m_o}$, γ is the damping rate and $E = \mathrm{Re}[E_o \exp(i(\omega t - \Phi))]$ is the external electric field. The above equation is that of a forced oscillator with solutions of the form $x(t) = X_o\mathrm{Re}[\exp(i(\omega t - \Phi'))]$[1]. Substituting the expressions for E and x in the above equation gives;

$$X_o = -\frac{q_e E_o/m_o}{(\omega_o^2 - \omega^2) - i\gamma\omega} \tag{4.10}$$

where the different phase factors are subsumed in the amplitudes X_o and E_o. This induces a time dependent dipole moment $p(t)(= -q_e x(t))$ in each atom. Considering there are N atoms per unit volume the net dipole polarisation density (P_{dip}) is given by,

$$P_{dip} = N\frac{q_e^2/m_o}{(\omega_o^2 - \omega^2) - i\gamma\omega}E \tag{4.11}$$

[1] only the time dependent part of E and x are considered

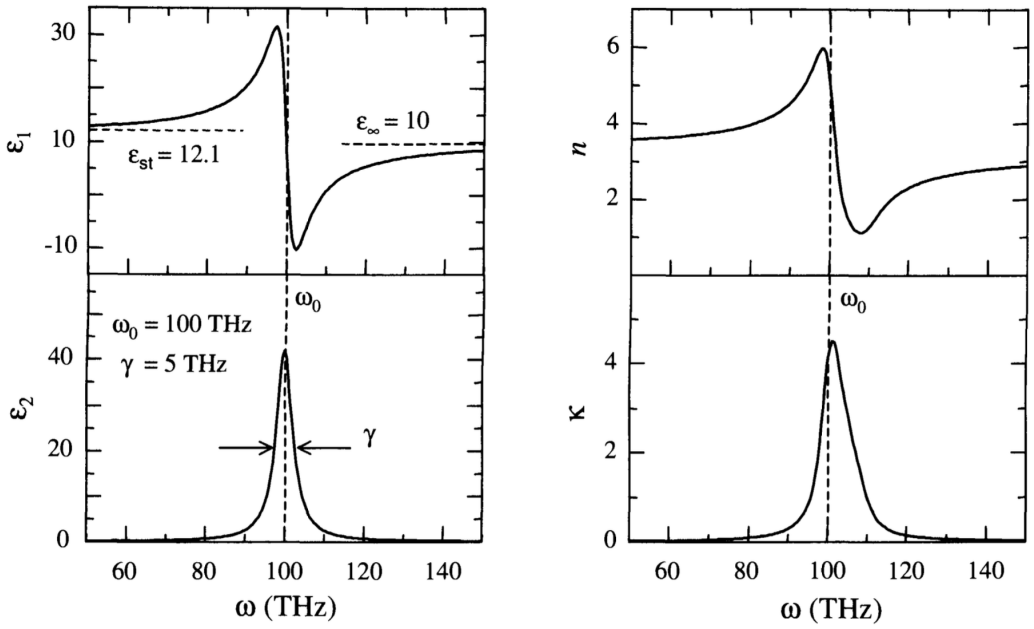


Figure 4.4: Frequency dependence of the real and imaginary parts of $\tilde{\epsilon}_r$ and $\tilde{n}$ of a single dipole oscillator at frequencies close to resonance. (Fox)

Here E is strictly not the external electric field but the net electric field including the field of the other dipoles $E_{net} = E_{ext} + E_{otherdipoles}$. The effect can be ignored assuming the dipole density is small[2]. The electric displacement $\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}$ and $\mathbf{P} = \mathbf{P}_b + \mathbf{P}_{dip}$, where $\mathbf{P}_b = \epsilon_0\chi\mathbf{E}$ is the background polarization of the media in which the dipoles are embedded. Together the relative permittivity ($\tilde{\epsilon}_r$) of the background media and the dipoles may be derived as;

$$\tilde{\epsilon}_r = 1 + \chi + \frac{Nq_e^2/\epsilon_0 m_o}{(\omega_o^2 - \omega^2) - i\gamma\omega} \tag{4.12}$$

$$\epsilon_1 = 1 + \chi + \frac{Nq_e^2}{\epsilon_0 m_o}\frac{(\omega_o^2 - \omega^2)}{(\omega_o^2 - \omega^2)^2 + (\gamma\omega)^2} \tag{4.13}$$

$$\epsilon_2 = \frac{Nq_e^2}{\epsilon_0 m_o}\frac{\gamma\omega}{(\omega_o^2 - \omega^2)^2 + (\gamma\omega)^2} \tag{4.14}$$

In the low and high frequency limits i.e. $\omega \to 0$ and $\omega \to \infty$ we get,

$$\tilde{\epsilon}_r(0) = 1 + \chi + \frac{Nq_e^2}{\epsilon_0 m_o \omega_o^2} \tag{4.15}$$

$$\tilde{\epsilon}_r(\infty) = 1 + \chi \tag{4.16}$$

$$\implies \Delta\tilde{\epsilon}_r = \tilde{\epsilon}_r(0) - \tilde{\epsilon}_r(\infty) = \frac{Nq_e^2}{\epsilon_0 m_o \omega_o^2} \tag{4.17}$$

Noting that the last term in the RHS of the eqns. 4.19 assume structure only near resonance and decay to zero elsewhere, we define $\delta\omega = \omega_o - \omega$, whence $(\omega_o^2 - \omega^2) \simeq 2\omega_o\delta\omega$. Thus close to resonance ($\omega \to \omega_o$) the real and imaginary $\tilde{\epsilon}_r$ are given by,

$$\epsilon_1 = \tilde{\epsilon}_r(\infty) + \Delta\tilde{\epsilon}_r\frac{2\omega_o\delta\omega}{4(\delta\omega)^2 + \gamma^2} \tag{4.18}$$

$$\epsilon_2 = \Delta\tilde{\epsilon}_r\frac{\gamma\omega_o}{4(\delta\omega)^2 + \gamma^2} \tag{4.19}$$

Fig. 4.4 plots the frequency dependence of the $\tilde{\epsilon}_r$ and $\tilde{n}$ of a single dipole oscillator at frequencies close to resonance.

[2]see Fox section 2.2.4

1. The plot for ϵ_2 shows a strongly peaked function that maximizes at $\omega = \omega_o$ with an FWHM $= \gamma$ and quantifies absorption of the dipole.

2. ϵ_1 saturates to $\tilde{\epsilon}_r(0)$ and $\tilde{\epsilon}_r(\infty)$ at the lower and higher extremes.

3. At resonance, $\delta\omega = 0$ and $\epsilon_1 = \tilde{\epsilon}_r(\infty)$

4. ϵ_1 displays a minima and maxima at $\omega \pm \gamma/2$.

5. The entire action of the dipole is restricted to frequencies in the range $\omega \pm \gamma$ demonstrating that damping causes line broadening. The plotted line shapes are known as the Lorentzian.

In general a medium will have several characteristic frequencies ω_{oi} with different damping γ_i and also different species. Again assuming that the dipole density is low and the dipoles DO NOT interact with each other, following the previous discussion the polarization and the dielectric constant may be written as,

$$P_{dip} = \sum_j N \frac{(q_e^2/m_o) f_j}{(\omega_{oj}^2 - \omega^2) - i\gamma_j\omega} E \tag{4.20}$$

$$\tilde{\epsilon}_r = 1 + \chi + \sum_j N \frac{(q_e^2/m_o) f_j}{(\omega_{oj}^2 - \omega^2) - i\gamma_j\omega} \tag{4.21}$$

f_i is a phenomenological constant that denotes the strength of individual oscillators i.e. excitations.

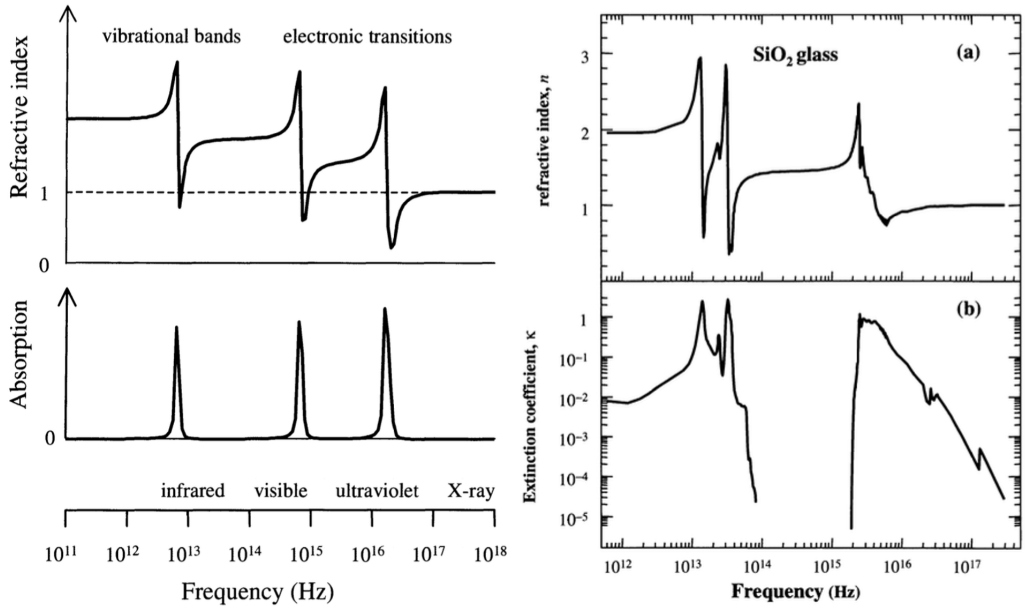


Figure 4.5: (left) Frequency dependence $\tilde{n}$ of a system with oscillators at 3 different resonance frequencies, calculated from eqn. **??** (right) $\tilde{n}$ of fused silica (SiO_2)(Fox)

Fig. 4.5(left) plots the complex $\tilde{n}$ for a system with dipole oscillators with 3 different ω_{oi} and γ_i, but same f_i.

