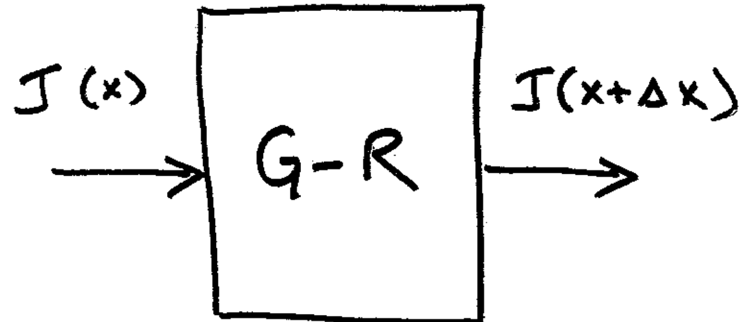
# Review of Semiconductor Device Equations

## Current Continuity

Considering an infinitesimal region of material, the rate of change in the carrier concentration is just the the difference between the flux in and flux out (the gradient in the flux) and the net generation rate (generation minus recombination).




Thus, we have two current continuity equations:

$$\frac{dn}{dt} = \frac{1}{q}\nabla \cdot J_n + G - R \tag{1}$$

$$\frac{dp}{dt} = -\frac{1}{q}\nabla \cdot J_p + G - R \tag{2}$$

## Current Density

In a homogeneous isotropic semiconductor, carrier transport depends on the electric field (drift) and concentration gradient (diffusion).

$$J_n = q\left(\mu_n n \mathcal{E} + D_n \nabla n\right) \tag{3}$$

$$J_p = q\left(\mu_p p \mathcal{E} - D_p \nabla p\right) \tag{4}$$

## Poisson's Equation

The final basic equation is Poisson's equation.

$$\nabla^2 \psi = -\frac{q(p - n + N_d^+ - N_a^-)}{K\epsilon_0}, \tag{5}$$

with $\mathcal{E} = -\nabla\psi$.

Equations (1) through (5) constitute a system of five equations in five unknowns ($n$, $p$, $J_n$, $J_p$, and $\psi$) and, given appropriate boundary conditions, can be used to analyze the carrier concentrations, currents and fields in an arbitrary device structure. However, we also need to specify the other parameters (mobility, diffusivity, net recombination rate, etc.) and how they depend on the material properties, carrier densities and local potential and field.

## Recombination/Generation

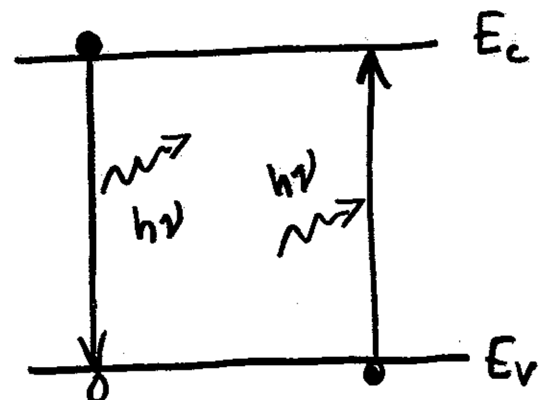
Electron/hole pairs are continually being generated and recombining.

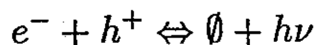
There are three basic recombination/generation mechanisms:

1. Band-to-band or direct generation/recombination.
2. Indirect, through intermediate trapping level.
3. Auger recombination – excess energy transferred to third carrier.

### Direct (Band to Band) Recombination

Conduction band electrons and valence band holes can recombine directly and electron/hole pairs can be generated via a valence band electron becoming excited into the conduction band.



$$e^- + h^+ \Leftrightarrow \emptyset + h\nu$$

Energy is conserved by generation or absorption of photons (radiative) or phonons (non-radiative).

Because momentum as well as energy must be conserved, direct recombination is generally only important in direct band-gap semiconductors (e.g., GaAs), since photons have large energy but very little momentum, while phonons have large momentum, but small energies.

Recombination is proportional to the number of electrons in the conduction band times that of empty sites in the valence band (holes).

$$R = Kpn \quad (K = \text{constant}) \tag{6}$$

In thermal equilibrium,

$$G_{th} = R_{th} = Kp_0n_0 \quad (n_0, p_0 \text{ are equilibrium values}) \tag{7}$$

Suppose we generate excess carriers at a rate $G_L$ (due to incident light for example), then

$$R = Kpn = K(p_0 + \Delta p)(n_0 + \Delta n) \tag{8}$$

where $\Delta p \equiv p - p_0$ and $\Delta n \equiv n - n_0$ are the excess carrier concentrations.

The net recombination rate for this process is

$$U = R - G_{th} = K(pn - n_i^2),$$

which in steady-state is just equal to $G_L$.

The effective lifetimes are defined as

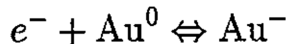
$$\tau_n \equiv U/(n - n_0), \qquad \tau_p \equiv U/(p - p_0) \tag{9}$$

For minority carriers under conditions of low-level injection (majority carrier concentration near equilibrium, e.g., $n \cong n_0$),

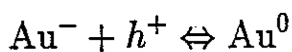
$$\tau_n = 1/Kp_0, \qquad \tau_p = 1/Kn_0.$$

## Indirect Recombination (through trapping levels)

Conduction band electrons and valence band holes can also recombine by sequentially being trapped in an intermediate deep level,

