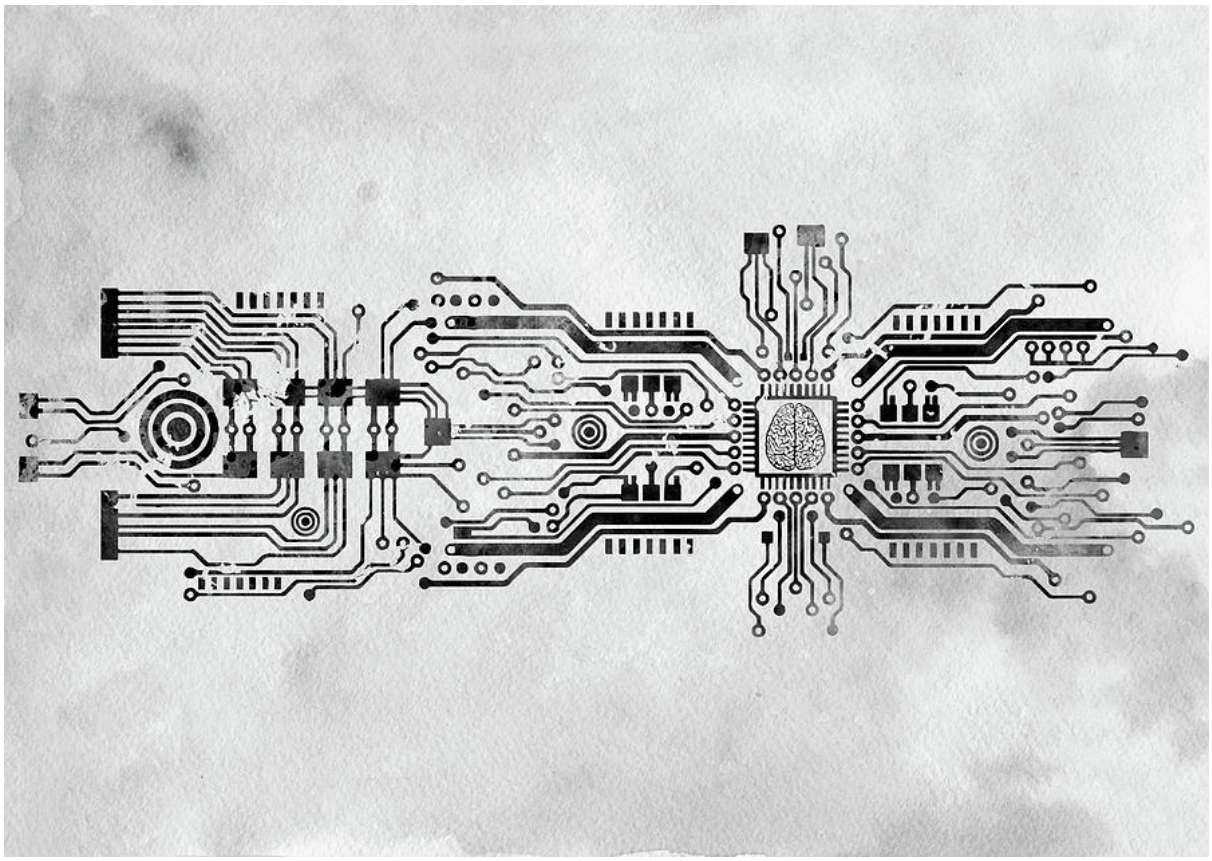
DEVICE PHYSICS

Debasish Panda 21D070021

February 18, 2024

Mentor - Jay Sonawane

Department of Electrical Engineering
Indian Institute of Technology, Bombay



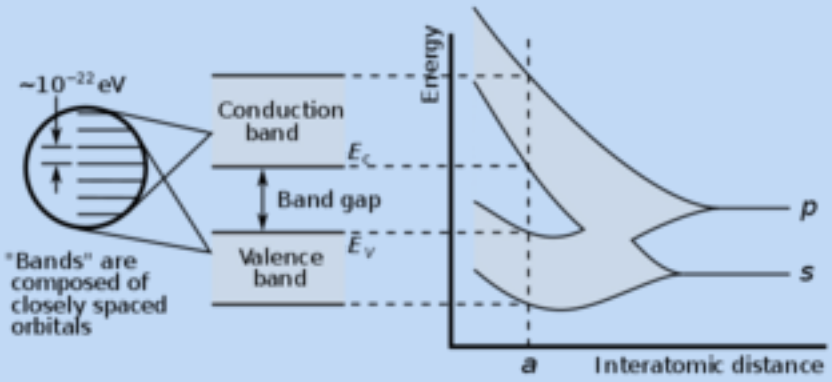
Contents

1 Charge Carriers in Semiconductors

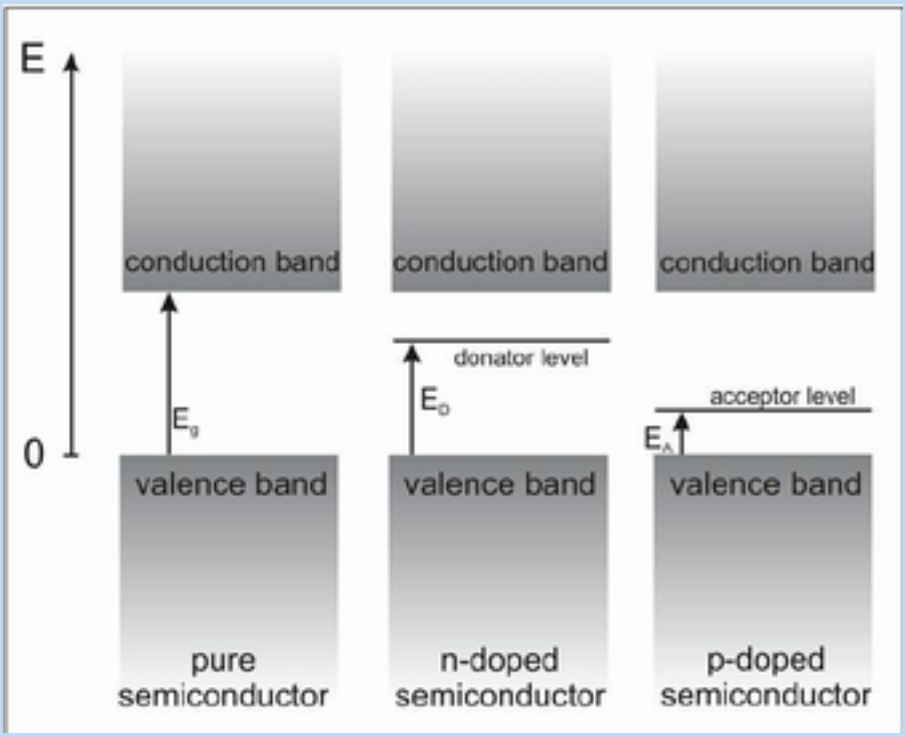
1.1 Energy bands

In order to fully understand the physics behind semiconductors, we need to get a clear understanding of the manner in which electrons are distributed within its crystal. Energy band diagrams serve this purpose by acting as an approximate visual guide to show how electrons are distributed in the various energy states available.

Consider two carbon atoms($1s^2 2s^2 2p^2$)initially separated by a large distance from each other. In such a state, the electrons in the atoms occupy discrete,well-defined energy levels. However, as the two atoms come closer, this view no longer holds true. At a finite distance(comparable to atomic sizes) between the two atoms, their wavefunctions overlap strongly and the discrete energy states change into a near-continuum energy state distribution. As more and more atoms get involved in the crystal, the overlapping of their wavefunctions ultimately leads to the formation of 'energy bands', that are composed of numerous discrete energy levels but appear as a continuous band due to the large number of atoms involved.



Energy band structure in a crystal of carbon atoms



Energy band diagram in semiconductors

When electrons are excited to the empty energy level, they can freely move about under external influence, and hence, conduct electricity. This band is therefore known as the conduction band. The energy band gap for semiconductors is intermediate between that of conductors and insulators-which explains their nomenclature.

The energy band gaps for some of the common semiconductors are: Si-1.1 eV, Ge-0.7 eV, GaAs-1.42 eV, etc. At room temperature of around 300 K, the kinetic energy of an electron is of the order of k_BT, where k_B represents the Boltzmann's constant. On calculating, this value turns out to be around 26 meV, far too less for an electron to be excited to the conduction band from the valence band as per the rules of classical physics. However, on the atomic scales, where the rules of quantum mechanics dominate, there is still a non-zero probability for the electron to be excited from the valence band to the conduction band.

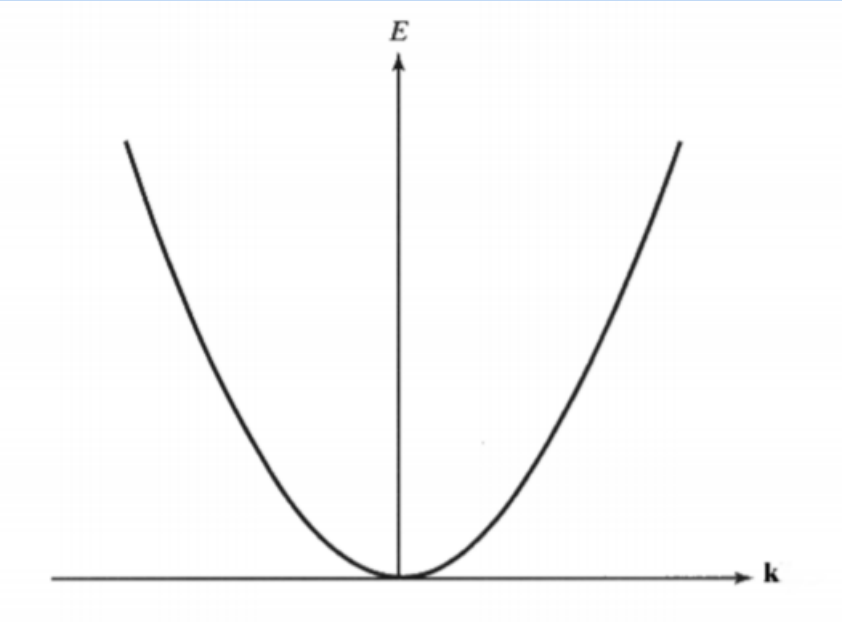
1.1.1 Physical model of Semiconductor lattice

Even though it may theoretically seem quite difficult to analyse the motion of an electron in the semiconductor lattice using quantum mechanics, it actually turns out that relatively simple physical models can appropriately describe the electron's motion to high degree of accuracy and serve nearly all our practical purposes.

Let us consider a free electron moving along the x-axis. The motion of the electron can be described by an E-k diagram which plots energy of the electron as a function of its wave-vector, **k** (The relation is established using De-broglie's hypothesis).

$$E = \hbar^2 k^2 / 2m_e$$

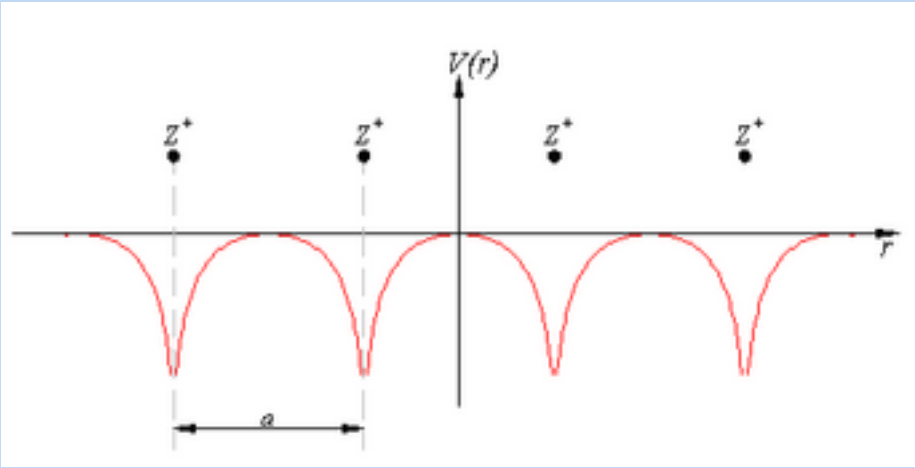
where m_e represents the rest mass of electron.



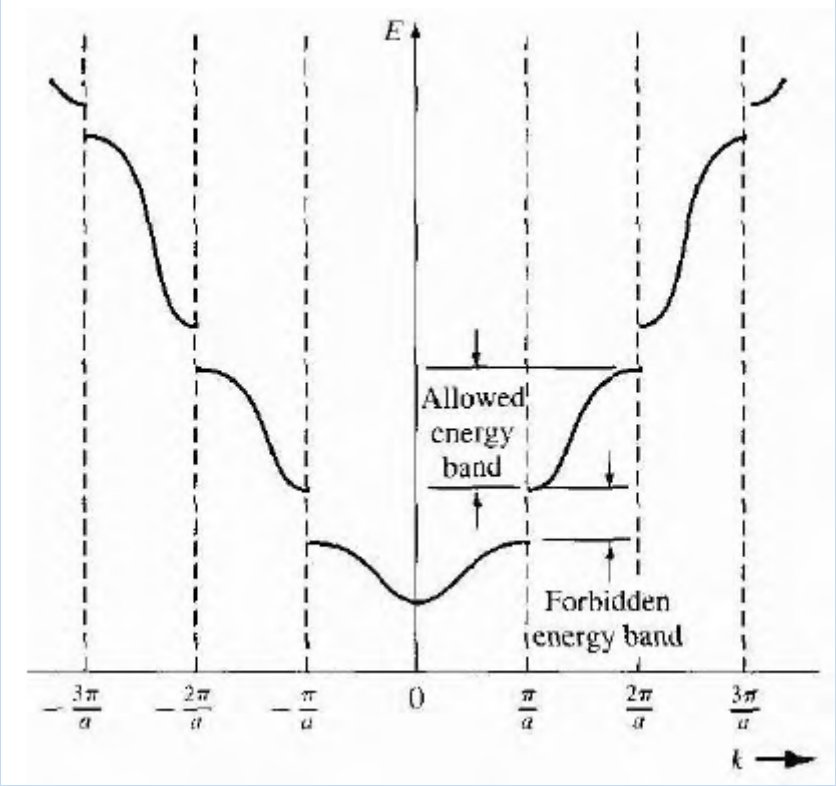
E-k diagram for a free electron

However, as we can see, inside a semiconductor lattice, the assumption of a free electron no longer holds true. The electron in a crystalline lattice can be thought of as experiencing a periodic potential because of the regular, organised arrangement of atoms. As of now, we will consider the 1-D case for simplicity and the 3-D case follows as a natural extension of our analysis.

The 1-D crystalline lattice can be imagined of being composed of a series of Dirac-delta potential wells, each centered at one of the atoms in the lattice and at a constant distance from each other, known as the lattice constant. Now, having established a feasible model for the lattice, we can solve for the wavefunction of electron inside this 1-D lattice using Bloch's Theorem. This model of particle trapped inside a one-dimensional lattice is also known as the Kronig-Penney model.



Periodic potential inside the crystal lattice



E-k diagram for an electron inside the lattice

As we can observe, there are several discontinuities in the graph at quantized values of **k**. However, for most of the practical semiconductor devices, the majority of electrons lie in the range of $-\pi/a < k < \pi/a$, otherwise known as the first Brillouin zone. The energy of an electron near the bottom of the Brillouin region can be expanded as a Taylor series:

$$E = E_C + [(\frac{dE}{dk})|_{k=0}]k + [(\frac{1}{2}\frac{d^2E}{dk^2}|_{k=0})]k^2 + HOTs$$

where HOTs refers to higher-order terms. The derivatives are evaluated at $\mathbf{k} = 0$, which gives us $\frac{dE}{dk} = 0$. Neglecting the HOTs gives us:

$$E = E_C + [(\tfrac{1}{2}\tfrac{d^2E}{dk^2}|_{k=0})]k^2$$

Note that we can justify the fact that we neglected the higher order terms as long as we stay close to the region wherein the E-k graph is parabolic in nature. In fact, in general, even $\frac{dE}{dk}$ also need not be zero. Thus, the effective mass(m^*) of the electron in this parabolic region can be expressed as:

$$m^* = \hbar^2 / \tfrac{\partial^2 E}{\partial k^2}$$

The need for this theoretical construct of *'effective mass'* arises since we would wish to apply the usual equations of electrodynamics to the electrons inside the semiconductor, which are in fact, quantum mechanical particles. In doing so, we encapsulate the quantum mechanical properties in the effective mass so that the electrons and holes can be treated as "quasi-free" carriers in most computations.

1.1.2 Concept of holes

When an electron is excited from the valence band to the conduction band, it leaves behind an "empty space" in the energy state it previously occupied. In such a situation, when an external electric field is applied, the electron adjacent to the hole moves forward to fill the empty slot. This process continues and thus creates an electric current due to the net motion of the electrons. Physically, this is equivalent to a positive charge of the same magnitude as the electronic charge, constituting a current. Such a quasi-particle has been assigned the name 'holes' and constitute another type of carrier found in a semiconductor.

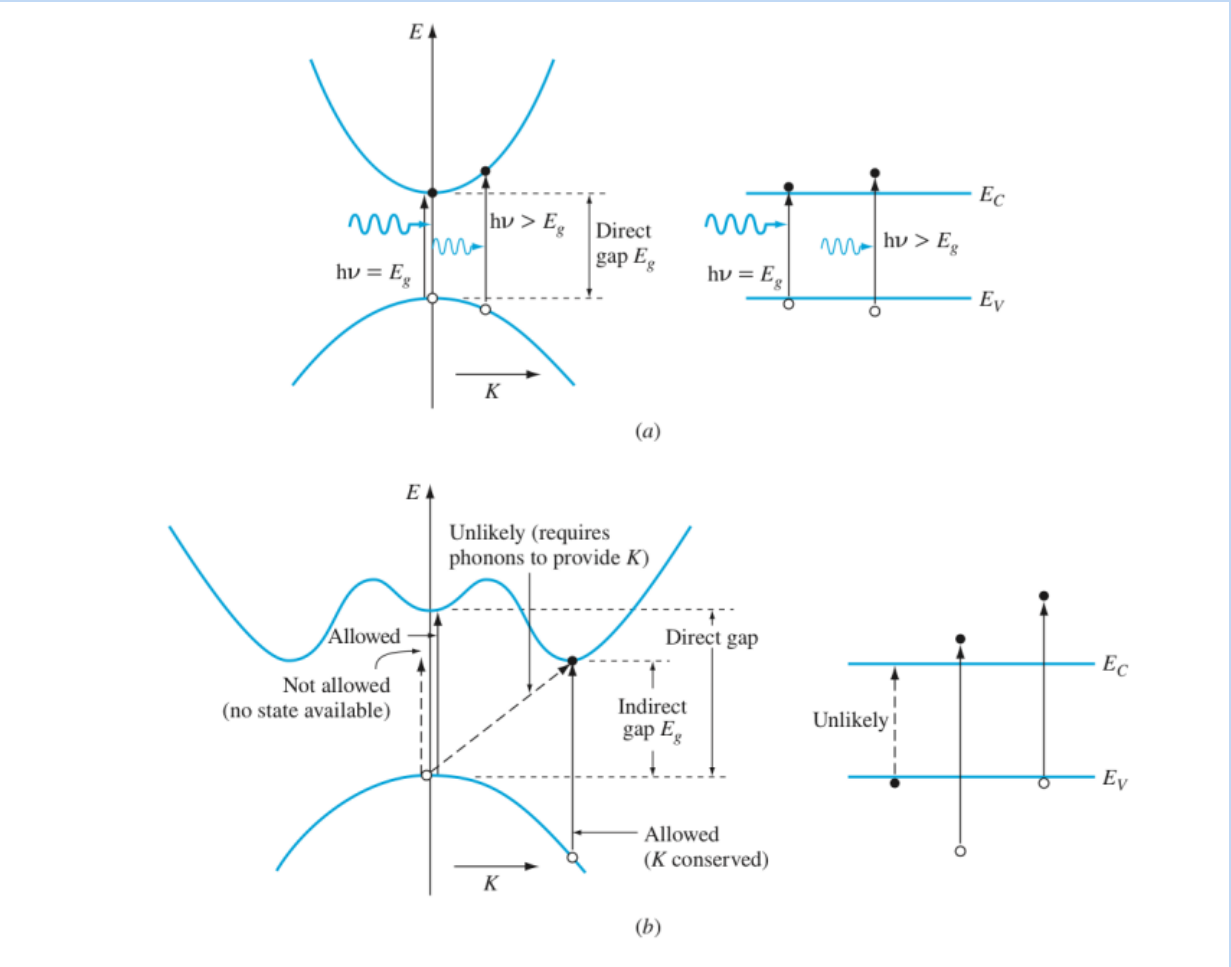
1.2 Direct and Indirect band-gap semiconductors

In general, the E-k diagram of an electron inside the semiconductor is a far more complex surface, which should be visualised in three dimensions- the reason being that $\mathbf{k}$ in general, represents the 3-D wave vector of the electron. Since the periodicity of most lattices is different in various directions, the E-k diagram must be plotted for the various crystal directions.