1. as labelled in the figure, typically resonances in the IR originate from vibrational (rotational) modes of excitation and those in the VIS and UV originate from electronic excitations

2. the system is lossy or absorbs significantly near the resonances and is transparent for other ω.

3. since $\tilde{\epsilon}_r(0) > \tilde{\epsilon}(\infty)$ the n decreases with increasing ω

4. individual oscillators may have different strengths which are accounted for by the parameter f_i. The classical model discussed here gives all $f_i = 1$.

Fig. 4.5(right) plots the complex $\tilde{n}$ for fused silica (SiO_2). Note that $\kappa << n$ for all ω other than close to the ω_os i.e. the material has high T except close to ω_os, where A is strong. Fig. 4.5(right) also shows that close to the ω_os, n may be less than 1. This is a special regime connected to a class of materials known as the *epsilon near zero* materials which display interesting optical phenomena[3].

Kramers–Kronig Relations

The mathematical forms of n and κ and their physical manifestation show that they are not independent quantities but are inter-related mathematically by what are known as the Kramers–Kronig relations. These are bidirectional mathematical relations, connecting the real and imaginary parts of any complex function that is analytic in the upper half-plane. The relations are used to compute the real part from the imaginary part (and vice versa) of linear response functions like $\tilde{n}$ in physical systems,

$$n(\omega) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} \frac{\kappa(\omega')}{\omega' - \omega}\, d\omega \tag{4.22}$$

$$\kappa(\omega) = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} \frac{n(\omega')}{\omega' - \omega}\, d\omega \tag{4.23}$$

Here $\mathcal{P}$ denotes the Cauchy principal value. The real and imaginary parts are thus not independent, allowing the full function to be reconstructed given just one of its parts. In effect, this also applies for the complex relative permittivity and electric susceptibility (χ and other response functions.

4.4 Interband Transitions

4.5 Excitons

4.6 Luminescence

4.7 Quantum Confinement

Optical properties of solids are generally independent of their size - till we make them very small. But why and how small is that "very small"? As an example see figure 4.6 that shows CdSe quantum dots (QD) of different sizes illuminated with UV radiation. CdSe is a semiconductor with a bulk band gap $E_g \sim 1.75$ eV $\equiv$ 708 nm. However the photoluminescence (PL) peak blue shifts with decreasing size of the QD. The observed colour originates from their individual photoluminescence, which is now a strong function of their QD size and a direct result of quantum confinement. Quantum confined

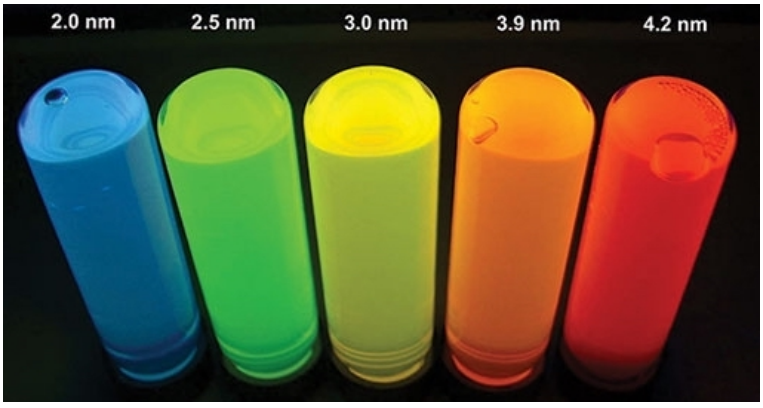


Figure 4.6: Suspension of quantum dots of CdSe of various sizes (increasing left to right) illuminated with UV light.

semiconductors are mostly artificially fabricated nanostructures where the erstwhile "*free*" electrons

[3] Adv. Photonics Research, 3, 2100153, 2022; J. Appl. Phys. 127, 043102, 2020

and holes are restricted in movement in one or more directions. These restrictions arise from physical confinement or size of the structures to be typically in the range of a few nanometers, typically ≤ 5 nm. To estimate this typical size below which the quantum confinement effects become realizable, we begin with the uncertainity principle;

$$\Delta p_x \sim \frac{\hbar}{\Delta x} \tag{4.24}$$

which says that if the particle's position is uncertain (or is confined) over a range Δx then the uncertainity in the conjugate momentum is Δp_x. Assuming that the "*confined*" particle is otherwise free within the range Δx i.e. PE is negligible, then this confiniement contributes to the particle energy which can be estimated as;

$$E_{con} = \frac{\Delta p_x^2}{2m} \sim \frac{\hbar^2}{2m(\Delta x)^2} \tag{4.25}$$

Now this confinement energy would be significant in magnitude iff it is comparable or greater than the thermal energy of the particle. i.e. $E_{con} \geq k_B T/2$. We can conclude that quantum confinement effects dominate when the particle size is;

$$\Delta x \lesssim \sqrt{\frac{\hbar^2}{m k_B T}} \tag{4.26}$$

Assuming $m = 0.1 m_e$ and T = 300 K, we can easily show that the RHS of the above equation 4.26 is $\sim$ 5 nm. The equation also tells us that quantum confinement effects become evident at lower temperatures since lower T would increase the Δx threshold. Remember that for free electrons the functional form of the DOS i.e. dependence on E of free electrons were dictated by the dimension of the system. Figure 4.7 shows the variations across systems with different dimensions. Needless to say that a 2D system has 2 DOF and is confined along 1 dimension, a 1D system has 1 DOF and is confined along 2 dimensions and in a 0D system i.e. a QD motion is confined along all 3 dimensions. Which means that motion of electrons and holes are quantized along all 3 directions in a 0D system, such that they are localised along all three directions, with no continuous band at all. In a 2D system motion is delocalised along the 2 directions orthogonal to the direction of confinement - along which the motion of electrons/holes are quantized. Confinement in any dimension and consequent quantization

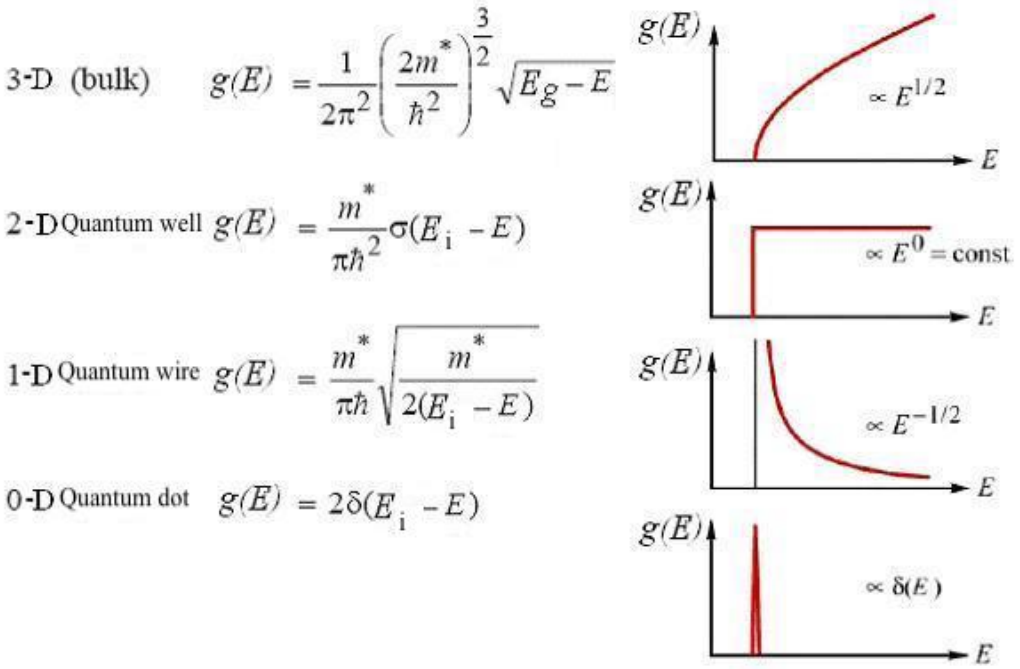


Figure 4.7: Variation of the functional form of the DOS and its plot with Energy, for various dimensions

of motion along that axis results in increase in energy of the quantum particles - by E_{con} and change in the functional form of the DOS. Figure 4.8 plots the DOS of a band gap semiconductor for which

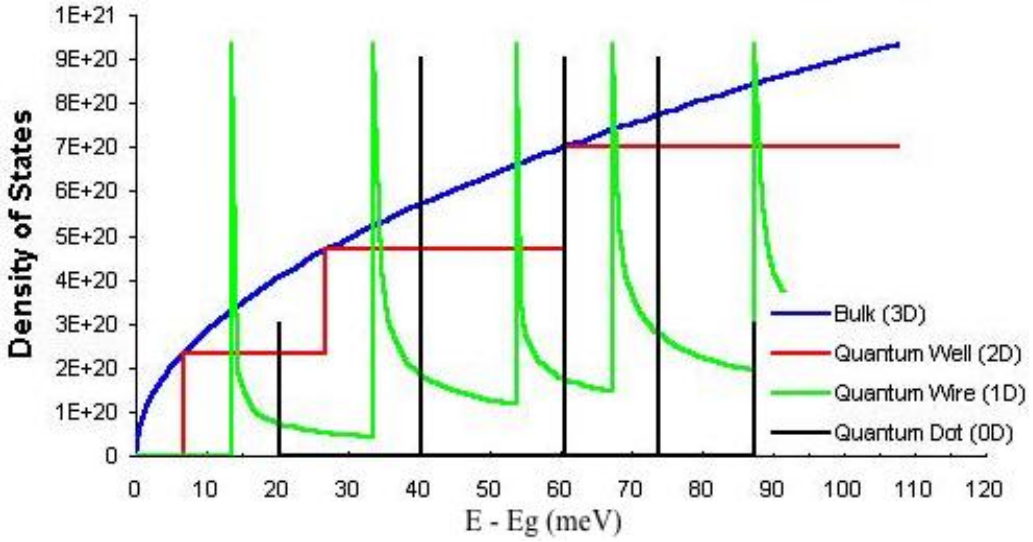


Figure 4.8: Evolution in DOS and its Energy dependence for a band gap $(E - g)$ semiconductor of increasingly restrictive dimensions

various number of dimensions are confined taking it from a bulk (3D) system to a QD (0D). Its worth noting that quantum confined systems, especially 2D electron gas systems and 0D quantum dots have very interesting optical and electrical properties that make them amenable for various interesting applications. Importantly, the general method of investigation introduced here are readily applicable to other quantum confined systems like carbon nanotubes or molecular wires etc.