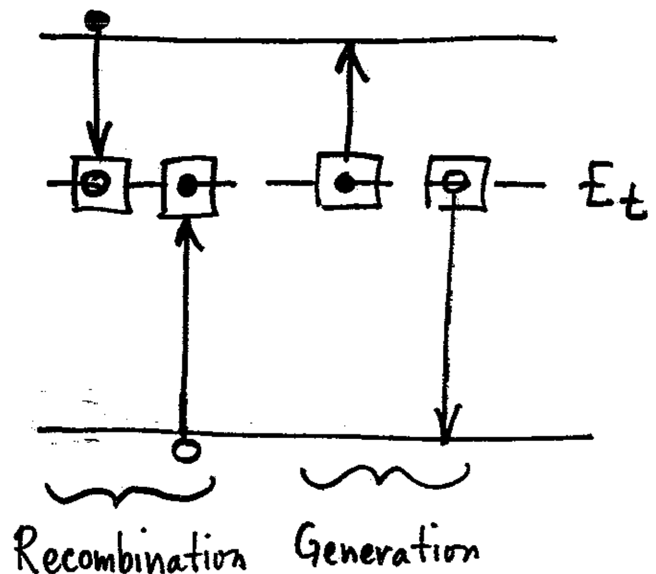


$$e^- + \mathrm{Au}^0 \Leftrightarrow \mathrm{Au}^-$$

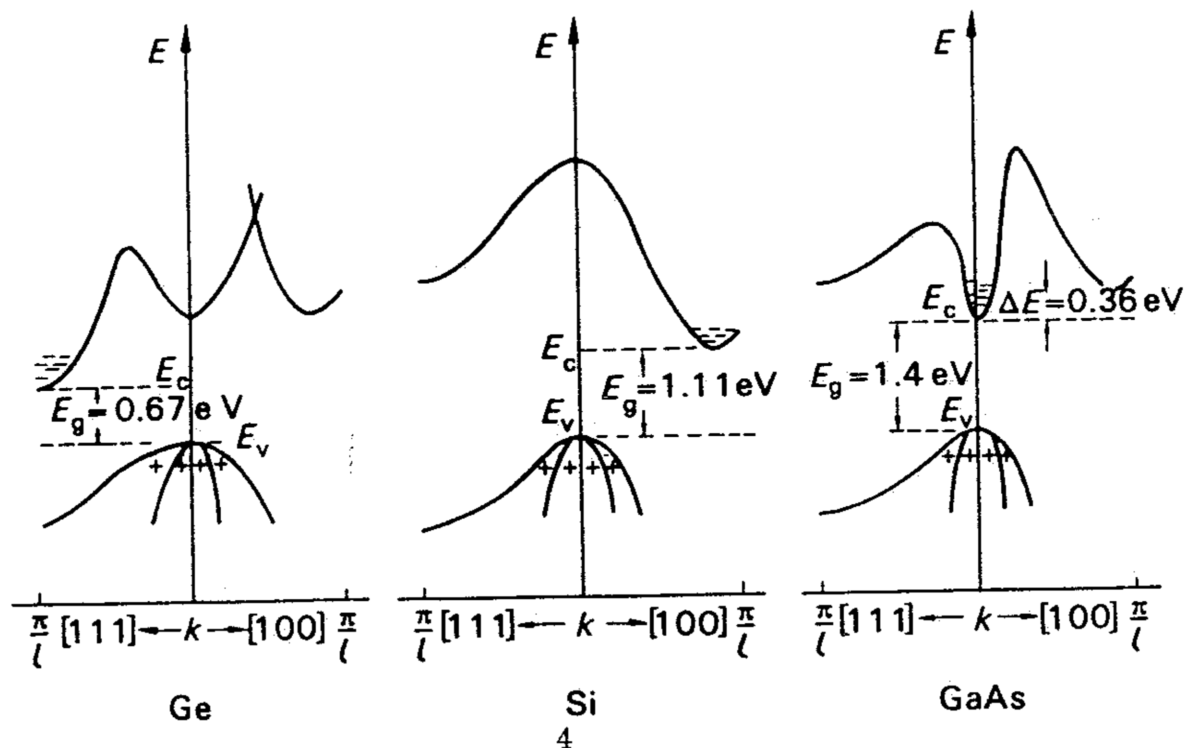


$$\mathrm{Au}^- + h^+ \Leftrightarrow \mathrm{Au}^0$$

with the reverse process (sequential emission) resulting in generation.




In silicon and germanium, this is usually the dominant mechanism since they are "indirect band semiconductors" which means that the minima of the conduction band does not occur at the same point in momentum space ($k$-space) as the maxima of the valence band.



Defect-assisted recombination is also known as Hall-Schockley-Read (HSR or SRH) recombination and the net rate is:

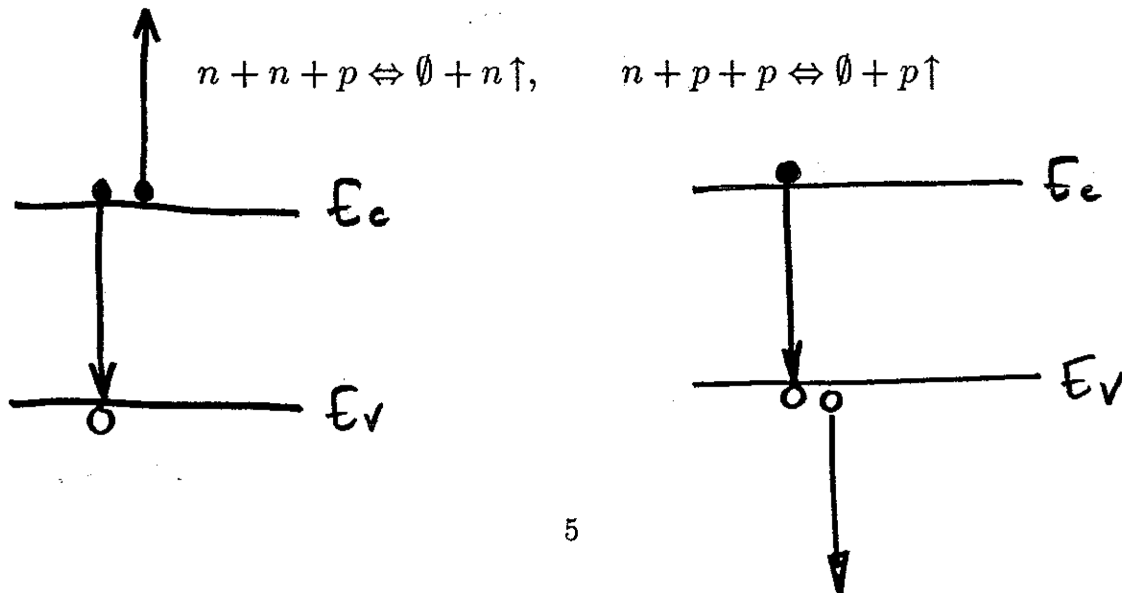
$$U = \frac{pn - {n_i}^2}{\tau_p \left[ n + n_i \exp \frac{E_t - E_i}{kT} \right] + \tau_n \left[ p + n_i \exp \frac{E_i - E_t}{kT} \right]}. \tag{10}$$

