PN Junctions

PN junctions form the basis of most semiconductor devices. Understanding their operation is basic to understanding most devices.

Fabrication

PN junctions are normally fabricated by solid state diffusion. The two "simple" impurity profiles that result from this process are the complementary error function (erfc) and Gaussian. These result from the boundary conditions of constant surface concentration (N_0) and constant dose (N' or Q) as shown below.

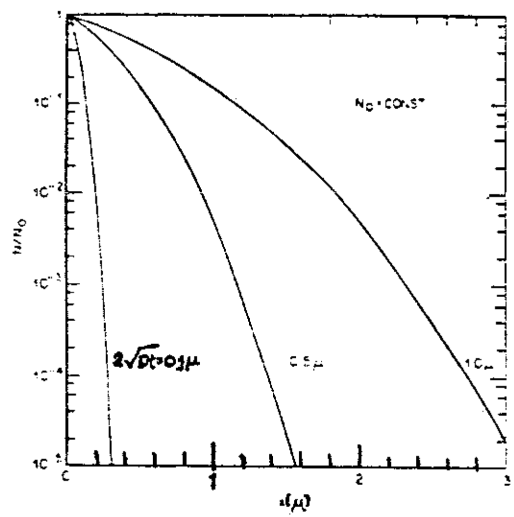


Fig. 5.21 Complementary error function profile

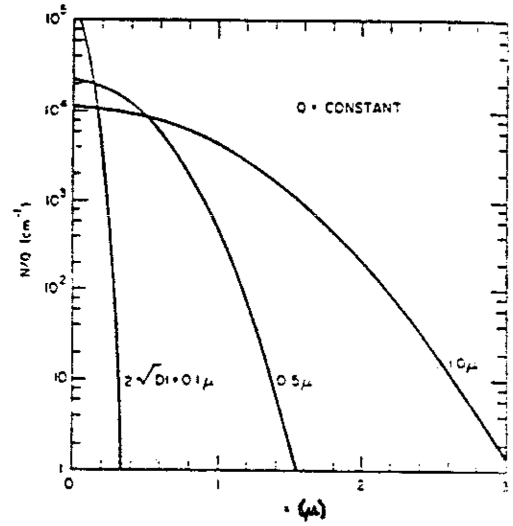
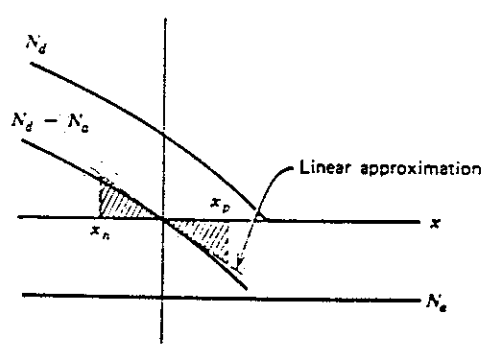
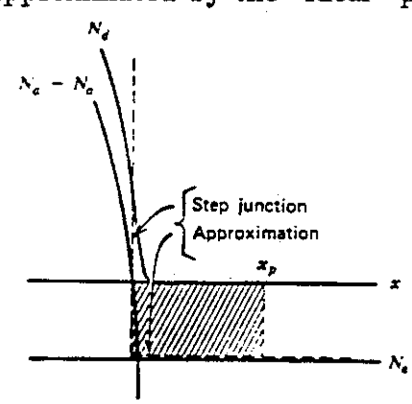


Fig. 5.22 Gaussian impurity profile

Often, to simplify the math, these profiles are approximated by the "ideal" profiles shown below.





Shallow diffusions: "Predeps" + short drive-in $\Rightarrow$ STEP JUNCTION

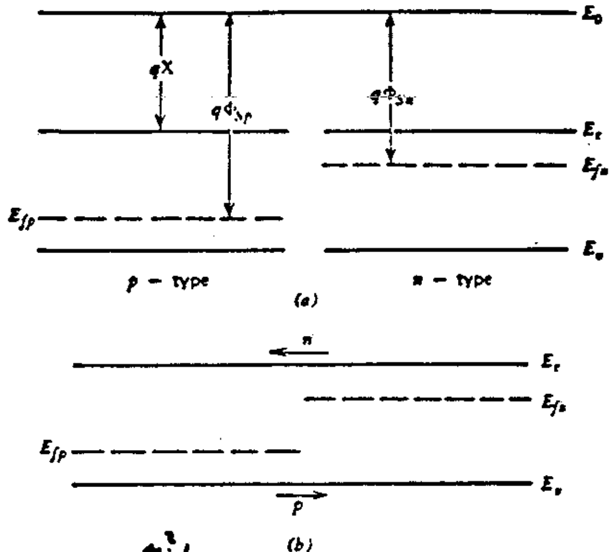
Deep diffusions: Dt(Drive-in) $\gg$ Dt(Predep) $\Rightarrow$ LINEARLY GRADED JUNCTION

(Cross over between the two occurs at $x_j = 2\mu$m as a rule of thumb). We shall consider these two cases in our analysis although real junctions are often in between the two.

Band Diagrams at Equilibrium

If N and P type materials are brought into contact, establishment of equilibrium is analagous to the Schottky diode we previously considered with the following fundamental differences:

- Both sides of the junction will generally be depleted.
- "Surface States" are usually not important (single crystal).
- Both electrons and holes usually move to establish equilibrium.



$\chi =$ electron affinity $\equiv E_0 - E_c$

$\phi_s =$ semiconductor work function

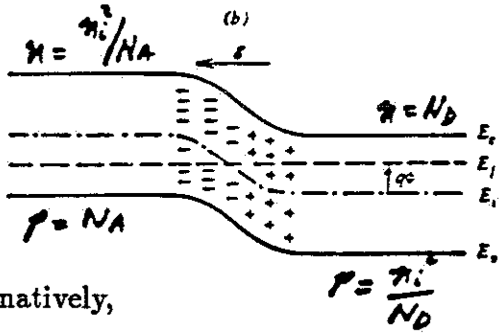
$\phi_{sp} = E_0 - E_{fp}$

$\phi_{sn} = E_0 - E_{fn}$

electrons flow: $N \to P$

holes flow: $P \to N$

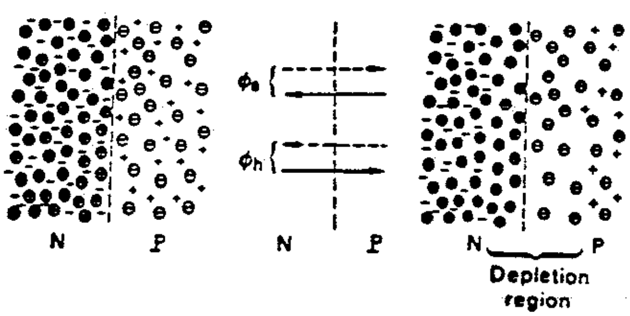
- **At Equilibrium:** $E_f =$ CONSTANT



Alternatively,

$$\text{Initially}\begin{cases}\text{electrons are in high concentration on N side}\\ \text{holes are in high concentration on P side}\end{cases}$$

Therefore, since diffusion of carriers occurs away from regions of high concentration, diffusion leaves behind uncompensated N_d^+ and N_a^-. Resulting space charge sets up electric field.



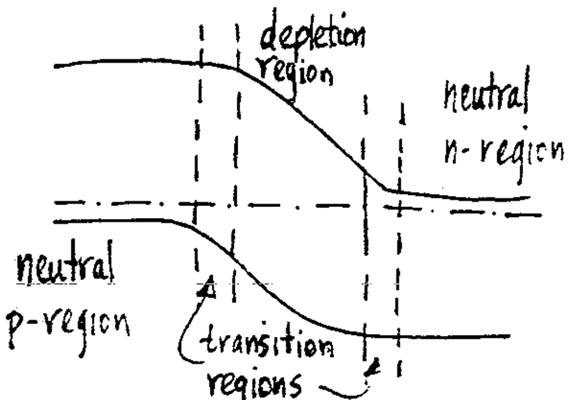
(a) (b) (c)

Electric field tends to pull electrons and holes back to their original positions. **Equilibrium when drift = diffusion.**

The general relationship between the charge distribution and the potential is given by Poisson equation:

$$\frac{\partial^2 V}{\partial x^2} = \frac{q}{K\epsilon_0}\left[(n-p)-(N_D-N_A)\right] \tag{1}$$

The electron and hole concentrations are given by



$$n = n_i \exp\left(\frac{E_f - E_i}{kT}\right) \tag{2}$$

$$p = n_i \exp\left(\frac{E_i - E_f}{kT}\right) \tag{3}$$

These equations are applicable in the neutral, depletion and transition regions under equilibrium (no applied bias, etc.).

Neutral regions:

N region: $n \cong (N_d^+ - N_a^-)$, $p \cong {n_i}^2/n$

$$n - p - N_D + N_A = 0 \quad \text{(no net charge)}$$

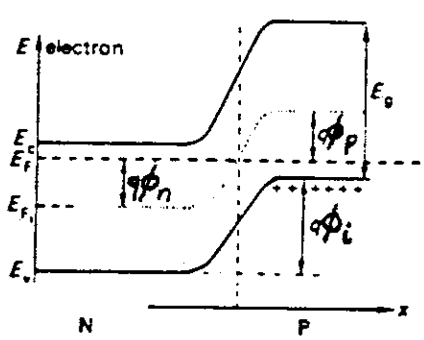
$$\frac{\partial^2 V}{\partial x^2} = 0 \quad \text{(the potential is constant in the N region)}$$

$$\frac{E_f - E_i}{q} \equiv \phi_n = \frac{kT}{q}\ln\frac{N_d}{n_i} \quad \text{(from (2))} \tag{4}$$

Similarly in the neutral P region,

$$\frac{E_i - E_f}{q} \equiv \phi_p = \frac{kT}{q}\ln\frac{N_a}{n_i} \tag{5}$$

Thus we have:



$$\phi_i \equiv \text{built in potential} = \phi_p + \phi_n$$

$$\phi_i = \frac{kT}{q}\ln\frac{N_d N_a}{{n_i}^2} \tag{6}$$

This is the built-in potential which exists across a PN junction with no applied bias (thermal equilibrium). Polarity: N(+), P(−).

Notes:

- An impurity doping change N^+ to N^- or P^+ to P^- results in a built in potential.
- The built in potential across a PN junction increases as N_d or N_a increases.

Depletion Regions

If this region is nearly completely depleted of carriers, then

$$\frac{\partial^2 V}{\partial x^2} \cong \frac{q}{K\epsilon_0}(N_a^- - N_d^+). \tag{7}$$

Solution of this equation requires some knowledge about the spatial variation of N_d and N_a. We will consider this shortly.

Transition Regions

In the transition regions between the depletion and neutral regions, neither our depletion nor neutrality approximations are valid and we have

$$\frac{\partial^2 V}{\partial x^2} = \frac{q}{K\epsilon_0}\left[(n-p)-(N_d-N_a)\right],$$

$$n = n_i \exp\left(\frac{E_f - E_i}{kT}\right),$$

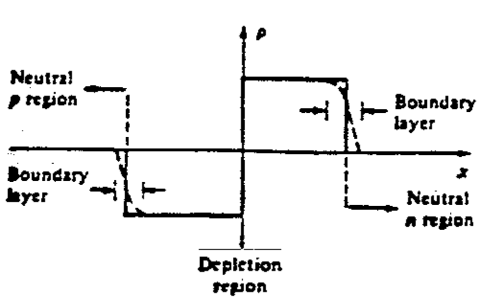
$$p = n_i \exp\left(\frac{E_i - E_f}{kT}\right).$$

General solutions across the transition regions require computer techniques. Using such techniques, the transition region is found to be $\approx 3L_D$ where L_D is the extrinsic Debye length and is given by

$$L_D \equiv \left[\frac{K\epsilon_0 kT}{q|N_d - N_a|}\right]^{1/2} \tag{8}$$

L_D may be thought of physically as the distance over which a fixed charge (N_d^+ or N_a^-) exhibits an influence over mobile carrier concentrations. In this instance it is a measure of the abruptness of the transition region between the depletion regions and the neutral regions.

Example: N_d or $N_a = 10^{16}\text{cm}^{-3}$ results in $L_D \approx 300\,\text{Å}$ and $3L_D \approx 900\,\text{Å}$. As long as the depletion region is wide compared to this length, the abrupt transition approximation will be approximately valid.



In summary, the charge distribution in the semiconductor is given approximately by

$\rho \approx 0$ outside depletion region.

$\rho \approx q(N_d^+ - N_a^-)$ within depletion region.

Boundary layer distance $\approx 3L_D$

Step Junctions

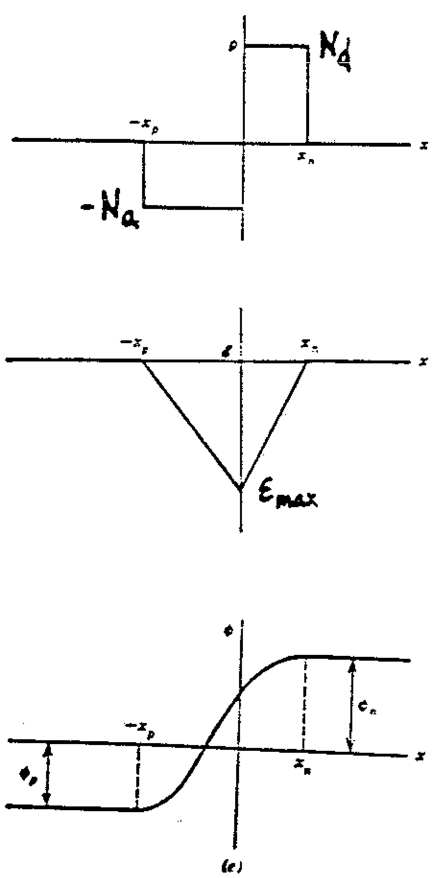
To proceed, we make some assumptions:

1. Step Junction (abrupt transition):

    - N_d constant in N material.
    - N_a constant in P material.