4.8 2D Quantum Well Structures

Semiconductor QW are artificially fabricated heterostructures where layers of different materials are grown on top of each other. Figure 4.9 shows the schematic of QWs made of GaAs and AlGaAs. The typical E_g(GaAs) = 1.4 eV and E_g(AlGaAs) $\rightarrow$ 1.4 - 2.16 eV, which can be controlled by changing the incorporated Al concentration.

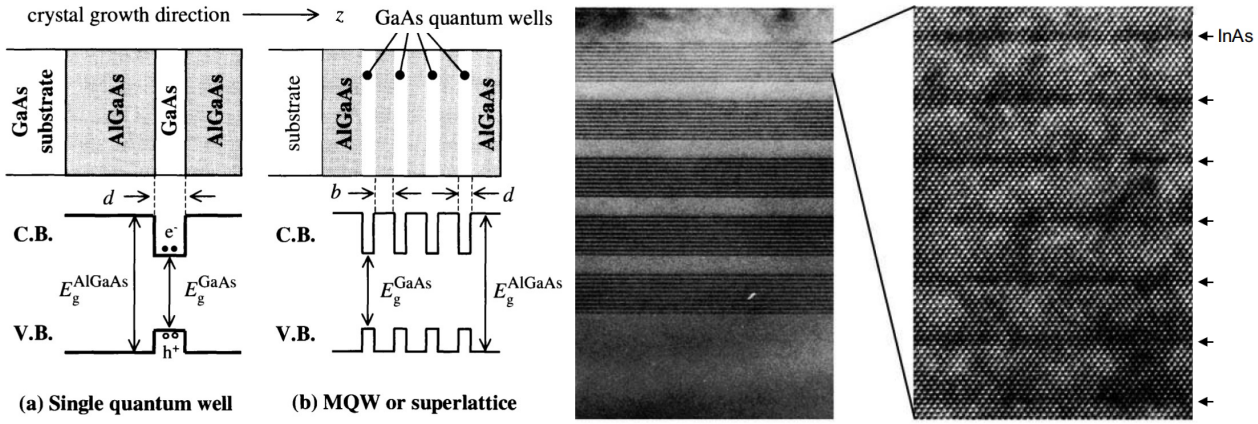


Figure 4.9: (a,b)Schematic of single QW and multiple QW or superlattice of GaAs and AlGaAs layers (c,d) Cross-sectional TEM of a MOVPE-grown superlattice of InAs in GaAsN. Individual InAs layers are indicated by arrows. [Fox, Grundmann]

4.9 *free* Electrons in Quantum Wells

- unbound in x, y directions - part of extended bands of the 2D electron system

- confined in z direction - quantized energy states

In QW heterostructures, semiconductors with different bandgaps are combined to fabricate the QW. Figure 4.10(a) shows the typical band alignment in a QW, where the CBE and VBE of the smaller band gap semiconductor E_{g1} lies completely within the band gap of the material with larger bandgap E_{g2}. Since the smaller E_{g1} material (e.g. GaAs) is bounded on either side with the larger E_{g2} material (AlGaAs), see figure4.9(a,b) the electrons and holes in E_{g1} will localise therein, trapped by the discontinuities in the band edges (potential barriers) on either side i.e. ΔE_C and ΔE_V in figure 4.10(a).These conduction- and valence-band discontinuities are given by;

$$\Delta E_C = (\chi_1 - \chi_2) \tag{4.27}$$
$$\Delta E_V = (\chi_1 + E_{g1}) - (\chi_2 + E_{g2}) \tag{4.28}$$

where χ_i are the electron affinity taken as the CB minima (E_{Ci}) of the respective materials and similarly the VB maximum (E_{Vi}), with ($E_{gi} = E_{Ci} - E_{Vi}$). These barriers quantize the states in the z direction, but the motion in the x, y plane is still free. We thus effectively have a 2D system in which the carriers are quantized in one direction and free in the other two. A general QW structure consists of a series of repeated QW of width d separated from each other by AlGaAs layers of thickness b. This type of structure is either called a multiple quantum well (MQW) or a superlattice, depending on the parameter b. MQW s have large b values, so that the individual quantum wells are isolated from each other, and the properties of the system are essentially the same as those of single QW. They are often used in optical applications to give a usable optical density. Superlattices, by contrast, have much thinner barriers. These QW are coupled together by tunnelling through the barrier, and new extended states are formed in the z direction. Superlattices have additional properties over and above those of the individual QWs. QW structures of the type shown above can only be made if the physical properties of the constituent compounds are favourable to the formation of the artificial crystals i.e. lattice constant of the semiconductor materials are close to each other e.g. the unit cell size of GaAs and AlAs (and hence also the AlxGal-xAs alloy) are almost identical. Further, the relative position of conduction and valence band (band alignment) is determined by the electron affinities should be favourable for the intended application. Figure 4.10(b) shows the band energies for various semiconductors. The design of heterostructures to fulfill a certain device functionality or to have certain physical properties is called 'bandgap engineering'.

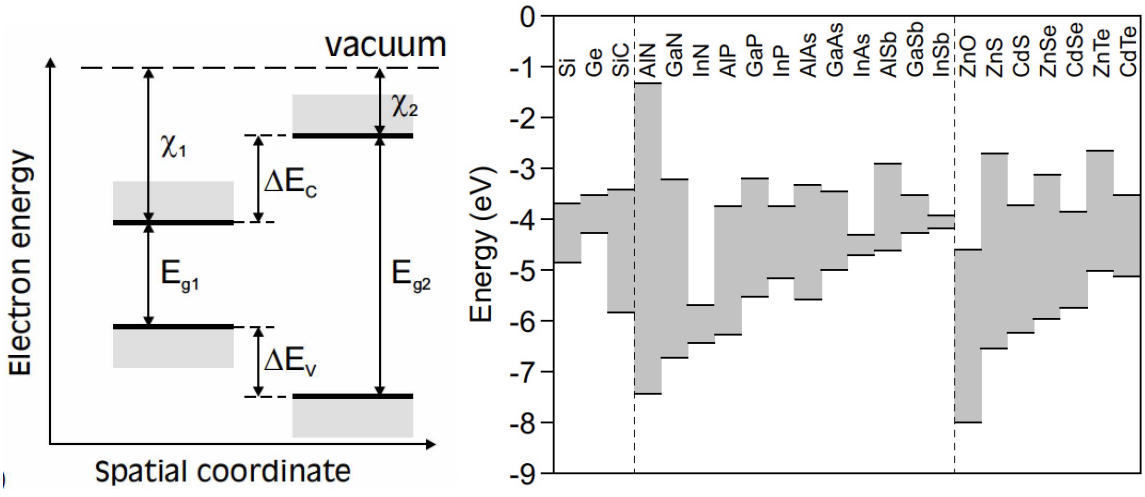


Figure 4.10: (a)Position of band edges (band alignment) in a typical QW heterostructure (b)Position of CB and VB edges for various semiconductors.[Grundmann]

4.10 Electronic States of the 2D Quantum Wells

The energy in a single quantum well of thickness Lz (along the growth direction z) can be calculated employing the quantum-mechanical particle-in-a-box model. The problem separates naturally between

the free motion in the xy plane and the quantized motion in the z direction. Thus, the time independent part of the electron wavefunction can be written as;

$$\Psi(x, y, z) = \psi_1(x, y)\psi_2(z) \tag{4.29}$$

evidently the delocalised motion in the xy plane may be described by a suitable wavevector $\vec{k}$ with that along the z direction may be described by a quantum number n. Assuming that the carrier is a free particle in xy plane the corresponding wave function may be written as;

$$\psi_1(x, y) = \frac{1}{A} e^{i\vec{k}.\vec{r}} \tag{4.30}$$

here $\vec{r}$ pans the xy plane and A is a normalisation constant and The energy corresponding to this motion is just the kinetic energy determined by the effective mass of the *free* particle, such that the total energy i.e. including the quantized energy is given by;

$$E_{total}(n, \vec{k}) = E_n + \frac{\hbar^2 k^2}{2m^*} \tag{4.31}$$

In order to obtain $\psi_2(z)$ we need to estimate the bounding potentials ΔE_C and ΔE_V that localise carriers within the smaller band gap semiconductor. As a first approximation lets assume that these binding potentials are infinitely high. That is the electron confinement states may be modelled by those of a 1D infinite well of width d.The Schrodinger equation within the well is given by;

$$-\frac{\hbar^2}{2m^*}\frac{d^2\psi_2(z)}{dz^2} = E\psi_2(z) \tag{4.32}$$

Applying the boundary conditions for the infinite potential well that $\psi_2(0) = \psi_2(d) = 0$ gives us;

$$\psi_2(z) = \sqrt{\frac{2}{d}}\sin(\frac{n\pi z}{d}) \tag{4.33}$$

with the energy of the n^{th} level given as;

$$E_n = \frac{\hbar^2}{2m^*}(\frac{n\pi}{d})^2 \tag{4.34}$$

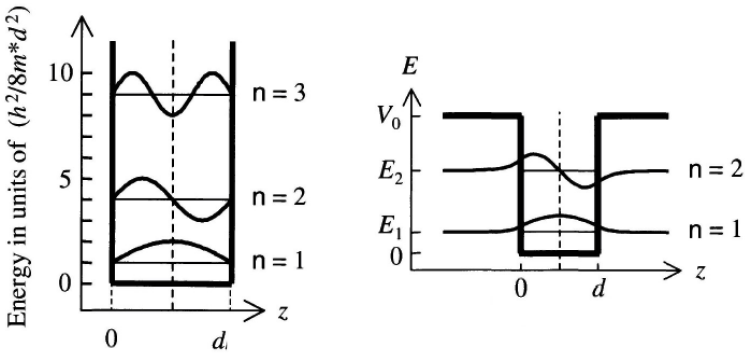


Figure 4.11: (a)Infinite 1D potential well. The first three energy levels and corresponding wave functions. (b)Finite 1D potential well with first 2 bound states and corresponding wave functions.[Fox]