In certain semiconductors, for e.g. GaAs, the minima in the conduction band and the maxima in the valence band occur both for the same value of $\mathbf{k}$, namely $\mathbf{k} = 0$. On the other hand, in semiconductors such as Si, the valence band maximum is at a different value of $\mathbf{k}$ than its conduction band minimum. Thus, an electronic transition from the

conduction band minima to the valence band maxima in the first case occurs without any change in **k** value, whereas the second class of semiconductors do require some change in **k** for the transition to be possible. Thus there are two classes of semiconductor energy bands; *direct* and *indirect* band-gap semiconductors.

The distinction between direct and indirect band-gap materials is evident from their respective applications: in a direct band-gap material like GaAs, the electron can directly fall from the conduction band minima to the valence band maxima, giving off the energy difference as photon energy. Hence, such materials are used in LEDs and devices requiring light output. On the other hand, in an indirect band-gap material like Si, the electron must undergo a change in **k**, and consequently, a change in its momentum and energy. In an indirect transition, part of the energy is eventually given up as heat to the lattice rather than being converted into photonic energy. So, such materials are unsuitable for light-emitting purposes.



Direct and indirect band-gap transitions

1.3 Carrier Statistics

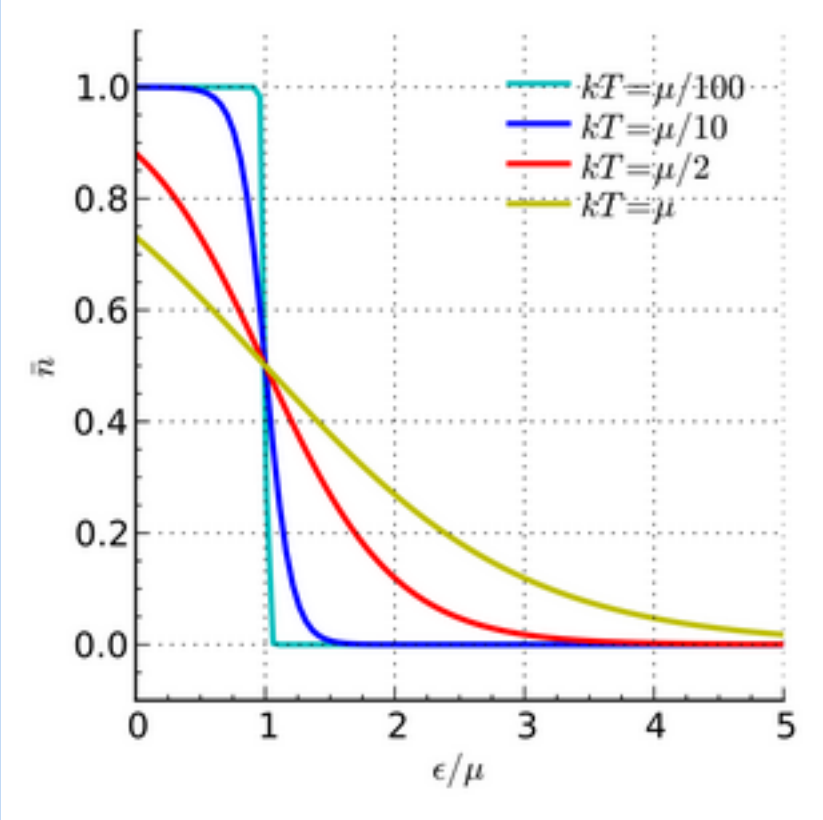
Electrons and holes are fermions- meaning they have half-odd-integer spin and follow Pauli's exclusion principle. The statistical rules governing the dynamics of an ensemble of such fermions are known as the Fermi-Dirac statistics.

1.3.1 Fermi-Dirac Statistics

The probability distribution governing Fermi-Dirac statistics is derived by using the fact that the multiplicity of the system should be maximized. On doing this, we obtain the probability distribution(as a function of energy) as:

$$\mathrm{f}(E_i) = (1 + exp(E_i - E_F/k_BT))^{-1}$$

where E_F represents the Fermi energy level.



Probability distribution graph of F-D statistics

In order to estimate the number of particles occupying a given energy state, we also need to estimate the density of states function, g(E) which represents the degeneracy

of a particular energy level. The density of states functions for the C.B. and V.B. of semiconductors is as follows:

$$\text{g(E)} = \tfrac{4\pi}{h^3}(2m_e^*)^{3/2}\sqrt{E - E_c}, \text{ for C.B.}$$

$$\text{g(E)} = \tfrac{4\pi}{h^3}(2m_h^*)^{3/2}\sqrt{E_v - E}, \text{ for V.B.}$$

where m_e^* and m_h^* represent the effective masses of the electron and hole, respectively.

Spoken in simpler terms, *degeneracy* refers to the number of available states that a particular energy level (E to E+dE) has to offer; while the *probability distribution function* refers to the fraction of states occupied out of the degenerate states at a particular energy level (again, E to E+dE). Note that we cannot explicitly (it being physically insensible) state that the number of states at the energy level E is such and such- thus the need for the energy gap, dE.

1.3.2 Carrier concentrations

Having obtained the density of states function and probability distribution function, we can easily obtain the total number of electrons and holes in the C.B. and V.B., respectively. The corresponding equations are given as:

$$n_e = \int_{E_c}^{\infty} g(E)f(E)\,dE$$

$$n_h = \int_{-\infty}^{E_v} g(E)f(E)\,dE$$

As a side note, the integrals $\int_{E_c}^{\infty} g(E)\,dE$ and $\int_{-\infty}^{E_v} g(E)\,dE$ gives us the total number of available energy states in the conduction band and the valence band, respectively.

These integrals are of the form of Fermi-Dirac integrals of 1/2 order and are expressed as:

$$n_e = N_c F_{1/2}(\tfrac{E_F - E_c}{k_B T})$$

$$n_h = N_v F_{1/2}(\tfrac{E_v - E_F}{k_B T})$$

where N_c and N_v are known as the effective conduction band density of states and $F_{1/2}$(x) represents the Fermi integral of order 1/2. Their formulae are given as follows:

$$N_c = 2(\frac{2\pi m_e^* k_B T}{h^2})^{3/2}$$

$$N_v = 2(\frac{2\pi m_h^* k_B T}{h^2})^{3/2}$$

Physically, these quantities represent the number of electrons just at the edge of the valence band or conduction band.

At equilibrium, if no external impurities have been added to the semiconductor lattice, it is known as an intrinsic semiconductor and in such cases, n_e and n_h are both equal to n_i, termed as the intrinsic carrier concentration. In a semiconductor, n_i represents the minimum number of charge carriers that are present in it without any sort of external influence.

In order to get an estimate of the value of n_i, certain approximation methods can be used to evaluate the Fermi-Dirac integral associated with n_e and n_h. If $E_c - E_F \geq 3k_BT$ or $E_F - Ev \geq 3k_BT$, we can approximate the integrals as:

$$F_{1/2}(\frac{E_F - E_c}{k_B T}) \approx exp(\frac{E_F - E_c}{k_B T})$$

$$F_{1/2}(\frac{E_v - E_F}{k_B T}) \approx exp(\frac{E_v - E_F}{k_B T})$$

For an intrinsic semiconductor, the value of E_G is considerably larger than $3k_BT$. So the above approximation holds true for quite a large number of commonly used semiconductor materials. Hence, to find n_i, approximate the expression as:

$$n_e \approx N_c exp(\frac{E_F - E_c}{k_B T})$$

$$n_h \approx N_v exp(\frac{E_v - E_F}{k_B T})$$

On multiplying the two expressions, one can observe that:

$$np = n_i^2 = N_c N_v exp(\frac{-E_G}{k_B T})$$

This is one of the most important relations governing carrier concentrations in semiconductor physics, otherwise known as the **Law of mass action**.

There might be a fleeting misconception in the reader's mind that the Fermi energy level, E_F is equal to $\frac{E_c + E_v}{2}$ for intrinsic semiconductors. However, the electron and hole effective masses need not necessarily be equal in the material due to which there is (usually) a small energy difference between the two energy levels.

$$n_o = p_o = n_i = N_c exp(\frac{E_F - E_c}{k_B T}) = N_v exp(\frac{E_v - E_F}{k_B T})$$

Also,

$$N_c/N_v = e^{\frac{E_c + E_v - 2E_F}{k_B T}}$$

Solving for E_F gives us

$$E_F = \frac{E_c + E_v}{2} + \frac{3}{4} k_B T ln(\frac{m_h^*}{m_e^*})$$

This means that E_F is offset from midgap by the term $\frac{3}{4} k_B T ln(\frac{m_h^*}{m_e^*})$, which is usually pretty small.

1.3.3 Kinetic energy of carriers

The potential energy of an electron (or a hole) in the conduction band (or in the valence band) is constant and equal to E_c (or E_v). Thus, the kinetic energy of the electron at energy level E (for simplicity's sake, we will consider only one kind of charge carrier and extend the result quite naturally) will be given by $E_K = E - E_c$. We can calculate the average kinetic energy of the electron and its average speed through the expectation values of E_K and $\sqrt{E - E_c}$, respectively. The average K.E. and the average speed are given by:

$$< E_K >= \frac{3}{2} k_B T$$

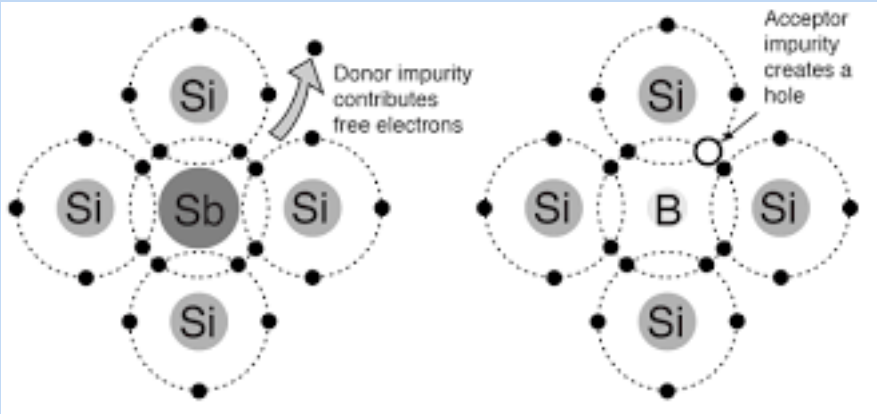
$$< v >= \sqrt{\frac{8 k_B T}{\pi m_e^*}}$$

Notice the strong analogy with the equivalent quantities in the gaseous state, with the only difference being that the inertial mass has been replaced by effective mass, m^*.

1.4 Doping of Semiconductors

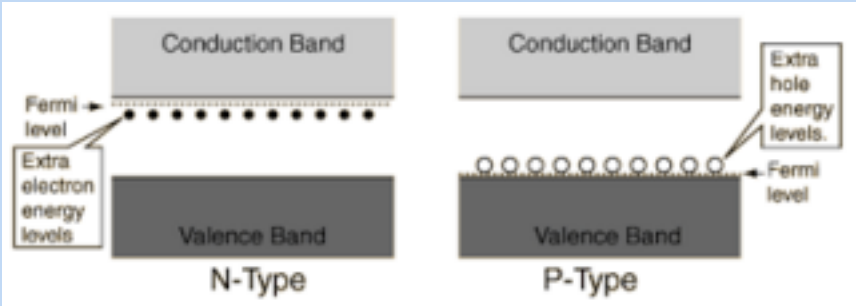
As we know, one of the most striking features of semiconductors is their ability to change their conductivities over a wide range of orders of magnitudes. This is achieved by doping of semiconductors. Doping essentially involves the addition of extraneous impurities into the semiconductor lattice which infuse the lattice with a particular kind of charge carrier and thus, decrease the concentration of the other.

Let us consider an example of n-type doping where phosphorus(P) atoms are added to a crystalline lattice of silicon(Si) atoms. Since P has one extra electron in its valence band as compared to Si, this electron cannot be accommodated under covalent bonding phenomena. Hence, the effective nuclear attraction experienced by the extra electron decreases and it now has a high probability of being excited into the conduction band through thermal excitation. This entire process effectively creates a positively charged, immobile ion centre at the site of the P-atom. The same goes for an electron-deficient material such as boron(B) added to the semiconductor lattice- except that this generates a hole instead of an electron. Through doping, one can reduce the concentration of minority charge carriers drastically. Consider a sample of Si crystal which is n-type doped using P atoms such that $N_D = 10^{17}/cm^3$ and $n_i = 1.50 \times 10^{10}/cm^3$. Here, the minority carrier concentration is given by $p = n_i^2/N_D$ (since $N_D >> n_i$, we can approximate majority carrier conc. as N_D) which is equal to $2.25 \times 10^3/cm^3$, about 10^{14} times smaller than the majority carrier concentration. However, despite this, there is a theoretical limit on the maximum levels of doping in a semiconductor crystal since higher levels of doping can lead to higher rates of recombination of charge carriers, thereby neutralising the high carrier concentrations.



n-type and p-type doped semiconductor crystal

In order to understand doping effects, we must also analyze the effects of external impurities on the band diagram of the semiconductor. It turns out that on doping, the Fermi level inside the semiconductor changes, which is a consequence of the deviation of carrier concentrations from the intrinsic carrier concentration, n_i. In an n-type doped material, the E_F gets closer to the conduction band whereas in a p-type doped material, E_F gets closer to the valence band. As a result, from the formulae for n_e and n_h, we can observe that $n_e > n_h$ for an n-type material and vice-versa for a p-type material as we expected physically too.



n-type and p-type doped semiconductor crystal

The abovementioned diagram illustrates the fact that on doping the material, we get additional dopant energy levels(E_D or E_A), which are quite close to the C.B. or V.B., respectively. Thus, these dopant energy bands can quite freely donate charge carriers even at room temperature, since the energy difference is so low as compared to k_BT.

The exact relation between n, p, N_A, N_D(without the assumption that $N_D >> n_e$ or $N_A >> n_h$) can be established using the concept of charge neutrality and the relation $n \cdot p = n_i^2$. Thus, the equation goes as follows:

$$n + N_A^- = p + N_D^+$$

Assuming complete ionization of donor and/or acceptor atoms, $N_A^- = N_A$ and $N_D^+ = N_D$. Now, substituting $n \cdot p = n_i^2$ in the equation,

$$\frac{n_i^2}{n} + N_D = n + N_A$$

$$n^2 + n(N_A - N_D) - n_i^2 = 0$$

On solving this quadratic equation, we obtain:

$$n = \frac{-(N_A - N_D) + \sqrt{(N_A - N_D)^2 + 4n_i^2}}{2}$$

Similarly,

$$p = \frac{-(N_D - N_A) + \sqrt{(N_D - N_A)^2 + 4n_i^2}}{2}$$

Earlier we have seen that under certain conditions, the Fermi-Dirac integral involved in n_i can be simplified. A relatively advanced approximation method also covers the cases when $E_c - E_F \leq 3k_BT$ or $E_F - E_v \leq 3k_BT$, also known as the Joyce-Dixon approximation. According to this,

$$ln(\frac{n}{N_c}) + \frac{1}{\sqrt{8}}\frac{n}{N_c} = \frac{E_F - E_c}{k_B T}$$

$$ln(\frac{p}{N_v}) + \frac{1}{\sqrt{8}}\frac{p}{N_v} = \frac{E_v - E_F}{k_B T}$$

A side note: From fundamental physical principles, we know that,

$$\sigma = nq\mu_n + pq\mu_p$$

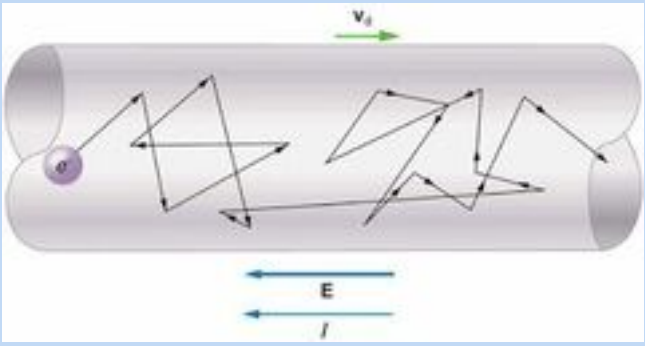
where σ is the conductivity of the material.

2 Drift and Diffusion processes of Carriers

The two major processes through which an electric current flows inside the semiconductor is by drift and diffusion processes. While the process of drift is fundamentally similar in both conductors and semiconductors, diffusion is something that is unique to semiconductors.

2.1 Drift and mobility of carriers

Electrons and holes that are undergoing random thermal motion inside the semiconductor cannot conduct electricity since their net displacement over a long interval of time is zero. Under application of an external electric field, the random thermal walk of electrons is superimposed with a net displacement due to the electric field's influence and thus, can constitute a current.



Electron undergoing collisions with atoms in the lattice

During the random thermal walk, electrons interact with the lattice in two manners- i) through collision with vibrating atoms in the lattice, ii) through electrostatic interaction with ionized impurities. The 'collisions' here merely refer to the interaction of the electron with a potential field and not to any form of physical collision taking place. Vibrating atoms interact with the electrons and holes in the lattice through particles known as 'phonons'.

Using the relation between drift velocity and applied electric field, we know that:

$$v = (\frac{q\tau}{m^*})E$$

where τ represents the average time between successive collision events and m^* represents effective mass of the electron. The quantity $\frac{q\tau}{m^*}$ represents 'mobility' of the carrier inside the semiconductor.

Despite the fact that drift velocity and electric field are directly proportional to each other, at very high field values the drift velocity saturates to a maximum possible velocity, v_{sat}. This is because with increase in field values, the frequency of collisions with lattice increases manifolds and this acts as a limiting factor for drift velocity. Typical mobility values for semiconductors are usually in the range of 200-400 $cm^2V^{-1}s^{-1}$. However, novel materials like graphene can have very high values of mobility, typically in the range of 50,000-100,000 $cm^2V^{-1}s^{-1}$.

The two factors affecting mobility- collisions with the lattice and coulombic interaction with ionized centres behave quite differently as the temperature varies.

(i) Atomic vibrations: As temperature increases, electrons collide more frequently with atoms, hence their mobility due to atomic vibrations decreases.

(ii) Ionized atoms: As temperature increases, the thermal velocity of the electrons increases and they find it easier to overcome the Coulombic barrier of ionized dopant atoms. Hence, the mobility due to ionized impurities increases with increasing temperature.

