Semiconductor Devices: Lectures 1&2

I. Carrier Transport in Semiconductors

To begin with we need to review some of the basic principles which govern charge transport due to the presence of (i) an applied electric field, (ii) a charge density gradient.

(1) Carrier drift in applied field

The electron drift velocity is:

$$v_n = \frac{q\tau}{m_n^*} E = \mu_n E$$

where τ is the mean time between scattering events, and μ_n is the electron *mobility*. There is a similar relation for holes. The mobility depends on field at high field due to inelastic scattering processes, and the drift velocity saturates; the maximum value of v_n is $\sim 10^5 \mathrm{ms}^{-1}$ in Si. The net current is:

$$J = J_n + J_p = q(n_0\mu_n + p_0\mu_p)E$$

(2) Carrier Diffusion

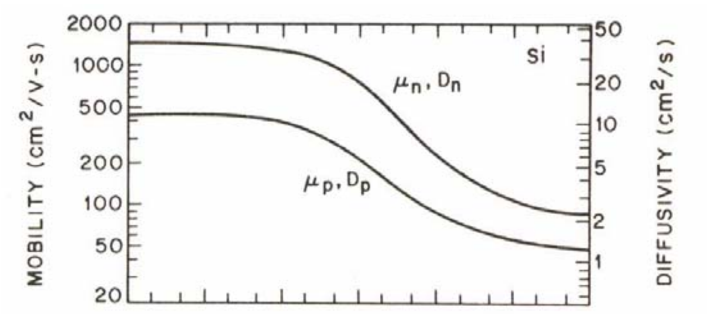
When there is a spatial variation in the electron density, there is a diffusion current. In one dimension:

$$J_n = -qD_n \frac{dn}{dx}$$

where D_n is the electron *diffusivity*. D_n and μ_n are related via the Einstein relation:

$$D_n = \frac{k_B T}{q}\mu_n = \frac{k_B T\tau}{m_n^*}$$

The hole mobility and diffusivity are usually much smaller than the electron values because of its heavier mass.



Mobility and diffusivity data for Si at 300K as a function of dopant density. The decrease at high density is due to scattering from impurities.

(3) Carrier Recombination

The recombination rate ***R*** is proportional to the product of the number of electrons in the conduction band and the number of holes in the valence band:

$$R = \beta np$$

In thermal equilibrium the recombination rate is equal to the generation rate:

$$G_{th} = R_{th} = \beta n_0 p_0$$

If there is an excess of carriers of a particular type, e.g. caused by illumination of a doped semiconductor, the excess carriers will recombine. If Δp holes are injected into n-type material:

$$\frac{dp}{dt} = -\beta n_0 p \quad \Rightarrow \quad p(t) = p_0 + \Delta p e^{-\beta n_0 t}$$

$(\beta n_0)^{-1}$ is the *minority carrier lifetime* τ_p.

(4) Continuity

In general, carrier populations vary in time and space due to the combined effects of diffusion and recombination. The continuity equations for electrons and holes are:

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} - R_n \qquad \frac{\partial p}{\partial t} = -\frac{1}{q}\frac{\partial J_p}{\partial x} - R_p$$

Under steady state conditions the minority carrier distribution is given by:

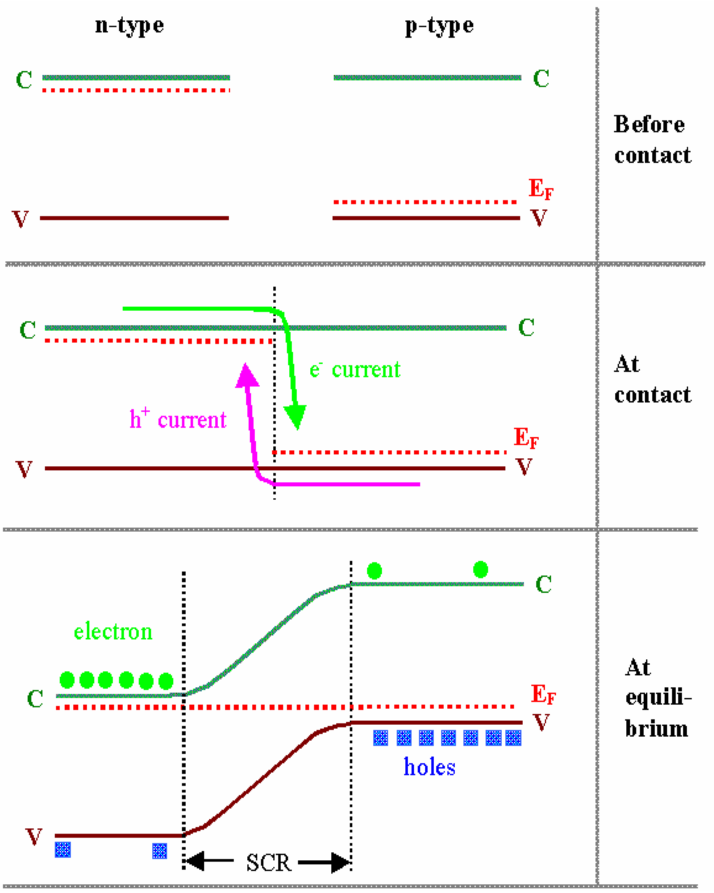
$$\frac{\partial p}{\partial t} = 0 = D_p \frac{\partial^2 p}{\partial x^2} - \frac{p - p_0}{\tau_p}$$

Using the boundary condition $p(x \rightarrow \infty) = p_0$ yields the solution:

$$p(x) = p_0 + [p(0) - p_0]e^{-x/L_p}$$

where $L_p = \sqrt{D_p \cdot \tau_p}$ is the *diffusion length*, and $(p(0)-p_0)$ is the excess hole population at x=0.

II. The PN-Junction



Now consider a semiconductor that contains both *p*-type and *n*-type regions. The large carrier concentration gradients at the junction between *n*-type and *p*-type regions cause carrier diffusion. This leaves positively charged donor ions on the *n*-side and negatively charged acceptor ions on the *p*-side. Consequently there is a negative space charge region (SCR) on the *p*-side and a positive SCR on the *n*-side of the junction.

The SCR creates an electric field which points from *n-to-p*. The hole drift current is clearly *n-to-p* whereas the hole diffusion current is *p-to-n*. The electron diffusion current is *p-to-n,* and the drift current is *n-to-p*. At equilibrium there is no net current across the junction, and the Fermi level is the same on both sides of the junction.

(5) P-N Junction at Equilibrium

The electrostatics of the junction at equilibrium are described by the Poisson equation:

$$\frac{d^2V}{dx^2} = -\frac{dE}{dx} = -\frac{\rho}{\varepsilon_0\varepsilon_r} = -\frac{q}{\varepsilon_0\varepsilon_r}\left(N_D - N_A + p - n\right)$$

where we have assumed that all donors and acceptors are ionized. Far away from the junction we have:

$$\frac{d^2V}{dx^2} = 0 \quad \text{so that} \quad \left(N_D - N_A + p - n\right) = 0$$

Assume the junction is abrupt. In the depletion region the free carriers are totally depleted, so that :

$$\frac{d^2V}{dx^2} = \frac{qN_A}{\varepsilon_0\varepsilon_r} \qquad -d_A \le x < 0$$