where $\tau_n = (\sigma_n N_t v_{thn})^{-1}$ and $\tau_p = (\sigma_p N_t v_{thp})^{-1}$, which are also the minority carrier lifetimes in low-level injection.

- Energy conservation is achieved by emission or absorption of phonons or photons.
- The numerator is proportional to the restoring force. The net recombination rate is always of the proper sign to move the system towards equilibrium ($pn = {n_i}^2$).
- The net recombination rate $U$ is maximum when the denominator is minimized. That occurs when $E_t = E_i$ for $\tau_n = \tau_p$. In other words, the trapping levels near midgap are the most efficient recombination centers.

## Auger Recombination

It is also possible for the excess energy to be given in the form of kinetic energy to another carrier (the inverse of impact ionization)



The net recombination rate is:

$$U = (K_p p + K_n n)(pn - ni^2) \tag{11}$$

The Auger process is most important when the carrier concentration is very high since the effective minority carrier lifetime is proportional to the square of the majority carrier concentration. For low-level injection,

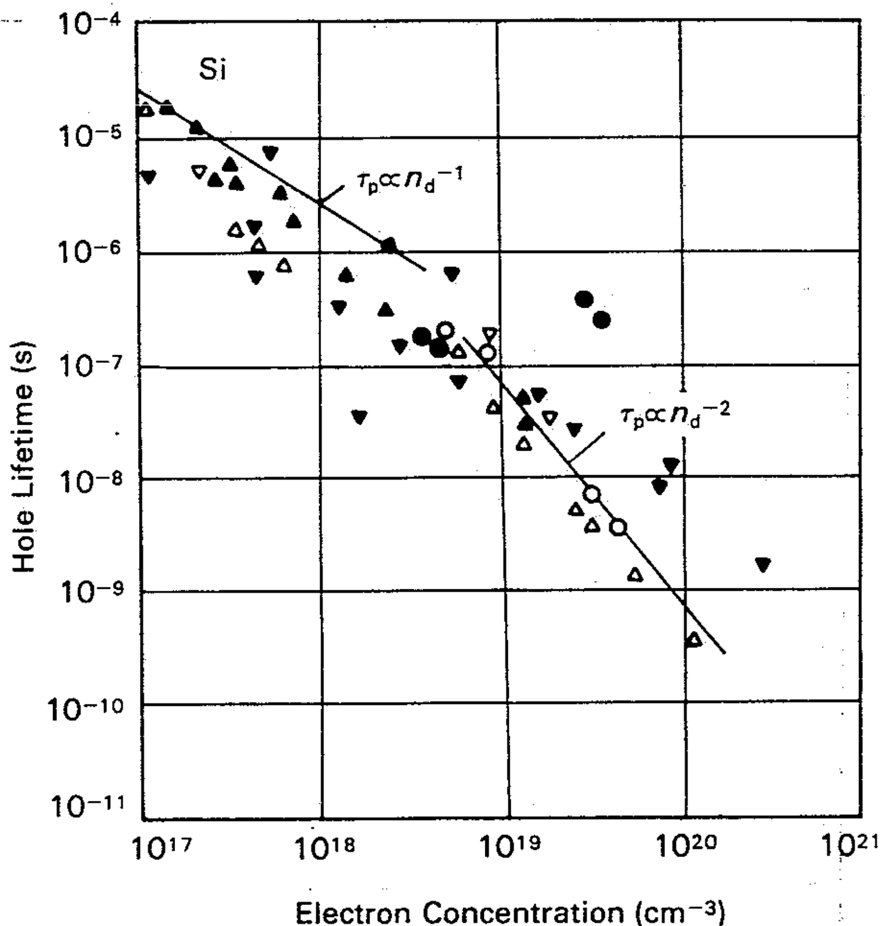
$$1/\tau_n = K_p p^2, \quad 1/\tau_p = K_n n^2$$

## Lifetime versus Doping

Considering all of the recombination processes acting in parallel, the overall recombination rate increases with doping level, while the recombination lifetime decreases correspondingly.

For low doping the lifetime is first constant and then decreases linearly with doping as the doping results in additional trapping centers.

For high doping, Auger recombination dominates and the lifetime drops as the doping squared.



## Surface Recombination

Recombination at surfaces is often enhanced relative to the bulk due to the presence of surface states which act as trapping centers.

