RECOMBINATION

In equilibrium $np = n_i^2$ and generation=recombination (detailed balance). G_{th} is thermal generation.

For direct bandgap — very simple mechanism. (see Fig)

It is reasonable to expect $R \sim n$ and $\sim p$, i.e. $R \sim \beta np$. In equilibrium $R_{th} = G_{th} = \beta n_{n0} p_{n0}$ (n-type).

Light absorption $\rightarrow$ carrier pairs generated at rate G_L and n$\uparrow$ and p$\uparrow$.

Recombination works on re-establishing equilibrium.

$G = G_L + G_{th}, R = \beta n_n p_n = \beta(n_{n0} + \Delta n)(p_{n0} + \Delta p)$.

$\Delta n = \Delta p$ for charge neutrality.

From cont. equation: $\frac{\partial p_n}{\partial t} = G_L + G_{th} - R$ (no J)

In steady state there is no change in hole conc. $G_L = R - G_{th} \approx \beta n_{n0} \Delta p = \frac{p_n - p_{n0}}{\tau_p}$

What happens when the light is turned off? Exponential decay with time constant τ_p

Indirect recombination - traps involved (see fig). Four basic processes: a) electron capture, b) el. emission, c) hole capture, d) hole emission.

Case a): the rate of el. capture is proportional to conc. of traps that are not occupied by electrons, as given by $N_t(1-F)$

$$F = \left(1 + e^{\frac{E_t - E_F}{kT}}\right)^{-1}, \qquad R_a \sim n N_t (1 - F) \tag{1}$$

The proportionality constant for R_a can be written as $v_{th}\sigma_n$ (thermal velocity and capture cross section) and it measures the effectiveness of capture.

Case b): $R_b = eN_tF$, e=emission probabiliity.

In equilibrium: $R_a = R_b$ and $e_n = v_{th}\sigma_n(1-F)/F$, or after simplification (use $n = n_i \exp(\frac{E_t-E_i}{kT})$)

$$e_n = v_{th}\sigma_n n_i e^{\frac{E_t-E_i}{kT}} \tag{2}$$

?: when are these rates $\uparrow$ or $\downarrow$?

Procedure for hole capture/emission is similar.

In equilibrium: $R_a = R_b$ and $R_c = R_d$. Out of equil.: $R_a - R_b = R_c - R_d$ so that net recombination rate

$$U = \frac{np - n_i^2}{(p_n + p_1)\tau_n + (n_n + n_1)\tau_p} \tag{3}$$

what happens in low injection? $(\tau_p = v_{th}\sigma_p N_t)^{-1})$

Minority carrier lifetimes τ_n, τ_p. Recombination limited by capture of minority carrier!

For qualitative dependence of lifetime on trap levels, see fig.

What about depletion regions? Main difference: both carriers are scarce ($p_n < n_i, n_n < n_i$.

Generation rate becomes: $\frac{n_i}{\tau_g}$. For small carrier conc. energy level of traps plays much more important role (see fig). With proper choice of E_t we can get both τ_g large and τ_r small (leakage currents and turn-off).

Other recombination mechanisms — interface associated. Surface recombination, heterojunctions etc. Can be treated via surface recombination velocity concept.

Si generally has acceptable lifetimes, but that depends on technology. GaAs has much shorter lifetimes.

High-Field Effects

Changes in v vs. El. field relationship. Not a simple linear proportionality! Velocity saturation effect (see fig). Scatterings become more frequent as carriers become more energetic - reduction in velocity increase. Empirical relationship

$$v_n, v_p = \frac{v_s}{\left(1 + \left(\frac{E_0}{E}\right)^{\gamma}\right)^{1/\gamma}} \tag{4}$$

In GaAs much more complex: intervalley transfer occurs at some el. field (see fig). Very important for TED operation, but can be troublesome otherwise. (clarify meaning of overshoot)

If field high enough: impact ionization. Carrier energetic enough to create additional el-hole pairs (see fig.) If those carriers continue generating more carriers we have avalance breakdown.