2. Depletion Approximation (ignores boundary or transition regions):

$$\rho = \begin{cases} -qN_a, & -x_p < x < 0 \\ qN_d, & 0 < x < x_n \\ 0 & x < -x_p,\ x > x_n \end{cases} \tag{9}$$



Therefore,

$$\frac{\partial^2 V}{\partial x^2} = -\frac{\rho}{K\epsilon_0} = \begin{cases} -\dfrac{qN_d}{K\epsilon_0}, & 0 < x < x_n \\ \dfrac{qN_a}{K\epsilon_0}, & -x_p < x < 0 \end{cases} \tag{10}$$

From neutrality,

$$N_a x_p = N_d x_n = Q/q \tag{11}$$

where the total depletion layer width is given by $x_d = x_n + x_p$.

Taking the N side of the junction as an example, we have

$$\mathcal{E} = -\frac{\partial V}{\partial x} = \int_{x_n}^{x} \frac{qN_d}{K\epsilon_0}\, dx$$

$$\mathcal{E} = \frac{qN_d}{K\epsilon_0}(x - x_n), \qquad 0 < x < x_n \tag{12}$$

Similarly, in the P side of the junction,

$$\mathcal{E} = -\int_{-x_p}^{x} \frac{qN_a}{K\epsilon_0}\, dx$$

$$\mathcal{E} = -\frac{qN_a}{K\epsilon_0}(x + x_p) \qquad -x_p < x < 0 \tag{13}$$

$$\begin{aligned} \mathcal{E}_{max} &= -\frac{qN_d x_n}{K\epsilon_0} \\ &= -\frac{qN_a x_p}{K\epsilon_0} \end{aligned} \tag{14}$$

$\mathcal{E}$ varies linearly between 0 and $\mathcal{E}_{max}$.

The total potential across the depletion region is simply the area under the $\mathcal{E}$ field curve since $\mathcal{E} = -\partial V/\partial x$. Therefore $V = -\int \mathcal{E}\, dx$. That is,

$$\phi_i = \frac{1}{2}\mathcal{E}_{max}(x_n + x_p) = \frac{1}{2}\mathcal{E}_{max}x_d \tag{15}$$

where x_d = total depletion layer width. Therefore,

$$x_d = \frac{2\phi_i}{\mathcal{E}_{max}} = \frac{2\phi_i}{\frac{qN_d x_n}{K\epsilon_0}} = \frac{2\phi_i}{\frac{qN_a x_p}{K\epsilon_0}}$$

Using the relationships for x_n and x_p and noting that $\phi_n + \phi_p = \phi_i$, which we know from Equation (7):

$$x_n = \left[\frac{2K\epsilon_0}{q}\phi_i\left(\frac{N_a}{N_d(N_a + N_d)}\right)\right]^{1/2}, \tag{16}$$

$$x_p = \left[\frac{2K\epsilon_0}{q}\phi_i\left(\frac{N_d}{N_a(N_a + N_d)}\right)\right]^{1/2}, \tag{17}$$

$$x_d = x_n + x_p = \left[\frac{2K\epsilon_0}{q}\phi_i\left(\frac{1}{N_a} + \frac{1}{N_d}\right)\right]^{1/2}. \tag{18}$$

The depletion region width depends most strongly on the doping on the lightly doped side and varies approximately as the inverse square root of that doping.

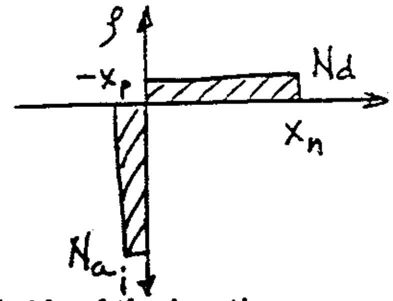
More generally, to calculate the potential drop across the junction (or built-in voltage ϕ_i), Equations (12) and (13) are integrated.

$$\phi_i = -\int_{-x_p}^{x_n} \mathcal{E}\, dx \tag{19}$$

$$\begin{aligned}\phi_i &= \frac{q}{K\epsilon_0}\left[N_a\int_{-x_p}^{0}(x + x_p)\, dx - N_d\int_0^{x_n}(x - x_n)\, dx\right] \\ &= \frac{q}{K\epsilon_0}\left[N_a\left(\frac{x^2}{2} + x_p x\right)_{-x_p}^{0} - N_d\left(\frac{x^2}{2} - x_n x\right)_0^{x_n}\right] \\ &= \frac{q}{K\epsilon_0}\left[N_a\left(\frac{x_p{}^2}{2}\right) + N_d\left(\frac{x_n{}^2}{2}\right)\right]\end{aligned} \tag{20}$$

We already know ϕ_i from Equation (7). Since the charges on either side of the junction must be balanced, $N_a x_p = N_d x_n$. Therefore, we have two equations ((15) and (16)) in two unknowns and can solve for x_p and x_n and therefore x_d with the same results as above.

Note that the zero of the potential curve is at the junction only if $N_a = N_d$. In fact for a one sided step junction ($N_a >> N_d$ or $N_d >> N_a$), we have:

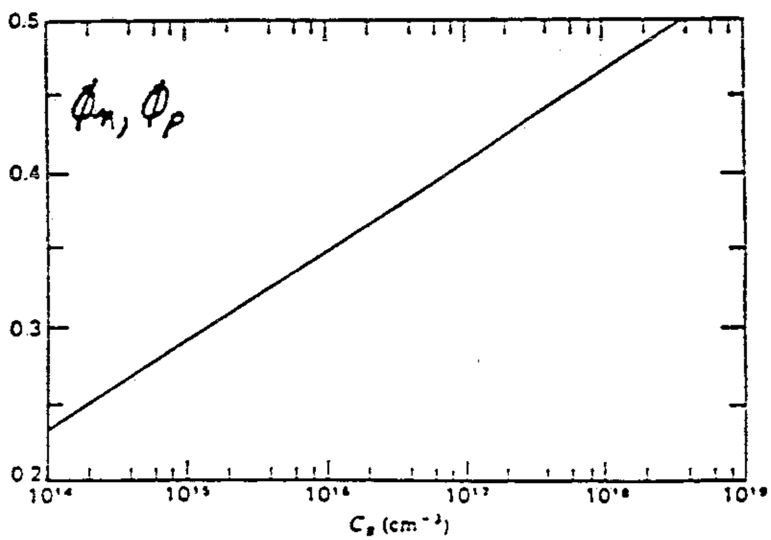


- The depletion region is primarily in the lightly doped side of the junction.
- Most of the potential change also occurs across the lightly doped side. In the limit for this example ($N_a \to \infty$), we approach a metal-semiconductor connection.

Notes:

- x_d is most strongly dependent on the doping in the lightly doping side of the junction.
- x_d varies inversely as the square root of the doping on the lightly doped side.

Summary for Step Junctions



$$\phi_n = \frac{kT}{q} \ln \frac{N_d}{n_i}$$

$$\phi_p = \frac{kT}{q} \ln \frac{N_a}{n_i}$$

$$\phi_i = \phi_n + \phi_p = \frac{kT}{q} \ln \frac{N_d N_a}{n_i^2}$$

$$x_d = \left[\frac{2K\epsilon_0}{q}\phi_i\left(\frac{1}{N_a} + \frac{1}{N_d}\right)\right]^{1/2}$$

ϕ_n and ϕ_p are on the order of 0.2 to 0.5 V. Therefore ϕ_i is on the order of 0.4 to 1.0 V. x_d is typically 0.1-1.0μm.

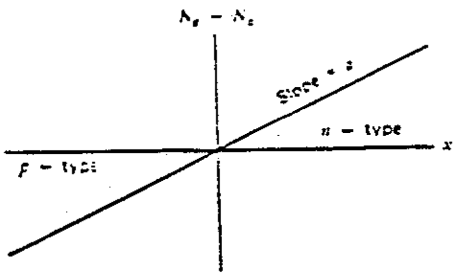
Example: Given a PN step junction with $N_d = 10^{16}\text{cm}^{-3}$ and $N_a = 4 \times 10^{18}\text{cm}^{-3}$.

$$\phi_i = \frac{kT}{q} \ln \frac{N_d N_a}{n_i^2} \cong 0.83\,\text{V}$$

$$x_d = \left[\frac{2K\epsilon_0}{q}\phi_i\left(\frac{1}{N_a} + \frac{1}{N_d}\right)\right]^{1/2} \cong \left[\frac{2K\epsilon_0\phi_i}{qN_d}\right]^{1/2} = 0.33\mu\text{m}$$

$$\mathcal{E}_{max} = -\frac{qN_d x_n}{K\epsilon_0} \cong -\frac{qN_d x_d}{K\epsilon_0} = -5 \times 10^4\text{V/cm}$$

Linearly Graded Junctions

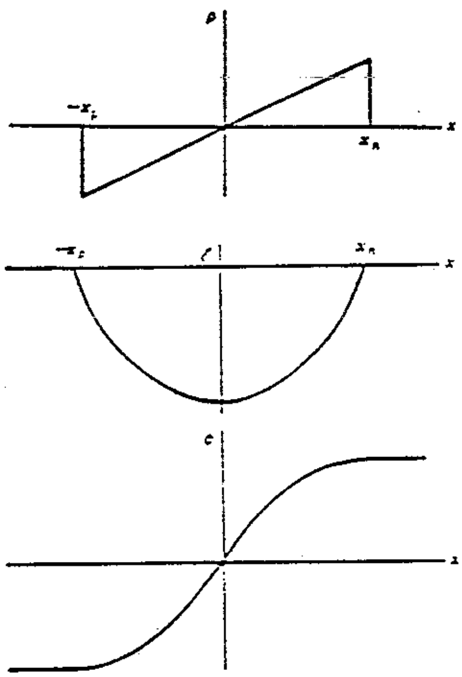


$N_a - N_d = -ax$ (typically used as an approximation for Gaussian profiles).

$$a = \text{slope} \quad (\text{cm}^{-4})$$

$$\frac{\partial^2 V}{\partial x^2} = -\frac{q}{K\epsilon_0}ax \tag{21}$$

Integrating as in the step junction approximation yields:



$$\mathcal{E}(x) = \frac{qa}{2K\epsilon_0}(x^2 - x_p{}^2) \tag{22}$$

$$\phi_i = \frac{2qa}{3K\epsilon_0}x_n^3 \qquad (x_n = x_p) \tag{23}$$

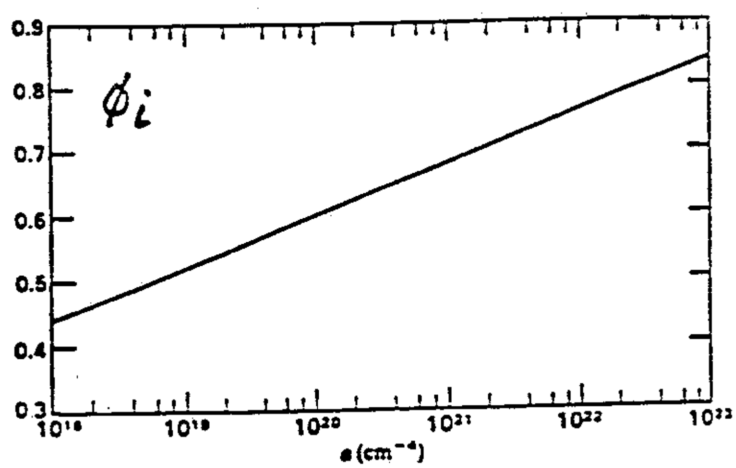
The built-in voltage is determined by the doping concentrations at the edges of the depletion region,

$$\phi_i = 2\frac{kT}{q}\ln\frac{ax_d}{2n_i} \tag{24}$$

$$2x_n = x_d = \left[\frac{12K\epsilon_0\phi_i}{qa}\right]^{1/3} \tag{25}$$

Note the 1/3 power dependence of x_d on voltage.

The built-in voltage may be calculated from (18) and (19) and is shown at the right. Again note that ϕ_i is in the range of 0.5 to 0.9 volts. This is approximately twice the value for the Schottky diodes considered earlier.