By analogy with bulk HSR recombination, we can model the net recombination surface rate in exactly the same manner with $\tau_n$ and $\tau_p$ replaced by $1/s_n$ and $1/s_p$, where we define

$$s_n \equiv v_{thn}\sigma_n N_{ts}, \qquad s_p \equiv v_{thp}\sigma_p N_{ts} \tag{12}$$

to be the surface recombination velocities (cm/sec).

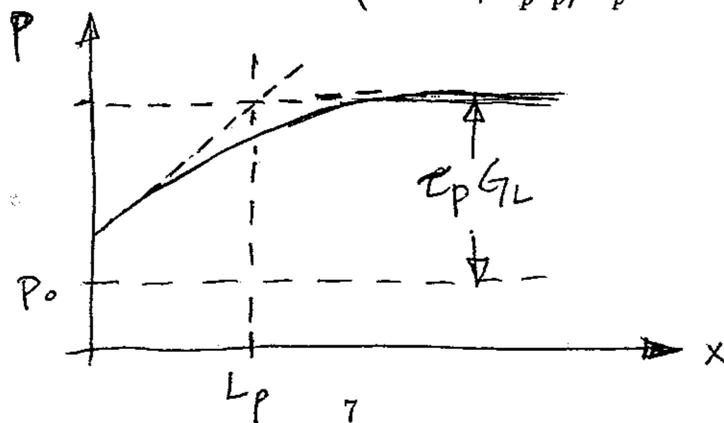
Thus in low-level injection in $n$-type material, the net recombination becomes,

$$U = s_p\left[p(0) - p_0\right] \tag{13}$$

The surface recombination rate acts as the boundary conditions for the continuity equation at semiconductor/insulator interfaces. Surface recombination is present at semiconductor/metal interfaces as well and must be considered in conjunction with the contact currents.

As an example, uniform generation ($G_L$) in the bulk of a uniformly-doped $n$-type wafer resulting in low-level injection, plus surface recombination leads to:

$$\Delta n(x) = \Delta p(x) = \tau_p G_L\left(1 - \frac{s_p\tau_p/L_p}{1 + s_p\tau_p/L_p}e^{-x/L_p}\right)$$



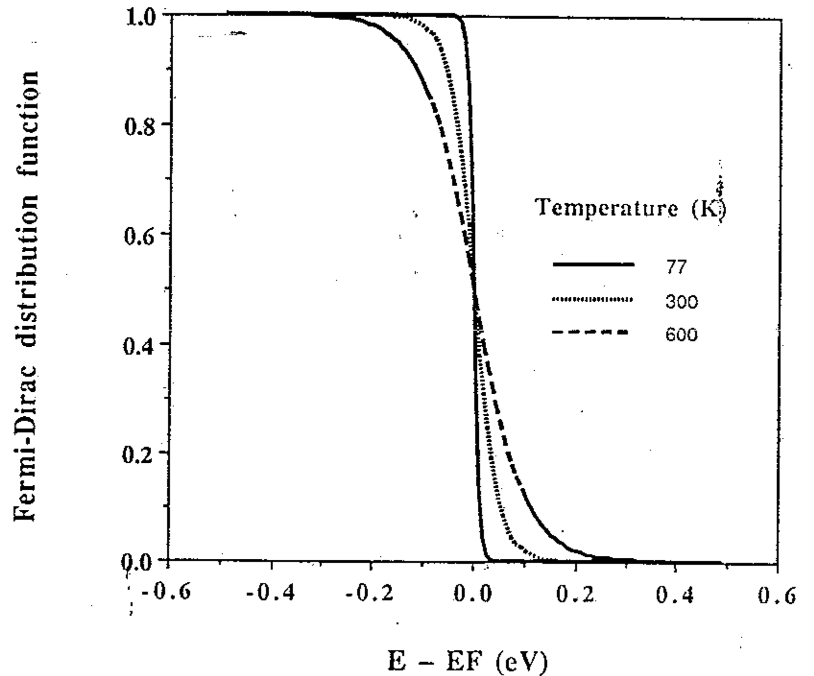
## Fermi Level

The Fermi level is the equilibrium electrochemical potential. It represents the expected value of the change in energy for one electron added to (or taken from) the system.

For any state with energy $E$, the probability of occupation for that state is given by the Fermi-Dirac distribution:

$$f_{FD}(E) = \frac{1}{1 + \exp\left(\frac{E - E_f}{kT}\right)} \tag{14}$$

The Fermi level represents the boundary between the mostly filled and mostly empty states in equilibrium, with the temperature determining how sharp that boundary is. Far above or below the Fermi level, the Fermi-Dirac distrution can be approximated by the Boltzmann distribution.




$$f_{FD}(E) \cong \begin{cases} \exp(E - E_f/kT) & \text{for } E > E_f + 3kT \\ 1 - \exp(E_f - E/kT) & \text{for } E < E_f - 3kT \end{cases} \tag{15}$$

To calculate the relation between Fermi level and carrier concentration, it is necessary to specify the density of states in the cionduction (or valence) band. For a single parabolic conduction band with

$$E = E_c + \frac{\hbar^2 k_x^2}{2m_x^*} + \frac{\hbar^2 k_y^2}{2m_y^*} + \frac{\hbar^2 k_z^2}{2m_z^*}, \tag{16}$$

the density of states is

$$\mathcal{N}(E) = \frac{\sqrt{m_x^* m_y^* m_z^*}}{2\pi^2} \left(\frac{2}{\hbar}\right)^{3/2} \sqrt{E - E_c} \tag{17}$$