Mostly try to avoid it, but some devices (e.g. IMPATTs) use the avalnche process for power generation.

Estimate of energy required fro I-I: start with $m_1 v_s^2/2$ and momentum $m_1 v_s$. After collision two additional particles — need to satisfy momentym and energy conservation. Simplification: three carriers have the same mass, kinetic energy and momentum (after I-I). Then the total energy $= 3/2mv_f^2$, and total momentum $= 3mv_f$.

The result: the minimum energy required to initate the I-I process is $E_0 = 1.5E_g$. Real values actually much larger: 3.2E_g for electrons, and 4.4E_g for holes in Si.

Ionization rate: number of e-h pairs generated by an electron per unit length traveled. (see fig.) For substantial α the el. field required is cca. 0.3 MV/cm in Si, and 0.4 MV/cm in GaAs.

Generation rates: $qG_A = \alpha_n|J_n| + \alpha_p|J_p|$. Use this in continuity equations.

p-n Junction

In equilibrium: Fermi levels have to line up. For n-type: E_F is close to conduction band edge E_C. For p-type: E_F is close to valence band edge E_V.

To reach equilibrium a redistribution of charge takes place. Very large gradients of concentrations exist as the p and n parts are brought together $\rightarrow$ large diffusion currents. The carriers go in opposite directions and as they leave the junction region, ionized dopants are left behind, resulting in the field creation. Final result: drift and diffusion current components cancel each other for each carrier separatelly.

How do we get fields and concentrations? For fields we use Poisson's equation.

$$\frac{d^2\psi}{dx^2}\left(= -\frac{dE}{dx}\right) = \frac{q}{\epsilon_s}\left(p - n + N_D^+ - N_A^-\right) \tag{5}$$

Far away from the junction we have charge neutrality, i.e. $p - n + N_D^+ - N_A^- = 0$. We can also define electrostatic potential in those regions via $q\psi_p = -(E_i - E_F) = -kT\ln(N_A/n_i)$.

Since the two potentials will be of different signs for n and p-type region, the potential difference will be the sum of the two: $V_{bi} = \psi_n - \psi_p = kT/q\ln(N_A N_D/n_i^2)$ (see figs)

Standard approximation: depletion layer approx. which enables us to treat the problem analytically (figs). Max. field and width:

$$E_m = \frac{qN_D x_n}{\epsilon_s} \qquad W = \sqrt{\frac{2\epsilon_S}{q}\left[\frac{N_A + N_D}{N_A N_D}\right]V_{bi}} \tag{6}$$

For one-sided junctions simplifications are possible. Depletion layer extends mostly onto the lighter doped side!

Depletion layer capacitance turns out to be very simple (see fig.)

$$C_j = \frac{\epsilon_s}{W} = \sqrt{\frac{q\epsilon_s N_B}{2(V_{bi} - V)}} \tag{7}$$

Forward bias $\rightarrow$ reduced width, increased capacitance.

Reverse bias $\rightarrow$ increased width, decreased capacitance. (see figs).

Current-Voltage Characteristics

Starting point is the behavior of minority carriers; their concentration increases (decreases) for forward (reverse) bias in exponential manner, e.g. $p_n = p_{n0}\exp(qV/kT)$. In the neutral regions, current is carried by diffusion, and once the concentration dependence on x is found, current is easily calculated, e.g.

$$J_p(x_n) = -qD_p \frac{dp_n}{dx}\bigg|_{x_n} = \frac{qD_p p_{n0}}{L_p}\left(e^{qV/kT} - 1\right) \tag{8}$$

where $L_p = \sqrt{D_p \tau_p}$ is the diffusion length of minority holes.

Many simplifications included here, but qualitative picture is OK. What about reverse bias? See fig for influence of recombination and temperatue.