Conclusions The first order theory that we have developed for the PN junction allows calculation of:

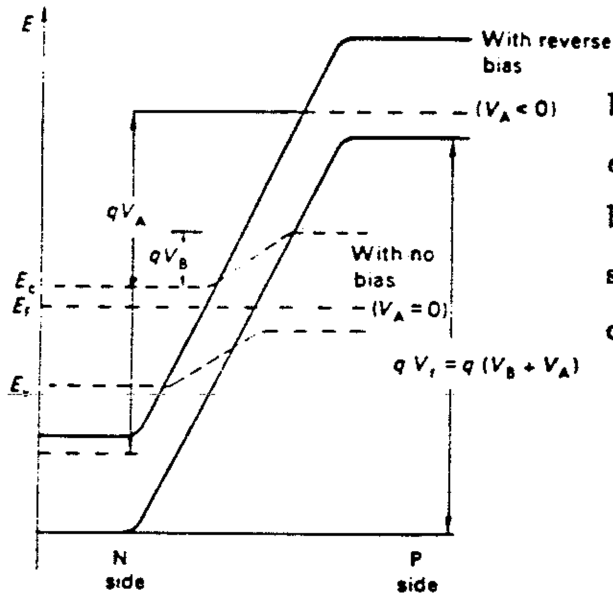
- ϕ_i
- x_d
- $\mathcal{E}$(x)

from fundamental material parameters (K, ϵ_0, q, n_i, etc.) and practical considerations (N_a, N_d).

- In general this theory is much better at predicting junction behavior than the first order Schottky theory we previously considered.
- The main reason for this is that the PN junction does not suffer from surface state problems (a perfect crystal forms the boundary).

PN Junction Capacitance Under Reverse Bias

Under equilibrium conditions we have,



A built in potential ϕ_i provides a barrier to majority carrier current (i.e. $e^- : N \rightarrow P,\ h^+ : P \rightarrow N$). If we bias the junction externally with the N side $+$, P side $-$, the barrier to majority carrier current flow is increased.

Results

1. Applied voltage appears essentially entirely across the depletion region.

2. x_d increases, mobile carriers pulled further away from junction.

3. Minority carrier flow is increased but I is small because they are low in number.

4. $\mathcal{E}$ field across depletion region increases.

Step Junction

$$x_d = x_p + x_n = \left[\frac{2K\epsilon_0}{q}\left(\frac{1}{N_a} + \frac{1}{N_d}\right)(\phi_i - V_a)\right]^{1/2} \tag{26}$$

$$\mathcal{E}_{max} = \frac{2(\phi_i - V_a)}{x_d} = 2\left(\frac{\text{potential across depletion region}}{\text{width of depletion region}}\right) \tag{27}$$

i.e. ϕ_i is replaced by $(\phi_i - V_a)$ (Note that V_a is negative, i.e. sign conventions are the same as with Schottky diodes, i.e.

Reverse bias $\Rightarrow \phi_i, V_a$ add

Forward bias $\Rightarrow \phi_i, V_a$ subtract

The small signal capacitance of the junction is given by

$$C' \equiv \frac{\partial Q'}{\partial V_a} = \frac{K\epsilon_0}{x_d} \tag{28}$$

(A derivation in the Muller and Kamins text shows that this latter result holds for an arbitrary doping profile. Given any profile, if we can calculate x_d, we can calculate the small signal capacitance C.)

By small signal capacitance we mean: (1) Apply a D.C. potential V_{bias} to establish x_d. (2) Measure the capacitance by superimposing a small AC signal ($V_{AC} \ll V_{bias}$) so x_d is not changed significantly.

Example: Consider a step junction

$$Q' = \frac{Q}{A} = qN_d x_n = qN_a x_p \tag{29}$$

$$C' \equiv \frac{\partial Q'}{\partial V_a} = qN_d \frac{\partial x_n}{\partial V_a} = qN_a \frac{\partial x_p}{\partial V_a} \tag{30}$$

But from (23), $x_p = (N_d/N_a)x_n$ and $x_d = x_n + x_p$. Therefore,

$$x_n = x_d - x_p = x_d - \frac{N_d}{N_a}x_n$$

and

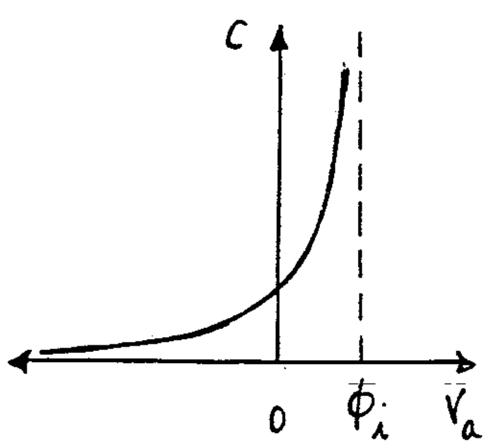
$$x_n = \frac{x_d}{1 + N_d/N_a} = \frac{\left[\frac{2K\epsilon_0}{q}\left(\frac{1}{N_a} + \frac{1}{N_d}\right)(\phi_i - V_a)\right]^{1/2}}{1 + \frac{N_d}{N_a}} \tag{31}$$

Therefore

$$\frac{\partial x_n}{\partial V_a} = \frac{1}{N_d}\left[\frac{K\epsilon_0}{2q\,(1/N_a + 1/N_d)\,(\phi_i - V_a)}\right]^{1/2}$$

$$C' = \left[\frac{qK\epsilon_0}{2\,(1/N_a + 1/N_d)\,(\phi_i - V_a)}\right]^{1/2} \tag{32}$$

just as would have been calculated based on Equation (28).



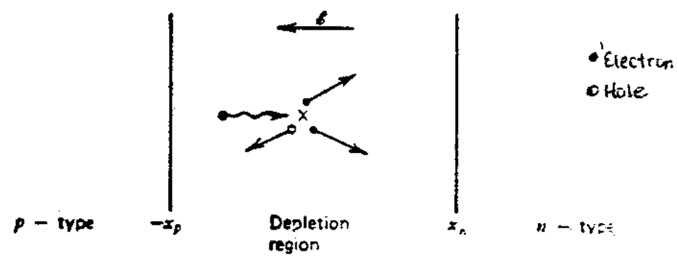
In the reverse direction C falls as the square root of V_a. In the forward direction, $C \to \infty$ as $V_a \to \phi_i$. (Actually in forward bias our analysis breaks down because mobile carriers in depletion region causes (29) to be invalid.) Equivalent expressions may be calculated for C for other impurity profiles. As in the Schottky diode case, measurement of C vs V_a can be used to extract information on the doping profile.

PN Junction Breakdown

As the voltage across a pn junction is increased, the depletion region widens and the $\mathcal{E}$ field increases. Eventually, a limit must be reached. When this happens, the diode begins to conduct large currents in the reverse direction. Further increase in the voltage will cause currents to be limited only by the external circuit. Three principal mechanisms are responsible for this breakdown.

1. **Avalanche Breakdown**

   Electrons enter the depletion region (or are generated there by G-R centers). If $\mathcal{E}$ is high enough, the free electrons can gain sufficient energy that when they collide with valence electrons they can break the covalent bonds and excite the valence electron to the conduction band, creating an electron-hole pair. This process is repeated as those electrons and holes in turn gain energy due to $\mathcal{E}$ and collide with other bound electrons and results in large currents.

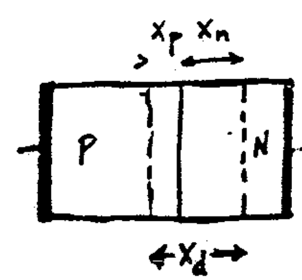


2. **Zener Breakdown**

   If the $\mathcal{E}$ field is large enough, it can directly pull an electron away from a covalent bond, creating an electron-hole pair. This process also results in large currents.

3. **Punch-Through**

   In a narrow diode, as the depletion region increases, it can reach all the way through to the metal contact which can supply a very large number of carriers. Large currents can result since no more voltage can be dropped across the depletion region. When $x_n = W_n$ or $x_p = W_p$, punch-through occurs. Punch-through almost always occurs on the lightly-doped side of the junction since almost all of the depletion region is on that side.



Avalanche Breakdown

In avalanche breakdown, if the field is not large enough, free carriers will scatter without generating new carriers. Therefore, a critical field is required for multiplication and avalanche breakdown.

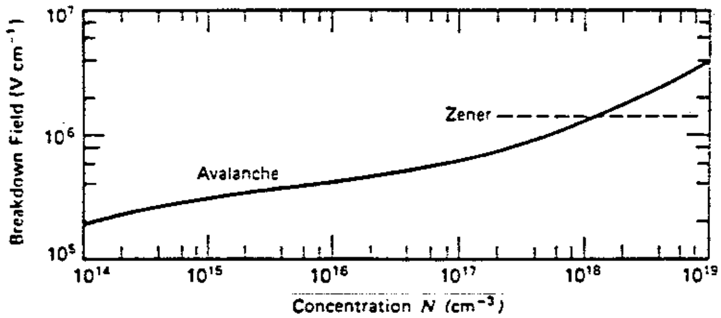


Figure 3.12 The critical electric fields for avalanche and Zener breakdown in silicon as functions of dopant concentration [1,2,3]

Note that the critical field increases for avalanche breakdown as the doping increases. This occurs because ionized impurity scattering becomes dominant with heavy doping and scattering becomes more frequent so carriers are unable to gain as much energy before being scattered. The critical field for zener breakdown is approximately constant and is the limiting factor for junctions which are heavily doped on both sides (p and n).

To analyze avalanche breakdown, we can make the following observations:

- Near the edges of the depletion or space charge region, $\mathcal{E} \ll \mathcal{E}_{cr}$ and multiplication will not occur.

- In the central region, near $\mathcal{E}_{max}$, if $\mathcal{E} > \mathcal{E}_{cr}$ avalanche multiplication can occur.

For electrons and holes within the region where $\mathcal{E} > \mathcal{E}_{cr}$, we define

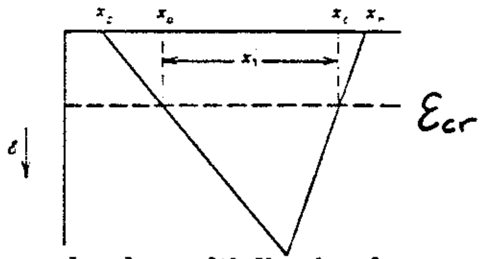
$$\alpha_n \equiv \text{ionization coefficient of } e^-$$

$$\alpha_n dx \equiv \text{probability } e^- \text{ creates an } e^-\text{-}h^+ \text{ pair within } dx$$

$$\alpha_p \equiv \text{ionization coefficient of } h^+$$

$$\alpha_p dx \equiv \text{probability } h^+ \text{ creates an } e^-\text{-}h^+ \text{ pair within } dx$$

Assuming $\alpha_n = \alpha_p = \alpha$, which is only approximately true, then for both holes and electrons



$$n_c - n_a = n_c \int_{x_a}^{x_c} \alpha \, dx$$

$$M \equiv \frac{n_c}{n_a} = \frac{1}{1 - \int_{x_a}^{x_c} \alpha \, dx}, \tag{33}$$

where $M \equiv$ avalanche multiplication factor.

Therefore, avalanche multiplication occurs when

$$\int_{x_a}^{x_c} \alpha \, dx = 1 \ \Rightarrow \ M \rightarrow \infty \tag{34}$$

Evaluation of this integral can be difficult because α is a function of $\mathcal{E}$ and therefore x.

We can consider a simplified case. Suppose we have a one-sided junction in which $N_d \gg N_a$. Most of the depletion region occurs on the lightly doped side.

$$\mathcal{E}_{max} = \frac{2(\phi_i - V_A)}{x_d} \quad \text{from (22)}$$

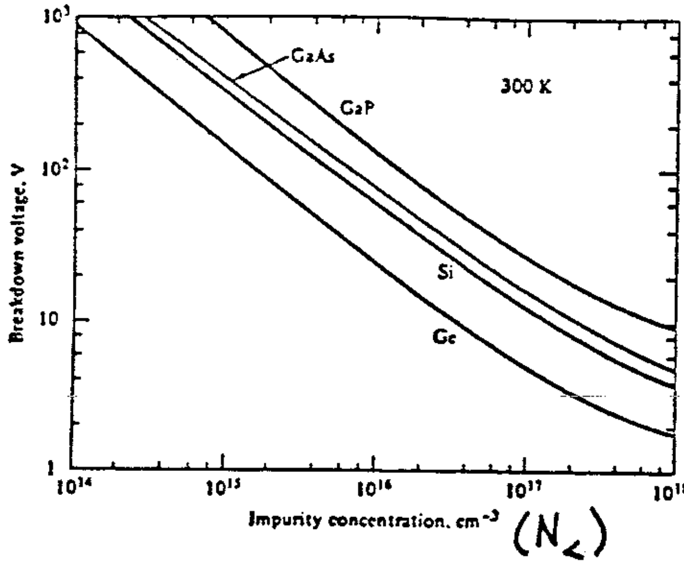
$$x_d \cong x_p \cong \left[\frac{2K\epsilon_s}{qN_a}(\phi_i - V_A) \right]^{1/2} \quad \text{from (21)}$$

$$\mathcal{E}_{max} \cong \left[\frac{2qN_a|V_R|}{K\epsilon_0} \right]^{1/2}$$

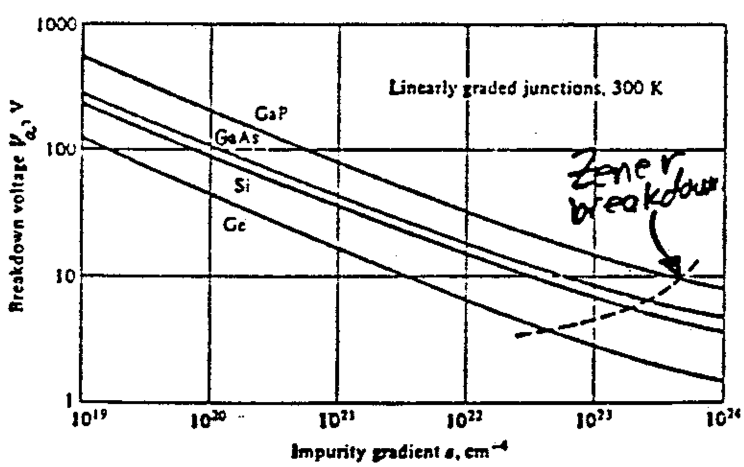
If we define the breakdown to occur when the maximum field reaches the critical field (this is a conservative estimate since a field greater than the critical field has to exist over a distance for breakdown to occur) then

$$BV \cong \frac{K\epsilon_0 {\mathcal{E}_{cr}}^2}{2qN_a} \tag{35}$$

Since the critical field varies slowly with doping, **the breakdown voltage varies approximately inversely as the doping on the lightly doped side of the junction.**



- For BV greater than a few volts, the inverse relationship holds reasonably well.
- At high doping levels, the BV flattens out because a new mechanism (Zener breakdown, to be discussed next) becomes dominant.
- BV increases for a given doping as the band gap of the material increases. This is to be expected since it is more difficult to create hole-electron pairs in wider band gap materials.
- Similar calculations for linearly graded junctions give similar results with larger gradient producing lower BV.