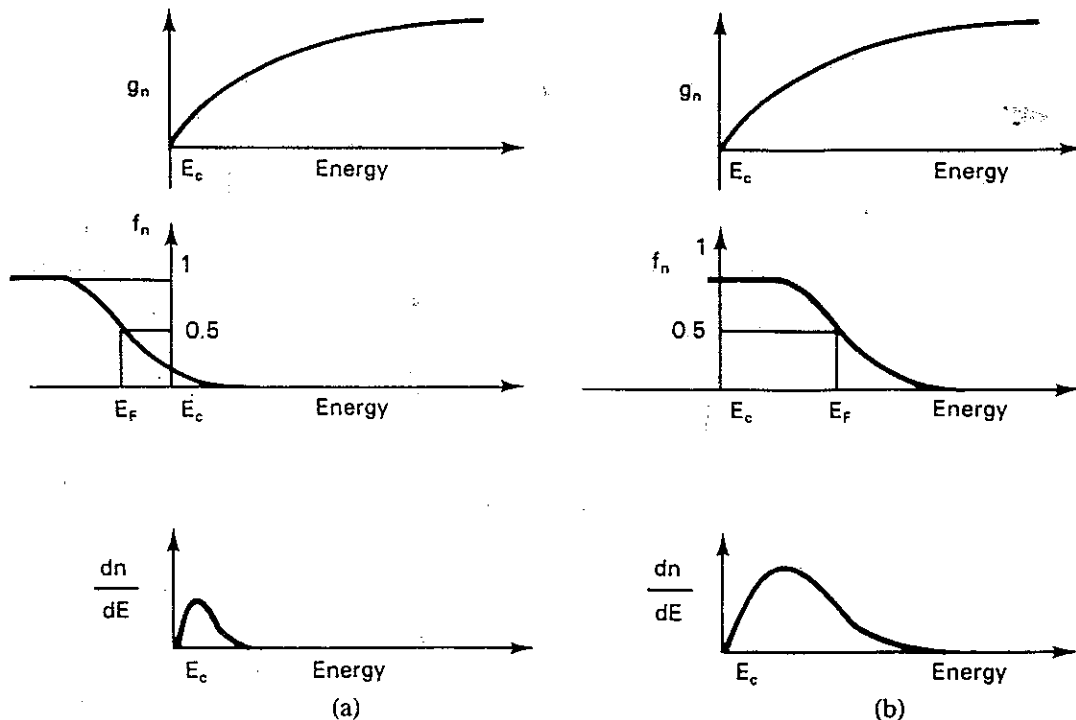
For a semiconductor with multiple band extrema (for example, silicon with 6 equivalent conduction band minima and 2 nonequivalent valence band maxima), the density of states for all the bands must be added together. The result is

$$\mathcal{N}(E) = \frac{1}{2\pi^2} \left(\frac{2m_{dn}}{\hbar}\right)^{3/2} \sqrt{E - E_c}, \tag{18}$$

where $m_{dn}$ is the density-of-states effective mass for the conduction band.

The number of electrons in the conduction band is given by integrating the Fermi-Dirac distribution times the density of states over energy.

$$n = \int_0^\infty f_{FD}(E) \mathcal{N}(E)\, dE \tag{19}$$



The result is

$$n = N_c F_{1/2}\left(\frac{E_f - E_c}{kT}\right) \tag{20}$$

where

$$N_c = 2\left(\frac{m_{dn}kT}{2\pi\hbar^2}\right) \tag{21}$$

and

$$F_n(\eta) = \frac{2}{\sqrt{\pi}}\int_0^\infty \frac{x^n\,dn}{1+\exp(x-\eta)} \tag{22}$$

Similarly for the valence band,

$$p = N_v F_{1/2}\left(\frac{E_v - E_f}{kT}\right) \tag{23}$$

Since $F_{1/2}(\eta) \cong \exp(\eta)$ for $\eta < -3$, the relations reduce to the more familiar forms for nondegenerate material ($E_v + 3kT < E_f < E_c - 3kT$):

$$n = N_c \exp(E_f - E_c/kT), \qquad p = N_v \exp(E_v - E_f/kT) \tag{24}$$

and

$$np = N_c N_v \exp(-\underbrace{(E_c - E_v)}_{E_g}/kT) = n_i^2 \tag{25}$$

For degenerate material, these equations no longer hold and, for the opposite extreme, when the Fermi level is well within the conduction or valence band, $F_{1/2}(\eta) \cong 4\eta^{3/2}/3\sqrt{\pi}$ for $\eta > 3$. However, under these conditions, other effects such as bandgap narrowing also come into play. We will consider these effects in more detail at a later point.

## Equilibrium Electrochemical Potential (Fermi Level)

Considering a nondegenerate material for simplicity, the Fermi level can be written in terms of the electron concentration:

$$E_f = E_c + kT \ln\left(\frac{n}{N_c}\right) = E_i + kT \ln\left(\frac{n}{n_i}\right) \tag{26}$$

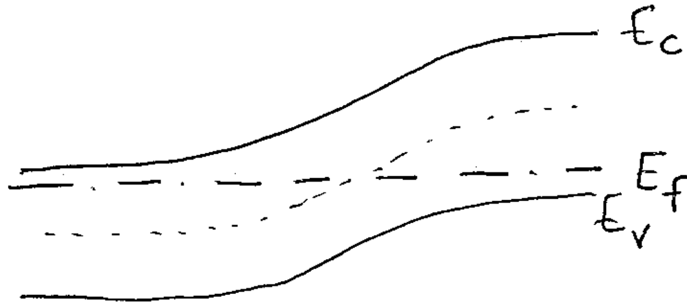
The electric field is defined as the negative gradient of the electrical potential $\psi$, but it can also be expressed in terms of the gradient in the conduction or valence band (no change in bandgap).

$$\mathcal{E} \equiv -\frac{d\psi}{dx} = \frac{1}{q}\frac{dE_c}{dx} = \frac{1}{q}\frac{dE_v}{dx} = \frac{1}{q}\frac{dE_i}{dx} \tag{27}$$

If we apply these equations plus the the Einstein relations ($\mu_n/q = D_n/kT$) to the electron current equation, it is possible to show (try it) that:

$$J_n = n\mu_n \frac{dE_f}{dx} \tag{28}$$

Thus in equilibrium, where $J_n = 0$, $E_f =$ **constant**. Another way to think about this fact is that if there was a change in electron potential with position, electrons could move around to reduce the energy of the system.