Although real semiconductor quantum wells have finite barriers, the infinite barrier approximation gives us a fair estimate of the concerned parameters and a viable model to discuss their properties. Evidently, the accuracy of the calculated parameters for this model will be highest for states with small quantization energies in material combinations that give rise to high barriers at the interfaces. Its worth noting the following important points. 1) Energy of quantised levels are (i) $\propto 1/m^*$ and (ii) $\propto 1/d$ which implies that thinner wells with lighter effective mass of particles have higher energy. (2) Since the energy depends on the effective mass, electrons, heavy holes and light holes will all have

different quantization energies i.e. the VB heavy holes will have the lowest energy and will form the ground state level.

Its imporatnt that we estimate the quantization energy given by equation 4.34. Assuming $m^* = 0.1m_e$ and d=10 nm we get the energies of the first two states as 38 meV and 150 meV respectively. Remember that these energies are above the band edge or E_C of the system. Evidently, the quantization energy is greater than the thermal energy at room temperature, $k_BT \sim 25$ me V, justifying the quantum description. Overall the infinite well model overestimates the quantization energy. In real quantum wells with finite barriers, the particles are able to tunnel into the barriers to some extent, and this allows the wave function to spread out further and thus reduces the confinement energy.

4.11 Finite Potential Well

The z confinement of a real QW is far better modelled using a finite 1D potential well, schematically shown in figure 4.11(b). Here the height of the confining potential well is $V_0(=\Delta E_C)$ and the potential inside the well is 0 $(=E_{C1})$. The Schrodinger equation within the quantum well $(-d/2 \leq z \leq d/2)$ is the same as before but beyond the equation is different, which are given as;

$$-\frac{\hbar^2}{2m_w^*}\frac{d^2\psi_{2w}(z)}{dz^2} = E\psi_{2w}(z) \tag{4.35}$$

$$-\frac{\hbar^2}{2m_b^*}\frac{d^2\psi_{2b}(z)}{dz^2} + V_0\psi_{2b}(z) = E\psi_{2b}(z) \tag{4.36}$$

Note that the effective mass of the particle is different inside (m_w^*) and outside (m_b^*) the well - which are obviously composed of different material and thus have differnt m^*. The corresponding wave functions are given by;

$$\psi_{2w} = C\sin(kz) or C\cos(kz) - - - - - - - (-d/2 \leq z \leq d/2) \tag{4.37}$$

$$\psi_{2b} = C'e^{-\kappa z} - - - - - - - - - (z \geq d/2) \tag{4.38}$$

$$\psi_{2b} = C'e^{\kappa z} - - - - - - - - - - (z \leq -d/2) \tag{4.39}$$

where $\hbar^2k^2/2m_w = E$ and $\hbar^2\kappa^2/2m_b = V_0 - E$. The solutions are oscillatory inside the well and decay outside. We know that for symmetric potentials, $V(x) = V(-x)$, the wave functions of the bound states must be either even or odd. Thus the bound state wave functions (equation 4.26) corresponding to the potential shown in figure 4.11(b) can be written as;

$$\psi_2(z) = \sqrt{\frac{2}{d}}\sin(\frac{n\pi}{d}z)(n = 2, 4, 6, 8, ...) \tag{4.40}$$

$$\psi_2(z) = \sqrt{\frac{2}{d}}\cos(\frac{n\pi}{d}z)(n = 1, 3, 5, 7, ...) \tag{4.41}$$

That is, the wave functions corresponding to odd quantum numbers $(n = 1, 3, 5, 7, ...)$ are symmetric, $\psi_2(z) = \psi_2(-z)$, and those corresponding to even numbers $(n = 2, 4, 6, 8, ...)$ are antisymmetric, $\psi_2(z) = -\psi_2(-z)$. Coupled with the boundary conditions that the wave functions and the particle flux are continous at the boundaries we get;

$$\psi_{2w}(0/d) = \psi_{2b}(0/d) \tag{4.42}$$

$$\frac{1}{m_w^*}\frac{\psi_{2w}}{dz}|_{(0/d)} = \frac{1}{m_b^*}\frac{\psi_{2b}}{dz}|_{(0/d)} \tag{4.43}$$

which gives rise to the conditions;

$$\tan(kd/2) = m_w^*\kappa/m_b^*k \tag{4.44}$$

$$\tan(kd/2) = -m_b^*k/m_w^*\kappa \tag{4.45}$$

In principle we can solve for the Energy once we have substituted the functional forms of k and κ, however there are no analytic solutions, which have to be arrived at numerically or graphically. See any standard Quantum Mechanics text for the graphical solutions. Figure 4.11(b) shows the wave

functions of a typical finite well with two bound states. The similarity between these wave functions and the first two states of the infinite well is apparent. The main difference is that the wave functions of the finite well spread out more by tunnelling into the barrier, whereas the wave functions of the infinite well stop abruptly at the boundary. The table 4.1 below gives the first few bound state energies for electrons and holes for a 10 nm GaAs QW in $Al_0.3Ga_0.7As$ for both the finite and infinite QWs.

Table 4.1: Bound states energies of a GaAs/ AlGaAs QW under finite and infinite well models. Energies are in meV

particle	Q. No.	Finite QW	Infinite QW
e	1	32	57
e	2	120	227
e	3	247	510
hh	1	7	11
hh	2	30	44
hh	3	66	100

Based on the above discussion its worth noting that;

- The spreading of the wave functions into the barrier by tunnelling reduces the quantum confinement energy compared to that of an infinite barrier well.
- Levels near the top of the well with Energies close to V_0 have a smaller decay constants and tunnel more into the barrier.
- The infinite QW overestimates the quantization energy. The difference increasing for higher quantum number of the states (n).
- Quantization energies of heavy holes are smaller than those of the electrons because of their heavier effective mass.
- Separation between successive electron levels is $> 3k_BT$ at RT, ensuring two-dimensional physics for the electrons at 300 K.

4.12 Optical Absorption in QWs

Here we consider absorption of electromagnetic radiation by a QW, where an electron in the VB of the well absorbs energy and is excited to an electronic state in the CB. The absorption rate calculated from applying Fermi's golden rule gives us;

$$W_{i-f} = \frac{2\pi}{\hbar}|M|^2 g(\hbar\omega) = \frac{2\pi}{\hbar}|\langle f| - e\vec{r}.\vec{E}_0|i\rangle|^2 g(\hbar\omega) \tag{4.46}$$

and

$$M \propto \langle f|\vec{r}|i\rangle = \int \Psi_f^*(\vec{r}) x \Psi_i(\vec{r}) \tag{4.47}$$

assuming that the electric field is along x direction. Since QWs are highly anisotropic in their potentional configuration, especially along the confinement direction z we note that;

$$\langle f|x|i\rangle = \langle f|y|i\rangle \neq \langle f|z|i\rangle \tag{4.48}$$

and remember in case of the QWs the wavefunction is given by;

$$\Psi(\vec{r}) = \psi_1(x,y) \times \psi_2(z) \tag{4.49}$$

$$= \frac{1}{A} e^{i\vec{k}_{xy}.\vec{r}_{xy}} \times \psi_2(z) \tag{4.50}$$

If the incident light is polarized along x axis and travels along the z direction and we consider electron transition from a bound state characterised by the quantum number n_i, in the VB to a a bound state characterised by the quantum number n_f, in the CB. The corresponding wavefunctions are given by;

$$\Psi_i \quad \propto u_v\, \psi_{2h}(n_i, z)\, e^{i\vec{k}_{xy}.\vec{r}_{xy}} \tag{4.51}$$

$$\Psi_f \quad \propto u_c\, \psi_{2e}(n_f, z)\, e^{i\vec{k'}_{xy}.\vec{r}_{xy}} \tag{4.52}$$

We note the following in the above equations;

1. The plane waves characterised by the term $e^{i\vec{k}_{xy}.\vec{r}_{xy}}$ spans the 2D xy plane.
2. The functions u_C and u_V are the 3D Bloch functions of the CB and VB commensurate with the periodic potential of the atomic lattice.
3. ψ_{2hn_i} and ψ_{2en_f} are the localised bound states of the QW
4. $\vec{k}_{xy} = \vec{k'}_{xy}$ since the photon momentum is nsgligibly small wrt the crystal momentum.