The net mobility (μ_{tot}), mobility due to phonon's interaction (μ_{phonon}) and mobility due to ionized impurities ($\mu_{I.Imp}$) are related as:

$$\mu_{tot}^{-1} = \mu_{phonon}^{-1} + \mu_{I.Imp}^{-1}$$

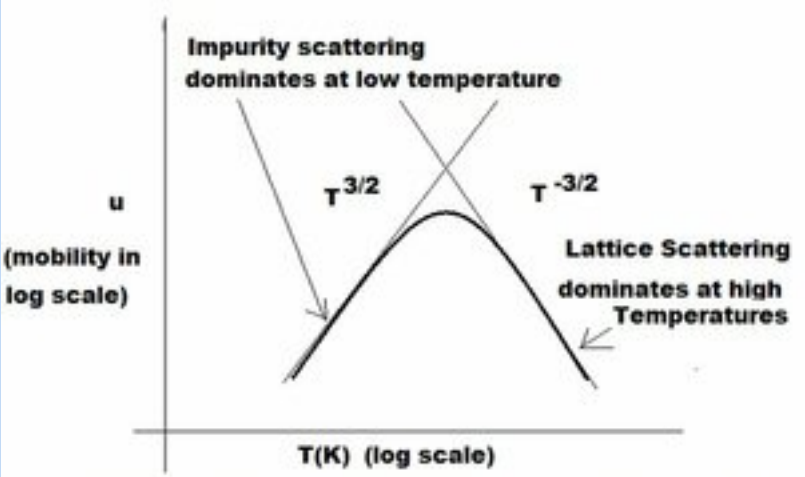


Figure 2.2.38.Temperature dependence of mobility in Si.

Variation of mobility with temperature

The drift process also depends upon the electric field inside the material- at high electric fields, the drift velocity tends to saturate.

$$v = \mu E = \frac{\mu_{lf}}{1+E/E_c} E$$

where μ_{lf} is the carrier mobility at low field magnitude, and E_c is known as the 'critical electric field'.

Taking into account both kinds of carriers, the net drift current, J_{drift} can be represented as:

$$J_{drift} = J_n + J_p = qnv_n + qpv_p = (qn\mu_n + qp\mu_p)E$$

2.2 Generation-Recombination processes

Generation refers to the process wherein we create an electron-hole pair (EHP) from excitation of an electron from the valence band to the conduction band. Recombination refers to the process wherein an electron from the conduction band is moved to the valence band, thus annihilating the EHP and releasing energy in the process.

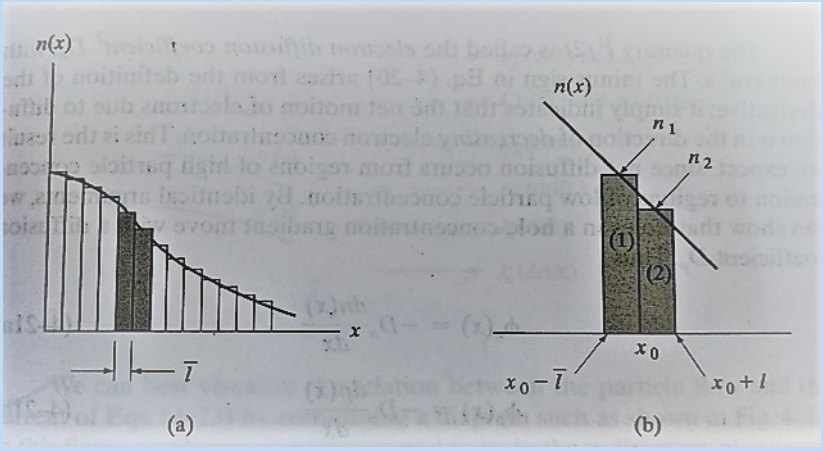
Although the processes of generation-recombination take place intrinsically in a semiconductor, their rates can be increased by external factors such as heat, light,etc. At equilibrium and in the absence of any external stimuli, the rates of recombination and generation are exactly equal to each other, and hence, there is a constant carrier concentration w.r.t time in the semiconductor throughout.

2.2.1 Direct recombination

Semiconductors can be broadly classified under two categories- direct band-gap semiconductors and indirect band-gap semiconductors. The essential difference between the two is that in direct band-gap materials, there are no intermediate energy levels (also known as *traps*) to wherein the electron may transition while being excited from the V.B. to C.B.. Direct band-gap materials are useful for light emission since nearly all the energy released during transition is in the form of photon energy, $h\nu$ (where $h\nu = E_G$). Hence most of the materials used in LEDs such as GaAs, GaN, InGaN, etc. are direct band-gap materials.

2.3 Diffusion of carriers

Diffusion of charge carriers constitutes the diffusion current in semiconductors. It occurs due to the non-uniform, space-varying concentration of charge carriers (which might otherwise be at equilibrium). This process of diffusion is described by Fick's law. In order to understand the process, consider the graph below depicting a non-uniform carrier concentration which varies along the x-direction (uni-dimensional diffusion).



Graph showing variation of carrier concentration with x-coordinates

Consider the graph to be made up of a number of smaller rectangles, each of width $\bar{l}$, where $\bar{l}$ represents the average distance covered by the electrons between subsequent collisions. Now, consider the two small rectangles (1) and (2) shaded out in fig.(a). Let the electron concentrations in (1) and (2) be n_1 and n_2, respectively. Let τ be the relaxation time (average time between successive collisions). Consider the common boundary between the two rectangular bars. The net flux of electron through this cross section is non-zero since the carrier concentration is different on both sides. At any moment of time, about half of the electrons enclosed within the rectangular regions

would be crossing the boundary from each side. Hence, the net flux through the surface can be expressed as:

$$\phi_n(x_o) = \frac{\bar{l}}{\tau}\frac{n_1 - n_2}{2}$$

Using fundamental ideas of calculus, it can be shown that:

$$n_1 - n_2 = \frac{n(x) - n(x+\Delta x)}{\Delta x}\bar{l}$$

Therefore,

$$\phi_n(x_o) = \frac{\bar{l}^2}{2\tau}\frac{n(x) - n(x+\Delta x)}{\Delta x}\bar{l}$$

In the limit of $\Delta x \to 0$, the expression can be further simplified as:

$$\phi_n(x_o) = -\frac{\bar{l}^2}{2\tau}\frac{dn(x)}{dx}$$

The quantity $\frac{\bar{l}^2}{2\tau}$ is also known as the *electron diffusion coefficient*, D_n. From the minus sign, we may infer that the flux of the electrons is from regions of higher concentration to regions of lower concentration. Similar equations can also be written for holes, with minor changes in variables.

$$\phi_n(x_o) = -D_n\frac{dn(x)}{dx}$$

$$\phi_p(x_o) = -D_p\frac{dp(x)}{dx}$$

Now that we have the equations for carrier flux through the boundary, one can easily define the current density at any given point. Therefore,

$$J_n = (-q)(-D_n\frac{dn(x)}{dx}) = qD_n\frac{dn(x)}{dx}$$

$$J_p = (q)(-D_p\frac{dp(x)}{dx}) = -qD_p\frac{dp(x)}{dx}$$

Therefore, the net diffusion current density is given as:

$$J_{diff} = J_n + J_p = qD_n\frac{dn(x)}{dx} - qD_p\frac{dp(x)}{dx}$$

Now that we have derived both the drift as well as diffusion components of the current, we are in a position to write the expression of net current in a semiconductor:

$$J_n = nq\mu_n E + qD_n \frac{dn(x)}{dx}$$
$$J_p = pq\mu_p E - qD_p \frac{dp(x)}{dx}$$

where J_n and J_p are the contributions of electrons and holes, respectively, to the net current density.

2.3.1 Einstein Relation

Einstein relation acts as a link between the drift and diffusion currents in a semiconductor at equilibrium through which no net current is flowing. As we know,

$$n(x) = N_c exp(\frac{E_F - E_c(x)}{k_B T})$$

$$E(x) = \frac{1}{q}(\frac{dE_c(x)}{dx})$$

where $E(x)$ represents the electric field inside the semiconductor due to the slope in the energy levels of C.B. and V.B.. Now, equating the sum of drift and diffusion currents of the individual carriers to zero, we obtain,

$$qD_n \frac{dn(x)}{dx} + q\mu_n n(x)E(x) = 0$$

From the expression for n(x), we obtain,

$$\frac{dn(x)}{dx} = N_c(-\frac{1}{k_B T})exp(\frac{E_F - E_c(x)}{k_B T})\frac{dE_c(x)}{dx} = n(x)(-\frac{1}{k_B T})\frac{dE_c(x)}{dx}$$

Substituting back into the original equation,

$$D_n n(x)(-\frac{1}{k_B T})\frac{dE_c(x)}{dx} + \mu_n n(x)\frac{1}{q}\frac{dE_c(x)}{dx} = 0$$

$$D_n = \frac{k_B T}{q}\mu_n, D_p = \frac{k_B T}{q}\mu_p$$

This relation between diffusion coefficients and mobility suggests that diffusion and drift processes are both interrelated.

Using the Einstein relation, we can more compactly express the electron and hole currents as:

$$J_n = q\mu_n(nE + \frac{k_B T}{q}\frac{dn(x)}{dx})$$
$$J_p = q\mu_p(pE - \frac{k_B T}{q}\frac{dp(x)}{dx})$$

2.3.2 Continuity equation

The continuity equation is one of the most fundamental equations in physics and engineering applications. While continuity equation for flow of currents inside conventional conductors is quite simple, in semiconductors, it is more involved since generation, and recombination processes may produce/annihilate charge carriers inside the semiconductor, so that current density may not be constant over space.

Consider a rectangular block of semiconductor slab and take a thin slice of it of thickness Δx, perpendicular to the current density vector. The rate of increase of conduction band electrons inside this differential volume is given by:

$$\frac{dn}{dt} = \frac{1}{q}\frac{dJ_n(x)}{dx} + (G_n - R_n)$$

where G_n is the electron generation rate and R_n is the electron recombination rate. The differential term expresses the difference in electron flux at either end of the region of length dx.

Let G_{th} be the rate of thermal generation of electrons (due to phonons), and G_{op} be the rate of optical generation of electrons (due to photons). Then the total generation rate, G_n is given by-

$$G_n = G_{th} + G_{op}$$

As discussed beforehand, there is a minority carrier lifetime, τ_n associated with recombination. The recombination rate, R_n is proportional to the number of electrons available for recombination, and is thus, given by:

$$R_n = \frac{n}{\tau_n} = \frac{n_o}{\tau_n} + \frac{\Delta n}{\tau_n}$$

Substituting this back into the equation gives us:

$$\frac{dn}{dt} = \frac{1}{q}\frac{dJ_n(x)}{dx} + (G_{th} + G_{op} - \frac{n_o}{\tau_n} - \frac{\Delta n}{\tau_n})$$

At equilibrium, the net current is zero, there are no external fields or temperature gradients, and no light is shining on the sample. For a sample at equilibrium, it then means that $\Delta n = 0$, $J_n = 0$, and $G_{op} = 0$. Thus the continuity equation leads us to a simple and quite intuitive result:

$$G_{th} = \frac{n_o}{\tau_n}$$

which means that for a semiconductor sample at equilibrium, the thermal rate of generation of carriers is exactly equal to the intrinsic rate of recombination of carriers. The continuity equation (for electrons) can then be finally written as:

$$\frac{dn}{dt} = \frac{1}{q}\frac{dJ_n(x)}{dx} + (G_{op} - \frac{\Delta n}{\tau_n})$$

Similarly, for holes,

$$\frac{dp}{dt} = -\frac{1}{q}\frac{dJ_p(x)}{dx} + (G_{op} - \frac{\Delta p}{\tau_p})$$

Substituting the current equation into the continuity equation gives us:

$$\begin{aligned}\frac{dn}{dt} &= n\mu_n\frac{dE}{dx} + \mu_n E\frac{dn}{dx} + D_n\frac{d^2n}{dx^2} + (G_{op} - \frac{\Delta n}{\tau_n}) \\ \frac{dp}{dt} &= -p\mu_p\frac{dE}{dx} - \mu_p E\frac{dp}{dx} + D_p\frac{d^2p}{dx^2} + (G_{op} - \frac{\Delta p}{\tau_p})\end{aligned}$$

Now, consider a semiconductor block, initially at equilibrium, which is now illuminated uniformly. The continuity equation yields:

$$\begin{aligned}\frac{d(\Delta n)}{dt} &= G_{op} - \frac{\Delta n}{\tau_n} \\ \Rightarrow \Delta n(t) &= G_{op}\tau_n(1 - e^{-t/\tau_n})\end{aligned}$$

Now if the light source is turned off at $t = t_o$, $G_{op} = 0$ and thus,

$$\Delta n(t) = \Delta n(t_o)e^{(-t/\tau_n)}$$

Let us now consider a semiconductor block at a steady state, with a non-zero current density and $G = 0$. We also assume that external electric field, $E = 0$. Then,

$$\frac{d(\Delta n)}{dt} = \frac{1}{q}\frac{dJ_n(x)}{dx} - \frac{\Delta n}{\tau_n}$$

At steady state, $\frac{d(\Delta n)}{dt} = 0$, which gives us,

$$\begin{aligned}\frac{1}{q}\frac{dJ_n(x)}{dx} &= \frac{\Delta n}{\tau_n} \\ \Rightarrow \frac{d^2\Delta n}{dx^2} &= \frac{\Delta n}{L_n^2}\end{aligned}$$

where $L_n = \sqrt{D_n\tau_n}$ is known as the *diffusion length* of the carrier inside the material.

The general solution to this differential equation is of the form-

$$\Delta n(x) = c_1 e^{\frac{x}{L_n}} + c_2 e^{-\frac{x}{L_n}},$$

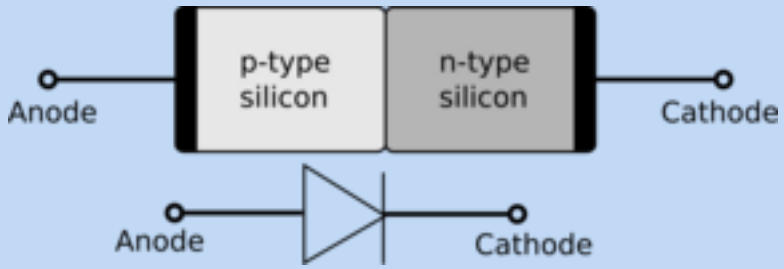
where the coefficients c_1 and c_2 will be determined as per the boundary conditions applicable to the situation. The equation can be further simplified for practical use if we take into consideration the fact that there are, generally, two possible cases- one where the diode length is much larger than the depletion length(long diode) and the other where the diode length is much shorter than the depletion length(short diode).

In a short diode, the exponential terms can be simplified to their linear approximations ($e^x = 1 + x$, for $x << 1$) since $x/L_n << 1$. So the excess carrier concentration in this case falls off linearly. In a long diode, the coefficient of the term $e^{\frac{x}{L_n}}$ must be zero since it would otherwise imply that the excess carrier concentration increases indefinitely on moving away from the origin, which is not possible theoretically. Hence, the excess carrier concentration falls off exponentially in this case.

The continuity equation can be expressed in many formats as we have seen above. Hence, it depends a lot on the situation as to what form of the equation should be used for our purpose.

3 P-N Junctions

p-n junctions are elementary constituents of many modern electronic devices such as diodes, solar cells, rectifiers, ICs, transistors, etc. This section will cover the part of *p-n homojunctions* only. The term *homojunction* means that the junction is between two regions of the same material. On the other hand, in a *heterojunction*, the junction is between regions of two different materials (e.g. Si and Ge).
There are several methods of fabricating a p-n junction, among which the most common method involves implanting donor atoms into a p-type Si substrate. With the appropriate doping levels, it inverts the p-type substrate into n-type Si, which is now in metallurgical contact with p-type Si.



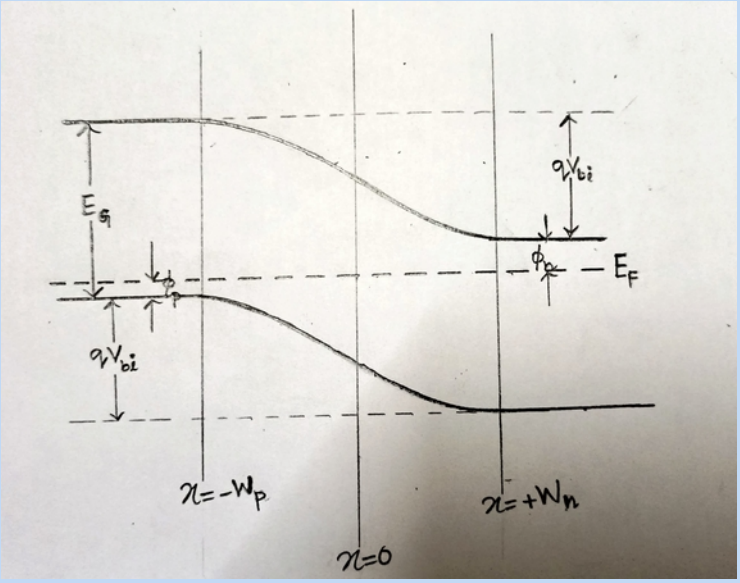
p-n junction diode and its symbol

Some of the assumptions that we will be taking while dealing with p-n junctions are as follows: (i) The material is light/moderately doped, (ii) Complete ionisation of dopant atoms, (iii) All calculations are made at ideal operating conditions, i.e. at room temperature.

3.1 p-n junction at equilibrium

Due to the concentration gradient across the junction, a diffusion current starts flowing across it. However, at the junction itself, electrons and holes capture each other and thus, recombine, resulting in the release of energy. This process of recombination depletes the region adjacent to the junction of electrons and holes, leaving behind an excess of immobile dopant atoms therefore, it is also known as the depletion region. These immobile atoms now create an in-built potential across the junction which is opposite to the flow of the diffusion current. This in-built potential steadily grows with time until the diffusion current can no longer flow across the junction. Hence, even without any application of external potential, there exists an in-built potential, say V_{bi} across the p-n junction.

When a p-n junction is under equilibrium, the Fermi level on both p- and n-side is *invariant* with position. This is a consequence of the fact that rates of generation and recombination exactly balance each other at equilibrium and can be proved in general. The C.B.s and V.B.s of the two sides merge with each other in a continuous fashion, keeping band-gap energy constant throughout, as shown in the figure below:



Energy band diagram of a p-n junction

Let the width of the depletion region be $W = W_n + W_p$ where W_n and W_p represent the widths of the depletion region on n- and p-sides respectively. Let ϕ_n and ϕ_p represent the energy gap between E_c and E_F on n-side and E_v; and E_F on p-side respectively. We know that,

$$\phi_n = k_B T ln(\tfrac{N_c}{N_D})$$

$$\phi_p = k_B T ln(\frac{N_v}{N_A})$$

From observation, one can reach the conclusion that:

$$qV_{bi} = E_G - \phi_n - \phi_p$$

$$qV_{bi} = E_G - k_B T ln(\frac{N_c N_v}{N_A N_D}) = E_G - k_B T ln(\frac{N_c N_v}{np})$$

Now,

$$n_i^2 = N_c N_v exp(-\frac{E_G}{k_B T})$$

Substituting this in the previous equation and eliminating E_G,

$$V_{bi} = \frac{k_B T}{q} ln(\frac{N_A N_D}{n_i^2})$$

which gives us an estimate for the built-in potential in terms of dopant concentrations. We can also derive a relation between V_{bi} and the width of the depletion region using Poisson's equation.

Since the sample must remain electrically neutral throughout the process, net charge must be zero within the depletion region. This implies that $N_A W_p = N_D W_n$. This also means that the higher the doping on a given side, the lower the depletion width on that side. Now, applying Poisson's equation to the n-side,

$$\frac{d^2 V}{dx^2} = -\frac{\rho}{\epsilon_o \epsilon_r}$$

This implies:

$$\frac{dE}{dx} = \frac{qN_D}{\epsilon_o \epsilon_r}$$

On integrating with the appropriate limits,

$$E = \frac{qN_D}{\epsilon_o \epsilon_r}(x - W_n), \text{ for } 0 < x \leq W_n$$

Similarly,

$$E = -\frac{qN_A}{\epsilon_o \epsilon_r}(x + W_p), \text{ for } -W_p \leq x < 0$$

Now, having found the electric field in the depletion region, we are in a position to determine the built-in potential difference across the junction.

$$V_{bi} = -(\int_{-W_p}^{0} E\,dx + \int_{0}^{W_n} E\,dx)$$

On solving this particular integral, we obtain,

$$V_{bi} = \frac{q}{2\epsilon_o\epsilon_r}(N_A W_p^2 + N_D W_n^2)$$

which on further simplification, yields:

$$W = \sqrt{\frac{2\epsilon_o\epsilon_r}{q} V_{bi}(N_A^{-1} + N_D^{-1})}$$

where W represents the width of the depletion region.