- For avalanche breakdown, BV increases with temperature since scattering lifetimes τ and therefore α decrease with temperature.

Zener Breakdown (Tunneling Breakdown)

For avalanche breakdown a narrow depletion region results in $\int_{x_a}^{x_c} \alpha\, dx$ being smaller because the field is near its maximum over only a short distance. Therefore, the breakdown voltage due to avalanche is increased.

Zener breakdown occurs when both sides of a junction are so highly doped that the depletion region is narrow enough to allow tunneling ($< 100\,\text{Å} = 10\text{nm}$) and there is a very large electric field across the depletion region. Zener breakdown occurs when both sides are so heavily doped to at least approach degeneracy. The situation is analogous to ohmic contacts in metal-semiconductor contacts. With an applied reverse bias, an electron can tunnel directly from the valence band on the p side to the conduction band on the n side.

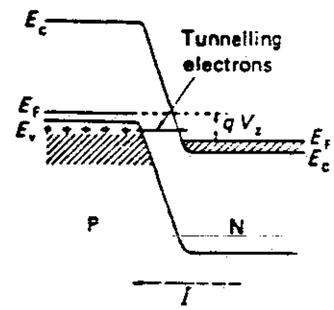


Figure 9.7. Energy bands of a junction in Zener breakdown

- For Zener breakdown, $BV < 5\,\text{V}$.
- BV decreases with temperature since E_g (the height of the barrier to be tunneled through) decreases with temperature.
- Can differentiate between Zener and avalanche breakdown by observing effect of T on BV

  $+ \Rightarrow$ avalanche

  $- \Rightarrow$ Zener

  $\sim 0 \Rightarrow$ mixed (4-6 V)

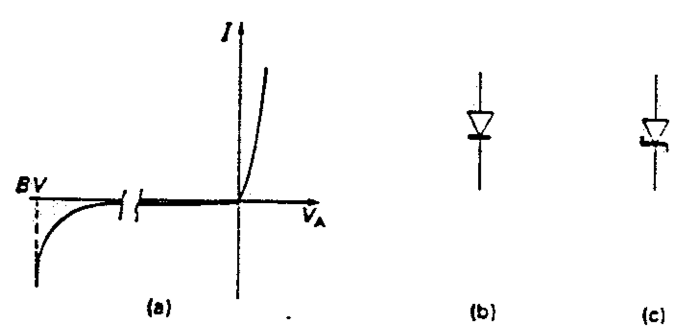


Figure 9.8. The real diode: (a) I–V characteristic; (b) standard symbol for a rectifier diode; (c) standard symbol for a voltage reference (Zener) diode.

Summary

All diodes operated in reverse breakdown are called Zener (or reference) diodes although most actually break down by avalanche. After breakdown, the amount of current which flows through the diode is limited by the external circuit. Most breakdown is reversible unless the external circuit allows enough current to flow that ohmic heating becomes destructive.

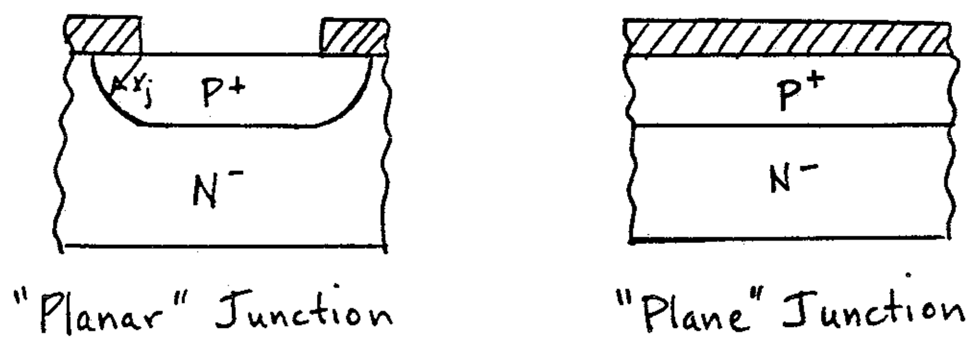
Avalanche: Most usual, $BV \propto (1/N_<)$. $BV > 5\,\mathrm{V}$.

Zener: High doping (nearly degenerate) on both sides, $x_d < 100\,\text{Å}$, $BV < 5\,\mathrm{V}$.

Punch-through: Narrow region on lightly-doped side, $x_d = W$.

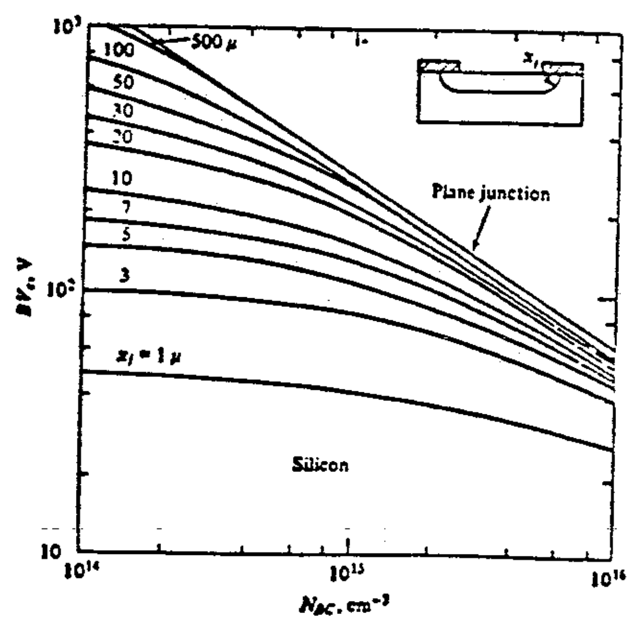
Practical Considerations

In any real junction made with planar technology, the calculations above which relate to **plane** (flat) junctions do not strictly apply.



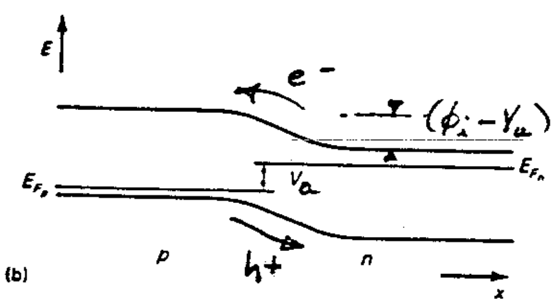
Curvature of the planar junction under the oxide edge due to lateral diffusion increases the $\mathcal{E}$ field in these regions resulting in lower BV than in the plane junction.

Computer calculations show how BV is reduced as the radius of curvature or x_j is reduced, with shallow junctions showing the greatest reduction. The use of guard rings such as discussed for Schottky diodes can help BV for planar junctions approach that for the ideal plane junction.



Current-Voltage Characteristics of PN Junctions

So far, we have described the behavior of PN junctions in equilibrium and under reverse bias, conditions where the current is small. We will now consider forward bias where significant currents can flow.



In thermal equilibrium $\phi_i = \phi_n + \phi_p =$ built-in potential. Under forward bias, the barrier to majority carrier flow is reduced.

electrons are injected: $N \to P$

holes are injected: $P \to N$

- Applied voltages will still be dropped across depletion region so $\phi_i \Rightarrow (\phi_i - V_a)$
- The barrier to majority carriers is thus reduced by V_a and we would expect, based on the Boltzmann approximation to Fermi-Dirac statistics, that the current would depend exponentially on V_a. (The number of majority carriers with sufficient energy to surmount barrier depends exponentially on energy).
- Once injected, the majority carriers become minority carriers on the other side. The behavior of minority carriers is of fundamental importance to the behavior of PN junctions.
- The injected minority carriers will recombine with time, while at the same time they diffuse away from the junction.

The number of injected carriers can be related to V_a. To begin we know that

$$\phi_i = \frac{kT}{q} \ln \frac{N_d N_a}{{n_i}^2} \tag{36}$$

We can define:

n_{p0} = number of electrons in P region in equilibrium

p_{n0} = number of holes in N region in equilibrium

n_{n0} = number of electrons in N region in equilibrium

p_{p0} = number of holes in P region in equilibrium

$$n_{n0} \cong N_d, \quad p_{n0} \cong \frac{{n_i}^2}{N_d} \tag{37}$$

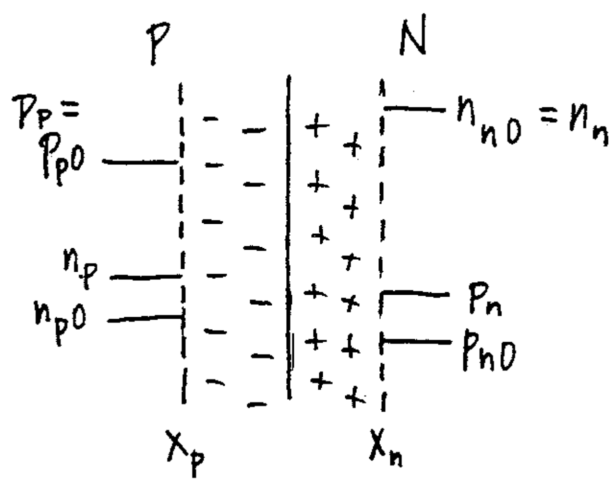
$$p_{p0} \cong N_a, \quad n_{p0} \cong \frac{{n_i}^2}{N_a} \tag{38}$$

$$\phi_i = \frac{kT}{q} \ln \frac{n_{n0}}{n_{p0}} = \frac{kT}{q} \ln \frac{p_{p0}}{p_{n0}} \tag{39}$$

Therefore,

$$n_{n0} = n_{p0} \exp \frac{q\phi_i}{kT} \tag{40}$$

$$p_{p0} = p_{n0} \exp \frac{q\phi_i}{kT} \tag{41}$$



In other words, the minority carrier concentrations on one side of the depletion region are related to the majority carrier concentrations on the other side by ϕ_i. Just as for reverse bias, we replace ϕ_i by $(\phi_i - V_a)$ and with the result that:

$$n_n(x_n) = n_p(-x_p) \exp \frac{q(\phi_i - V_a)}{kT}, \tag{42}$$

$$p_p(-x_p) = p_n(x_n) \exp \frac{q(\phi_i - V_a)}{kT}, \tag{43}$$

(the subscript 0 has been eliminated because we are no longer in equilibrium.)

Low Level Injection

We will first assume that $n_n = n_{n0}$ and $p_p = p_{p0}$, the low level injection condition, as for light generated carriers, is that the majority carrier concentrations are unaffected by the injection process.

Using the above equations then,

$$n_{p0} \exp \frac{q\phi_i}{kT} = n_p(-x_p) \exp \frac{q(\phi_i - V_a)}{kT}, \tag{44}$$

$$n_p(-x_p) = n_{p0} \exp \frac{qV_a}{kT}, \tag{45}$$

Similarly for holes:

$$p_n(x_n) = p_{n0} \exp \frac{qV_a}{kT}, \tag{46}$$

These equations provide boundary conditions for the minority carrier concentrations at the edge of the depletion region which can be used in calculating current densities.

Summarizing, under low level injection:

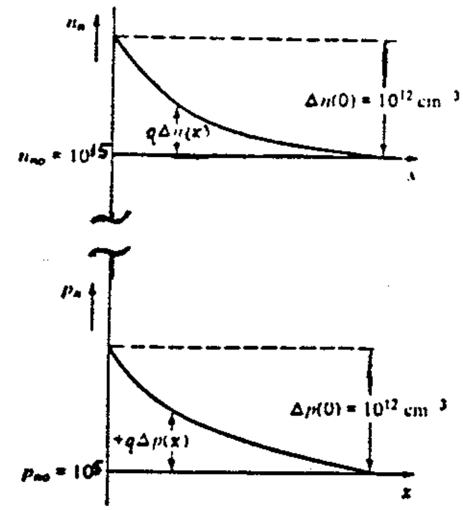
- n_n and p_p are unaffected by injection.
- n_p and p_n are related exponentially to n_{p0} and p_{n0} as $\exp(qV_a/kT)$.
- n_{p0} and p_{n0} are related exponentially to n_{n0} and p_{p0} as $\exp(-q\phi_i/kT)$.