Qualitative features of temperature dependence:

$$\frac{I_{diff}}{I_{recom}} = 2\frac{n_i L_p \tau_r}{N_D W \tau_p} e^{qV/2kT} \sim \exp\left[-\frac{E_g - qV}{2kT}\right] \tag{9}$$

$$J_s \approx \frac{qD_p p_{n0}}{L_p} \sim n_i^2 \sim \exp\left[-\frac{E_g}{kt}\right] \tag{10}$$

The activation energy obtained from the slope of a plot of J_s vs. 1/T correspond to the energy bandgap E_g. For reverse bias I_{diff}/I_{recom} is proportional to n_i and eventually diff. current dominates.

Charge Storage

Any current flow implies some charge storage, but here we are primarily considering diffusive flow of minority carriers.

$$Q_p = q\int_{x_n}^{\infty}(p_n - p_{n0})dx = qL_p p_{n0}\left(e^{qV/kT} - 1\right) = \frac{L_p^2}{D_p}J_p(x_n) \qquad (11)$$

The stored charge is a source of an additional capacite component: diffusion capacitance

$$C_d = A\frac{dQ_p}{dV} = \frac{Aq^2L_p p_{n0}}{kT}e^{qV}kT \qquad (12)$$

Main contribution from the charge on lower doped side of the junction. What can be done about it?

Why do we care? Depletion and diffucion capacitances contribute to the total capac. in a device and this in turn determines frequency response, e.g. $f_T = g_m/2\pi C_{in}$, and $f_{max} = \sqrt{f_T/8\pi r_b C_c}$. Determination of C_{in} is not simple.

MOS Diode

For energy band diagram see fig. The energy difference between the metal work function $q\phi_m$ and the semiconductor work function $q\phi_s$ is zero and $q\phi_{ms} = 0 = q\phi_m - (q\chi + E_g/2 + q\psi_B)$. No additional charges in oxide or at surface and there is no carrier transport in this direction. The last one results in constant (quasi-) Fermi levels.

What happens with applied voltage on the gate? (see fig) bending upwards for $V < 0$ and downwards for $V > 0$.

Since the bulk material is p-type, $V < 0$ will result in increase of hole concentration above its bulk value $\rightarrow$ accumulation case. For $V < 0$ holes are depleted $\rightarrow$ depletion case. This goes only so far, until the surface inversion layer is formed.

Once we have inversion layer, the bulk is effectively shielded from any changes on the gate and any changes in gate voltage will result in changes of the inversion charge. $Q_s = Q_n + Q_{sc} = Q_n - qN_AW_m$, where W_m is the max. width of the depl. region.

How to define such point? For onset of strong inversion we use $\psi_s = 2\psi_B = 2kT/q \ln(N_A/n_i)$ and we can find the max. depletion layer width $W_m = (\frac{2\epsilon_s \psi s(inv)}{qN_A}$.(see fig.)

Few things to remember: in absence of any workfunction difference, the applied voltage appears partialy across the oxide and partialy across semiconductor: $V = V_o + \psi s$. The potential accross oxide is $V_o = E_o d = |Q_s|d/\epsilon_{ox} = |Q_s|/C_o$

The total capacitance of the MOS structure is

$$C = \frac{C_o C_j}{C_o + C_j} = C_o / \sqrt{1 + \frac{2\epsilon_{ox}^2 V}{q N_A \epsilon s d^2}} \tag{13}$$

where $C_j = \epsilon_s / W$ is the same as for abrup p-n junction.

Threshold voltage can also be found

$$V_T = \frac{q N_A W_m}{C_o} + \psi_s(inv) \approx \frac{\sqrt{2\epsilon_s q N_A 2\psi_B}}{C_o} + 2\psi_B \tag{14}$$

Once in strong inversion, capacitance remains fixed $C_{min} = \epsilon_{ox}/(d + W_m \epsilon_{ox}/\epsilon_s)$. Changes with frequency, however (see fig.)

Some changes in realistic Si structures: there is a finite work function difference, and there are interface and oxide charges that need to be taken into account.