3.2 Quasi-Fermi levels

Before we move into the analysis of p-n junctions under bias, we need to understand the concept of quasi-Fermi levels. Quasi-Fermi levels are used to calculate the carrier concentrations in a semiconductor sample when it is under a state of quasi-equilibrium.

The Fermi level in a semiconductor can estimate the carrier concentrations at thermal equilibrium. As we know, this state of thermal equilibrium is fairly dynamic in nature- though the carrier population remains constant over time, an individual carrier simply doesn't remain at rest in a particular energy band. Due to thermal generation of EHPs, some electrons are excited into the conduction band from the valence band. At the same time, an equal number of holes from C.B. recombine with electrons from V.B., keeping carrier populations constant over time. After an electron has been excited into a higher energy level, it undergoes rapid transitions (phonon interactions) to reach the lowest energy level in C.B. Similarly, a hole in lower regions of V.B. undergoes rapid transitions to reach the top of V.B. (Since the energy bands are w.r.t electrons, hole energy goes the other way round). These intra-band transitions take about 10^{-12} to 10^{-13} seconds. In contrast, the inter-band transitions take about 10^{-8} to 10^{-9} seconds. Such a wide range of differences in the transition times is what makes possible a quasi-equilibrium state in the semiconductor.

The need for the introduction of quasi-Fermi levels is clearly understood through a physical analogy of a water tank and pump.

Let us say that the energy level, E_c and E_v can be represented by two tanks of water where the water level in each represents the electronic concentration. The thermal generation process can be thought of as a thermal pump operating between the two tanks

and pumping water from the lower tank (E_v) to the upper tank (E_c) at a given rate for a particular temperature. The recombination process can be visualised as a leaky upper tank so that some of the water pumped into the upper tank leaks into the lower tank, thereby decreasing the water level of the upper tank. At T = 0K, the thermal pump doesn't operate, so there's no change in water level (electrons) in either of the tanks (the upper tank is empty as of now). At a finite temperature, the thermal pump starts operating and pumping water into the upper tank. Some of the water leaks out into the lower tank due to the holes in the upper tank and thus, they help establish an equilibrium water level in both tanks.

Now, suppose that due to external factors, there is an additional pump operating between the two tanks. Now, at the same temperature, the equilibrium water level would be different in both tanks. The water level in the lower tank falls while that in the upper tank rises. This is analogous to electrons being pumped into the C.B. by external factors, which is the same situation as in a p-n junction under bias. Thus, it can be inferred that the equilibrium Fermi level can no longer account for this new population since the temperature remains *unchanged* throughout the process. This is where the quasi-Fermi levels are taken into consideration.

Instead of a single Fermi level, we now define two different quasi-Fermi levels such that,

$$n = N_c exp(\frac{E_{Fn} - Ec}{k_B T})$$

$$p = N_v exp(\frac{E_v - E_{Fp}}{k_B T})$$

Consequently,

$$np = n_i^2 exp(\frac{E_{Fn} - E_{Fp}}{k_B T})$$

Hence, the value of $E_{Fn} - E_{Fp}$ determines how much the deviation is from the equilibrium position in the semiconductor crystal. At thermal equilibrium, the two quasi-Fermi levels merge together to form a single Fermi level.

3.3 p-n junction under bias

Suppose that we apply an external voltage, V_a, across the p-n junction. The device itself consists of three regions:
(i) The quasi-neutral region on the n-side,
(ii) The depletion region,
(iii) The quasi-neutral region on the p-side
The resistivity is inversely proportional to the number of free carriers in a semiconductor-

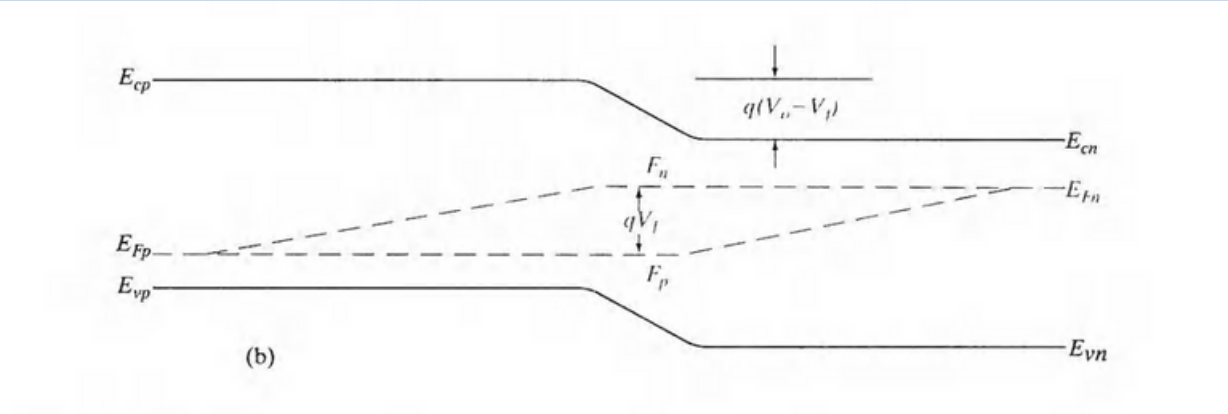
this implies that nearly all the applied bias voltage drops across the depletion region (which has virtually no free carriers since they have been swept away). In general, the junction voltage, V_j will be given by-

$$V_j = V_{bi} - V_a$$

3.3.1 Forward bias

Let us first consider the case of a p-n junction diode under forward bias. In a forward-biased p-n junction diode, the majority carriers from one side are pumped to the other side due to the applied voltage. For example, electrons from the n-side are pumped to the p-side, becoming the minority carriers. Now, these minority carriers on each side recombine with majority carriers and start to decay, creating a concentration gradient in the respective regions they occupy. This concentration gradient starts a diffusion current, which can be attributed to minority carriers. In order to estimate this diffusion current, first, we need to evaluate the concentration profile. And in order to do so, we need to use the concept of quasi-Fermi levels and how these quasi-levels behave under the application of an external bias.

Under forward bias, the quasi-Fermi levels inside the diode are separated by a gap of qV_a, where V_a is the applied bias voltage. Away from the junction, this energy difference gradually reduces and eventually merges to a single quasi-Fermi level on both sides, i.e. at a large enough distance from the junction, the carrier concentrations are restored to their equilibrium concentrations. This is depicted in the energy band diagram shown below:



Quasi-Fermi levels in a p-n junction diode under forward bias

Let the dopant concentration on n-side be N_D and the equilibrium minority carrier

concentration be given by p_{n0} , where $p_{n0} = n_i^2/N_D$. Let $p_n(x)$ represent the minority carrier distribution function. Then, $p_n(W_n)$ is given by-

$$p_n(W_n) = N_v exp(\frac{E_{vn}-F_p}{k_BT})$$

$$p_n(W_n) = N_v exp(\frac{E_{vn}-E_{Fn}}{k_BT}) exp(\frac{qV_a}{k_BT})$$

Therefore,

$$p_n(W_n) = p_{n0} exp(\frac{qV_a}{k_BT})$$

Hence, excess minority carrier concentration at the left-hand boundary of the n-side is given by-

$$\Delta p_n = p_{n0}(exp(\frac{qV_a}{k_BT}) - 1)$$

We now have two boundary conditions for the excess minority carrier concentration on the n-side - at a large distance from the junction, $\Delta p_n = 0$ and at the junction boundary, $\Delta p_n = p_{n0}(exp(\frac{qV_a}{k_BT}) - 1)$. From the continuity equation(for a long diode), we can obtain the excess minority carrier concentration profile as follows-

$$\Delta p_n(x) = \frac{n_i^2}{N_D}(exp(\frac{qV_a}{k_BT} - 1))exp(-\frac{x-W_n}{L_p})$$

Similarly, for the p-side,

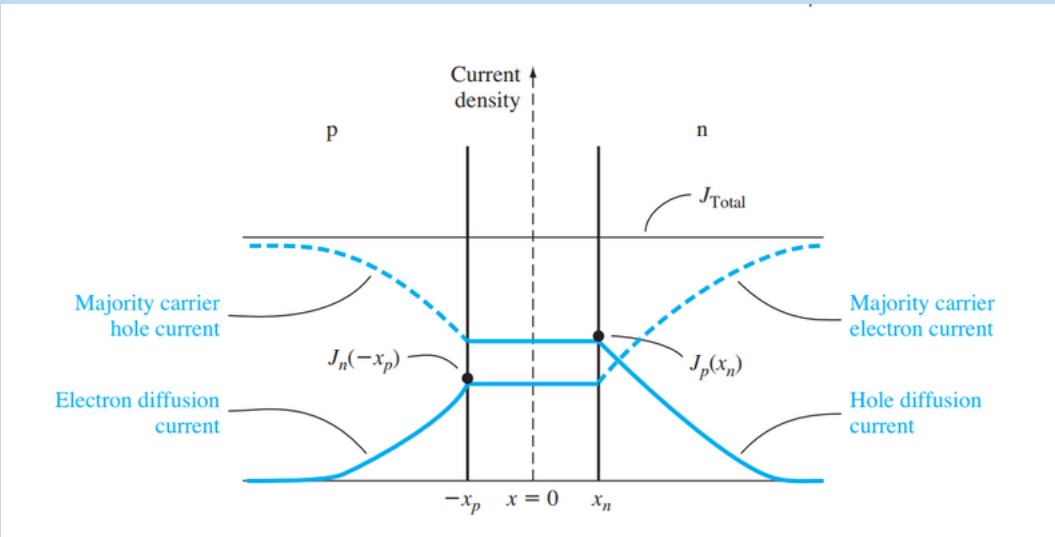
$$\Delta n_p(x) = \frac{n_i^2}{N_A}(exp(\frac{qV_a}{k_BT} - 1))exp(\frac{x+W_p}{L_n})$$

The diffusion currents can now be determined directly as:

$$J_n(x) = qD_n\frac{d\Delta n(x)}{dx} = -q\frac{D_n}{L_n}(exp(\frac{qV_a}{k_BT} - 1))exp(\frac{x+W_p}{L_n})$$

$$J_p(x) = -qD_p\frac{d\Delta p(x)}{dx} = -q\frac{D_p}{L_p}(exp(\frac{qV_a}{k_BT} - 1))exp(-\frac{x-W_n}{L_p})$$

However, an important point has been overlooked during the calculations- the net current is not constant with position, since we have not taken drift current into consideration. Ideally, due to the high resistance of the depletion region(since its carrier concentration is very low), it is assumed that all of the forward bias voltage drops across it. However, due to the finite, but small, resistance of the quasi-neutral regions, there is a small electric field in this region. Coupled with a high majority carrier concentration, this small electric field can create a drift current equivalent in magnitude to the diffusion currents of minority carriers. Therefore, the current density profile is as shown below:



Current density profile

At the junction boundary($x = W_n^-$ and $x = -W_p^+$), the electric field is practically zero as well as the majority carrier concentration is very low. Hence, the drift current magnitude is zero at the junction boundaries. The net current, therefore, is given by the sum of the magnitudes of the two diffusion currents at the respective junction boundaries.

$$J_{net} = q\frac{D_p}{L_p}p_{n0}(exp(\frac{qV_a}{k_BT}) - 1) + q\frac{D_n}{L_n}n_{p0}(exp(\frac{qV_a}{k_BT}) - 1)$$

$$J_{net} = J_0(exp(\frac{qV_a}{k_BT}) - 1),$$

where $J_0 = q\frac{D_p}{L_p}p_{n0} + q\frac{D_n}{L_n}n_{p0}$ and is known as the 'reverse saturation current'. At room temperature of 300K, k_BT is approximately only 26 meV while qV_a is typically much larger than this. So effectively, the current equation becomes an exponential relation between the current and applied forward voltage bias.

3.3.2 Reverse bias

Mathematically, the situation in reverse bias is similar to forward bias- in the current equation, we can replace V_a by $-V_a$ in order to evaluate the reverse bias current. Physically, what happens under reverse bias is that the minority carrier injection drastically falls due to the large potential difference existing between the two sides. In fact, we can say that minority carriers are '*extracted*' instead of being '*injected*' across the p-n junction. This large decrease in minority carrier profile results in exponentially lower diffusion currents and, consequently, very low net current. Hence, a p-n junction practically blocks any current under reverse bias. Due to this unique property, a p-n junction is used as a

'rectifier' in circuits. For V_a ¡ -$3k_BT$, to good approximation the diode reverse current is -J_o.

Due to the high electric field inside the depletion region under reverse bias situation, several interesting phenomena occur on the application of high reverse voltages, notable among them being the Zener breakdown and the avalanche breakdown. Due to these breakdown phenomena, after a certain voltage limit(known as the breakdown voltage), the current in the diode increases tremendously, even on a minuscule change in the voltage, as illustrated by the I-V characteristics.

Zener breakdown occurs when the high electric field inside the depletion zone enables quantum tunnelling of electrons across the depletion region, which leads to a large increase in charge carriers. Avalanche breakdown, as the name suggests, proceeds like a chain reaction. A high-energy carrier can collide with an atom in the depletion zone and, thus, generate an extra carrier contributing to the current flow. This newly generated carrier can further collide with atoms, releasing more carriers. Thus, the process becomes a chain reaction(like an avalanche), leading to a large current increase.

3.4 Short-Base Diodes

In the above discussion, we assumed that the diode length was significantly larger than the diffusion lengths of the charge carriers. If this is not so, and the diode length is comparable or, in some cases, much smaller than the diffusion lengths, the diode is said to be a *short-base diode*. In this section, we derive the current equation for the short-base diode.

Let us assume that the length of the n-side is W_B, and that of the p-side is W_A (both are much smaller than L_n and L_p, respectively). The excess minority carrier concentrations at the edge of the band are given by:

$$\Delta p(W_n) = p_{n_o}(exp(\frac{qV_a}{k_BT}) - 1)$$

$$\Delta n(-W_p) = n_{p_o}(exp(\frac{qV_a}{k_BT}) - 1)$$

The minority carrier concentration as a function of position inside the quasi-neutral regions will be given by:

$$\Delta n(x) = c_1 + c_2\frac{x}{L_n}$$

$$\Delta p(x) = c_1^{'} + c_2^{'}\frac{x}{L_p}$$

where c_1, $c_1^{'}$, c_2 and $c_2^{'}$ are some constants. On applying the appropriate boundary conditions along with the expression for diffusion current, we obtain the following:

$$J_n(x) = q\frac{D_n}{W_A}n_{p_o}(exp(\frac{qV_a}{k_BT}) - 1)$$

$$J_p(x) = q\frac{D_p}{W_B}p_{n_o}(exp(\frac{qV_a}{k_BT}) - 1)$$

Observe that while the exponential relation of voltage with current stays the same, the value of the *reverse saturation current*, J_o, has changed.

3.5 Generation and Recombination Currents

While analysing the p-n junction diode, we assumed that generation-recombination processes take place only in the quasi-neutral regions of the diode, and we ignored such processes in the depletion region. However, carrier generation-recombination does occur in the depletion region and constitutes the *generation-recombination (G-R) current*. In fact, it turns out that the G-R current is much larger than the reverse saturation current in a reverse-biased diode (G-R current is typically three orders larger than the reverse saturation current). The net recombination rate (under certain assumptions) inside the semiconductor is given by-

$$R - G = \frac{np - n_i^2}{\tau_o(n+p+2n_i)}$$

At equilibrium, $np = n_i^2$, and therefore, the net recombination rate is zero everywhere inside the material. In a non-equilibrium steady state situation, np need not be equal to n_i^2; therefore, there is either net recombination or generation inside the material.

3.5.1 Generation-Recombination in reverse bias

In reverse bias, the recombination rate is very nearly zero since most of the minority carriers inside the depletion region have been 'extracted' out. Thus, $n \approx p \approx 0$ and thus,

$$R = 0, G = \frac{n_i}{2\tau_o}$$

$$J_G = -qGW = -\frac{qn_iW}{2\tau_o}$$

3.5.2 Generation-recombination in forward bias

In forward bias, within the transition region,

$$np = n_i^2 exp(\frac{qV_a}{k_BT})$$

Due to this reasonably high product value, np, we can argue that recombination processes dominate in the depletion region. Hence,

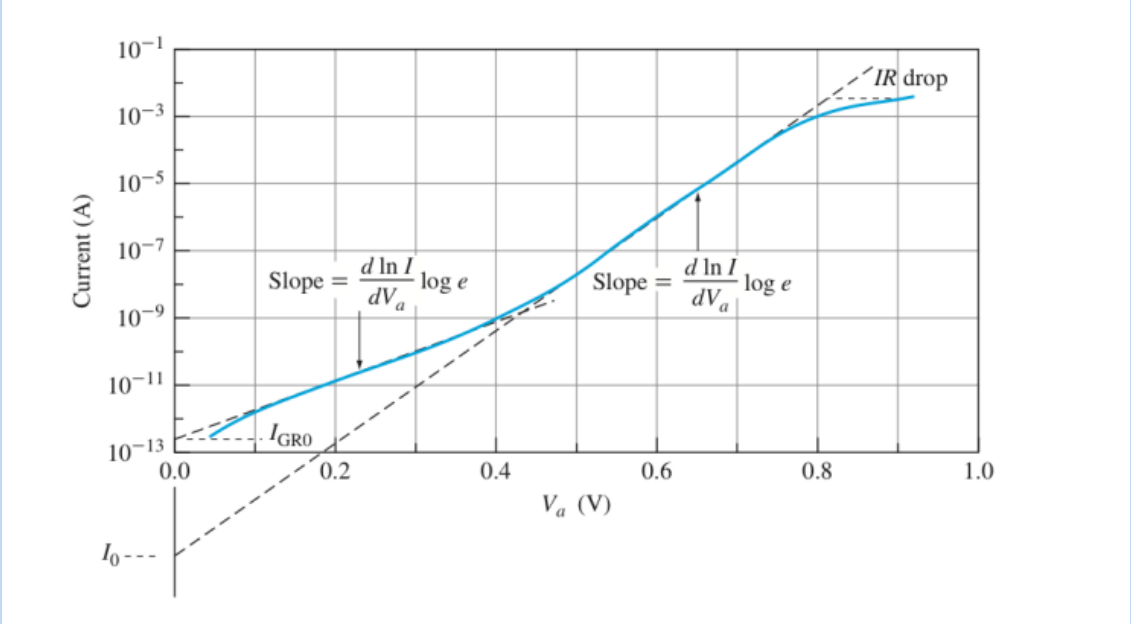
$$G = 0, R = \frac{n_i^2 exp(\frac{qV_a}{k_B T})}{\tau_o(n+p)}$$

The G-R current density is approximately given by:

$$J_{GR} = J_{GR0}(exp(\frac{qV_a}{2k_B T}) - 1)$$

The total current density is then given by:

$$J = J_{GR} + J_{diff} = J_{GR0}(exp(\frac{qV_a}{2k_B T}) - 1) + J_o(exp(\frac{qV_a}{k_B T}) - 1)$$



I-V characteristics after taking G-R current into consideration

At small V_a, recombination current dominates since $J_{GR0} >> J_o$. At higher values of V_a, the diffusion current predominates. At still higher values of V_a, the line deviates from linearity because of the IR drop inside the semiconductor (which was negligible at low-to-moderate voltages). In general, the diode current is expressed as-

$$J = J_s(exp(\frac{qV_a}{nk_B T}) - 1)$$

where J_s is a function of J_{GR0} and J_o, and the *ideality factor*, n, depends upon the applied bias voltage ($1 < n < 2$).