Due to the injected carriers:

1. The injected carriers set up an $\mathcal{E}$ field in the both the N and P regions.
2. This field causes majority carriers, which are present in much larger numbers and thus respond more strongly to electric fields, to move in order to re-establish neutrality and eliminate the $\mathcal{E}$ field.
3. At the same time, the excess injected minority carriers diffuse away from the junction into the neutral regions.
4. These minority carriers recombine with majority carriers with the recombination process taking place over some distance.

At right is pictured the carrier concentrations in the N region as a function of distance. The majority carrier concentration is essentially unchanged ($n_{n0} = N_d = 10^{15}\text{cm}^{-3} \gg 10^{12}\text{cm}^{-3} = p_n$).

The injected hole concentration decays to its equilibrium value over some distance.



The $\mathcal{E}$ field which exists in the regions where there is an excess of injected carriers causes a drift current which in the N region is given by

$$I_{\text{drift}} = \begin{cases} q\mu_n n_n \mathcal{E} & \text{for majority carriers } (e^-) \\ q\mu_p p_n \mathcal{E} & \text{for minority carriers } (h^+) \end{cases} \tag{47}$$

Since $n_n \gg p_n$, the drift of minority carriers is neglible. Therefore, the minority carriers move primarily by diffusion (proportional to gradient not concentration) while for majority carriers the drift and diffusion terms nearly balance.

The minority carriers control the behavior of PN junctions. This is in contrast to Schottky diodes where the majority carriers dominate the behavior.

We will assume that the injected minority carriers move away from the junction by diffusion alone. This is known as the **Diffusion Approximation** and is only valid for low level injection.

In the quasi-neutral N region:

$$I_p \cong -qAD_p \frac{dp_n}{dx} \tag{48}$$

$$\frac{\partial p_n}{\partial t} = D_p \frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{n0}}{\tau_p} \tag{49}$$

The second equation is just the continuity equation for holes.

In the quasineutral P region:

$$I_n \cong qAD_n \frac{dn_p}{dx} \tag{50}$$

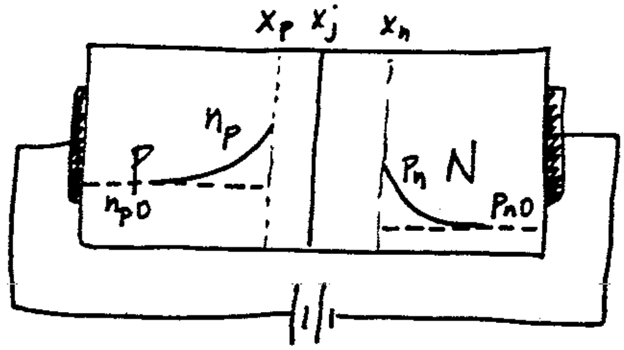
$$\frac{\partial n_p}{\partial t} = D_n \frac{\partial^2 n_p}{\partial x^2} - \frac{n_p - n_{p0}}{\tau_n} \tag{51}$$

Ideal Diode I-V Characterisics

By solving equations (48) - (51) with appropriate boundary conditions, we can find the current flowing in a diode with an applied V_a.

Long Base Diode

Recombination reduces p_n and n_p to equilibrium values before ohmic contacts.



In steady state:

$$D_p \frac{\partial^2 p_n}{\partial x^2} - \frac{\Delta p}{\tau_p} = \frac{dp_n}{dt} = 0$$

$$D_n \frac{\partial^2 n_p}{\partial x^2} - \frac{\Delta n}{\tau_n} = \frac{dn_p}{dt} = 0$$

The general solutions of these equations are given by

$$\Delta p = K_1 \exp \frac{-x}{\sqrt{D_p \tau_p}} + K_2 \exp \frac{x}{\sqrt{D_p \tau_p}} \tag{52}$$

$$\Delta n = K_3 \exp \frac{-x}{\sqrt{D_n \tau_n}} + K_4 \exp \frac{x}{\sqrt{D_n \tau_n}} \tag{53}$$

The appropriate boundary conditions are:

$$p_n = \begin{cases} p_{n0} \exp \frac{qV_a}{kT}, & x = x_n \\ p_{n0}, & x = \infty \end{cases}$$

$$n_p = \begin{cases} n_{p0} \exp \frac{qV_a}{kT}, & x = -x_p \\ n_{p0}, & x = -\infty \end{cases}$$

Thus we have

$$\Delta p = p_{n0}\left(\exp\frac{qV_a}{kT} - 1\right)\exp -\frac{x - x_n}{L_p}, \qquad x > x_n \tag{54}$$

$$\Delta n = n_{p0}\left(\exp\frac{qV_a}{kT} - 1\right)\exp\frac{x + x_p}{L_n}, \qquad x < -x_p \tag{55}$$

where

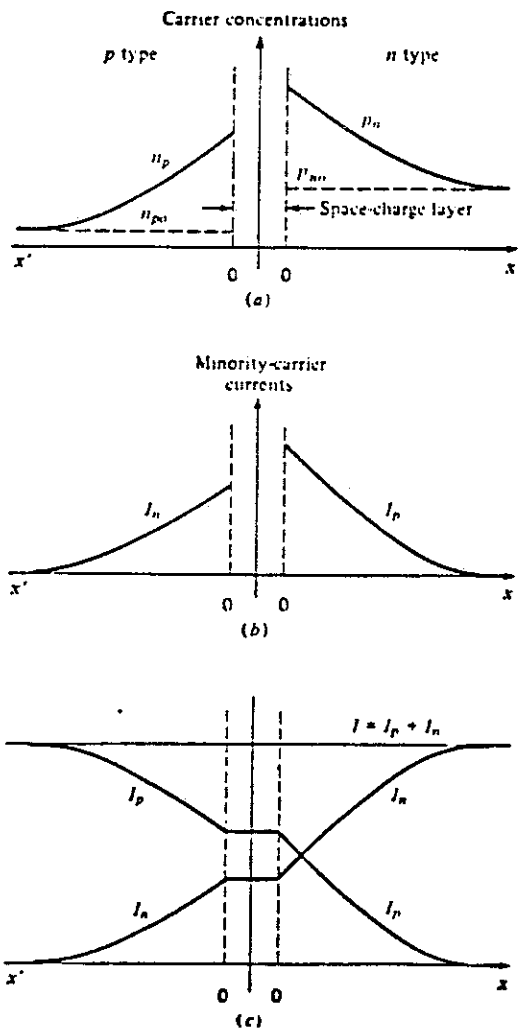
$$L_p \equiv \sqrt{D_p\tau_p} = \text{hole diffusion length}$$

$$L_n \equiv \sqrt{D_n\tau_n} = \text{electron diffusion length}$$

The hole and electron currents are given by substitution into (48) and (50)

$$I_p = qA\frac{D_p p_{n0}}{L_p}\left(\exp\frac{qV_a}{kT} - 1\right)\exp -\frac{x - x_n}{L_p}, \qquad x > x_n \tag{56}$$

$$I_n = qA\frac{D_n n_{p0}}{L_n}\left(\exp\frac{qV_a}{kt} - 1\right)\exp\frac{x + x_p}{L_n}, \qquad x < -x_p \tag{57}$$



Note that:

The minority carriers decrease with distance

From continuity of I, the majority currents must therefore increase with distance (i.e. minority current is continuously transformed into majority carrier current).

The total current flowing across the junction is thus given by the sum of I_p and I_n at the edges of the depletion region ($x = x_n$ and $x = -x_p$ respectively).

$$I_{tot} = I_p|_{x=x_n} + I_n|_{x=-x_p}$$

or

$$I = qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(\exp\frac{qV_a}{kT} - 1\right) \tag{58}$$

$$I = I_0\left(\exp\frac{qV_a}{kT} - 1\right) \tag{59}$$

where

$$\begin{aligned} I_0 &= qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right) \\ &= qAn_i^2\left(\frac{D_p}{N_d L_p} + \frac{D_n}{N_a L_n}\right) \end{aligned} \tag{60}$$

• This last result illustrates particularly well that it is the doping level on the **lightly doped** side of the junction that determines the current.

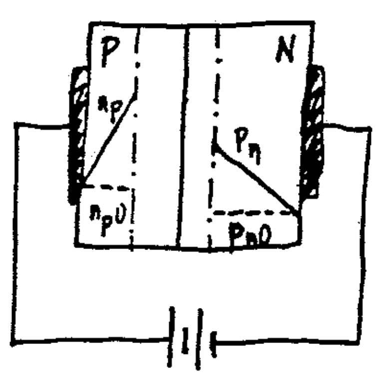
Example: N side is lightly doped, therefore

$$p_{n0} = \frac{n_i^2}{N_d} \gg n_{p0} = \frac{n_i^2}{N_a}$$

therefore $p_n >> n_p$ and N side controls amount of current. **Current is mostly due to injection into lightly doped side of junction.**

Short Base Diode

A second limiting case of considerable practical importance occurs when the widths of the neutral regions are $\ll L_p$ and/or L_n. A particularly important example is the base region of a bipolar transistor.



The surface recombination velocity of the ohmic contacts is $\approx \infty$ because of the large number of traps available there. Therefore $p_n(x = W_n) = p_{n0}$ and $n_p(x = -W_p) = n_{p0}$. If we assume that no recombination occurs within W_p, then the minority carrier current densities (due to diffusion) are given by

$$J_p = qD_p \frac{dp}{dx} \tag{61}$$

$$J_n = qD_n \frac{dn}{dx} \tag{62}$$

But these currents must be constant (no change in majority carrier current flows because no recombination). Therefore

$$\frac{dp}{dx} \text{ and } \frac{dn}{dx} = \text{constants}$$

and hole and electrons concentrations decrease linearly with distance as shown.

Under these conditions,

$$\begin{aligned} I_p =& qAD_p \frac{dp}{dx} = qAD_p \frac{\Delta p}{W_n} \\ =& qAD_p \frac{p_{n0}}{W_n} \left(\exp \frac{qV_a}{kT} - 1\right) \\ =& qAD_p \frac{n_i^2}{N_d W_n} \left(\exp \frac{qV_a}{kT} - 1\right) \end{aligned} \tag{63}$$

Similarly

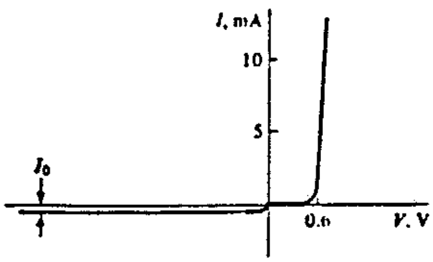
$$I_n = qAD_n \frac{n_i^2}{N_a W_p} \left(\exp \frac{qV_a}{kT} - 1\right) \tag{64}$$

The total current is again given by the sum of the two minority carrier currents, so that

$$I = qAn_i^2 \left[\frac{D_p}{N_d W_n} + \frac{D_n}{N_a W_p}\right] \left(\exp \frac{qV_a}{kT} - 1\right) \tag{65}$$

• Comparing the form of (65) with the equivalent expression for the long base diode (58), we see that the two equations are identical except for the characteristic length appropriate to the geometry ($L_{p,n}$ or $W_{n,p}$).

Summarizing we have:



$$I_0 = \begin{cases} qAn_i^2\left[\dfrac{D_p}{N_dL_p} + \dfrac{D_n}{N_aL_n}\right] & \text{Long Base} \\[2ex] qAn_i^2\left[\dfrac{D_p}{N_dW_n} + \dfrac{D_n}{N_aW_p}\right] & \text{Short Base} \end{cases}$$

$$I = I_0\left(\exp\frac{qV_a}{kT} - 1\right)$$

Actual diodes may be intermediate cases, in either case, it is usually the lightly doped and/or narrow side of the junction which mostly determines I.

Assumptions

- Ohmic drops in neutral region are neglible. Applied voltage is dropped across depletion region. Generally true except under high level injection.

- Quasi-equilibrium assumes that injected currents are small relative to the normal compensating drift and diffusion currents across the depletion region. We required this assumption to calculate $p_n(x_n)$ and $n_p(-x_p)$. This assumption is also generally valid except under high level injection.

Space Charge Recombination Currents

We have assumed to this point that all the minority carriers which enter the depletion region cross to the other side. This is not true in practice since some recombine through trapping centers in the depletion region. From previous notes for $\tau = \tau_n = \tau_p$:

$$U = \frac{pn - n_i^2}{\tau \left[n + p + 2n_i \cosh\left(\frac{E_t - E_i}{kT} \right) \right]} \tag{66}$$

Under reverse bias, these centers act as generation sites since $np < n_i^2$ and increase the reverse saturation current over the ideal value calculated earlier. Under forward bias $(np > n_i^2)$ the centers act as recombination sites increasing the forward current.

In the depletion region, both n and p are small. The generation or recombination rate is maximized when $n = p$. The the G-R process is dominated by that portion of the depletion region where that relation approximately holds. With an applied voltage,

$$pn = n_i^2 \exp \frac{qV_a}{kT}$$

so for maximum $|U|$,

$$p = n = n_i \exp \frac{qV_a}{2kT} \tag{67}$$

and if $E_t = E_i$,

$$U_{max} = \frac{n_i \left(\exp \frac{qV_a}{kT} - 1 \right)}{2\tau_0 \left(\exp \frac{qV_a}{2kT} + 1 \right)} \tag{68}$$

where τ_0 is the lifetime of carriers in the depletion region. (As we saw in the homework, $2\tau_0$ could be replaced by $\tau_n + \tau_p$ in the depletion region if the lifetimes are different.)