$$M \propto \langle f|\vec{r}|i\rangle = \int u_c^*\, \psi_{2e}^*(n_f, z)\, e^{-i\vec{k'}_{xy}.\vec{r}_{xy}}\, x\, u_v\, \psi_{2h}(n_i, z)\, e^{i\vec{k}_{xy}.\vec{r}_{xy}} d^3\vec{r} \tag{4.53}$$

$$= \int u_c^*\, x\, u_v d^3\vec{r} \int \psi_{2e}^*(n_f, z)\, \psi_{2h}(n_i, z) dz \tag{4.54}$$

$$M_{CV} = \int u_c^*\, x\, u_v d^3\vec{r} \tag{4.55}$$

$$M_{n_i n_f} = \int \psi_{2e}^*(n_f, z)\, \psi_{2h}(n_i, z) dz \tag{4.56}$$

$$M = M_{CV} \times M_{n_i n_f} \tag{4.57}$$

The term M_{CV} is akin to the allowed electronic transition between VB to CB in bulk systems i.e. GaAs the QW layer. What about the term $M_{n_i n_f}$? Obviously to calculate that term we'll need the exact functional forms of the integrands made up of the wavefunctions. Assuming the infinite QW model we can write;

$$M_{n_i n_f} = \frac{2}{d} \int_0^d \sin(\frac{n_f \pi}{d} z)\, \sin(\frac{n_i \pi}{d} z) dz \tag{4.58}$$

$$= \frac{2}{2d} \int_0^d \cos(\frac{(n_f - n_i)\pi}{d} z) - \cos(\frac{(n_f + n_i)\pi}{d} z) dz \tag{4.59}$$

$$= \frac{1}{d} [-\sin(\frac{(\Delta n)\pi}{d} z) + \sin(\frac{(n_f + n_i)\pi}{d} z)]_0^d \tag{4.60}$$

Evidently, second term is always zero. And the overlap integral is equal to zero if $\Delta n = 0$ and unity otherwise. In finite QWs e and h wave functions with different quantum numbers are not always orthogonal to each other. However, even then $\Delta n \neq 0$ transitions have a weak transition rate and forbidden if $\Delta n =$ odd, since transition between odd and even parity states is zero. Figure 4.12 shows the band diagram along z along with the $E - k_{xy}$ dispersion relation for a QW. As we increase the photon energy ($\hbar\omega$) from zero, no transitions will be possible until even $\hbar\omega = E_g$ since there are no states at the bulk band edge. $\hbar\omega$ has to now cross the threshold for exciting electrons from the ground state of the valence band (the $n_i = 1$ heavy hole level) to the lowest conduction band state (the $n_f = 1$ electron level). This $\Delta n = 0$ transition is allowed with the energy balance equation as, $\hbar\omega = E_g + E_{e1} + E_{hh1}$. Thus the optical absorption edge of the QW has shifted to a higher energy compared to the band edge of the bulk parent material as shown in figure 4.13. Importantly since the shift $E_{e1} + E_{hh1}$ is determined by the well width d we can relatively easily tune the absorption edge. As shown in Figure 4.12the $E - k_{xy}$ bands, representing the free electron nature in xy plane, are parabolic in nature. The energy conservation equation of the depicted transition is given by;

$$\hbar\omega = E_g + (E_{e1} + \frac{\hbar^2 k_{xy}^2}{2m_e^*}) + (E_{hh1} + \frac{\hbar^2 k_{xy}^2}{2m_h h^*}) \tag{4.61}$$

$$= E_g + E_{e1} + E_{hh1} + \frac{\hbar^2 k_{xy}^2}{2\mu^*} \tag{4.62}$$

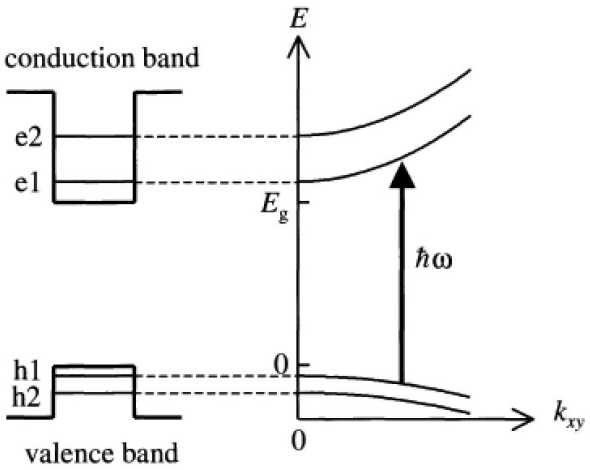


Figure 4.12: Electronic transition in a QW at finite k_{xy}[Fox]

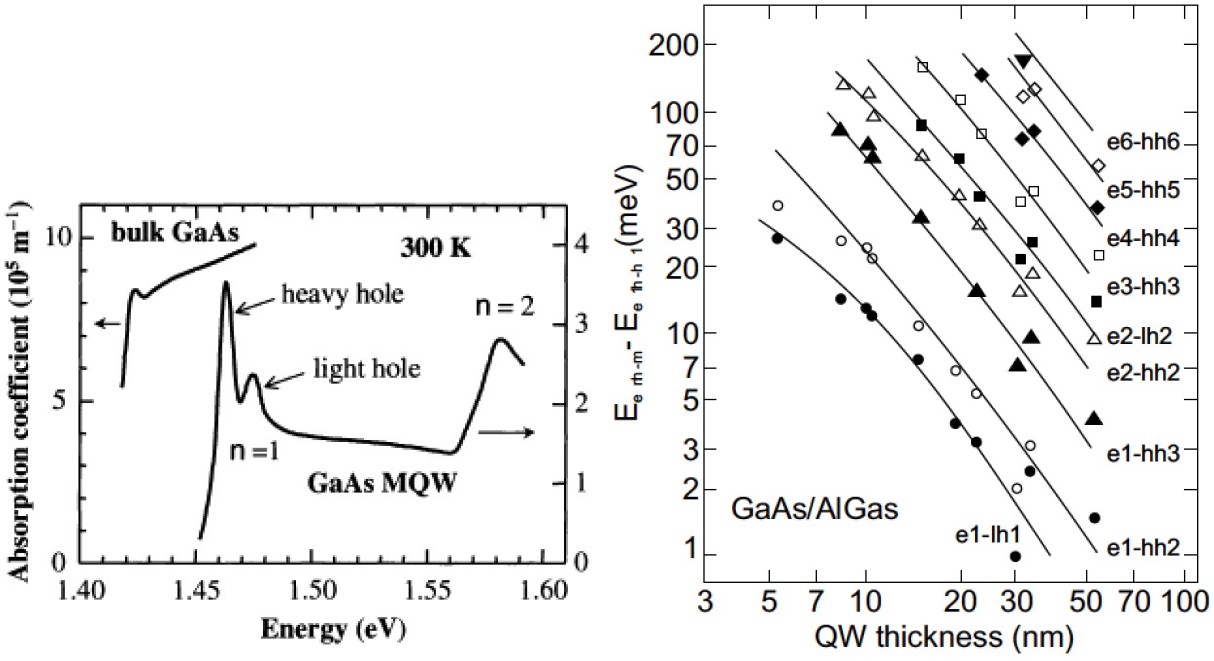


Figure 4.13: (a)RT absorption spectra of GaAs/AlGaAs QW of d=10 nm along with that for bulk GaAs.(b) Variation of $e-h$ transition energies in GaAs/AlGaAs QW of varying thickness.[Fox,Grundmann]

which means that most transitions matching $\hbar\omega$ happen for k_{xy} =0. Figure 4.14 shows the absorption coefficient for an infinite quantum well of width d compared to the equivalent bulk semiconductor. The dashed line shows the variation of α with energy as $\alpha \propto \sqrt{E}$ originating from the energy dependence of the electronic DOS. Here the joint DOS of the 2D material is independent of energy and behaves like a step function as shown in figure 4.7 for a 2D system given by $g(E) = \mu/\pi\hbar^2$. Thus the α has a step like nature with each absorption step located at energy given by $\hbar\omega - E_g = E_{ei} + E_{hhi}$. Which for an infinite QW given below and plotted in Figure 4.14.

$$\hbar\omega - E_g = \frac{\hbar^2 n^2}{2\pi^2 d^2}\left(\frac{1}{m_e^*} + \frac{1}{m_h^*}\right) \tag{4.63}$$

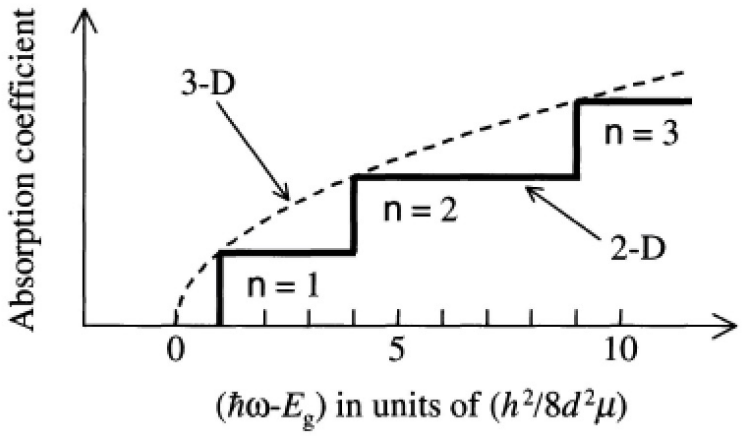


Figure 4.14: Absorption coefficient of bulk semiconductor and its Quantum well.[Fox]

4.13 Quantum Well Superlattice

In superlattices, the barrier thickness b (figure 4.9) is so small that the localised carrier wavefunctions leak out and can tunnel into neighboring QWs and there exists significant wavefunction overlap between adjacent wells. This leads to an engineered artificial band structure similar to the Kronig–Penney model. These superlattice the bands are called minibands, the gaps are called minigaps. Figure 4.15 shows the calculated bands of a GaAs/AlGaAs superlattice. We will come back to superlattices in Chapter 6 when we study emission from superlattices and QW lasers.