3.6 Carrier Multiplication and Tunnelling

3.7 Diode Capacitance

The p-n junction diode capacitance can be divided into two distinct parts- the small-signal capacitance, known as *junction(differential) capacitance* and the large-signal capacitance, known as *stored-charge capacitance*.

3.7.1 Junction Capacitance

Let us bear the obvious fact in mind that the width of the depletion region, and by extension, the charge due to immobile carriers bound within the depletion region, depends on the magnitude of the biasing voltage across the diode. Let us consider a small ac signal riding on a larger bias voltage applied across the p-n junction. For a small change dV_a in applied voltage, the space charge on one side of the junction changes by $+dQ$ and on the other side by -dQ. Observe that this dQ charge appears as a sheet of charge across the depletion region, and therefore, the capacitance is given by:

$$C_j = |\frac{dQ}{dV_a}| = \frac{\epsilon A}{W}$$

This capacitance is often called the *depletion capacitance*.

3.7.2 Stored Charge Capacitance

The stored charge capacitance is attributable to the change in minority carrier density on either side of the junction, as minority carriers are either injected or extracted at the junction edges, with changing bias.

Consider a p-n junction under forward bias. In the steady state, the minority carrier concentration on the p-side is given by:

$$\Delta n_p(x) = \Delta n_p(x_p) e^{\frac{-(x-x_p)}{L_n}}$$

This distribution is maintained at the steady state, even though the individual electrons and holes undergo dynamic changes. The net "stored charge" due to excess minority carriers in the p-region is given by-

$$Q_s = -qA \int_{x_p}^{\infty} \Delta n_p(x_p) e^{\frac{-(x-x_p)}{L_n}} \, dx = qA\Delta n_p(x_p) L_n$$

Note that on varying the bias voltage, this distribution does not change instantaneously. Instead, the sheet charge at x = x_p is dragged down while the distribution inside the bulk of the semiconductor

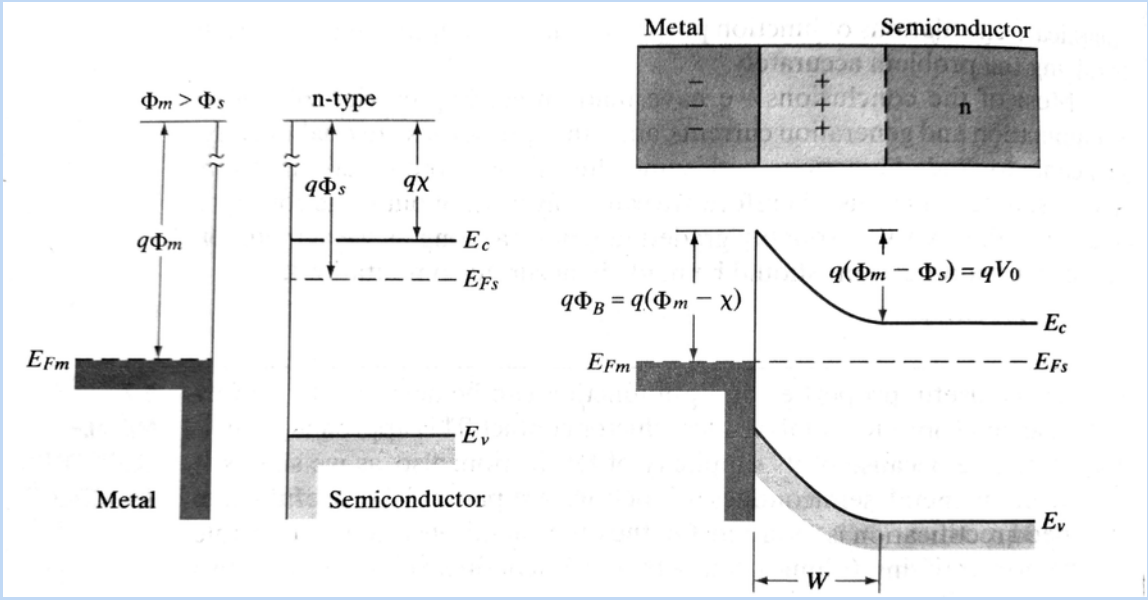
4 Metal-Semiconductor Junctions

Metal contacts are a quintessential part of semiconductor devices-most of the final packaged products in the industry involve numerous metal-semiconductor junctions. Apart from this, many measurement devices, cables, etc., involve metal contacts. So, it is important to understand the characteristics of metal-semiconductor junctions in order to understand real-life semiconductor devices.

The two types of metal-semiconductor junctions are the 'Schottky contacts' and the 'Ohmic contacts'. As the name suggests, ohmic contacts obey Ohm's law, while the Schottky contact's I-V characteristics resemble that of a p-n junction. That whether a given metal contact behaves as a Schottky contact or an ohmic one largely depends on the work function of the metal for a given semiconductor sample.

4.1 Schottky contacts

Let us assume that the semiconductor sample under consideration is n-type doped, and its Fermi level is above the Fermi level of the metal. When the metal and semiconductor sample are brought into mutual contact, electrons start to flow from the n-type material into the metal in a bid to equalize the two distinct Fermi levels at equilibrium. At equilibrium, the gap between E_c and E_F as well as that between E_v and E_F is non-uniform in the depletion region but tapers down to the initial value inside the bulk of the semiconductor.



Schottky contact energy band diagram

Let the Schottky contact be characterised by built-in potential V_{bi}, depletion width W_d, and Schottky barrier Φ_B. Let the work functions of the metal and semiconductor be ϕ_m and ϕ_s, respectively. Observe that,

$$qV_{bi} = \phi_m - \phi_s$$

Proceeding in a similar manner as in the case of p-n junctions, we can show that inside the semiconductor, Poisson's equation follows as

$$\frac{d^2V(x)}{dx^2} = -\frac{qN_D}{\epsilon_o\epsilon_s}$$

Which on further simplification along with the boundary condition $E(W_d^+) = 0$ yields,

$$\frac{dE(x)}{dx} = \frac{qN_D}{\epsilon_o\epsilon_s}$$

$$E(x) = \frac{qN_D}{\epsilon_o\epsilon_s}(x - W_d)$$

This yields the built-in potential as:

$$V_{bi} = \frac{qN_DW_d^2}{2\epsilon_o\epsilon_s}$$

This expression for the built-in potential is in many ways functionally similar to the one derived for the p-n junction diode. The difference between conduction band and Fermi band inside the bulk of the semiconductor is given by-

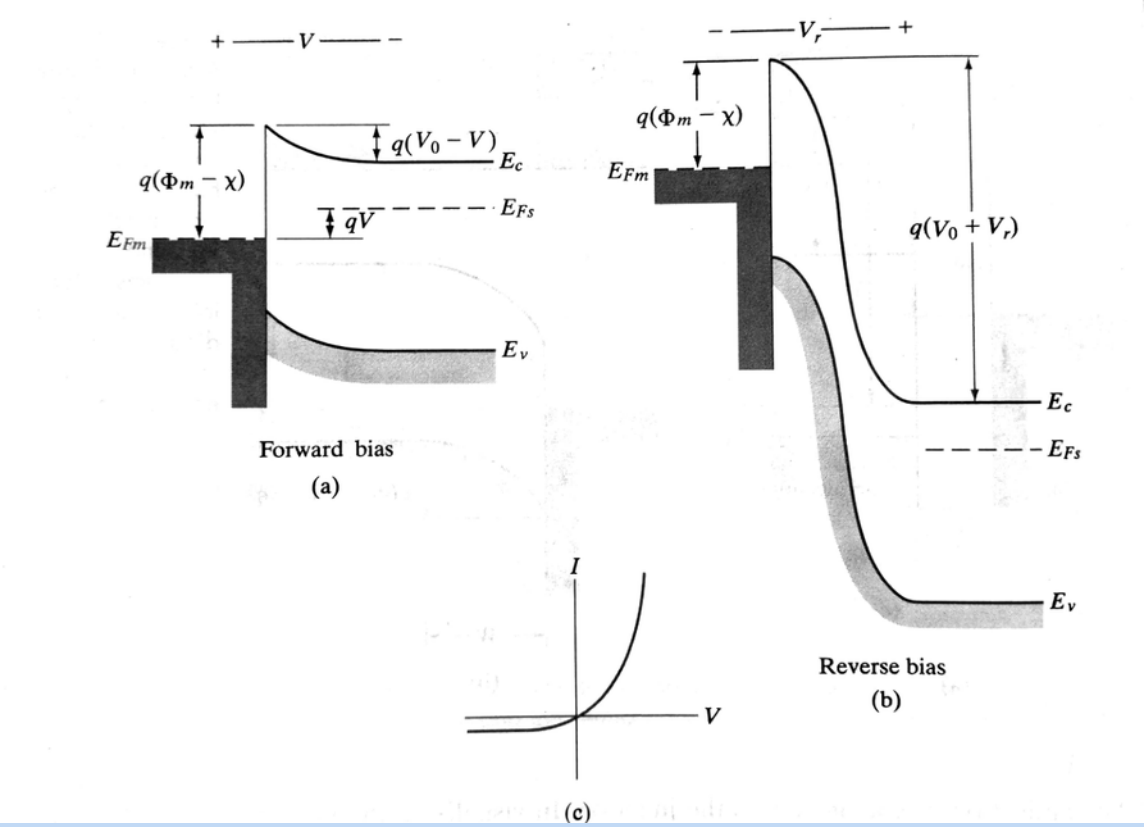
$$E_c - E_F = \frac{k_BT}{q}ln(\frac{N_c}{N_D})$$

The Schottky barrier, ϕ_B, preventing the injection of electrons from metal to the semiconductor is given by-

$$\phi_B = qV_{bi} + (E_c - E_F)$$

$$\phi_B = \frac{q^2N_DW_d^2}{2\epsilon_s} + \frac{k_BT}{q}ln(\frac{N_c}{N_D})$$

The Schottky contact's current equation is of the same functional form as the p-n junction diode. Therefore the Schottky contact acts as a rectifying contact like the p-n diode. In fact, in high-frequency applications the Schottky contact is preferred over the p-n junction diode.



Schottky contacts under bias

The forward current equation in a Schottky contact is given by-

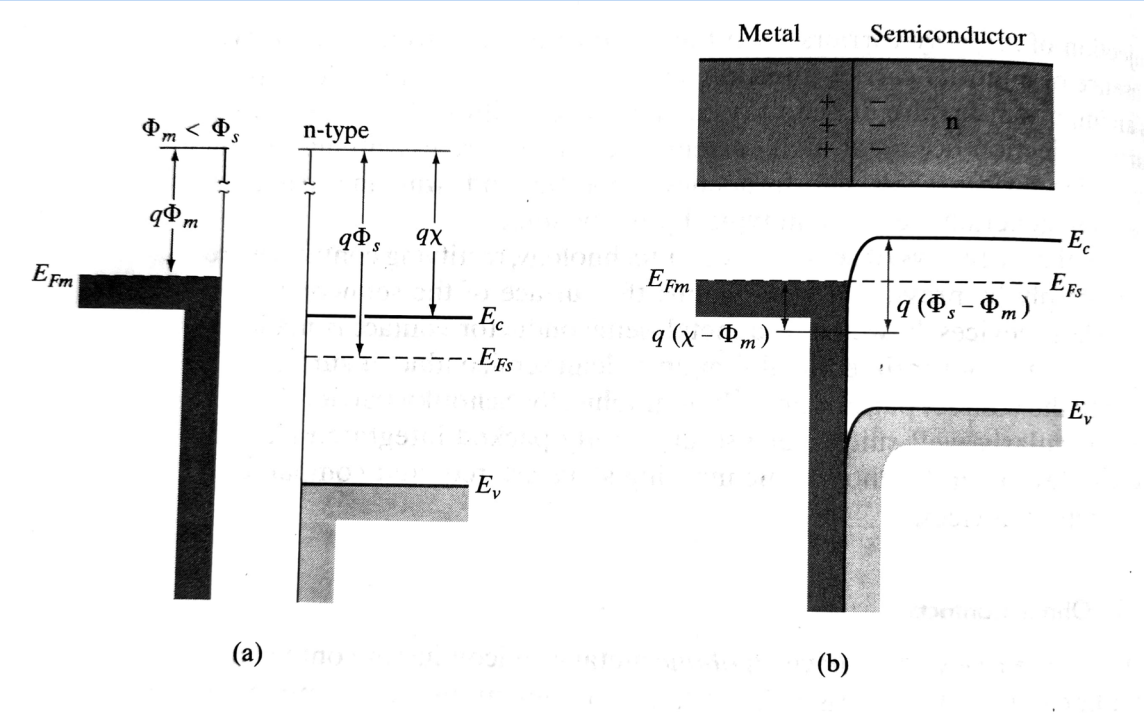
$$J = BT^2 exp(-\frac{q\phi_B}{k_B T}) exp(\frac{qV}{\eta k_B T})$$

where the constant B depends upon junction parameters and $1 < \eta < 2$. This equation is functionally similar to the current due to thermionic emission.

4.2 Ohmic contacts

When the Fermi level of the n-type material is below the metal Fermi level, it results in an inflow of electrons from the metal to the semiconductor, resulting in an accumulation

of electrons near the junction. This can also be understood in terms of energy bands as shown through the figure below:



Ohmic contact

The junction behaves as an 'Ohmic' contact because of the characteristic metal-like property associated with it - the Fermi level lying above the conduction band near the junction.

The cases considered till now were those of n-type semiconductors. For p-type materials, the cases are reversed - when the semiconductor Fermi level lies below the metal Fermi level, it forms a Schottky contact, and when it's above the metal Fermi level, it forms an Ohmic contact. These facts can be easily verified by drawing the relevant energy band diagrams.

5 Metal-Oxide-Semiconductor Field-Effect Transistors(MOSFETs)

MOSFETs and related transistors are by far the most extensively used semiconductor devices in the modern age, because of their uses in logic and memory devices. Although

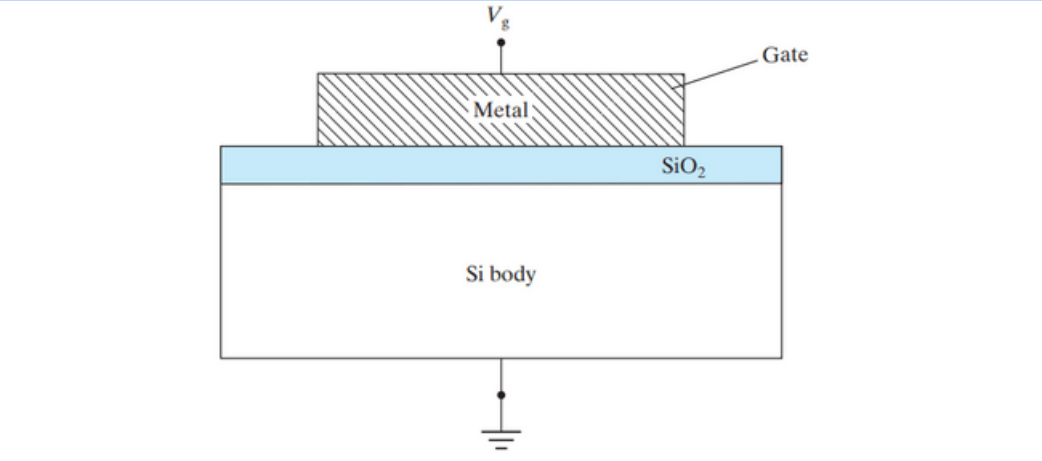
MOSFETs were introduced much later, they eventually overtook BJTs in their applications, primarily due to their very low power consumption. Microprocessors and memory chips include billions of MOSFETs and the number of MOSFETs included in these electronic devices is expected to double every two years, as per the well-known Moore's Law.

Initially, the gate material in MOSFETs was a metal (aluminium, to be specific). Present-day transistors, however, use degeneratively doped polycrystalline Si (poly-Si), which is highly conductive, as the gate material. Silicon dioxide has been historically used to create the oxide layer.

5.1 MOS Capacitor(MOSC)

An nMOS capacitor basically consists of a n^+-Si layer (the gate material) separated by a thin oxide layer (insulating) from the p-type substrate.

5.2 MOS Capacitor

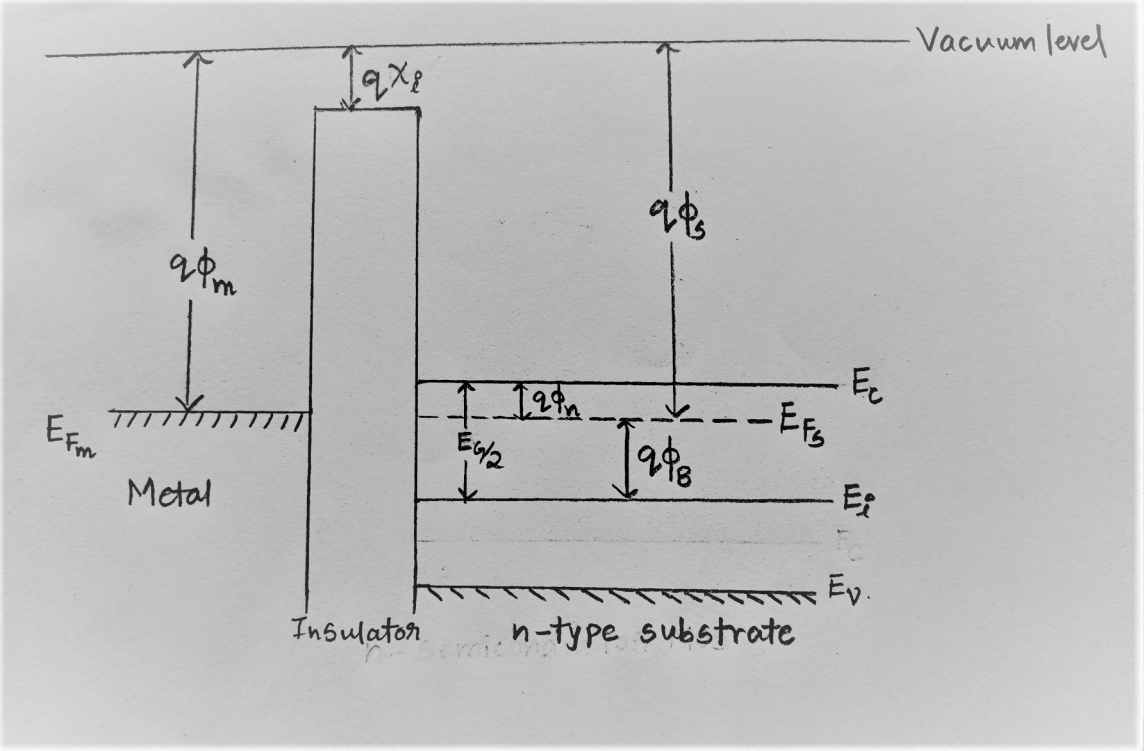


MOS Capacitor

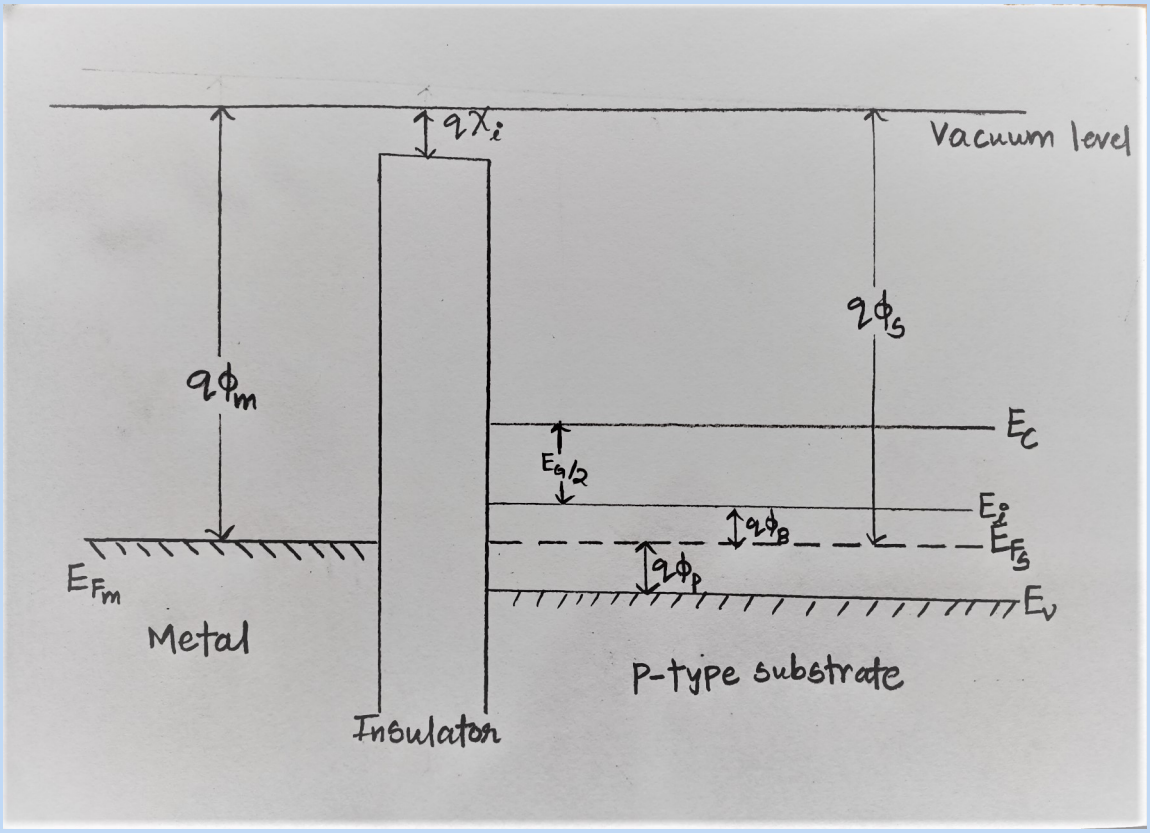
The MOS capacitor essentially consists of a metal gate in contact with an insulator/oxide medium, which further is in contact with a p- or n-type substrate as shown in the above figure. The insulator ensures that no current flows between the metal and semiconductor.