## Quasi-Fermi Levels

Under thermal equilibrium, we useful equations relating the Fermi level to the electron and hole concentrations. For example,

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

However, if we have injection ($np > n_i^2$) or extraction ($np < n_i^2$) then we are not in equilibrium and therefore cannot use these relationships because $E_f$ is no longer meaningful.

To replace $E_f$, we define two new quantities called quasi-Fermi levels:

$$n = n_i \exp\left(\frac{E_{fn} - E_i}{kT}\right) = N_c \exp\left(-\frac{E_c - E_{fn}}{kT}\right) \tag{29}$$

where $E_{fn}$ is the quasi-Fermi level for electrons, while for holes,

$$p = n_i \exp\left(\frac{E_i - E_{fp}}{kT}\right) = N_v \exp\left(-\frac{E_{fp} - E_v}{kT}\right) \tag{30}$$

The quasi-Fermi levels are mathematical tools and their values are chosen so that we can extend our familiar equilibrium equations to nonequilibrium situations (even for degenerate statistics).

Out of equilibrium, $E_{fn} \neq E_{fp}$, and $|E_{fn} - E_{fp}|$ is a measure of how far removed the semiconductor is from equilibrium.

$$pn = n_i^2 \exp\left(\frac{E_{fn} - E_{fp}}{kT}\right)$$

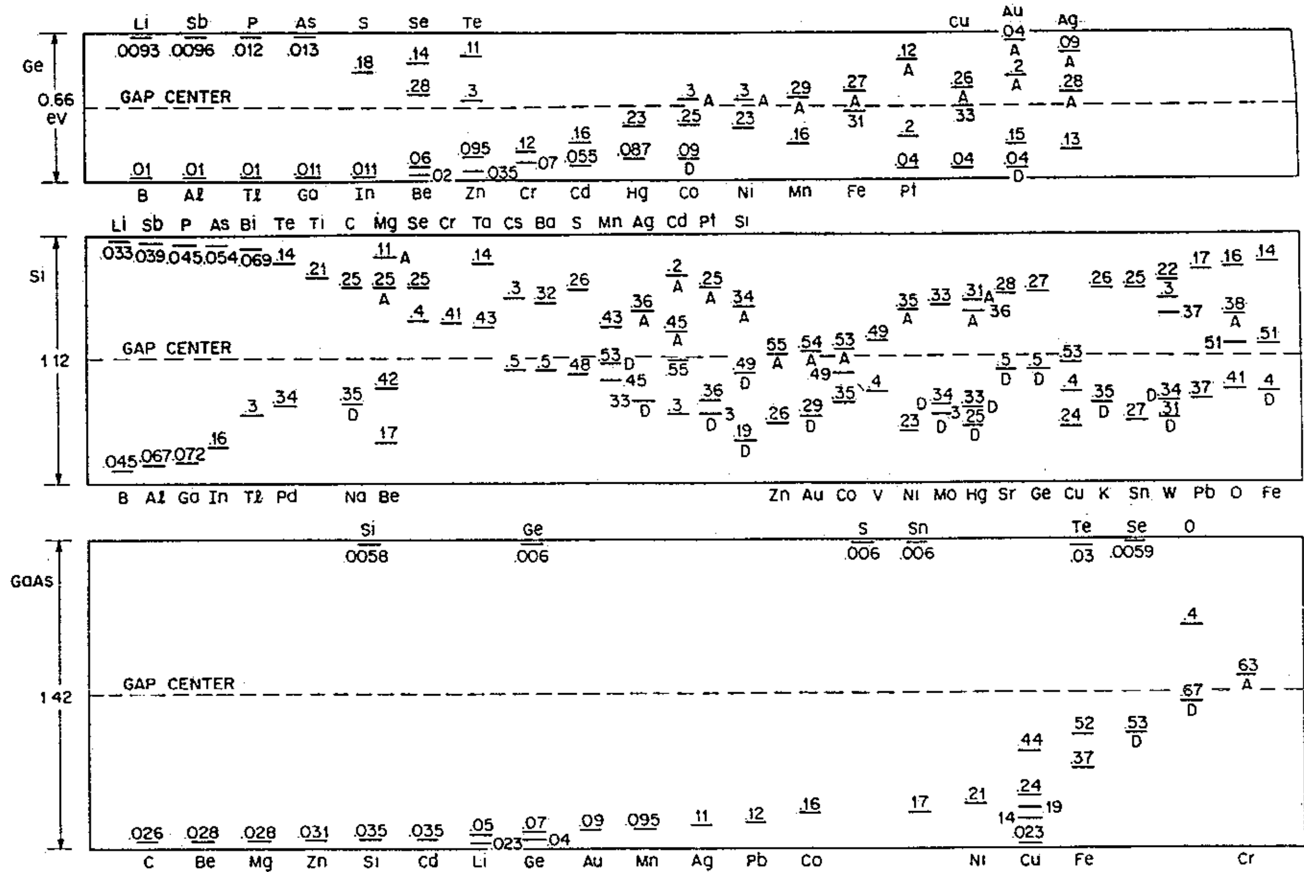
In addition, the gradients in the quasi-Fermi levels are the driving forces for carrier fluxes. Thus, as for $E_f$ in equilibrium,

$$J_n = n\mu_n \frac{dE_{fn}}{dx} \tag{31}$$

$$J_p = -p\mu_p \frac{dE_{fp}}{dx} \tag{32}$$

## Donors and Acceptors

Impurities and other defects result in defect levels within the band-gap. The diagram below illustrates the location of these levels within the band gap for Ge, Si and GaAs at 300°K.