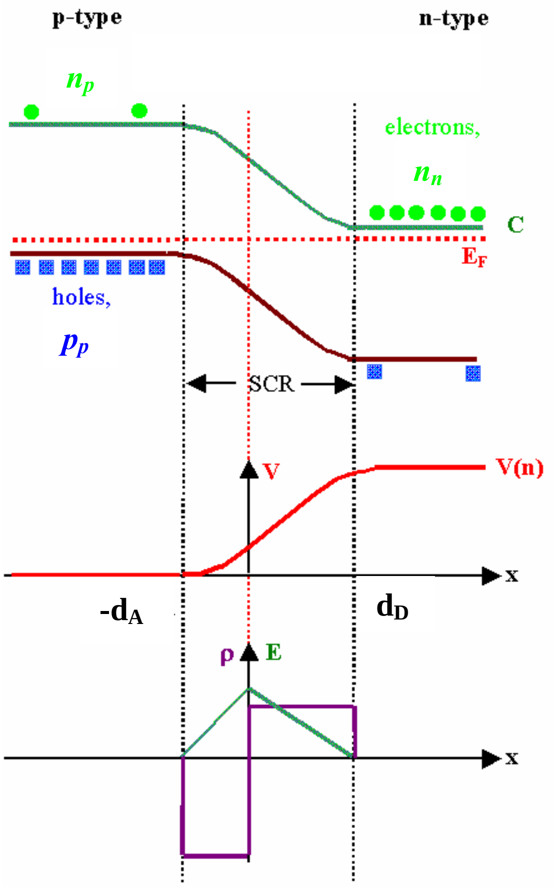
$$\frac{d^2V}{dx^2} = -\frac{qN_D}{\varepsilon_0\varepsilon_r} \qquad 0 < x \le d_D$$

The electric field is then:

$$E(x) = -\frac{qN_A(x + d_D)}{\varepsilon_0\varepsilon_r} \qquad -d_A \le x < 0$$

$$E(x) = \frac{qN_D x}{\varepsilon_0\varepsilon_r} - E_M \qquad 0 < x \le d_D$$

where $E_M = \dfrac{qN_D d_D}{\varepsilon_0\varepsilon_r} = \dfrac{qN_A d_A}{\varepsilon_0\varepsilon_r}$



This figure illustrates the variation of the energy bands across the junction together with the field E, potential V and charge density ρ.

(6) Depletion Layer at a *p-n* junction

The potential difference across the junction is:

$$V_C = -\int_{-d_A}^{d_D} E(x)dx = \int_{-d_A}^{0} \frac{qN_A(x + d_A)}{\varepsilon_0\varepsilon_r}dx - \int_{0}^{d_D} \frac{qN_D(x - d_D)}{\varepsilon_0\varepsilon_r}dx$$

$$= \frac{q}{2\varepsilon_0\varepsilon_r}(N_A d_D^2 + N_D d_D^2) \; = \frac{1}{2}E_M W \quad \textit{(i.e. the area under E)}$$

where W is the width of the junction:

$$W = d_A + d_D = \sqrt{\frac{2\varepsilon_0\varepsilon_r(N_A + N_D)V_C}{qN_A N_D}}$$

V_C can be obtained from the condition that the drift current balances the diffusion current:

$$-n(x)q\mu_e\frac{dV}{dx} + qD_e\frac{dn(x)}{dx} = 0$$

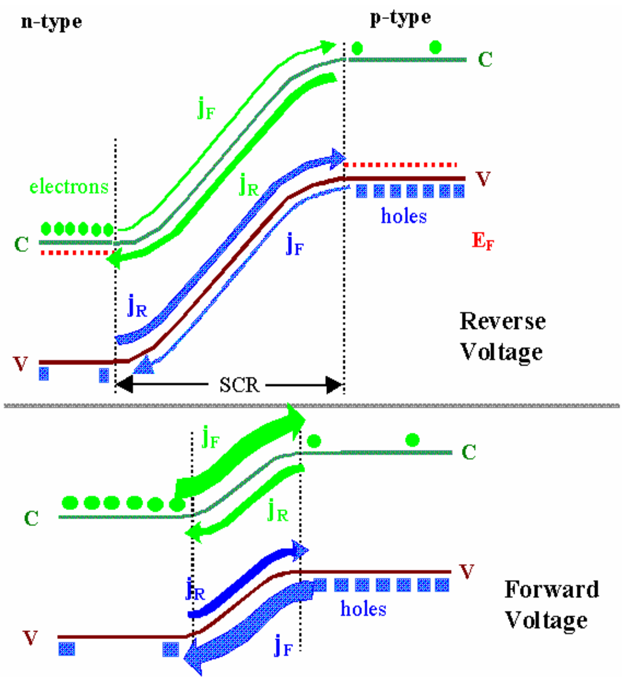
Integrate this from $-d_A$ to d_D to obtain:

$$V_C = V_n - V_p = \frac{k_B T}{q} ln(\frac{n_n}{n_p})$$

where n_n is the electron density in the *n*-region, and n_p is the electron density in the *p*-region.

N.B. $n_p = n_i^2 / N_A$, where n_i is the intrinsic carrier density.

III. P-N Junction under bias



A voltage applied to a p-n junction changes the balance between the drift and diffusion currents. Under ***forward*** bias the applied voltage reduces the electrostatic potential across the junction; the drift current is reduced in comparison with the diffusion current, giving minority carrier injection – holes into the *n*-region and electrons into the *p*-region. Under ***reverse*** bias the electrostatic potential is increased which dramatically reduces the diffusion currents, leaving a small reverse current.

(7) Voltage-Current Relation for a PN Diode

The ideal diode is assumed to operate under the following conditions:

(i) the depletion layer is abrupt, and there is charge neutrality outside of the layer,
(ii) the charge densities at the boundaries are given by the electrostatic potential,
(iii) the minority carrier injection is weak, much less than the majority densities,
(iv) there are no generation or recombination currents in the depletion layer.

The relation for V_C can be re-written:

$$n_n = n_p e^{qV_C/k_BT} \quad \text{and} \quad p_p = p_n e^{qV_C/k_BT}$$

We expect these relations to hold for non-equilibrium carrier densities when a bias V is applied:

$$\hat{n}_n = \hat{n}_p e^{q(V_C-V)/k_BT} \quad \text{and} \quad \hat{p}_p = \hat{p}_n e^{q(V_C-V)/k_BT}$$

Since $\hat{n}_n \approx n_n$ we then have :

$$\hat{n}_p - n_p = n_p(e^{qV/k_BT} - 1) \quad at \quad x = -d_A : \hat{p}_n - p_n = p_n(e^{qV/k_BT} - 1) \quad x = d_D$$

In the neutral *n*- and *p*-regions there is no electric field, so the continuity equations yield:

$$\frac{d^2\hat{p}_n}{dx^2} = \frac{\hat{p}_n - p_n}{D_p\tau_p} \quad \rightarrow \quad \hat{p}_n - p_n = p_n(e^{qV/k_BT} - 1)e^{-(x-d_D)/L_D}$$

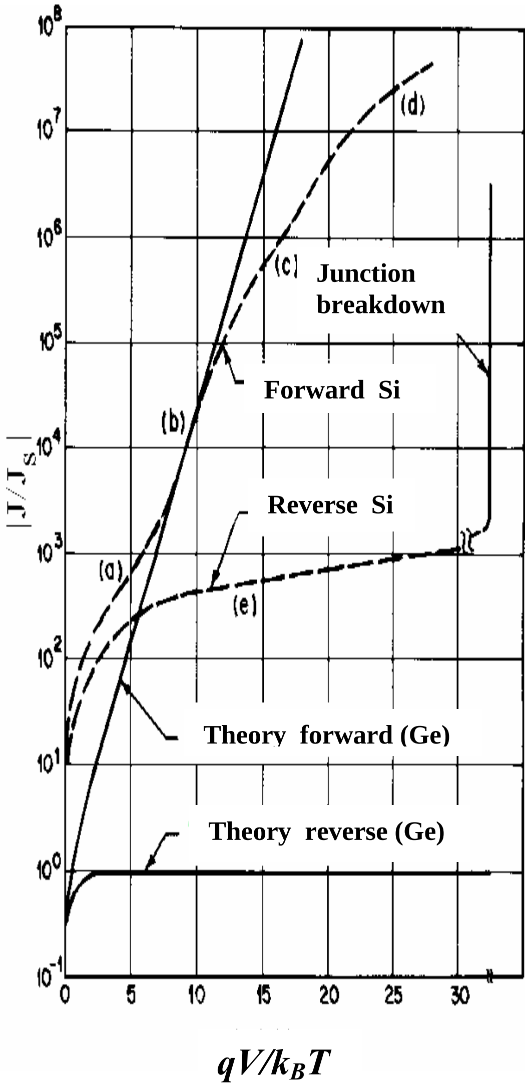
$$\frac{d^2\hat{n}_n}{dx^2} = \frac{\hat{n}_p - n_p}{D_n\tau_n} \quad \rightarrow \quad \hat{n}_p - n_p = n_p(e^{qV/k_BT} - 1)e^{(x+d_A)/L_A}$$