There are some basic assumptions that we shall put into use while analyzing MOS capacitors (later on, we will address them). One assumption is that there is no energy difference between the metal and semiconductor Fermi levels even before joining them,

i.e. flat band condition exists. The other assumption is that there are no free charges present in the bulk of insulator or the semiconductor. Under such assumptions, the energy band diagrams of the MOS capacitor are as follows:



Energy band diagram of a n-type MOS capacitor

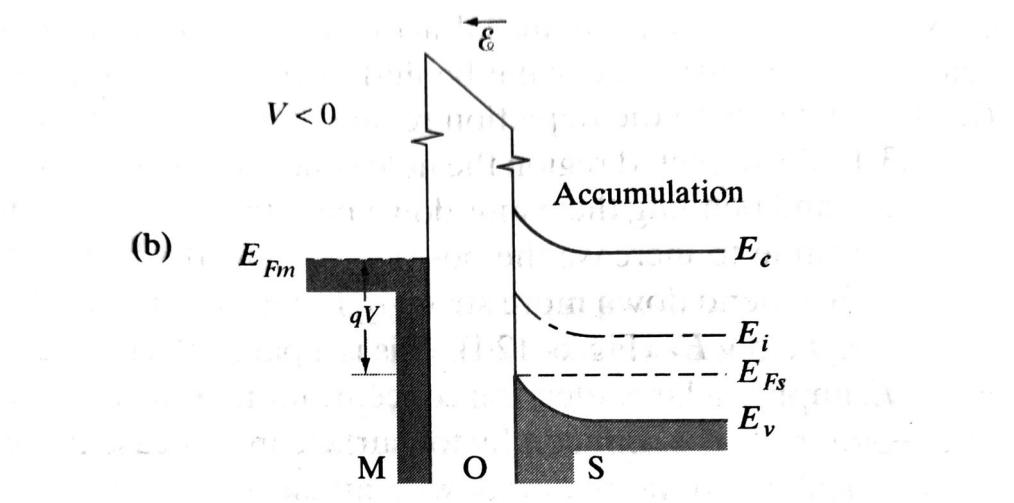


Energy band diagram of a p-type MOS capacitor

When a gate voltage, V_{GS} is applied, there are three possibilities that may arise- accumulation, depletion or inversion. All these modes have their unique properties and associated capacitances, which have to be analyzed separately. Let us consider a p-type substrate for the associated analysis. All the results derived as such will also hold for a n-type substrate.

5.2.1 Accumulation mode

On applying a negative gate voltage, holes are drawn from the substrate to the insulator-substrate junction and consequently, a negative charge density is also induced on the metal. This process results inn band-bending of the energy levels near the junction as shown:



Energy band diagram at accumulation mode

Since holes (majority carriers) are *accumulated* near the junction, this is known as the 'accumulation' mode of operation. The gate voltage is given by-

$$V_{GS} = \phi_s + V_{ox}$$

Generally, ϕ_s is negligible for moderate gate voltages as compared to V_{ox}, so we can simplify the expression further-

$$V_{GS} \approx V_{ox}$$

In MOS capacitor theory, unlike the electrostatic theory, we always use the substrate charge inside the semiconductor while calculating capacitances. Therefore, V_{ox} is given by-

$$V_{ox} = -Q_{sub}/C_{ox}$$

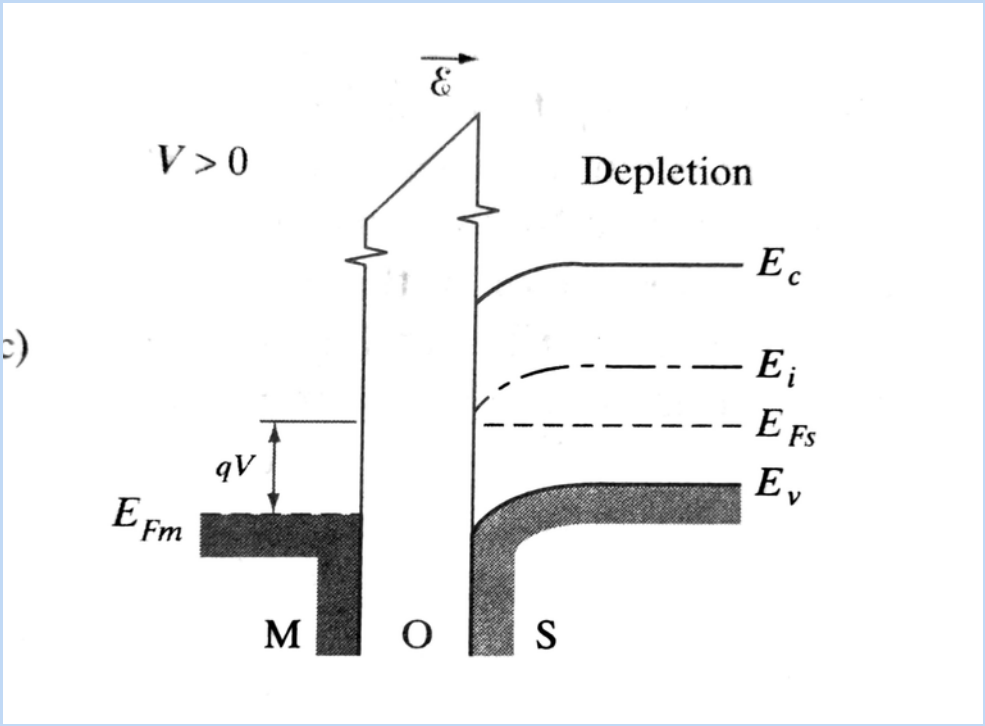
this implies,

$$Q_{sub} = Q_{acc} = -C_{ox}V_{GS}$$

5.2.2 Depletion mode

When a positive gate voltage is applied across the capacitor, the energy bands reverse their bending. The carrier population is now in quite a precarious situation- neither the

hole nor the electron concentration is considerable enough to contribute to any potential difference. The holes that have been pushed away from the insulator-substrate junction leave behind a depletion region, consisting of immobile acceptor atoms, similar to the depletion region of a p-n junction diode. Hence, this mode of operation is known as the 'depletion mode'.



Energy band diagram at depletion mode

Q_{sub} is now effectively equal to Q_{dep}, i.e. the charge enclosed within the depletion region formed. Q_{dep} is given by-

$$Q_{dep} = -qN_A W_d,$$

where W_d is the depletion width. Now, V_{ox} is given by-

$$V_{ox} = -\frac{Q_{sub}}{C_{ox}} = \frac{qN_A W_d}{C_{ox}}$$

The relation between ϕ_s and W_d is the same as in the case of the p-n junction diode(since the relation is a purely electrostatic one, it holds whenever there's no current density present). Therefore,

$$\phi_s = \frac{qN_A W_d^2}{2\epsilon_s}$$

The final expression for V_{GS} is-

$$V_{GS} = \frac{qN_AW_d^2}{2\epsilon_s} + \frac{qN_AW_d}{C_{ox}}$$

One of the most important stages of the depletion mode of operation is the *threshold condition*. The threshold condition represents a transition stage between the depletion and inversion modes of operation. Under this condition, the intrinsic energy level is as much below the Fermi level as it is above it. let $q\phi_B$ represent the gap between intrinsic and Fermi energy levels in the bulk of the semiconductor. Then, ϕ_B is given by-

$$\phi_B = \frac{k_BT}{q} ln(\frac{N_A}{n_i})$$

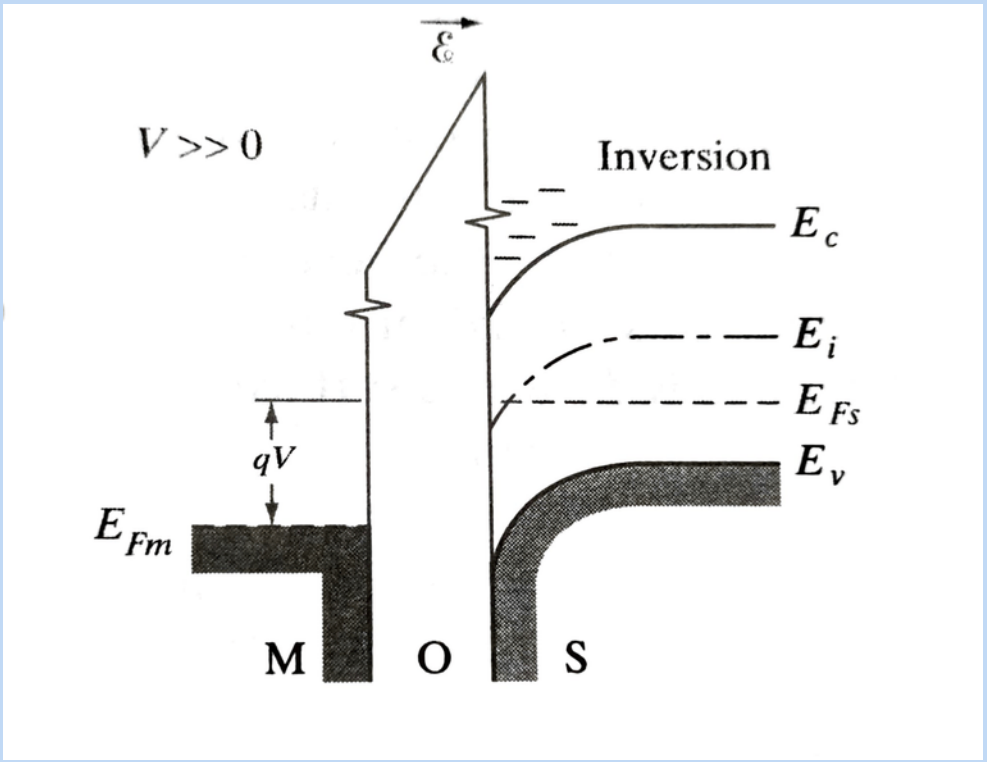
ϕ_s at threshold is thus given by $2\phi_B$. The gate voltage is given by,

$$V_{GS} = 2\phi_B + \frac{\sqrt{2qN_A\epsilon_s}}{C_{ox}}\sqrt{2\phi_B}$$

This particular value of gate voltage is a fundamental quantity for a MOS capacitor (and for that matter, a MOSFET too) and is known as the threshold voltage, V_T.

5.2.3 Inversion mode

The inversion mode of operation is when $V_{GS} > V_T$. Beyond this point, further increase in gate voltage leads to significant accumulation of electrons at the insulator-substrate junction. Thus, the substrate nearabout the junction is effectively 'inverted' from a p-type material to a n-type material. It is important to realize that this sheet of electrons shield off much of the electric field and therefore, any further increase in gate voltage does not affect the depletion region a lot. Rather, the excess of gate voltage from the threshold voltage only increases the inversion charge density.



Energy band diagram at inversion mode

The expression for V_{GS} is now given by-

$$V_{GS} = 2\phi_B - \frac{Q_{dep}}{C_{ox}} - \frac{Q_{IN}}{C_{ox}}$$

$$V_{GS} = V_T - \frac{Q_{IN}}{C_{ox}}$$

That is,

$$Q_{IN} = -C_{ox}(V_{GS} - V_T)$$

In theory, it is good enough to assume that the electrons in the inversion region are readily supplied by the p-type substrate. Practically, however, due to low minority carrier concentration, it can be upto several minutes before the electrons are generated thermally to be supplied. The MOSFET overcomes this problem by having n^+-type material attached to the p-type substrate to provide the electrons readily.

Although we have ignored the presence of inversion charges in the neighbourhood of threshold voltage, it is not entirely correct to do so. It is quite true that the effects of the inversion charge density are significant only after the threshold voltage. However, it

is also true that the inversion charge density should also be present at voltages below the threshold voltage (albeit of very low magnitude). This free charge is responsible for *subthreshold conduction* in the MOSFET. It is essential to consider the subthreshold mode of operation in very low-power applications, where we want the turn-on time of the electronic device to be very fast. The subthreshold mode of operation is what controls this time.

5.3 Flat-band condition and flat-band voltage

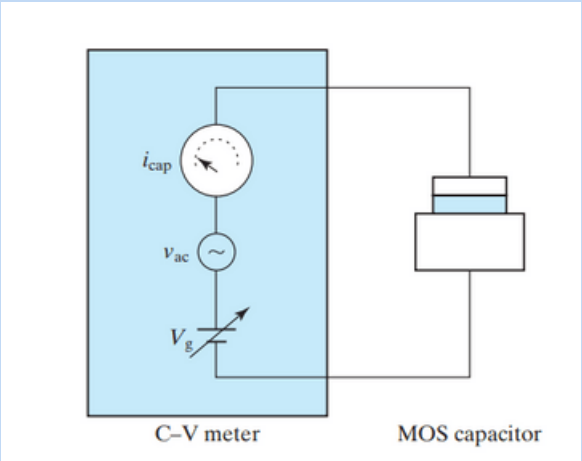
Until now, we have assumed that the Fermi levels of the metal and semiconductor are inherently equal. However, this assumption is not true practically. When the Fermi levels of the metal and substrate are not equal, on joining them, we get a built-in potential across the oxide layer because of the band-bending that occurs when the semiconductor Fermi level tries to align itself with the metal Fermi level.

In order to restore the bands to equilibrium, we need to apply a gate voltage equal to the Fermi-level energy difference, ϕ_{ms} (i.e. $\phi_m - \phi_s$)-. So, our previous equations have to be modified accordingly to take this factor into account. Since the effective gate voltage reduces by ϕ_{ms} due to band-bending, we can simply replace gate voltage V_{GS} by $V_{GS} - V_{FB}$, where V_{FB} is known as the flat-band voltage and $V_{FB} = \phi_{ms}$.

5.4 MOS C-V characteristics

The MOS C-V characteristics are measured w.r.t to a small-signal AC source, which is superimposed on a much larger DC bias voltage. The capacitance is thus, defined as-

$$C = \frac{dQ_{GS}}{dV_{GS}} = -\frac{dQ_{sub}}{dV_{GS}}$$



Setup for the C-V measurement

(a) Accumulation mode

In accumulation mode, the substrate charge is given by-

$$Q_{acc} = -C_{ox}(V_G - V_{FB})$$

Going by the formula then, we have,

$$C = C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$$

Note that the capacitance under consideration is *capacitance per unit area*.

(b) Depletion mode

In depletion mode, the substrate charge is given by-

$$Q_{sub} = -qN_AW_d$$

The gate voltage is given by-

$$V_{GS} = V_{FB} + \frac{qN_AW_d^2}{2\epsilon_s} + \frac{qN_AW_d}{C_{ox}}$$

Thus, we obtain-

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{W_d}{\epsilon_s} = \frac{1}{C_{ox}} + \frac{1}{C_{dep}}$$

A bit of algebraic manipulation yields-

$$\frac{1}{C} = \sqrt{\frac{1}{C_{ox}^2} + \frac{2(V_{GS}-V_{FB})}{q\epsilon_s N_A}}$$

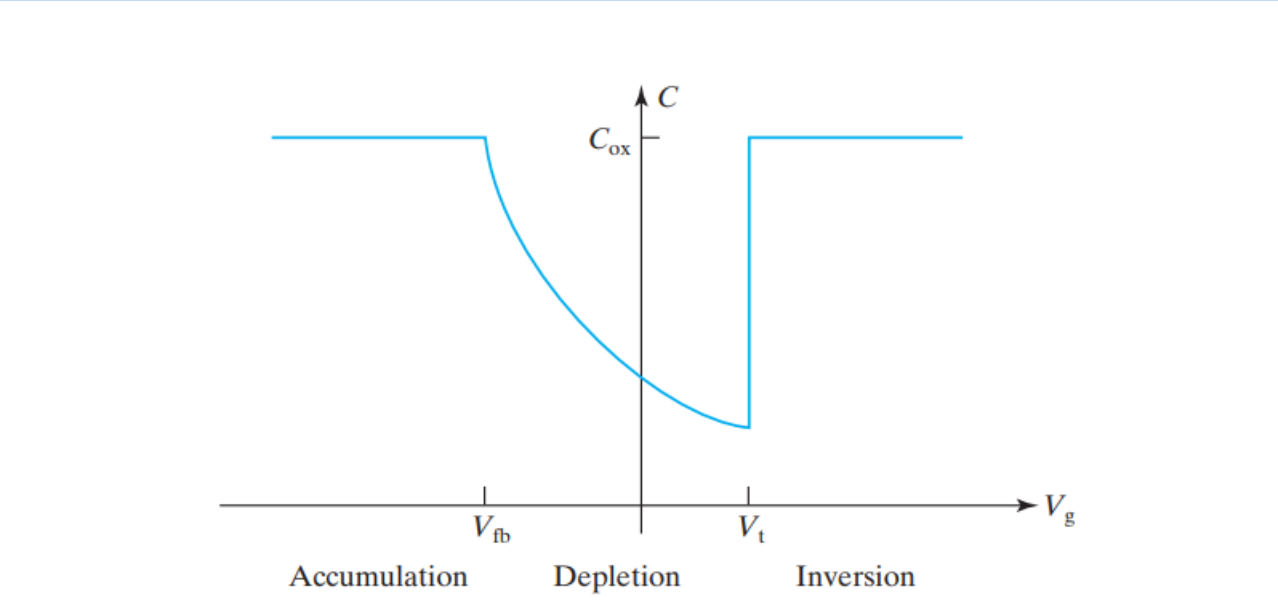
(c) Inversion mode

In inversion mode, as we have seen earlier, the inversion charge is slow to change in a MOS capacitor. So, it is only under the influence of a low-frequency signal that the inversion charge layer can oscillate with time. In such a case, the capacitance of the system will be equal to the oxide capacitance (since $Q_{IN} = -C_{ox}(V_{GS} - V_T)$). At high-frequency, the inversion charge remains constant whereas the charge inside the depletion region oscillates instead. In this case, the capacitance is given by-

$$C = -\frac{dQ_{dep}}{dV_{GS}}$$

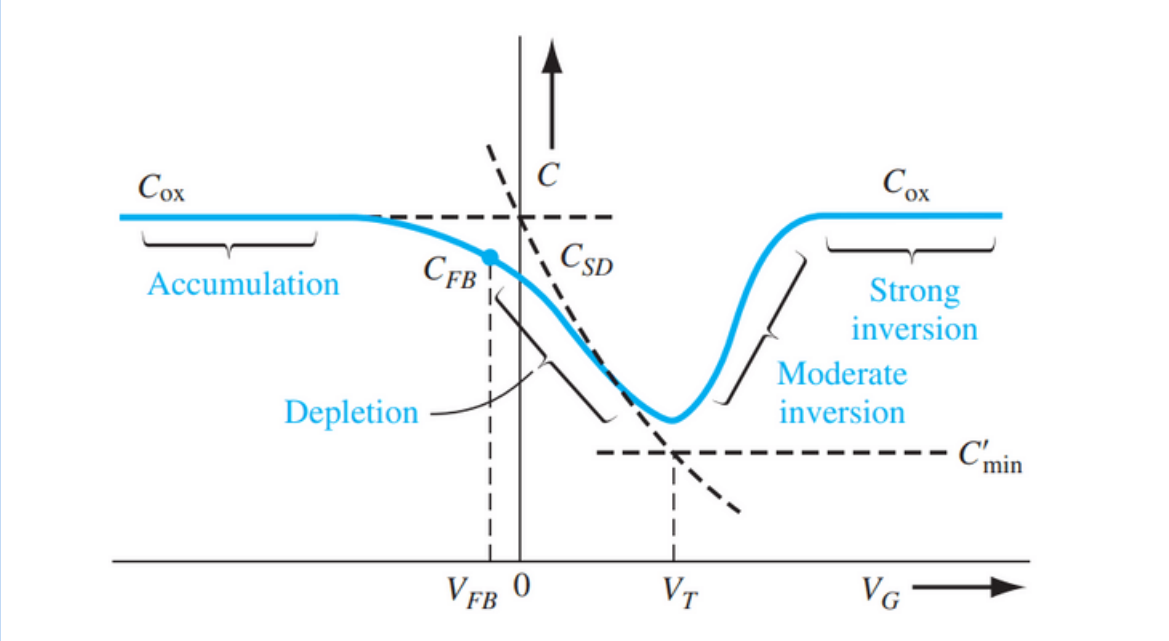
The capacitance in this case is the same as the depletion-mode capacitance at $V_{GS} = V_T$ (since we know that the depletion width is constant beyond V_T).