4.14 Triangular QWs

Now we consider a single heterointerface between n-doped materials e.g. n-AlGaAs/n-GaAs. Figure shows the formation of a triangular potential well. The Fermi energy is higher in AlGaAs compared to GaAs and thus electrons transfer from the formar to the latter to establish thermodynamic equilibrium. At equilibrium the system must have a constant Fermi level across the junction. This results in the formation of a triangular potential well in the GaAs close to the interface. A 2D electron gas is trapped in this potential triangular QW (figure 4.16). The charge transfer adjusts the band bending and the charge density (quantized levels in the well) in such a way that they are self-consistent. The Poisson equation and the Schrodinger equation are simultaneously fulfilled. Numerically, both equations are iteratively solved and the solution is altered until it is self-consistent, i.e. it fulfills both equations. Importantly, if the region of the 2DEG is not doped, the electron gas exists without any dopant atoms and ionized impurity scattering no longer exists. This is known as modulation doping. Absence of impurities and defects increases carrier mobility with mobilities up to 3.1 10^7 cm^2/Vs have been realized. Figure 4.17 shows the various milestones in high carrier mobility achieved in GaAs over the years.

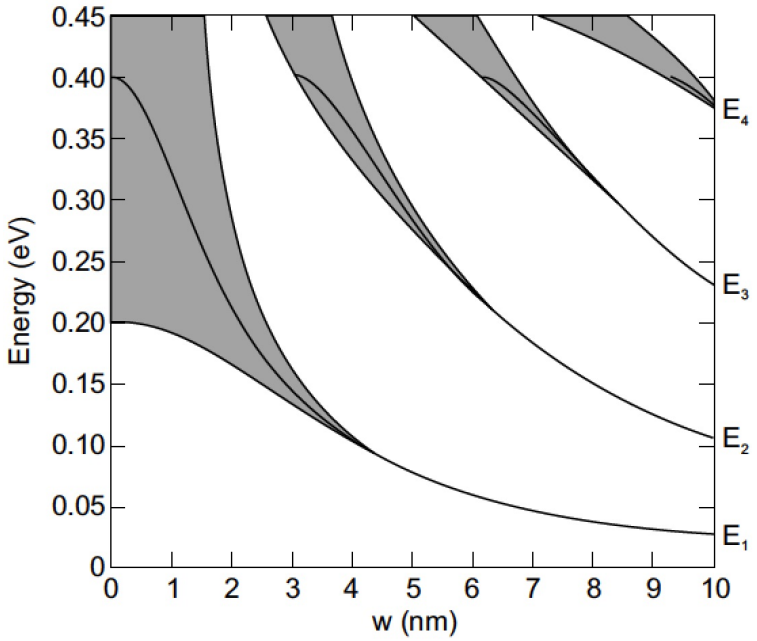


Figure 4.15: Calculated bands of a superlattice as a function of QW width w. [Grundmann]

4.15 Problems

1. Crown glass has a refractive index of 1.51 in the visible spectral region. Calculate the reflectivity of the air-glass interface, and the transmission of a typical glass window.

2. The complex dielectric constant of the semiconductor CdTe is given by $\epsilon_r = 8.92 + i2.29$ at 500 nm. Calculate the phase velocity of light and the absorption coefficient.

3. Sea water has a refractive index of 1.33 and absorbs 99.8 % of red light of wavelength 700 nm in a depth of 10 m. What is its complex dielectric constant at this wavelength?

4. Consider a flat, rectangular slab of material with complex refractive index. Show that R between the medium and air, at normal incidence, is given by;

$$R = \frac{(n-1)^2 + \kappa^2}{(n+1)^2 + \kappa^2} \tag{4.64}$$

5. If the thickness of the slab $t >> l_c$, the coherence length of incident light, i.e. interference effects may be neglected, show that

$$T = \frac{(1-R_1)(1-R_2)e^{-\alpha t}}{1 - R_1 R_2 e^{-2\alpha t}} \tag{4.65}$$

   All symbols denote parameters as denoted in the text.

6. Interference effects can not be neglected if $t << l_c$. Assuming $R_1 = R_2 = R$ show that;

$$T = \frac{(1-R)^2 e^{-alphat}}{1 - 2Re^{-\alpha t} cos\Phi + R^2 e^{-2\alpha t}} \tag{4.66}$$

   Φ is round trip phase shift within the slab.

7. Show that the uncertainity relation $\Delta x \Delta p$ for a 1D infinite wellis given by;

$$\Delta x \Delta p = \frac{\hbar}{2}(\frac{n^2\pi^2}{3} - 2) \tag{4.67}$$

8. Calculate the energy of the first electron bound state in a GaAs/AlGaAs QW with $d = 10$ nm and $V_0 = 0.3$ eV. Take $m_w^* = 0.067$ and $m_b^* = 0.092$. Compare this value to the one calculated for an infinite quantum well. *Ans: Finite well E_1=31.5 meV, Infinite well E_1= 57 meV*

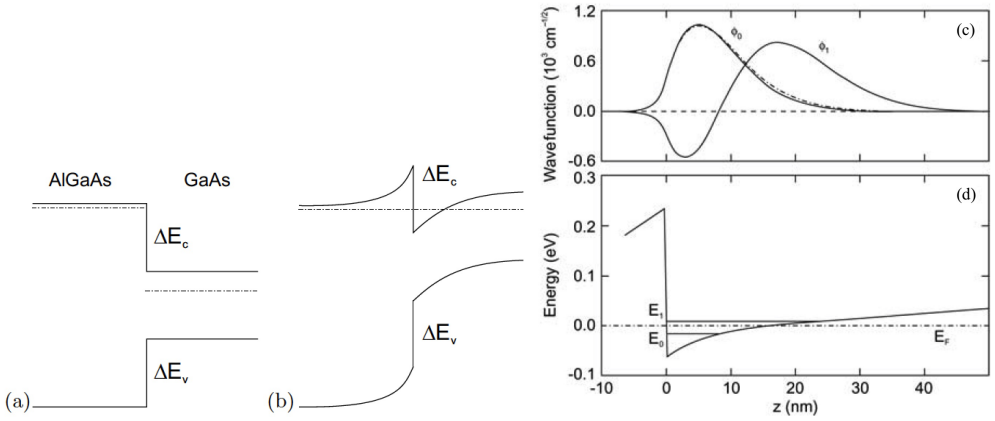


Figure 4.16: (a) GaAs and AlGaAs not in contact, showing different E_F (b)GaAs and AlGaAs in contact, showing equillibrium E_F, band bending and formation of triangular potential well(c)Wavefunctions of the first 2 states of the triangular potential well (d)Energy states of the triangular potential well. [Grundmann]

9. Estimate the difference in the wavelength of the absorption edge of a 20 nm QW of GaAs at 300 K. *Ans:* E_{e1}, E_{hh1} *are 14 and 2 meV, shift = 16 meV*

10. What is the step height of the calculated DOS at E_n?

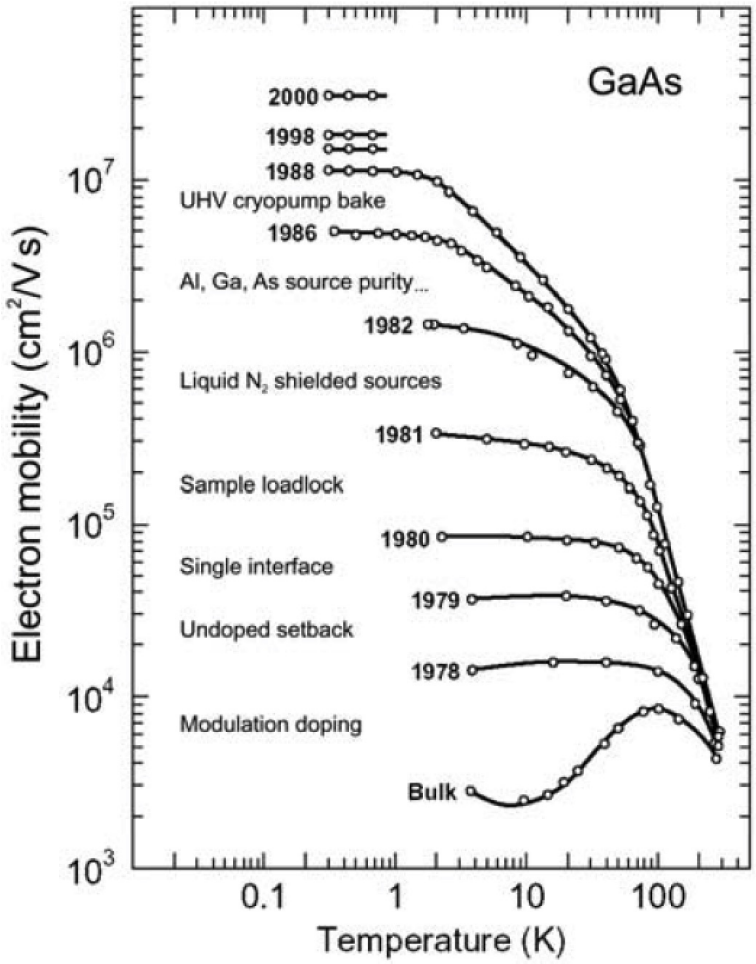


Figure 4.17: Progress in electron mobility in GaAs, with the technical innovation responsible for the improvement [Grundmann]

Crystal	E_g (eV) (0 K)	E_g (eV) (300 K)	Δ (eV)	m_e^*	m_{hh}^*	m_{lh}^*	m_{so}^*
GaAs	1.519	1.424	0.34	0.067	0.5	0.08	0.15
GaSb	0.81	0.75	0.76	0.041	0.28	0.05	0.14
InP	1.42	1.34	0.11	0.077	0.6	0.12	0.12
InAs	0.42	0.35	0.38	0.022	0.4	0.026	0.14
InSb	0.24	0.18	0.85	0.014	0.4	0.016	0.47

Figure 4.18: Various parameters of Group III V semiconductors[Fox]

5

Screening in 3D and 2D Electron Gas

5.1 Polarisation and Screening

Simplest case of $e-e$ interactions. In a dielectric, that has only bound charges but no free charges, the dipole moment per unit volume is related to field $\vec{E}$ by;

$$\nabla.\vec{P} = -\rho_{ind}\vec{P} = \epsilon_o\chi\vec{E} \tag{5.1}$$

Additionally, if there are free charges then they too would reorganise in response to the applied external field $\vec{E}$ field. To analyse the given situation consider the case where field due to an external charge, i.e. charge associated with an ionised impurity or a gate potential, is placed in a $free$ electron sea (CB electrons).
For such a case we can calculate the local potential from;

$$\nabla^2 V = -\frac{\rho_{ext} + \rho_{ind}}{\epsilon_o\epsilon_r} \tag{5.2}$$

But how do we estimate ρ_{ind}?