The total current flowing is:

$$J = J_p(d_D) + J(-d_A) = \left[-qD_p\frac{d\hat{p}_n}{dx}\right]_{d_D} + \left[qD_n\frac{d\hat{n}_p}{dx}\right]_{-d_A}$$

$$= J_S(e^{qV/k_BT} - 1), \qquad where \quad J_S = \frac{qD_p p_n}{L_p} + \frac{qD_n n_p}{L_e} = qn_i^2\left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_e N_A}\right)$$

Thus the current increases exponentially under forward bias, but saturates at $-J_S$ for negative bias.



The model described above applies more or less to Ge *p-n* diodes, but not for Si or GaAs diodes. The reason for this is that Ge has a relatively high intrinsic carrier density because of its small bandgap:

n_i(Ge) = $2.5\text{x}10^{13}\text{cm}^{-3}$
n_i(Si) = $1.45\text{x}10^{10}\text{cm}^{-3}$
n_i(GaAs) = $1.79\text{x}10^{6}\text{cm}^{-3}$

and at room temperature the diffusion current dominates. However, if n_i is small, the generation current in the depletion region can dominate.

The large asymmetry in the current-voltage characteristics of a p-n junction makes it an ideal device for rectifying *ac* currents. There are many other applications, e.g. photodiodes, solar cells, light-emitting diodes and diode lasers.

Junction breakdown occurs at large reverse bias voltage by two different mechanisms:

(a) ***tunnelling***: a valence electron in the p-region makes a transition to the conduction band of the n-region. This requires a field of $>10^8\text{Vm}^{-1}$, which requires high dopant density $\sim 5\text{x}10^{23}\text{m}^{-3}$.

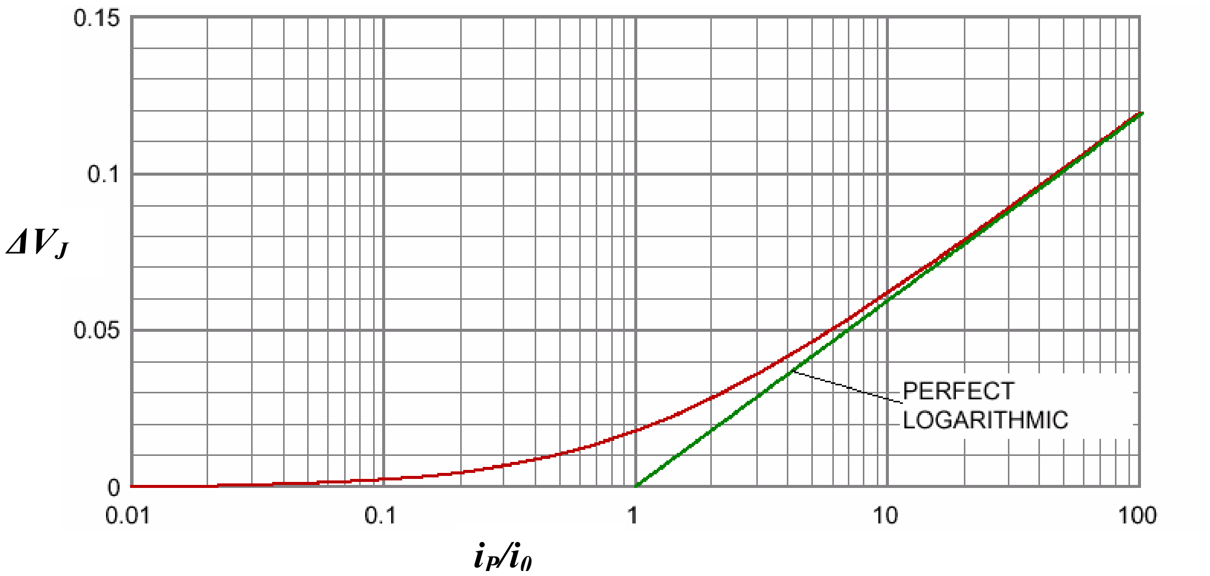
(b) ***avalanche***: a thermally generated electron gains enough kinetic energy in the electric field ($>qE_G$) to cause impact ionisation in which an electron-hole pair are generated. Multiplication factors ~100 can be achieved with several hundred volts of reverse bias.