For any of these defect states, the ratio of filled to unfilled states depends on the Fermi level with

$$\frac{N_d^0}{N_d^+} = g_d \exp\left(\frac{E_f - E_d}{kT}\right) \tag{33}$$

$$\frac{N_a^-}{N_a^0} = \frac{1}{g_a} \exp\left(\frac{E_f - E_a}{kT}\right) \tag{34}$$

The factors $g_d$ and $g_a$ are degeneracy factors. Accounting for the two possible spins of the unpaired electrons in the neutral donor or acceptors, as well as the two degenerate valence bands, $g_d = 2$ and $g_a = 4$ for substitutional group III and V impurities in silicon.

The sum of defects in each of the possible charge states is just the total defect concentration.

$$N_d^{\text{total}} = \sum_i N_d^i \tag{35}$$

Thus, for a donor with just two possible states (+ and 0),

$$N_d^+ = \frac{1}{1 + g_d \exp(E_f - E_d/kT)} \tag{36}$$

while for an acceptor,

$$N_a^- = \frac{1}{1 + g_a \exp(E_a - E_f/kT)} \tag{37}$$

For shallow dopants, at low concentration it is reasonable to assume that all of the dopants are ionized. However, for doping concentrations that approach $N_c$ or $N_a$. A significant fraction of the dopants will remain unionized. In a uniformly-doped material, the neutrality condition $p(E_f) - n(E_f) = N_d^+(E_f) + N_a(E_f)$ can be solved to determine the Fermi level position and thus occupation of states and bands.

Also, at low temperatures, the carriers have insufficient thermal energy to become excited into the conduction (or valence) band, reducing the density of ionized impurities. This phenomena is known as freeze-out.

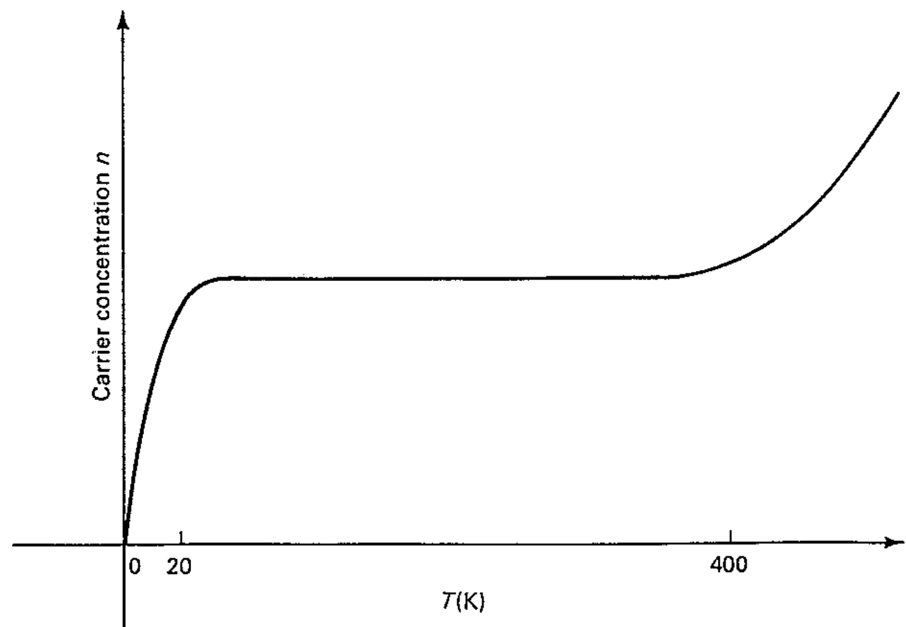



**Figure 6.7** *n* vs. *T* for a doped semiconductor.

## Mobility

Carrier mobilities depend on the scattering lifetimes and effective masses.

$$\mu_n = \frac{q\tau_n}{m_n^*}, \qquad \mu_p = \frac{q\tau_p}{m_p^*} \tag{38}$$

Scattering processes act in parallel so that:

$$\frac{dt}{\tau} = \sum_i \frac{dt}{\tau_i}, \qquad \frac{1}{\mu} = \sum_i \frac{1}{\mu_i} \tag{39}$$

Transport is primarily limited by the process that produces the shortest scattering time or lowest mobility.

Lattice (phonon) scattering dominates at low doping, while ionized impurity scattering dominates for high doping reducing the mobility as $N_d$ and $N_a$ increase.