The following graph sums up the MOS C-V characteristics in different modes of operation:



Ideal MOS C-V characteristics

The graph obviously depicts an ideal MOS capacitor. Practically, the graph is more complicated since many more factors have to be taken into account.

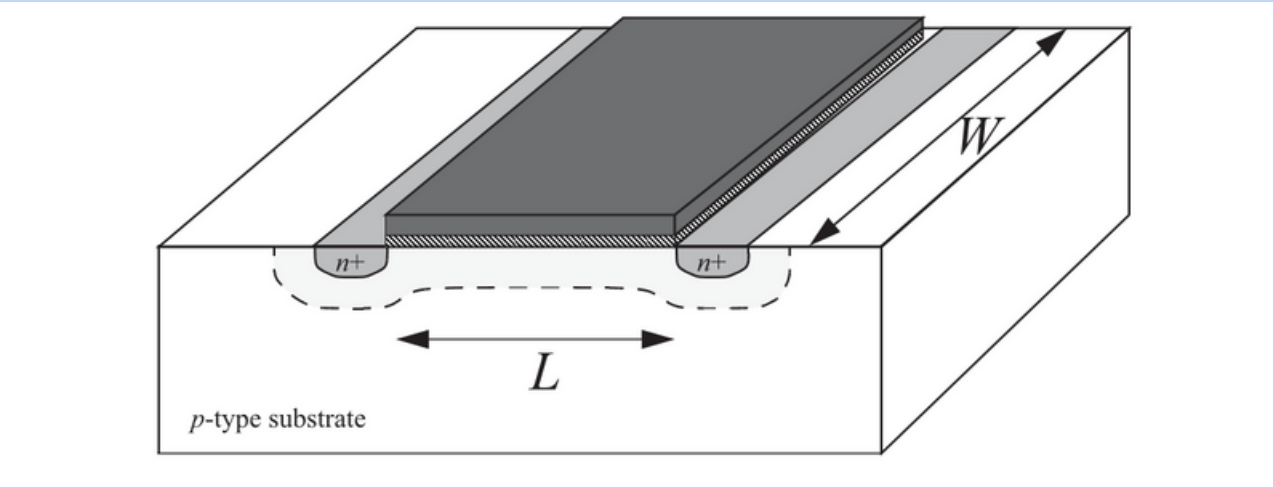


Experimental MOS C-V characteristics

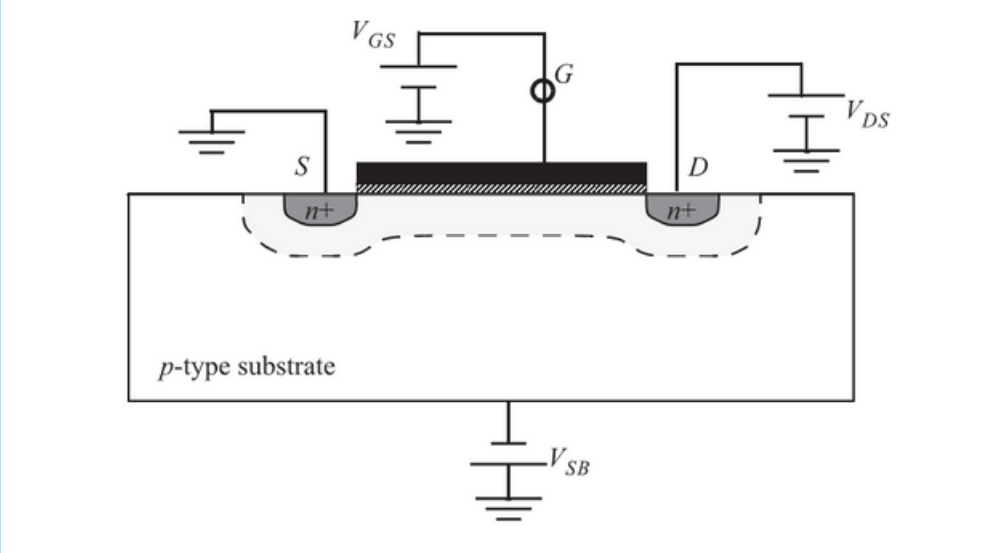
Here is a highly instructive gif from Wikipedia depicting experimental variation of MOS capacitor with gate voltage against varying oxide thickness.

5.5 MOSFET transistor

According to the definition, a transistor is a semiconductor device capable of amplifying or switching electrical signals and power. In this light, the MOS capacitor clearly provides us an opportunity to create a transistor since we can turn on or off a current (which is attributable to the inversion charge), by simply modulating the gate voltage, V_{GS}.



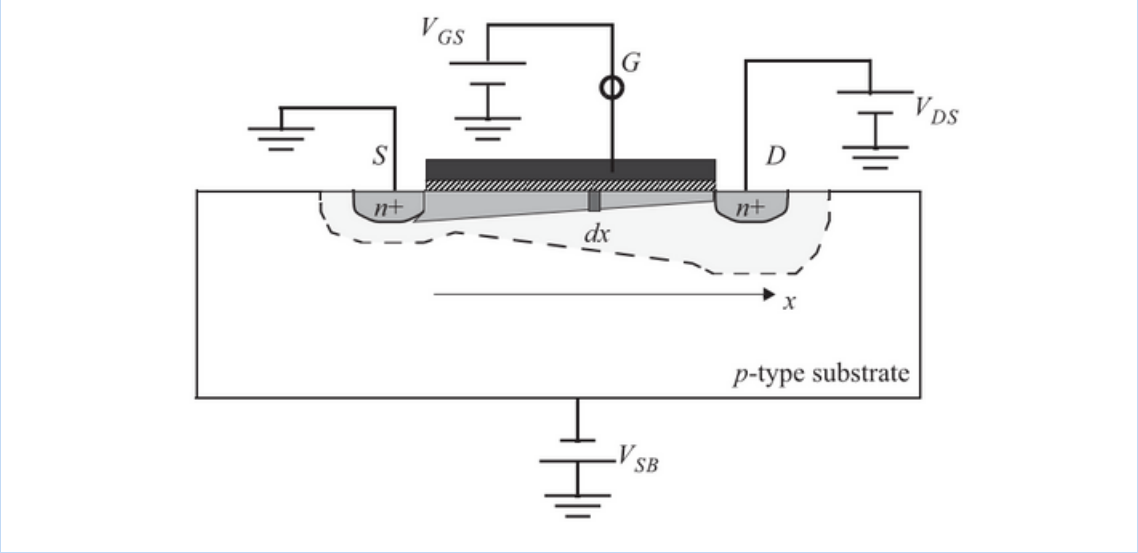
Layout of a MOSFET



MOSFET structure

Some of the basic differences between a MOS capacitor and a MOSFET are that while the MOS capacitor has no external source of minority carriers, the MOSFET has 'source' and 'drain' n^+ regions, which readily supply electrons. Thus, the MOSFET operates much faster in the inversion mode as compared to the MOS capacitor. Another important difference is the presence of an additional voltage source V_{SB}, which allows us to change the source-bulk voltage independent of the drain voltage or the gate voltage. This extra source is used to modulate the width of the depletion region and hence, the threshold voltage V_T.

The MOSFET I-V characteristics are derived by the Gradual Channel Approximation (GCA). The GCA states that as compared to the voltage variation along the y-axis (the direction perpendicular to the gate in the shown figure), the voltage variation along the x-axis (the direction from the sourcse to the gate) is quite slower. The inversion charge density acts like a sheet of charge along the x-axis and is also known as 'channel'. In a MOSFET, what happens physically is that on application of a drain voltage, V_{DS} (while keeping $V_{GS} > V_T$), the electrons in the channel move from the source to the drain by drift processes (since the voltage changes very slowly along the channel, it also implies that the electron density also changes very slowly - hence the diffusion component can be ignored). The problem of current conduction through the channel can be broken up into two parts - an electrostatic one where inversion and depletion charges are accumulated along the x-direction and a dynamic one where V_{DS} facilitates transport of electrons through channel.



The inversion charge distribution in the channel is such that the surface charge (w.r.t the oxide capacitance) can be expressed as $Q_{IN}(x) = \mathrm{n(x)}\Delta y$, where the volume charge density behaves as a kind of Dirac delta function. In the light of this argument, consider an elemental strip along the y-axis as shown in the above figure. The resistance of the , dR, is given by-

$$dR = -\frac{dx}{\mu_n Q_{IN}(x) W}$$

where W is the width of the channel, as shown in the structure of the MOSFET and μ_n is the mobility of electrons. The total channel charge, $Q_S(x)$ is given as $Q_D(x) + Q_{IN}(x)$. As per the voltage equations,

$$V_{GS} = V_{FB} - \frac{Q_S(x)}{C_{ox}} + \phi_s(x)$$

The term for ϕ_s now varies with position since there is an additional voltage drop in the channel now from drain to source due to the presence of V_{DS} which drives the current. This implies,

$$\phi_s(x) = 2\phi_B + V(x)$$

Therefore,

$$Q_D(x) = -\sqrt{2\epsilon_s q N_A (2\phi_B + V(x))}$$

Applying Ohm's Law to the differential strip under consideration,

$$dV = I_D dR$$

$$I_D dx = -Q_{IN}(x) \mu_n W dV(x)$$

On integrating,

$$I_D = \frac{\mu_n C_{ox} W}{L}([V_{GS} - V_T - \frac{V_{DS}}{2}]V_{DS} - \frac{2}{3}\frac{\sqrt{2\epsilon_s q N_A}}{C_{ox}}[(2\phi_B + V_{DS})^{3/2} - (2\phi_B)^{3/2}])$$

The term $\frac{\sqrt{2\epsilon_s q N_A}}{C_{ox}}$ is very small as compared to the other values and can practically be ignored without much loss of accuracy. The I-V characteristic then comes down to-

$$I_D = \frac{\mu_n C_{ox} W}{L}([V_{GS} - V_T - \frac{V_{DS}}{2}]V_{DS})$$

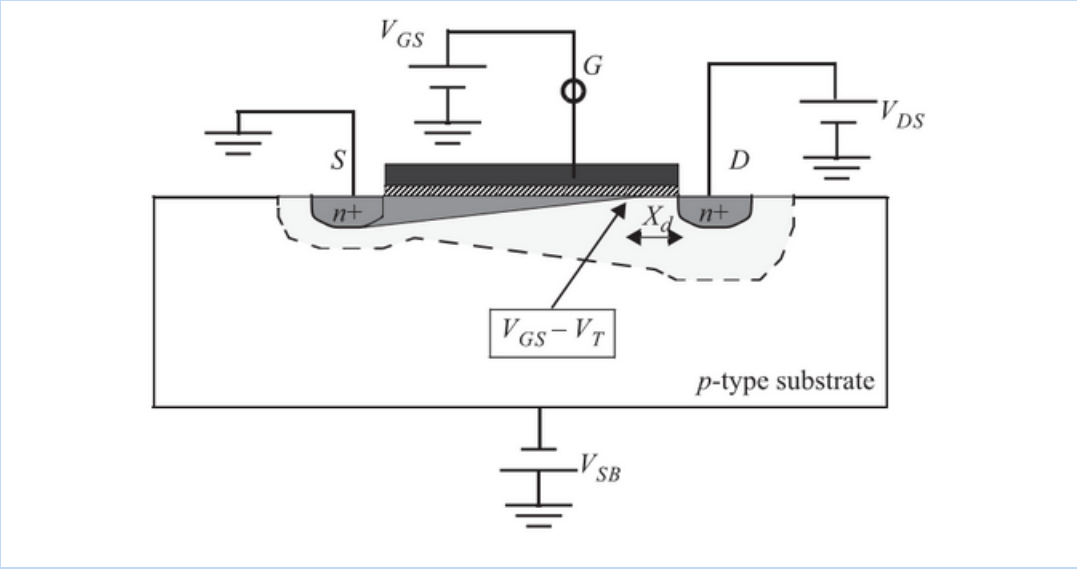
When the drain voltage is much smaller than $V_{GS} - V_T$, the I-V relation becomes linear-

$$I_D = \frac{\mu_n C_{ox} W}{L}([V_{GS} - V_T]V_{DS})$$

This represents the linear mode of operation of a transistor and the channel effectively behaves as an ohmic resistor. At moderate values of V_{DS}, we cannot neglect its value in the I-V relation and have to use the exact relation itself. At still higher values of drain voltages, we encounter the situation of what is known as the 'saturation' region of operation.

5.5.1 Saturation mode

As the gate voltage increases, at one point of time, it exceeds $V_{GS} - V_T$. In such a case, the inversion region in the vicinity of drain no longer exists and the channel is broken at that point. This point is known as *pinch-off* mode of operation. At still higher drain voltages, the channel ends at some point close to the drain, but is not extended all the way to it as shown:



Pinch-off mode of operation

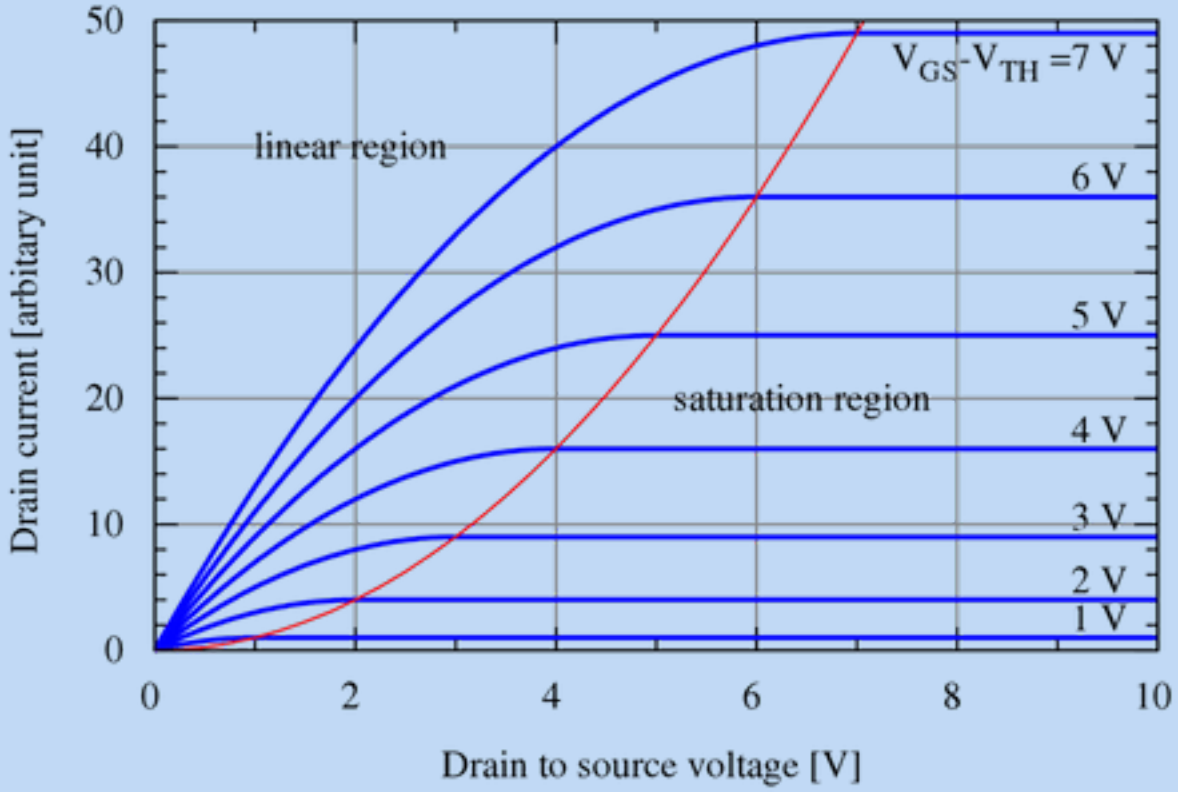
In saturation mode, the inversion region no longer exists after the pinch-off point. Despite this, it does not prevent the carriers from conducting current in this region. The conduction in this region happens through the depletion region surrounding the conductive channel and drain and source regions. The resistance of this region is much higher than the channel and is quite similar to that of silicon (since there are very few carriers in the depletion region). Hence, most of the drain voltage drops across this region as compared to the highly conductive channel. Now, the distance of the drain from the pinch-off point is proportional to the drain voltage. So, the resistance offered by this region is thus, proportional to the drain voltage - which implies that the saturation current is *independent* of the drain voltage, as observed experimentally. The saturation current is given by the value of drain current at the onset of saturation, i.e. when $V_{DS} = V_{GS} - V_T$.

$$I_{D,sat} = \frac{\mu_n C_{ox} W}{2L}(V_{GS} - V_T)^2$$

At saturation mode, GCA no longer holds true. However, the basic equation governing I-V characteristics remain unchanged and they can be solved in the same manner as we

followed while deriving I-V characteristics for the MOSFET (for more clarity on this, refer to Problem 3(f)).