IV. Photodiodes

Photodiodes are convenient devices for measuring optical signals. There are two modes of operation:

(i) Photovoltaic mode: When a p-n junction is illuminated by above-bandgap light electron-hole pairs are generated which immediately come under the effect of the junction electric field: electrons are swept to the positively charged *n*-region and holes to the negatively charged *p*-region, thus reducing the voltage below its equilibrium value. The current generated by a beam of light with power $\boldsymbol{P}$ and photon energy $\boldsymbol{h\nu}$ is: $\boldsymbol{i_P = \frac{q\eta P}{h\nu}}$, where $\boldsymbol{\eta}$ is the fraction of photons that generate e-h pairs. The new drift current is $\boldsymbol{i_0 + i_P}$, and so the diffusion current must also increase in order to bring the junction back to equilibrium. If the change in junction voltage is $\boldsymbol{\Delta V_J}$ the diffusion current becomes: $\boldsymbol{i_D = i_0 \, exp(\frac{q\Delta V_J}{k_B T})}$.

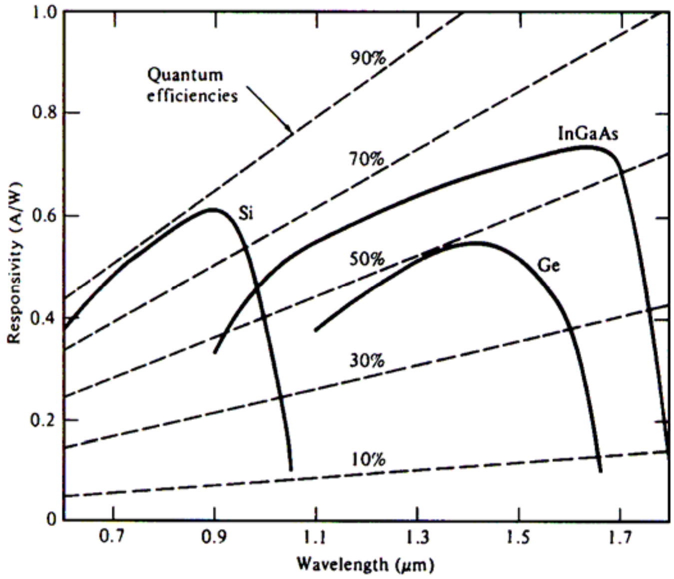
Consequently, $\boldsymbol{i_0 + i_P = i_0 \, exp(\frac{q\Delta V_J}{k_B T})}$ and $\boldsymbol{\Delta V_J = \frac{k_B T}{q} ln\left(\frac{i_P}{i_0} + 1\right)}$. Thus the open circuit voltage increases logarithmically with power at high power.