5.2 3D Electron Gas

Consider a metal or a degenerately doped semiconductor with $free$ electrons. At equillibrium the E_f is uniform across the system, that means at places higher e^- potential energy corresponds to lower e^- concentration and vice-versa. Figure 5.1 below shows the E_F in the CB, with electrons filling all energy states below it. Note that regions having higher CB minima (higer PE) has lower local density of electrons. Remember the case of Schottky and pn junctions where close to the junction

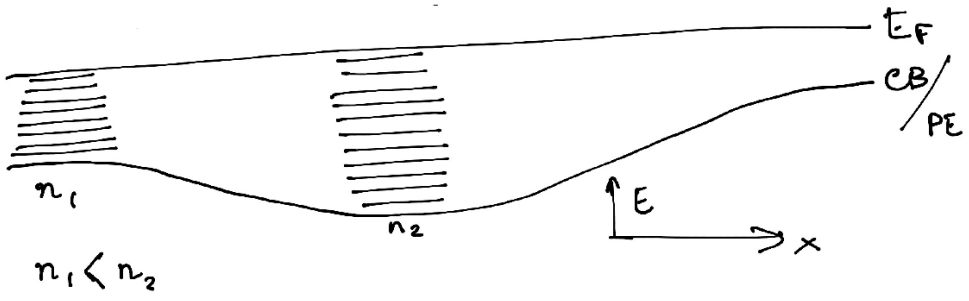


Figure 5.1: Spatial variation of PE and corresponding local number density.

on the n-type side the raising of CB minimum (local electron potential energy) resulted in depletion in free electrons. Now for an ideal $free$ electron gas (neglecting background potentials) at very low temperatures $T \sim 0K$ we can write the Fermi energy as;

$$E_F^o = -\frac{\hbar^2}{2m}(3\pi^2 n_o)^{2/3} \tag{5.3}$$

Now if we apply an electrostatic potential such that electron energy changes then the local chemical potential is given by;

$$\mu = E_F(x) - eV(x) \simeq -\frac{\hbar^2}{2m}(3\pi^2 n(x))^{2/3} - eV(r) \tag{5.4}$$

Here $E_F(x)$ is the "local" Fermi energy and $n(r)$ is the local free electron density. The above approximation is valid if $V(x)$ varies much slowly compared to $\lambda_F (= 2\pi/k_F)$.

$$\begin{aligned}
E_F(n) &\simeq E_F(n_o) + \frac{\partial E_F}{\partial n}|_{n_o}.\delta n && (5.5)\\
\delta E_F = E_F(n) - E_F(n_o) &\simeq \frac{\partial E_F}{\partial n}|_{n_o}(n(r) - n_o) = eV(r) && (5.6)\\
\frac{\partial E_F}{\partial n}|_{n_o} &= \frac{2E_F^o}{3n_o} && (5.7)\\
\Rightarrow n(x) - n_o = \delta n &= \frac{3}{2}n_o\frac{eV(r)}{E_F^o} && (5.8)\\
\Rightarrow \delta n &= D(E_F) \times eV(r) && (5.9)
\end{aligned}$$

where $D(E_F)$ is the electron density of states (DOS) at E_F and the local change in the free carrier density is proportional to the local DOS and the local potential.

$$\begin{aligned}
\Rightarrow \rho_{ind} &= -e \times \delta n = -e^2 D(E_F) \times eV(r) && (5.10)\\
\Rightarrow \nabla^2 &= -\frac{\rho_{ext} + \rho_{ind}}{\epsilon_o} && (5.11)\\
&= -\frac{\rho_{ext}}{\epsilon_o} + \frac{e^2 D(E_F)V(r)}{\epsilon_o} && (5.12)\\
\Rightarrow \nabla^2 V(r) - q_{TF}^2 V(r) &= -\frac{\rho_{ext}}{\epsilon_o} && (5.13)
\end{aligned}$$

where $\vec{q}_{TH}$ is the Thomas Fermi wave vector and $|q_{TF}| = \sqrt{\frac{e^2 D(E_F)}{\epsilon_o}}$. Fourier transform of the equation 5.13 gives us;

$$\Rightarrow -q^2 V(q) - q_{TF}^2 V(q) = -\frac{\rho_{ext}(q)}{\epsilon_o} \tag{5.14}$$

such that,

$$\begin{aligned}
V(r) &= \tfrac{1}{(2\pi)^3}\int d^3q V(q) e^{-i\vec{q}.\vec{r}} && (5.15)\\
\rho(r) &= \tfrac{1}{(2\pi)^3}\int d^3q \rho(q) e^{-i\vec{q}.\vec{r}} && (5.16)
\end{aligned}$$

Thus we can write;

$$\begin{aligned}
V(q) &= \frac{\rho_{ext}(q)}{\epsilon_o(q^2 + q_{TF}^2)} && (5.17)\\
&= \frac{q^2\rho_{ext}(q)/\epsilon_o}{(1 + q_{TF}^2/q^2)} && (5.18)\\
&= \frac{V_{ext}(q)}{(1 + q_{TF}^2/q^2)} && (5.19)
\end{aligned}$$

We can define a dielectric constant for the polarised free electron gas which is a function of q and thus the spatial coordinate r.

$$\epsilon_r(q) = (1 + q_{TF}^2/q^2) \tag{5.20}$$

Thus for the free electron gas in general, in the presence of an external non zero potential, $V(\omega, q)$ we can define its relative permittivity as $\epsilon_r(\omega, q)$, where

$$\begin{aligned}
\epsilon_r(\omega, 0) &= (1 - \omega_P^2/\omega^2) && (5.21)\\
\epsilon_r(0, q) &= (1 + q_{TF}^2/q^2) && (5.22)
\end{aligned}$$

As an example consider the case where the external potential is created by a single point charge of magnitude e, such that $\rho_{ext}(r) = e\delta(r)$. Thus we can write $V(q)$ as;

$$V(q) = \frac{e}{\epsilon_o(q^2 + q_{TF}^2)} \tag{5.23}$$

The FT of delta function at r = 0 is a constant, 1. Transforming back to real space via inverse Fourier transform we can show that $V(r) \sim \frac{e}{r}e^{-\vec{q}_{TF}.\vec{r}}$. This is also known as the **screened Coulomb potential** or **Yukawa potential**. Evidently, $1/q_{TF}$ gives a typical (decay) length scale of the potential. {Hint: To work out the above you will need to compute a simple contour integral. The only pole that will contribute to the integral will be at $z = +iq_{TF}$}. Further, $\because \nabla V(r) = \rho/\epsilon_o$, calculate the $\rho(r)$ corresponding to the above potential and show that it is the same as that of a positive charge $q = |e|$ at $r = 0$, surrounded by a screening charge density of opposite sign.

What is the typical magnitude of q_{TF}? We know from above that $1/q_{TF}$ gives the decay length scale of the negative charge distribution surrounding the positive point charge. Now, $q_{TF} = \sqrt{\frac{e^2 D(E_F)}{\epsilon_o}}$;

$$\begin{aligned} q_{TF}^2 &= \frac{e^2 D(E_F)}{\epsilon_o} && (5.24)\\ &= \frac{e^2}{\epsilon_o 2\pi^2}(\frac{2m}{\hbar^2})^{3/2} E_F^{1/2} && (5.25)\\ E_F &= \frac{\hbar^2}{2m}k_F^2 && (5.26)\\ \Rightarrow q_{TF}^2 &= \frac{e^2}{\epsilon_o 2\pi^2}(\frac{2m}{\hbar^2})k_F && (5.27)\\ & && (5.28) \end{aligned}$$

This gives the magnitude of $q_{TF} \sim 0.55$ Å for a typical electron density of $10^{22}/cc$. Which means that any effect of localised charge distributions are effectively screened by free electrons in metals within a length scale of half angstroms. Thus the absence of any band bending in the metal side of Schottky junctions.

5.3 Problems

1. Revelant Problems in file Practice Problems.pdf
2. Derive expressions for the relative permittivity, $\epsilon_r(q)$ for 2D and 1D electron gas systems.
3. Due to the $D(E_f)$ factor, the Thomas-Fermi wavevector depends on the dimensionality of the system. Show that in 3D it is given by the equation below. (a_B) is the Bohr radius.

$$q_{TF}^2 = \frac{4k_F}{\pi a_B} \tag{5.29}$$

4. What is the expression for q_{TF}^2 for 2D systems?

References:

1. Chapter 17 Solid state physics, N. W. Ashcroft and D. Mermin
2. Chapter 9 Physics of low-dimesnional semiconductors J.H. Davies

6

Semiconductor Optoelectronic Devices

Semiconductor optoelectronic devices can be classified into two broad categories dependent on the function they perform (i) light to electric voltage or current conversion - photovoltaics and (ii) electric voltage or current to light conversion - electroluminescence. Electroluminescence (EL) is phenomenon where a material/device emits light due to application of electric voltage or passing electric current. EL is quite differnt from incandescence or black body radiation where emission results from the heat content of the body. Importantly, EL have significantly low power consumption compared to other lighting technologies, such as tungsten filament lamps, or neon or even fluorescent lamps.