Forward Bias: If $V_a > 6kT/q$ then

$$U_{max} \cong \frac{n_i}{2\tau_0} \exp \frac{qV_a}{2kT} \tag{69}$$

The recombination current will increase as the exponential of $qV_a/2kT$.

$$I_{rec} = qA \int_{-x_p}^{x_n} U(x)\, dx \tag{70}$$

If we assume that the recombination rate is near its maximum throughout the depletion region,

$$I_{rec} \cong \frac{qAn_i x'_d}{2\tau_0} \exp \frac{qV_a}{2kT} \tag{71}$$

The total current is the sum of the diffusion current already calculated and the recombination current.

$$I_{forward} \cong I_0 \exp \frac{qV_a}{kT} + \frac{qAn_i x'_d}{2\tau_0} \exp \frac{qV_a}{2kT} \tag{72}$$

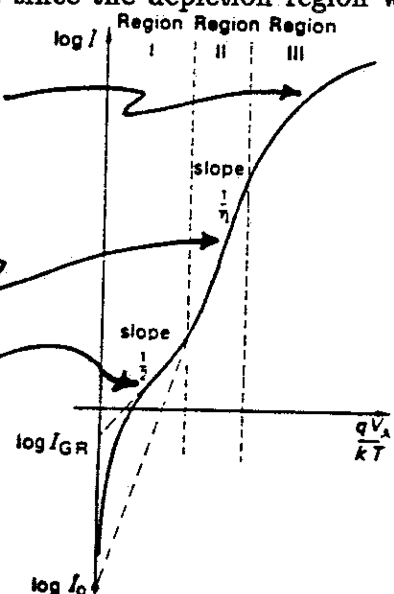
Therefore, the recombination current will dominate for small forward biases, while for larger forward biases the ideal diode current will dominate.

Reverse Bias: Under reverse bias, I_{rec} generally dominates the ideal reverse leakage current I_0. Under reverse bias $p, n \ll n_i$ so

$$U \cong -\frac{n_i}{2\tau_0} = -\frac{n_i}{\tau_n + \tau_p} \tag{73}$$

$$I_{rec} \cong \frac{qAn_i x'_d}{2\tau_0} \tag{74}$$

This current increases as the reverse bias increases since the depletion region width increases.



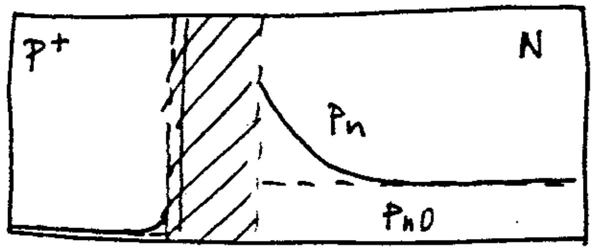
Practical Implications

- Recombination currents at low levels result in β roll-off in bipolar transistors at low current levels.
- Adding gold to switching diodes or transistors to reduce τ (improve speed) will cause larger reverse currents (higher leakage currents).

Charge Storage in PN Diodes

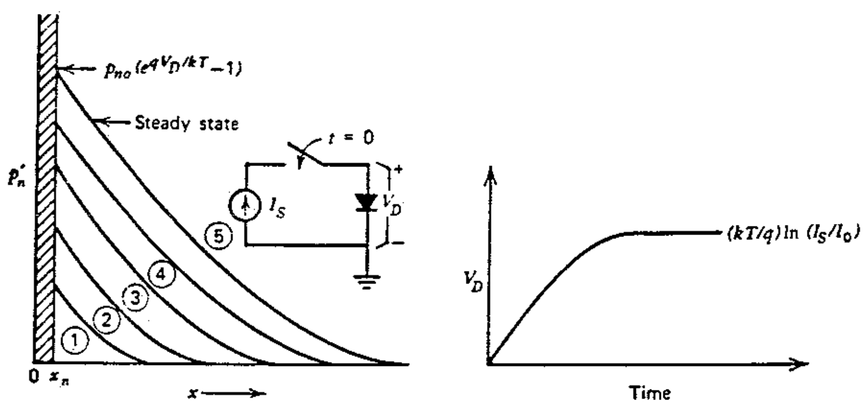
Consider the long base diode shown below (one sided)

almost no injection on p^+ side since $p^+ >> N$

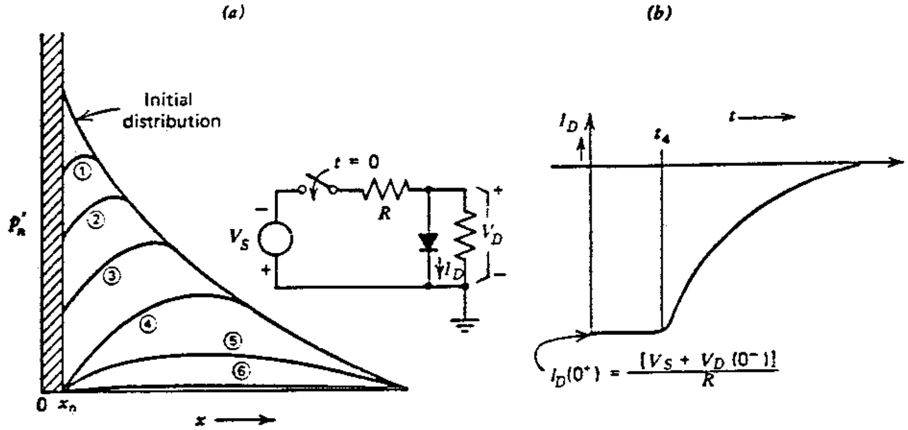


If the forward bias is removed at $t = 0$ and replaced by a reverse bias, the stored holes will flow back across the depletion region resulting in a reverse current until they are all removed.

The driving force (V_f) for injected holes is removed at t=0. The concentration decreases to equilibrium by flow backwards across the barrier.



(a) (b)



(a) (b)

The total charge stored in the N side of the diode is given by

$$Q_s = qA \int_0^{W_n} \Delta p \, dx \tag{75}$$

The continuity equation for holes is given by

$$\frac{1}{q}\frac{\partial J_p}{\partial x} + \frac{\Delta p}{\tau_p} = -\frac{\partial p}{\partial t} \tag{76}$$

Integrating from $x = 0$ to $x = W_n$, we obtain

$$\frac{1}{q}\int_0^{W_n} \frac{\partial J_p}{\partial x}\, dx + \int_0^{W_n} \frac{\Delta p}{\tau_p}\, dx = -\int_0^{W_n} \frac{\partial p}{\partial t}\, dx$$

Using (75), we obtain

$$\frac{1}{q}\left[J_p(W_n) - J_p(0)\right] + \frac{Q_s}{qA\tau_p} = -\frac{1}{qA}\frac{dQ_s}{dt}$$

or

$$I_p(0) - I_p(W_n) = \frac{dQ_s}{dt} + \frac{Q_s}{\tau_p} \tag{77}$$

This is the charge control equation and relates current to stored charge.

In the long base diode, $I_p(W_n) = 0$ and under **steady state,** forward bias conditions, $dQ_s/dt = 0$, so that

$$I_f = I_p(0) = \frac{Q_s}{\tau_p} \tag{78}$$

- The stored charge is directly proportional to I_f and to τ_p.

If we reverse the current direction, equation (77) becomes

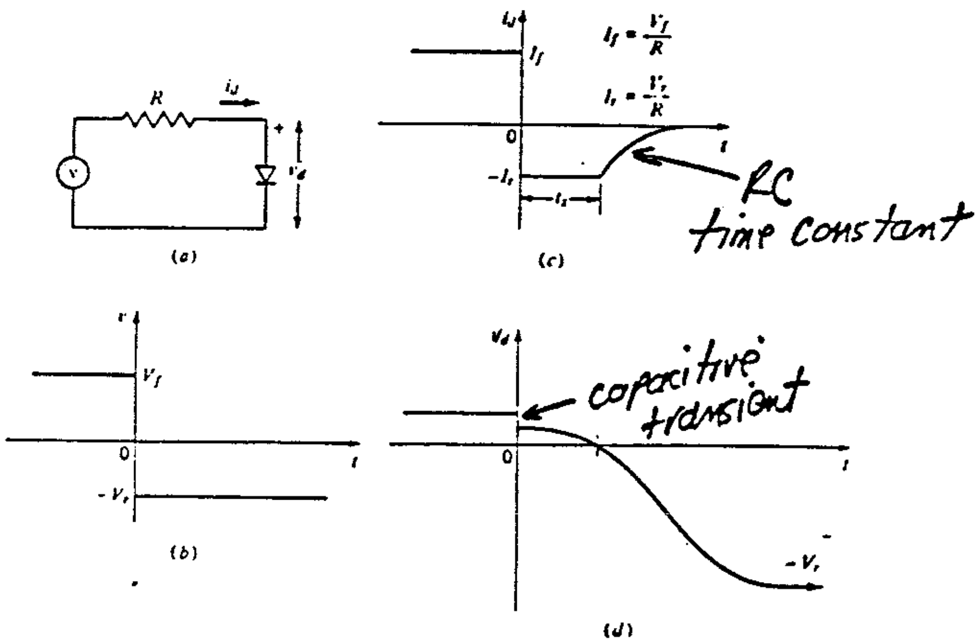
$$-I_R = \frac{dQ_s}{dt} + \frac{Q_s}{\tau_p} \tag{79}$$

This yields as a solution

$$Q_s(t) = \tau_p\left[-I_R + (I_f + I_R)\exp -\frac{t}{\tau_p}\right] \tag{80}$$

The stored charge goes to 0 when $Q_s(t) = 0$, so that

$$t_s = \tau_p \ln\left(1 + \frac{I_f}{I_R}\right) \tag{81}$$



Note:

1. The storage time is proportional to τ_p. Therefore, Au doping or damage to reduce τ_p reduces storage time.

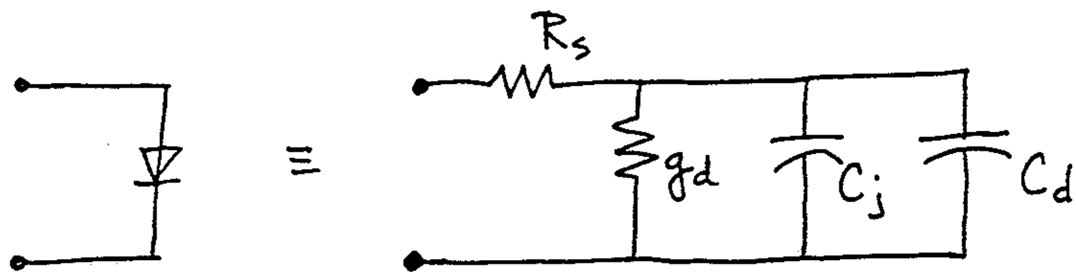
2. Diode supports full reverse voltage after t_s and an RC decay takes place due to diode capacitance and circuit-diode R.

3. Increasing reverse current decreases t_s.

4. Higher forward bias leads to longer t_s since more charge is stored.

Circuit Models for Diodes

A simple circuit model for the PN junction is as follows:



Note that this is a small signal model because all parameters are $f(V)$.

R_s = diode series resistance (ohmic contacts, bulk resistance)

$$g_d = \text{diode conductance (1/diode resistance)} = \frac{dI}{dV} = \frac{q}{kT} I_0 \exp \frac{qV_a}{kT} \cong \frac{q}{kT} I \tag{82}$$

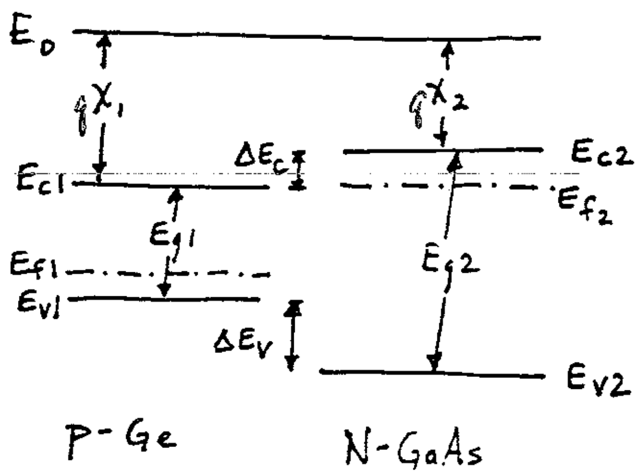
$$\frac{C_j}{A} = \text{depletion region capacitance} = \left[\frac{qK\epsilon_0}{2\left(\frac{1}{N_a} + \frac{1}{N_d}\right)(\phi_i - V_a)} \right]^{1/2} \quad \text{step junction} \tag{83}$$

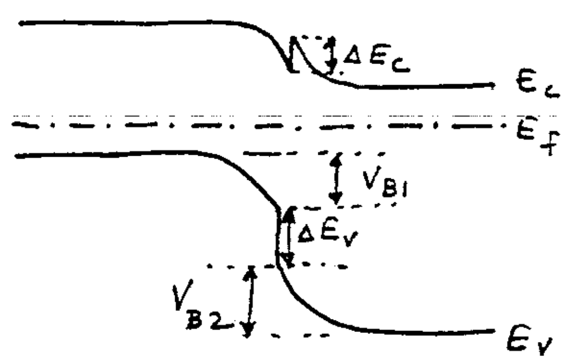
$$\frac{C_d}{A} = \text{diffusion capacitance (due to stored minority charge)} = \frac{qI_0\tau_p}{kT} \exp \frac{qV_a}{kT} \tag{84}$$

All of the parameters except R_s depend on V_a. For large signal calculations, piecewise linear models can be used or computer calculations.