(ii) Photoconductive mode: In this case a reverse bias is applied to the junction, which has the effect of reducing the drift and diffusion currents to very low values, and increasing the width of the depletion layer. If photons are now absorbed in the depletion layer a "generation" current is produced which is proportional to the absorbed power. This is the normal mode of operation of a photodiode.
The sensitivity of the device (responsivity R_O) is the photocurrent per unit power, i.e.

$$R_O = \frac{i_p}{P} = \frac{q\eta}{h\nu} = \frac{\eta\lambda}{1.24} \text{ amps watt}^{-1}$$

The quantum efficiency of a typical Si photodiode is ~80%, so that R_O~0.55AW^{-1}. The responsivity shows a broad peak centred near the bandgap energy: it falls off at longer wavelengths due to reduced optical absorption in the depletion layer, and at shorter wavelengths as light is absorbed in the surface layer. The responsivity can be improved by incorporating a relatively thick intrinsic semiconductor layer between the two doped regions (to form a *p-i-n* diode) so that more of the incident light is absorbed – although this will depend on the wavelength.



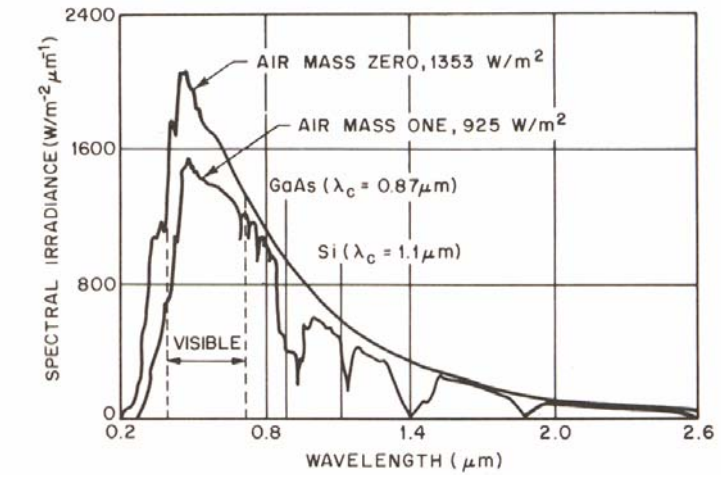
The speed of a photodiode detector depends on various factors:

a) *the time of drift of carriers across the depletion layer*: given an electron saturation velocity $v_s \sim 5x10^4 ms^{-1}$, depletion layer width W~2.5μm gives a response time ~50ps, and so a bandwith of ~20GHz.

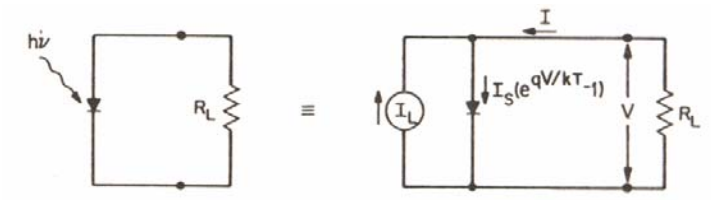
b) *diffusion through the p- and n-regions*: since diffusion is slower than drift, these layers must be much thinner than the depletion layer, and 0.5μm is typical.

c) *the RC time constant of the circuit*: to utilise the intrinsic speed of the diode, the load resistance for a typical junction capacitance of 25pF must be ~0.3Ω. For incident power P~1μW, $i_P \sim 10^{-7}$A, and so the voltage is ~30nV; this is rather low, and noise would be a severe problem. An impedance matching 50Ω load would be a better solution; the bandwidth would drop to <1GHz, but this is adequate for most purposes.

V. Solar Cells



Solar cells use p-n junctions to generate power from solar radiation. which provides about $1kWm^{-2}$ at the Earth's surface.



Solar cell equivalent circuit.