6.1 Light Emitting Diodes

Light-emitting diodes (LEDs) are the most ubiquitous example of semiconductor devices showing EL. EL primarily results from of radiative recombination of injected electrons and holes in a semiconductor device. Prior to recombination, electrons and holes may be separated either by doping the material to form a *pn* junction or through excitation by impact of high-energy electrons accelerated by a strong electric field (as with the phosphors in electroluminescent displays). In the former, the recombination process leading to light emission can be of intrinsic in nature i.e. band to band de-excitation and recombination or extrinsic i.e. recombination of impurity bound excitons. LEDs may

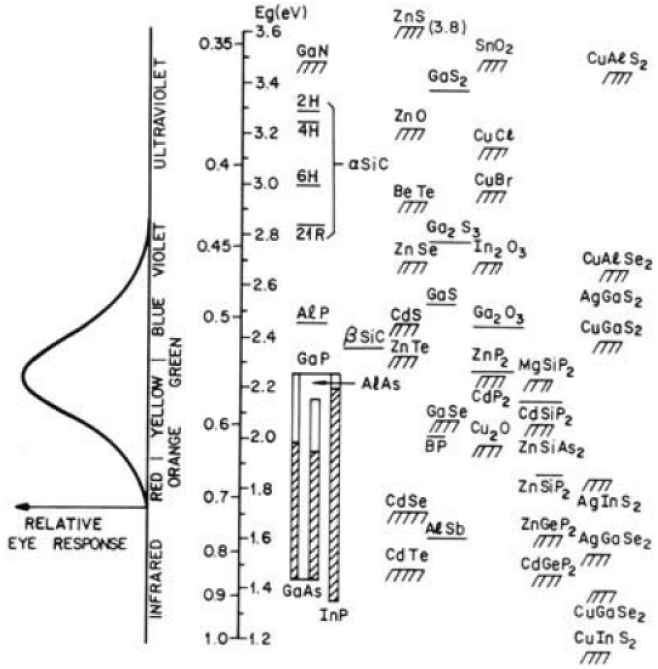


Figure 6.1: Spectral ranges of various materials [Grundmann]

be classified based on their emission window and relevant applications as shown in the table 6.1.

Figure 6.1 plots the spectral range of emission from various materials between the UV to the IR.

Table 6.1: LEDs Emission Ranges, Applications and Materials

Emission	Applications	Materials
infrared ($\lambda > 800$ nm)	remote controls, optocouplers	GaAs/AlGaAs
visible 400 nm $\leq \lambda \leq$ 800 nm	white and coloured LEDs, lighting	
red-yellow		GaAsP/GaAs/AlInGaP/
yellow-green		GaP:N
green-blue		SiC/GaN/InGaN
violet		GaN
ultraviolet ($\lambda < 400$ nm)	pump for white LEDs, biotechnology, sanitisation	AlGaN

The total quantum efficiency (QE), η is the number of photons emitted from the device per injected electron hole pair. It is given by the product of the internal quantum efficiency η_{int} and the external QE η_{ext}, i.e.

$$\eta = \eta_{int} \times \eta_{ext} \tag{6.1}$$

The internal QE is the number of photons generated inside the semiconductor per injected electron hole pair and is dependent on material quality, defect density and trap concentration. The external QE is the number of photons leaving the device divided by the total number of generated photons and is dependent on the geometry of the LED. Due to the large index of refraction of semiconductors ($n_s \sim 2.5$ - 3.5), light can leave the semiconductor only under a small angle to the surface normal due to total internal reflection. Against air ($n \sim$1), the critical angle is $\theta_c = \sin^{?1}(1/n_s)$, which is 16deg for GaAs and 17deg for GaP. Additionally, a portion of the photons that do not suffer total reflection is reflected back from the surface with reflectivity $R = \{(n_s - 1)/(n_s + 1)\}^2$ [1], which is about 30% or GaAs/air interface. These two parameters, i.e. θ_c and R decide the external QE, which for GaAs, is $0.7 \times 4\% \simeq 2.7\%$. Thus, typically only a small fraction of generated photons can leave the device and contribute to the light emission.

6.2 Semiconductor Laser

Recall the fundamentals of a two-level atomic or molecular laser. Applying the Einstein relations that we encountered before we can write the following set of equations for the two level system and the related emission;

$$\hbar\omega = \epsilon_1 - \epsilon_0 \tag{6.2}$$

$$\frac{dN_1}{dt} = -A_{10}N_1 - B_{10}D(\omega)N_1 \tag{6.3}$$

$$\frac{dN_0}{dt} = -B_{01}D(\omega)N_0 \tag{6.4}$$

where $D(\omega)$ is the photon density of states and N_1 and N_0 are the occupancy of the energy states. At equillibrium we can write $B_{01}g_0 = B_{10}g_1$ and that $B_{10} = c^3\pi A_{10}/2\omega^2$. The main difference between a electroluminescent diode and a laser is that the latter emits coherent radiation. Hence, a Fabry-Perot resonator is an essential part of a laser. In a crystal, two cleaved parallel faces serve for such a resonator. If we denote by R as the reflectivity of a cleaved surface assuming $L \gg \lambda$, the feedback condition is given by;

$$R \times exp \int_0^L (g - \alpha)dz \tag{6.5}$$

where g is the gain of the lasing media, $\approx (n_1 - n_0)$ and α is the absorption coefficient of the media, arising due to nonradiative decay in the system. In general the gain coefficient obviously depends on the photon flux which is also function of position z. For simplicity we neglect this dependence and the integration yields

$$g = \frac{1}{L}ln(\frac{1}{R}) + \alpha \tag{6.6}$$

The above analysis allows us to model the various components of lasing that is possible from a semiconductor diode. But the far more interesting application comes from the superlattices.

[1] valid for vertical incidence

6.3 Superlattice Lasers

Consider as GaAs/AlGaAs superlattice contained 60 periods of alternating GaAs and Gao.65Alo.35As layers, terminated by 60 nm of Gao.65Alo.35As on each side. These undoped structures constitute the intrinsic regions of $p^+ - i - n^+$ diodes grown by molecular-beam epitaxy. The well thicknesses were about 4 nm and the barrier thicknesses about 3 nm. An uniform E field in the superlattice was produced by applying an external voltage on the p^+ region (relative to the then + side). Excitation of electrons i.e. population inversion, may be achieved by lowpower radiation matching the fundamental gap of GaAs.

6.4 Quantum Cascade Lasers

The quantum cascade laser (QCL) is a special kind of semiconductor laser, usually emitting midIR light. These lasers operate on intersubband electronic transitions of a semiconductor structure. Figure 6.2 shows what happens to an electron injected into the gain region: in each period of the structure, it undergoes a first transition (black arrow) from sublevel 3 to 2 of a quantum well (which is the laser transition on which stimulated emission occurs), then a non-radiative transition (slanted arrow) to the lowest sublevel, before tunneling (dashed arrow) into the upper level of the next quantum well. By using several tens or even 100 quantum wells in a series or cascade), a higher optical gain and multiple photons per electron are obtained at the expense of a higher required electrical voltage. The operation voltage can easily be of the order of 10 V, whereas few volts are sufficient for ordinary laser diodes. As the transition energies are defined not by fixed material properties but rather by design

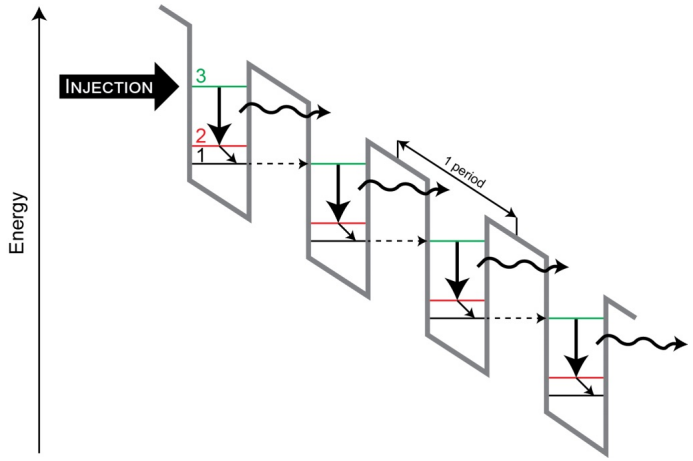


Figure 6.2: Schematic of the gain region of a QCL. The diagram shows the electron energy versus position in the structure, which contains four quantum wells. The overall downward trend of energy towards the right-hand side is caused by an applied dc electric field. In reality, each gain region must be divided into an active region and an injector. [rp-photonics.com]

parameters (particularly by well and barrier thickness values of quantum wells), quantum cascade lasers can be designed for operating wavelengths ranging from a few microns to well above 10 ?m, or even in the terahertz region. The quantum well structure is embedded in a waveguide, and the laser resonator is mostly of DBR or DFB type. There are also external-cavity lasers, where a wavelength tuning element such as a diffraction grating is part of the resonator. Whereas continuously operating room-temperature devices are normally limited to moderate output power levels in the milliwatt region (although more than a watt is possible), multiple watts are easily possible with liquid-nitrogen cooling. Even at room temperature, watt-level peak powers are possible when using short pump pulses. The power conversion efficiency of QCL is typically of the order of a few tens of percent. Recently, however, devices with efficiencies around 50% have been demonstrated although only for cryogenic operation conditions. Most QCLs emit mid-infrared light. However, quantum cascade lasers can also be made for generating terahertz radiation. Such devices constitute very compact and simple sources of terahertz radiation. Recently, even room temperature terahertz generation has been

achieved via internal difference frequency generation. Perhaps the most important applications for quantum cascade lasers will be in the area of laser absorption spectroscopy of trace gases, e.g. for detecting very small concentrations of pollutants in air. In addition to the suitable wavelength range, QCLs usually feature a relatively narrow linewidth and good wavelength tunability, making them very suitable for such applications.

6.5 Problems

1. Problems X, X, X in file Practice Problems.pdf

2. Derive expressions for the relative permittivity, $\epsilon_r(q)$ for 2D and 1D electron gas systems.

References:

1. Chapter 17 Solid state physics, N. W. Ashcroft and D. Mermin
2. Chapter 9 Physics of low-dimesnional semiconductors J.H. Davies