The I-V characteristics of the MOSFET are summed up by this graph-

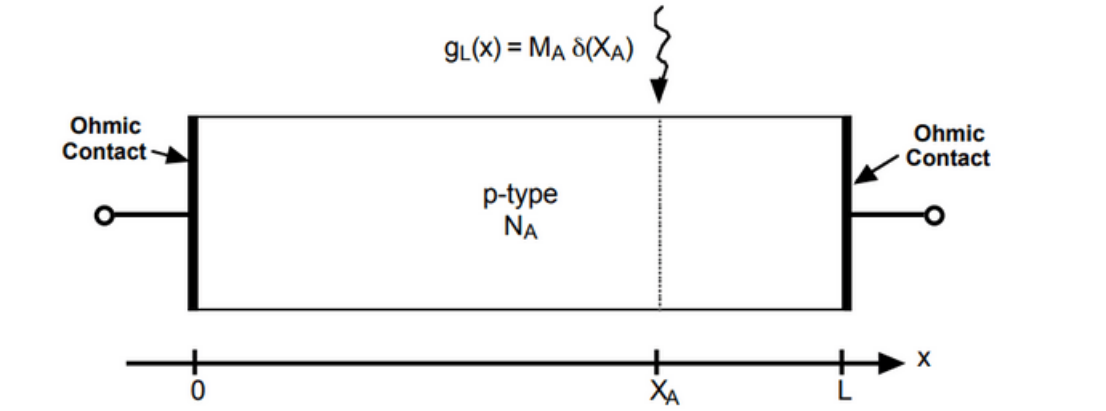


MOSFET I-V characteristics

6 Selected Problems

6.1 Problem 1

A p-type semiconductor sample with acceptor concentration N_A, and length L, illustrated below, has ohmic contacts at both its ends. A light source generates M_A electron-hole pairs/cm^2-s in the plane at x = X_A,i.e. $g_L(x) = M_A\delta(X_A)$. Assume low-level injection and quasi-neutrality everywhere in the bar.



The general equation governing the excess minority carriers in a uniformly doped material is

$$\frac{d^2n'(x)}{dx^2} - \frac{n'(x)}{L_e^2} = -\frac{1}{D_e}g_L(x)$$

(a) What boundary condition is imposed on the excess minority carriers n' at x = 0 and x = L ?

(b) We now make the assumption that the minority carrier lifetime is very long, which simplifies the general equation to:

$$\frac{d^2n'(x)}{dx^2} \approx -\frac{1}{D_e}g_L(x)$$

What quantitative restriction is placed on the minority carrier lifetime, τ_e, for this assumption to be valid ?

(c) Using the long-lifetime approximation in part (b), determine two constraints(i.e. boundary conditions) on the excess minority carriers at x = X_A, i.e. relating $n'(X_{A^-})$ to $n'(X_{A^+})$.

(d) Sketch the excess minority carrier concentration, $n'(x)$ and the minority carrier diffusion current, $J_{e,diff}(x)$ everywhere inside the semiconductor.

(e) A second light source is added illuminating a single spot along the semiconductor at x = X_B, where $X_B > X_A$, and generating electron-hole pairs at a rate M_B, so that $g_L(x)$ is now

$$g_L(x) = M_A\delta(X_A) + M_B\delta(X_B)$$

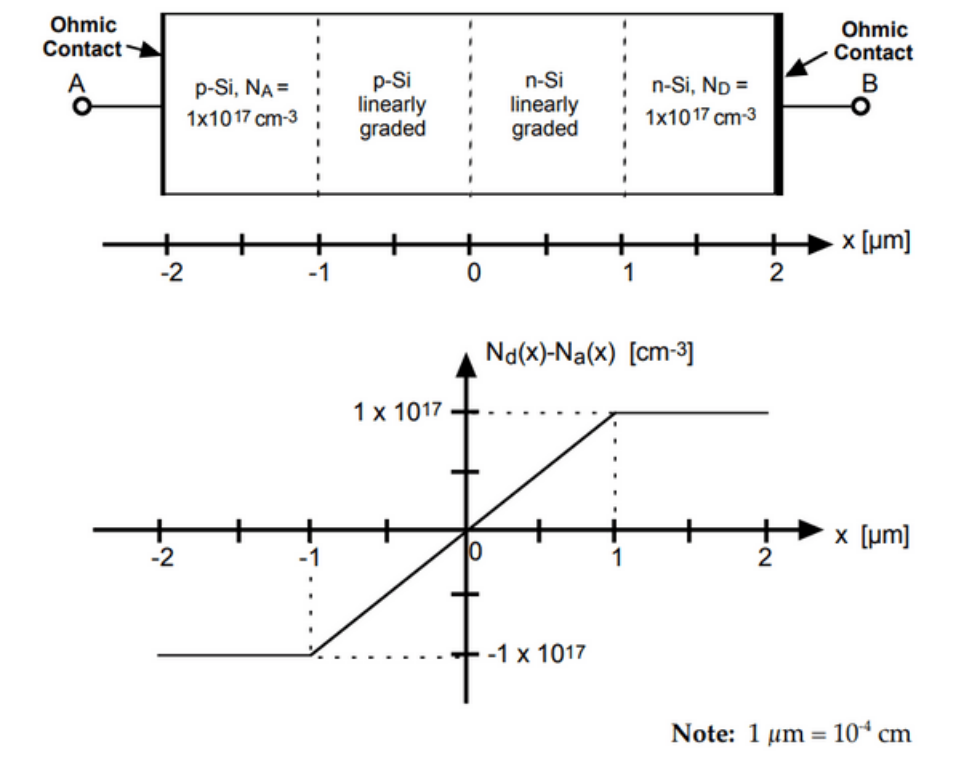
Find $n'(x)$ and $J_{e,diff}(x)$ under this new illumination condition.

[Courtesy : MIT OCW]

Solution

6.2 Problem 2

Consider the silicon diode pictured below. It is $4\mu m$ long, with ohmic contacts at each end, and it is uniform p-type with N_A = 1 x 10^{17} cm^{-3} for 1 μm on its far left end and uniform n-type with N_D = 1 x 10^{17} cm^{-3} on its far right end. In between these two uniformly doped regions, the net concentration, $N_d(x) - N_a(x)$, slowly grades linearly over a distance of 2 μm from -1 x 10^{17} cm^{-3} on the left to 1 x 10^{17} cm^{-3} on the right, as shown in the lower figure.



(a) In thermal equilibrium, what is the electrostatic potential, $\phi(x)$, in the left-hand quasi-neutral region at x = -1.5 μm, and what is the electrostatic potential, $\phi(x)$, in the right-hand quasi-neutral region at x = +1.5 μm, and what is the built-in potential step, $\Delta\phi_b$, seen transiting from x = -1.5 μm to x = +1.5 μm ?

(b) For the rest of this problem, a bias voltage, V_{AB}, is applied to this diode resulting in a total depletion region width of 1 μm, and $x_N = |x_P| = 0.5$ μm. Sketch and label the net charge density, $\rho(x)$ and the electric field, E(x), for -2$\mu m < x < 2\mu m$.

(c) What is the change in potential, $\Delta\phi$, transiting the depletion region when the bias is the same as in Part b, i.e. what is ϕ(0.5 μm) – ϕ(-0.5 μm) ?

(d) What is the change in potential, $\Delta\phi$, in transiting the quasi-neutral n-type graded region between x = 0.5 μm and x = 1.0 μm, i.e. what is ϕ(1.0 μm) – ϕ(0.5 μm) ?

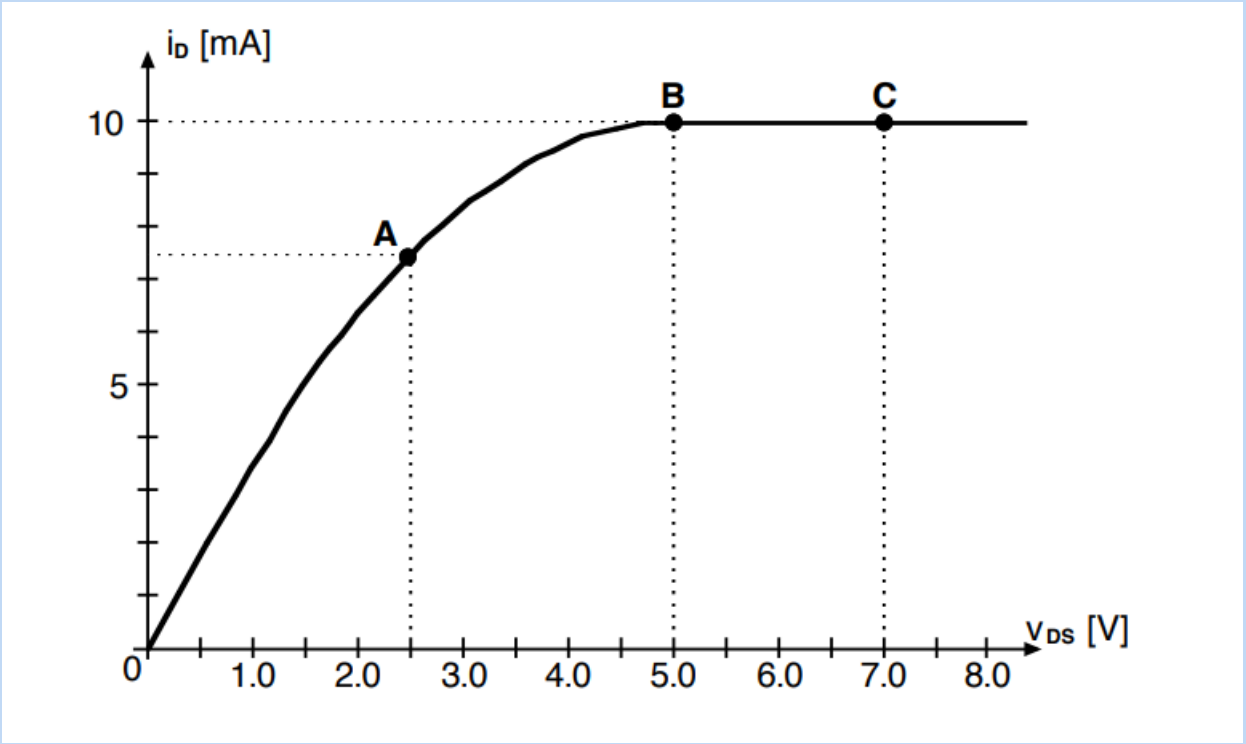
(e) What is the applied bias voltage, V_{AB} ?

[Courtesy : MIT OCW]

Solution

6.3 Problem 3

The $I_D - V_{DS}$ plot for an ideal n-channel MOSFET is shown below. The substrate bias, V_{BS}, is 0 V, the saturation current, I_{Dsat}, is 10 mA, and the saturation voltage, $V_{DS,sat}$,is 5 V. For this device t_{ox} = 10 nm, ϵ_{ox} = 3.5 x 10^{-13} F/cm, W = 50 μm, and L=10 μm.



(a) Given that V_T = 1 V, what is the gate voltage V_{GS} that must be applied to obtain the characteristic shown above?

(b) What is the slope, $\frac{dI_D}{dV_{DS}}$ of the characteristic at $V_{DS} = 0$ V ? Make sure you provide a formula as well as a value so that your answer is independent of the correctness of your Part (a).

(c) Assuming that V_T is independent of position in the channel, calculate the inversion layer sheet charge density, $q_N^*(y)$ corresponding to Bias Point A (i) adjacent to the source (the source end, y = 0) and (ii) adjacent to the drain (drain end, y = L).

(d) Calculate the electron drift velocity, $s_{e-Drift}$, at the (i) source end and (ii) drain end of the channel at Bias Point A.

(e) The transistor enters saturation at Bias Point B and simple theory suggests that the drain-end charge has become 0 while the drain-end velocity is infinite, so that the I_{Dsat} can flow in that part. Now, assuming instead that the electrons at the drain end move at their saturation velocity, $s_{sat} = 10^7$ cm/s, what is the channel charge density that must exist there to support the I_{Dsat} ?

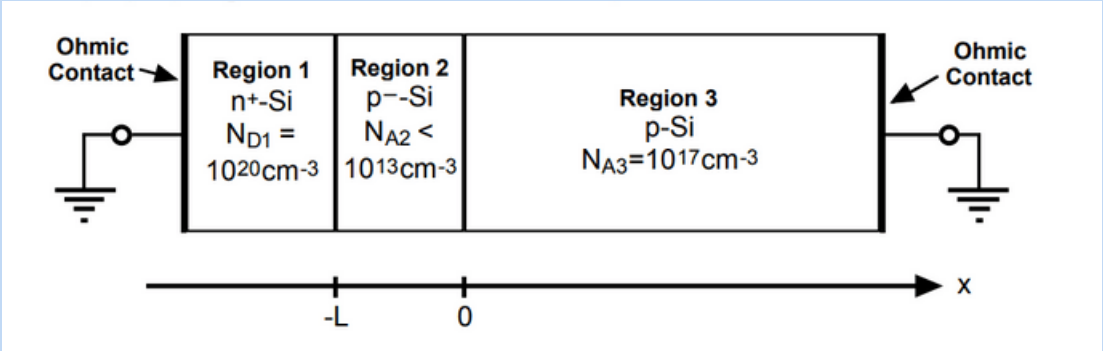
(f) Assuming V_{CS}(y) is the voltage that a hypothetical voltmeter would measure between the inversion layer at position y along the channel and the source. Derive an expression that could be solved for V_{CS}(L/2), i.e. at distance L/2 from the source in a device biased at Bias Point C, in terms of the transistor parameters and I_{Dsat}.

[Courtesy : MIT OCW]

Solution

6.4 Problem 4

Consider the $n^+ - p$ diode shown below in thermal equilibrium. The n^+ doping in Region 1 is high enough that it is degenerate, with $\phi_{n^+} = 0.55$ V. Assume that the depletion region on the n^+ side of the junction is negligibly small so that you may also assume negligible change in potential across it. Note too that Region 2, the p^- Si region, is very lightly doped (i.e. $N_{A2} < 10^{13}\ cm^{-3}$). Finally, L = 0.2 μm (= 2 x 10^{-5} cm). [ϕ_{n^+} refers to the energy difference between intrinsic band and Fermi level in Region 1]



(a) Calculate $\Delta\phi_{13}$, the built-in potential difference between the QNR (quasi-neutral region) in Region 1 (i.e. where x << -L) and the QNR in Region 3 (where x >> 0).

(b) Sketch and label $\rho(x)$, the net charge density, and E(x), the electric field, in the structure. You should label the depletion region width in Region 3 as x_D (you are not expect to find a value for x_D). You may assume that the width of the depletion region in Region 1 is negligibly small (so that there is essentially an impulse of charge at x = -L).

(c) Set up an equation with x_D, the width of the depletion layer in Region 3, being the only unknown, that you could use to find x_D.

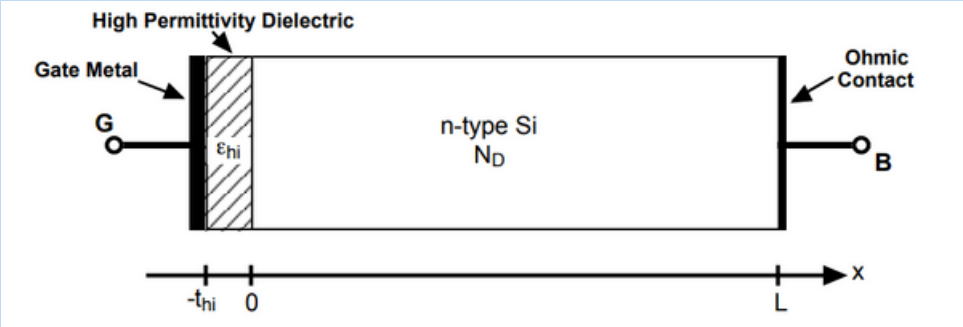
(d) Give an expression for C^*_{dp}, the depletion capacitance per unit area of this device at zero bias, V = 0 volts. x_D can appear as a parameter in your expression.

[Courtesy : MIT OCW]

Solution

6.5 Problem 5

Alice is a process engineer experimenting with a new high-permittivity dielectric material with a dielectric constant, ϵ_{hi}, that is 5 times as large as the dielectric constant of SiO_2, i.e. $\epsilon_{hi} = \epsilon_{ox}$. She chooses to test the material by fabricating p-MOS capacitors on n-type silicon, and to use a metal for the gate and contact for which ϕ_m = - 0.5 ϕ_n. Her structure is illustrated below; it also includes an adjacent p^+ region shorted to the substrate (not shown in the figure) to supply holes when an inversion layer is formed.



(a) Sketch the electrostatic potential, $\phi(x)$, through the device from G to B (i.e. starting in the gate metal and going into the ohmic contact metal) in flatband when $V_{GB} = V_{FB}$. Label all relevant features on your plot, including values for $\phi(0)$, depletion region width, and potential drop across the oxide. Finally, derive an expression for the flatband voltage, V_{FB}.

(b) At flatband, i.e. with $V_{GB} = V_{FB}$, what are the electron and hole concentrations, n(x = 0^+) and p(x = 0^+), at the silicon-dielectric interface?

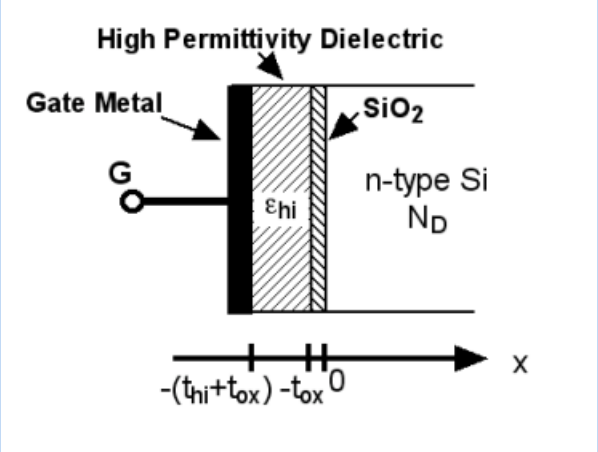
(c) Sketch the electrostatic potential, $\phi(x)$, and the charge distribution, $\rho(x)$, through the device from G to B at the onset of inversion, i.e. when $V_{GB} = V_T$. Label all relevant features on your plots, including values for $\phi(0)$, depletion region width, and potential drop across the oxide. Finally, derive an expression for the threshold voltage, V_T.

(d) At the onset of inversion, i.e., when $V_{GB} = V_T$,what are the electron and hole concentrations, n(x = 0^+) and p(x = 0^+), at the silicon-dielectric interface?

(e) A practical problem with depositing a dielectric other than SiO_2 directly on silicon is that new energy states and/or fixed sheet charge are introduced at the interface.

Imagine that the latter occurs, and that there is a fixed positive sheet charge density, σ_i, at the interface. Assuming that this charge can be modeled as an impulse of charge of intensity σ_i at x = 0, i.e. $\rho(x) = \sigma_i \delta(x)$, calculate the changes in the flatband voltage, V_{FB}, and in the threshold voltage, V_T, resulting from the presence of this charge.

(f) To eliminate the interface charge, a very thin layer of silicon dioxide, SiO_2, can be grown on the silicon before the high permittivity dielectric is deposited, as illustrated in the figure below. How much is the gate dielectric capacitance, C_G, changed relative to its original value in Part (a) by the addition this SiO_2 layer if $t_{ox} = 0.2t_{hi}$?



[Courtesy : MIT OCW]

Solution

7 Formula Sheet

Keeping in view the difficulty in dealing with the large number of formulae involved in this project, here is a formula sheet compiled by me to serve as a quick reference.

8 References

1. Ben G. Streetman, Sanjay Kumar Banerjee. Solid State Electronic Devices, Pearson Publications, 2018.

2. Tak H. Ning, Yuan Taur. Fundamentals of Modern VLSI Devices, Cambridge University Press, 2009.

3. Chenming Calvin Hu. Modern Semiconductor Devices for Integrated Circuits, Pearson Publications, First Edition.

4. Prof.Digbijoy N. Nath. NPTEL course-Fundamentals of semiconductor devices. IISc Bangalore.

5. Prof.Clifton Fonstad. MIT OpenCourseWare course - Microelectronic Devices and Circuits.

6. Prof.Ali Hajimiri. CHIC, Caltech

7. Wikipedia.org

8. Hyperphysics

9. Physics Stack Exchange

10. Electrical Engineering LibreTexts