The current through the cell is given by:

$$I = I_S(exp(qV / k_B T) - 1) - I_L$$

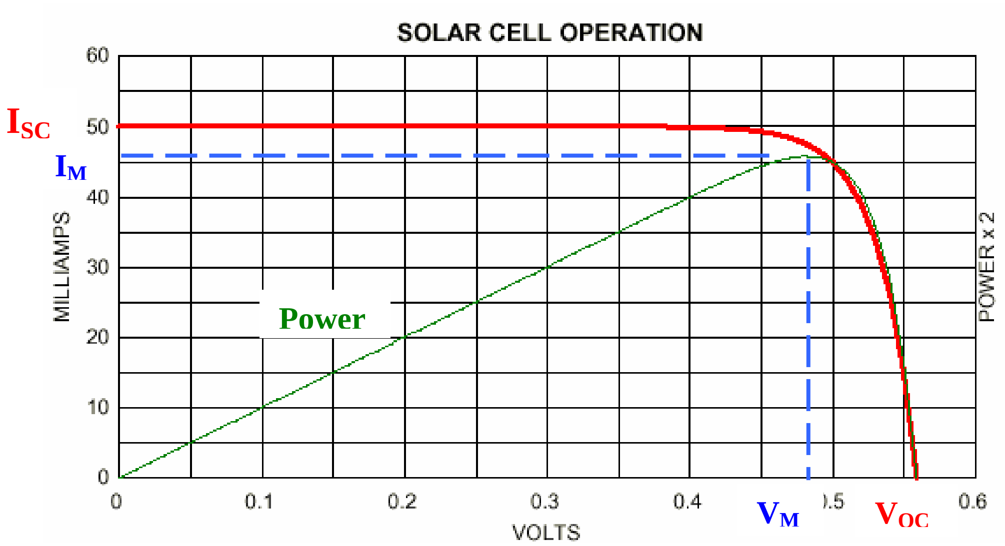
where I_L is the light-induced current, and V is the voltage across the diode.

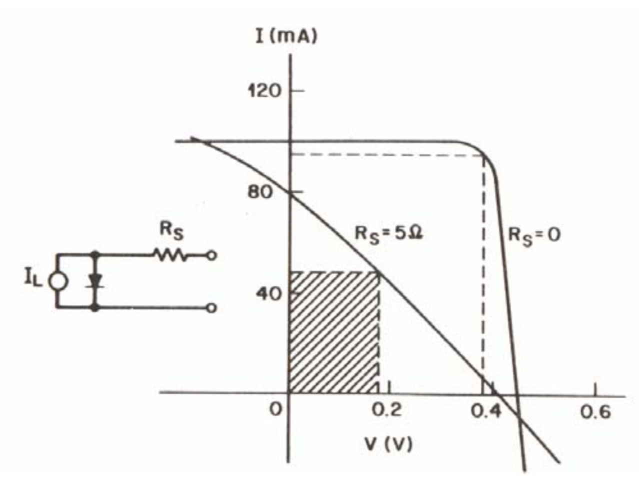
The maximum power that can be extracted is close to the value of the product of the short circuit current I_{SC} times the open circuit voltage V_{OC}, and depends on the choice of load resistor. When $I = 0$

$$V_{OC} = \frac{k_B T}{q} ln(\frac{I_L}{I_S} + 1)$$ and the output power is : $$P = IV = I_S V(e^{\frac{qV}{k_B T}} - 1) - I_L V$$

The maximum output power is:

$$P_M \cong I_L \left[V_{OC} - \frac{k_B T}{q} ln(1+\frac{qV_M}{k_b T}) - \frac{k_B T}{q} \right]$$





When there is a series resistance R_S - arising for example at the electrical contacts to the diode - the output voltage is: $V_{out} = V\text{-}IR_S$, and the power available is significantly lower. The area of the contacts can be increased in order to minimise R_S, but this reduces the effective optical apperture and so I_L. So once again we see that device design requires some degree of compromise.

.