Heterojunctions

Heterojunctions are interface between two different semiconductors. They are grown epitaxially and require an approximate lattice fit between the layers in order to grow single crystal materials. They can be isotype (n-N or p-P) or anisotype (p-N or n-P) where the capital letters signify the material with the larger bandgap.



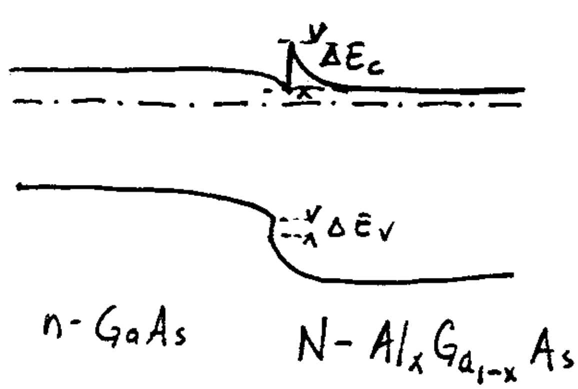


$$\Delta E_c = (\chi_1 - \chi_2)\, q$$

$$\Delta E_v = \Delta E_g - \Delta E_c$$

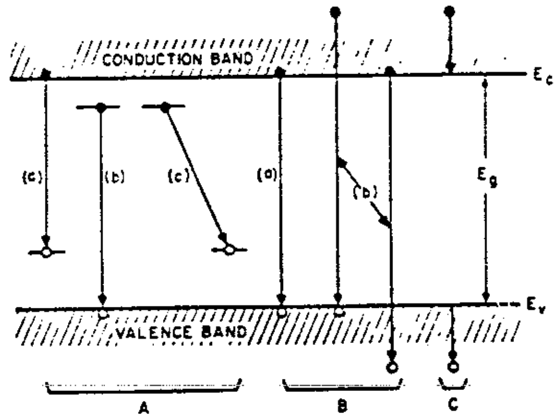
The amount of band bending in each material can be calculated using Poisson's Equation with different dielectric constants.

- The barrier for hole $\neq$ the barrier for electrons.



Light Interactions with PN Junctions

Absorption and Emission of Photons



$$E_{ph} = h\nu = h\frac{c}{\lambda} = \frac{1.24}{\lambda(\mu\text{m})}\,\text{eV}$$

$$\lambda_{max} = \frac{1.24\,\text{eV}}{E_g(\text{eV})} \qquad \text{for band to band interactions (w/o defects)}$$

Semiconductor material	λ_{max} (μm)	Working temperature (°K)
ZnS	0.35	300
CdS	0.52	300
GaP	0.55	300
GaAs	0.88	300
Si	1.2	300
Ge	1.8	300
InAs	3.0	77
InSb	5.5	77
Hg CdTe	8–14	77
Pb/SnTe	8–14	77
Ge:Cu	30	18

Absorption

Let ϕ = photon flux. The generation rate of hole-electron pairs then is

$$g(x) = -\frac{d\phi}{dx} = \alpha(\lambda)\phi(x) \tag{85}$$

where $\alpha(\lambda)$ is the absorption coefficient.

$$\phi(x) = \eta\phi_0 \exp(-\alpha(\lambda)x) \tag{86}$$

where $\phi_0 = P/E_{ph} = P/h\nu = P\lambda/hc$ is the incident photon flux and η is the percent of incident photons which are not reflected.

$$g(x) = \eta\alpha\phi_0 \exp(-\alpha x) \tag{87}$$

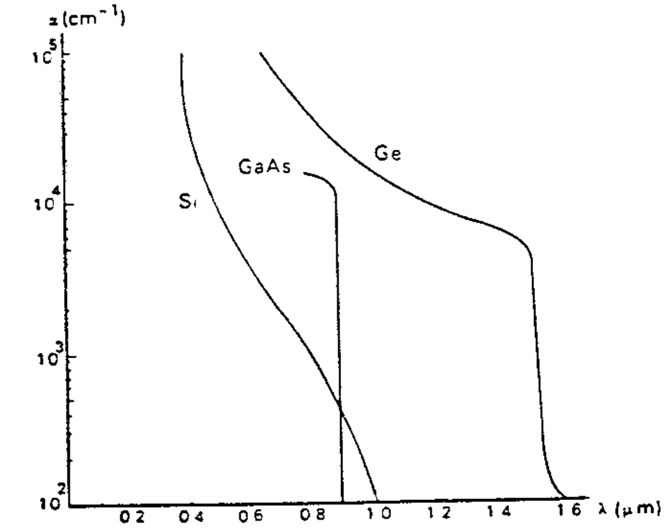
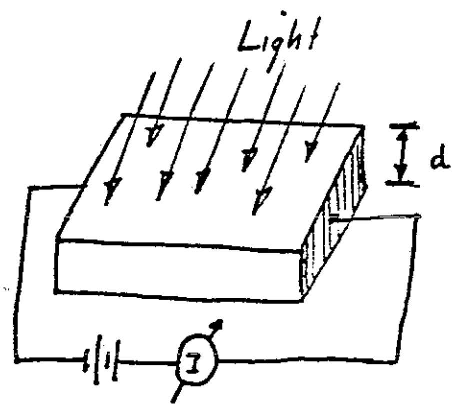


Figure 10.1. Absorption coefficient $\alpha(\lambda)$ for Si, Ge and GaAs at 300°K.

For $\lambda > \lambda_{max}$, $\alpha \to 0$. For $\lambda < \lambda_{max}$, α becomes very large for direct band gap materials. For indirect band gap materials like silicon however, for $E_{ph} \sim E_g$ phonon interactions are required so α remains small until $E_{ph} \sim E_{direct}$.



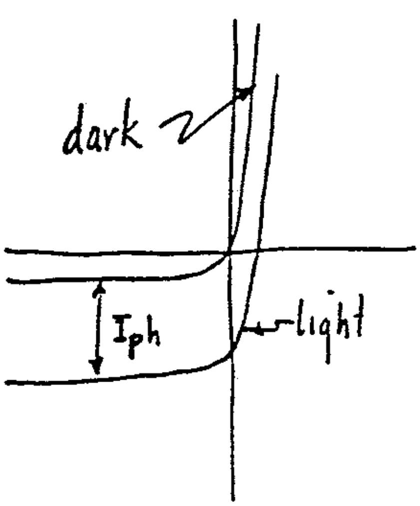
~~For a photoconductor, what is important is the total amount of~~ generation.

$$g_t = \int_0^d g(x)\,dx = \eta\phi_0\,[1 - \exp(-\alpha d)] \tag{88}$$

For $d \to 0$, $g_t \to 0$. For $d \to \infty$, $g_t \to \eta\phi_0$. The optimum thickness is $d \cong 1.25/\alpha$ since for a very thin material, very little light is absorbed, while for a thick material, too much of the material is unaffected by the light so the percentage change in conduction is small.

If light shines on a diode, the carriers created in the depletion region result in I_{GR}. Minority carriers generated within a diffusion length of the depletion region also have a good chance ending up as current by diffusing to the interface before recombining

$$p(\text{carrier} \to \text{current}) = \exp - \left(\frac{\Delta x}{L_n}\right)$$

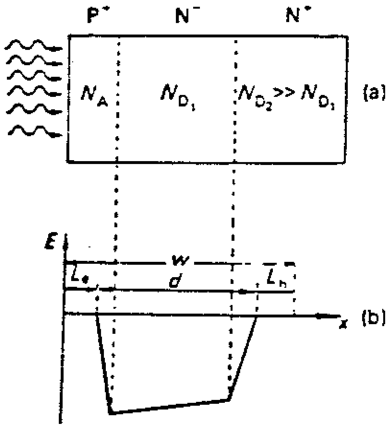


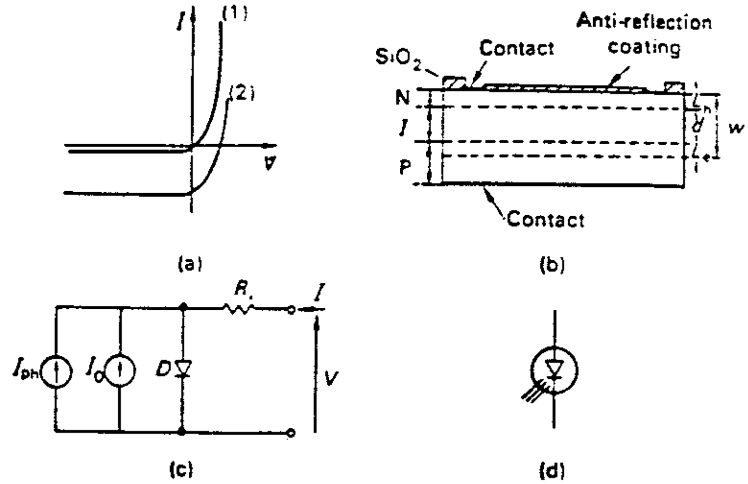
Therefore, the I-V characteristic is shifted by the photocurrent due to generation in or near the depletion region.

$$\begin{aligned} I_{ph} &= \int_\lambda qA \int_{x_1}^{x_2} g(x,\lambda)\,dx\,d\lambda \\ &= qA \int_\lambda \phi_0(\lambda)\eta(\lambda)\,[\exp(-\alpha(\lambda)x_1) - \exp(-\alpha(\lambda)x_2)]\,d\lambda \end{aligned} \tag{89}$$

$$I = I_0(\exp\frac{qV_a}{kT} - 1) - I_{ph} + I_{GR} \tag{90}$$

PIN Photodetector





A diode can be used as a photodetector by operating it under reverse bias. In order to get maximum current for a given incident flux, x_1 should be made as small as possible and x_2 as large as possible. x_1 can be reduced by making the top region narrow ($W_n < L_p$) so that only a small percentage of the photons are absorbed in that region and most of those that do diffuse to the interface before recombining. If, however, the light is absorbed too strongly near the surface, the large recombination velocity of the interface will cause a loss of carriers. In order to make x_2 large, the depletion region should be made as wide as possible. This can be accomplished by having a very lightly doped region (nearly intrinsic) between the P and N regions (pin structure). For these limiting cases

$$I_{ph} \cong qA \int_{\lambda} \phi_0(\lambda)\eta(\lambda)\,[1 - \exp(-\alpha(\lambda)W)]\; d\lambda \tag{91}$$

where $W = W_n + x_d + L_n$.

The response time of a diode photodetector is limited by the transit time across the depletion region, which for the pin structure is almost all in the intrinsic region.

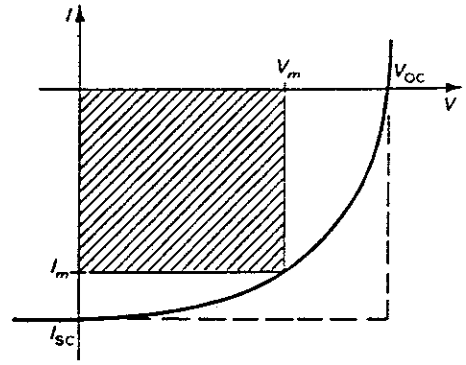
$$t_{tr} \sim \frac{x_i}{\mu \mathcal{E}} \cong \frac{x_i{}^2}{\mu V}$$

A photodetector can also be operated at the edge of avalanche, causing avalanche multiplication to magnify the current but such diodes are also much more noisy and behave nonlinearly. The minimum detectable signal is approximately the reverse leakage current (I_0) plus current due to background radiation.

An alternative method by which to increase sensitivity is by using a phototransistor, light shining on the base creates hole elecron pairs which as as base current and become multiplied by the gain of the transistor.

Photovoltaic Effect

A photodiode can also be operated in the fourth quadrant. In this case the illuminated diode acts as a source rather than a sink. The short circuit current is given by $I_{SC} = -I_{ph}$. Under open circuit conditions $I = 0$ so



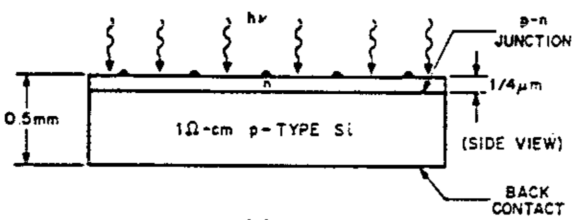
$$I_0\left(\exp\frac{qV_{OC}}{kT} - 1\right) = I_{ph},$$

and the open circuit voltage is given by

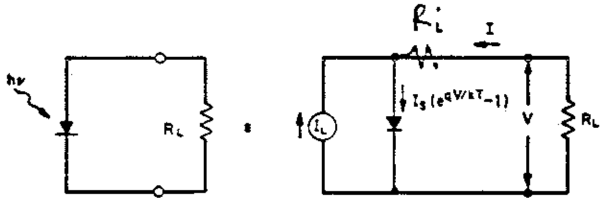
$$V_{OC} = \frac{kT}{q}\ln\left(\frac{I_{ph}}{I_0} + 1\right).$$

Solar Cell

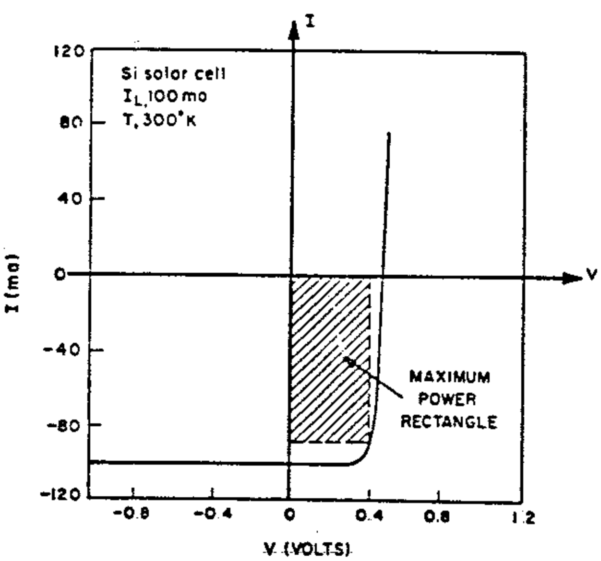
A photodiode can also be operated with a load to produce power from the incident radiation.



(a)



(b)



$$V_L = -IR_L = V_f + IR_i$$

$$V_f = \frac{kT}{q}\ln\left(\frac{I_{ph}+I}{I_0} + 1\right).$$

$$V_L = \frac{kT}{q}\frac{R_L}{R_i + R_L}\ln\left(\frac{I_{ph}+I}{I_0} + 1\right).$$

$$P = -IV_L = -CI\ln\left(\frac{I_{ph}+I}{I_0} + 1\right). \qquad (92)$$

where

$$C = \frac{kT}{q}\frac{R_L}{R_i + R_L}$$

In practice, $I_{ph} + I \gg I_0$ so

$$P \cong -CI \ln\left(\frac{I_{ph}+I}{I_0}\right). \tag{93}$$

The power output is maximized when $dP/dI = 0$ or

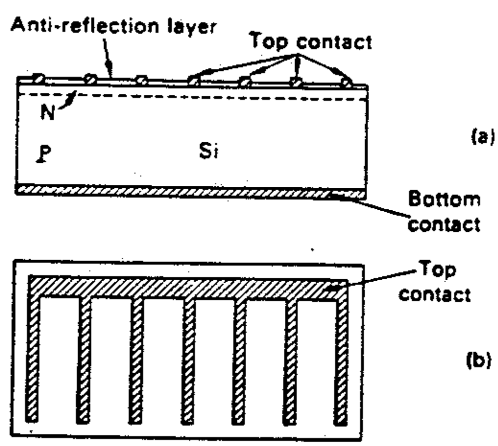
$$\ln\left(\frac{I_{ph}+I}{I_0}\right) = -\frac{I}{I_{ph}+I} \tag{94}$$

Equation (94) must be solved iteratively. If the solution is I_m then

$$P_{max} = CI_m \frac{I_m}{I_{ph}+I_m} = \frac{kT}{q}\frac{R_L}{R_i+R_L}\frac{{I_m}^2}{I_{ph}+I_m} \tag{95}$$

High power requires low I_0 (larger V_f) and low R_i (lower losses in neutral regions). Heavy doping reduces both I_0 and R_i, but also reduces depletion region width and eventually reduces minority carrier lifetime, mobility and diffusion length and thus I_{ph}.

Practical solar cells use shallow, heavily doped n-regions ($\sim 10^{18}$) on top of wide, more lightly doped p-regions ($\sim 10^{16}$). "Comb" contact is used to reduce R_i due to shallow n-region. Antireflective coatings are use to absorb maximum amount of light. Maximum achievable efficiencies are on the order of 15% with $V_{OC} \sim 0.6\,V$. Parallel and series connections are used to produce higher current and voltage.



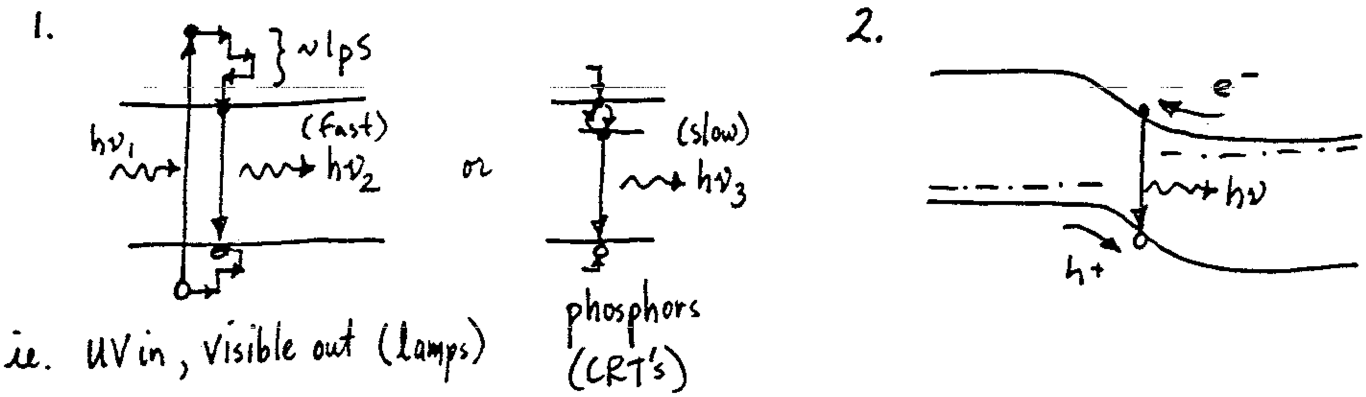
Light Emitting Diodes

For light emission we need excited carriers. These can be provided by two different injection mechanisms:

1. Generation by absorbing photons or high energy electrons

2. Injection of carriers from nearby regions (ie. forward biased diode)

Recombination will only result in radiation if radiative (direct) recombination dominates.



Optical Coupling

A combination of a LED and a photodetector can provide transfer of information while keeping the source and detector electrically isolated (no unwanted feedback). Optical communication links using optical fibers can transmit large quantities of information due to the small wavelength of light.

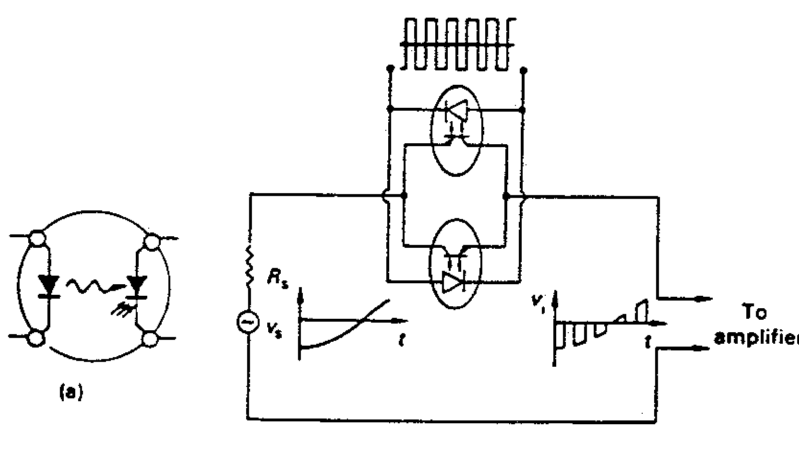


Figure 10.8. Optical coupling: (a) photon-coupled isolator; (b) use of two isolators to obtain a signal-chopping circuit.

Semiconductor Laser

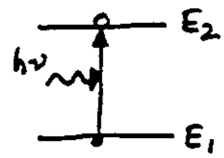
There are several requirements which much be fulfilled for lasing to occur.

1. Need direct band gap for recombination to occur predominately by photon emission.

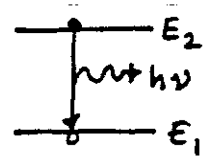
2. Need population inversion.

   There are three interactions that can occur between light and a semiconductor. $h\nu = E_2 - E_1$

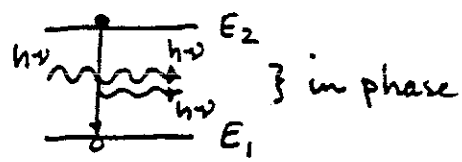
   (a) Absorption



   (b) Spontaneous Emission



   (c) Stimulated Emission



Based on quantum mechanics, it is known that the rate constant for stimulated emission is the same as for absorption. Which process dominates depends on the relative numbers of full and empty states.

$$p(\text{absorption}) \propto p(\text{full state at } E_1)p(\text{empty state at } E_2 = E_1 + h\nu)$$

$$p(\text{stimulated emission}) \propto p(\text{empty state at } E_1)p(\text{full state at } E_2)$$

$$W_{\text{absorption}} = K \int N_c(E)\,[1 - f_c(E)]\,N_v(E - h\nu)f_v(E - h\nu)\,dE \tag{96}$$

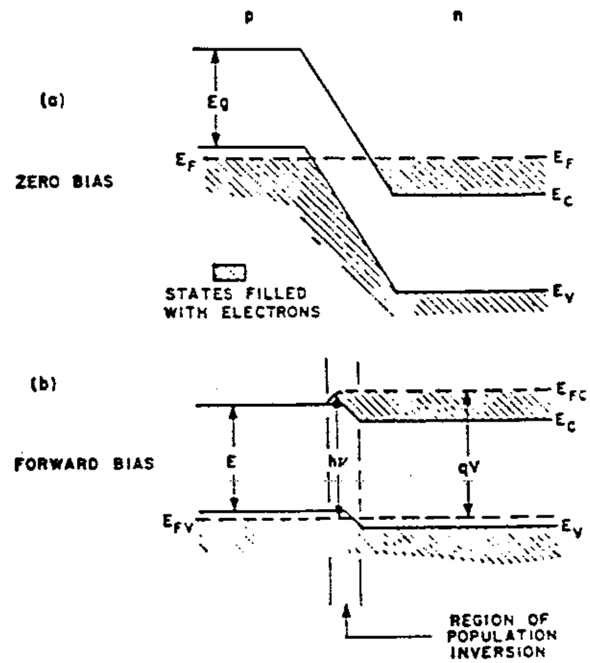
$$W_{\text{emission}} = K \int N_c(E)f_c(E)N_v(E - h\nu)\,[1 - f_v(E - h\nu)]\,dE \tag{97}$$

where

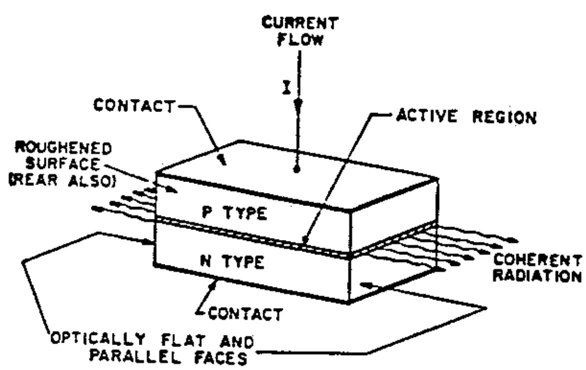
$$f_c = \frac{1}{1 + \exp\left(\dfrac{E - E_{fn}}{kT}\right)}$$

$$f_v = \frac{1}{1 + \exp\left(\dfrac{E - E_{fp}}{kT}\right)}$$

In a normal semiconductor, the valence band is mostly full and the conduction band is mostly empty and most photon interactions will result in absorption. Under conditions of very high level injection the quasi-Fermi levels can move into the bands so that the bottom of the conduction band is mostly full while the top of the valence band is mostly empty. In that case, known as population inversion, most photon interactions will result in stimulated emission rather than absorption. For this to occur $E_g < h\nu < (E_{fn} - E_{fp})$. Therefore, similar to the case of avalanche breakdown, each incident photon can create more photons all with the same phase. The gain is given by $g(\nu) \propto W_{\text{emission}} - W_{\text{absorption}}$.



3. The final requirement is confinement. In other words, the number of photons created in the active region must be more than the number lost to the rest of the semiconductor where absorption dominates. Lasers are made with cavities that confine the light. The most common type of cavity is reflective on two sides and roughened on the other two, in order to get only a single operating mode. The light is magnified in the direction with the reflective sides.



FABRY-PEROT CAVITY

In order for lasing to occur, light must completely traverse cavity without attenuation.

$$R\exp\left((g-\alpha)L\right) > 1$$

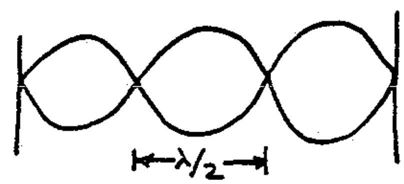
where R is the reflectivity of the end facets, α is the attenuation factor due to free carrier absorption and scattering and L is the length of the cavity. The minimum gain required then is

$$g_{min} = \alpha + \frac{1}{L}\ln\frac{1}{R}.$$

The size of the cavity determines the exact wavelength at which the laser will operate since reflection can only occur for constructive interference of the incident and reflected waves. Therefore, the length of the cavity must be an integral number of half-wavelengths.

$$L = m\left(\frac{\lambda}{2n_r}\right)$$

where n_r is the index of refraction.



m = 3

Heterojunction Lasers

The most efficient lasers are made using heterojunctions. The varying composition can both concentrate the recombination process in a small region while also containing the photons by the differences in refractive indices of the layers.

These lasers can be modulated very quickly (up to 2 GHz) and by altering the composition of the layers, the output wavelength can be controlled. Since the active region is very small relatively small current levels can sustain the lasing action (10-20 mA).