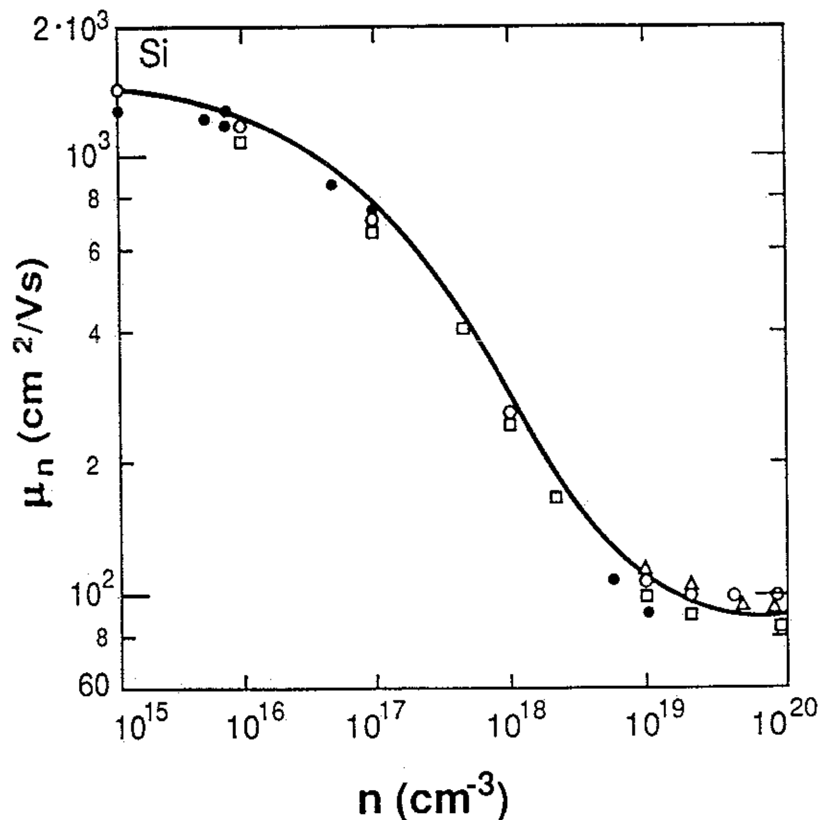



**Fig. 1-9-4.** Measured and calculated electron mobility in silicon versus doping at room temperature (Reprinted with permission from T. I. Tosic, D. A. Tjapkin, and M. M. Jevtic, *Solid State Electron.*, 24, p. 577 (1981) Pergamon Press p/c). The solid line shows calculations and the symbols represent measured data.

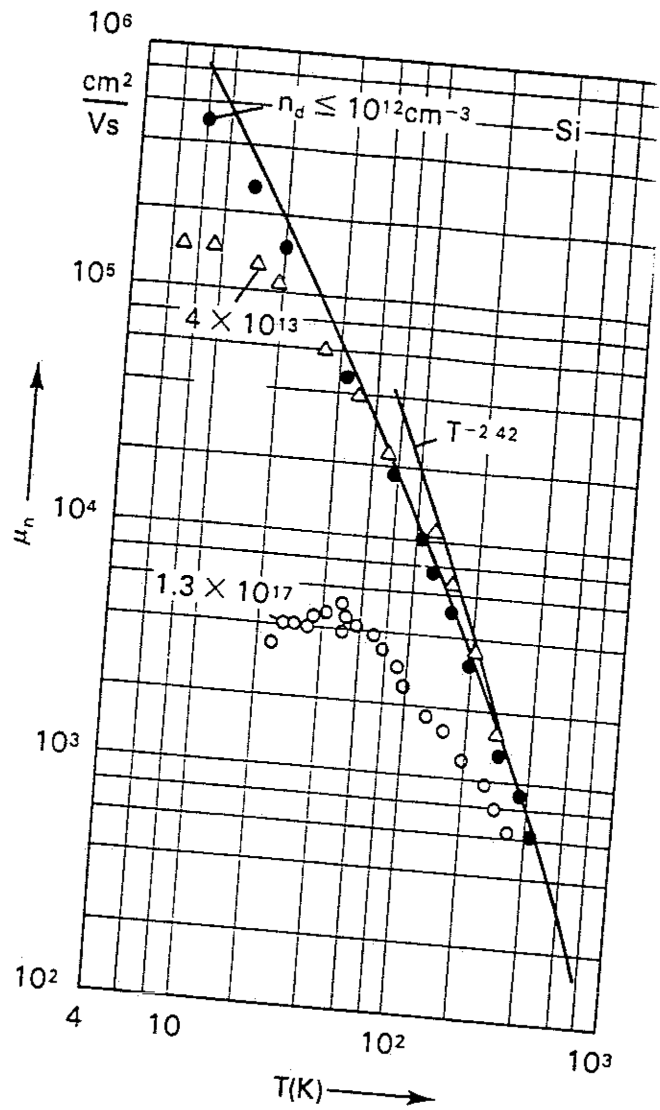
As temperature rises, lattice vibrations increase, increasing the amount of lattice scattering and reducing the mobility,

$$\mu_{\text{lattice}} \propto T^{-n}, \quad n \sim 2.5$$

In contrast, impurity scattering is reduced at high temperature since carriers move faster and remain near impurities for shorter times.

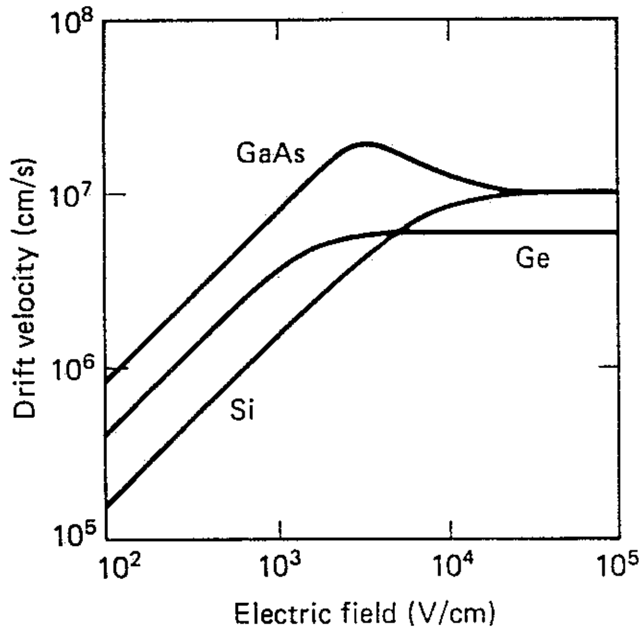
$$\mu_{\text{impurity}} \propto T^{n}, \quad n \sim 1.5$$

Therefore, at higher temperatures lattice scattering tends to dominate over impurity scattering.



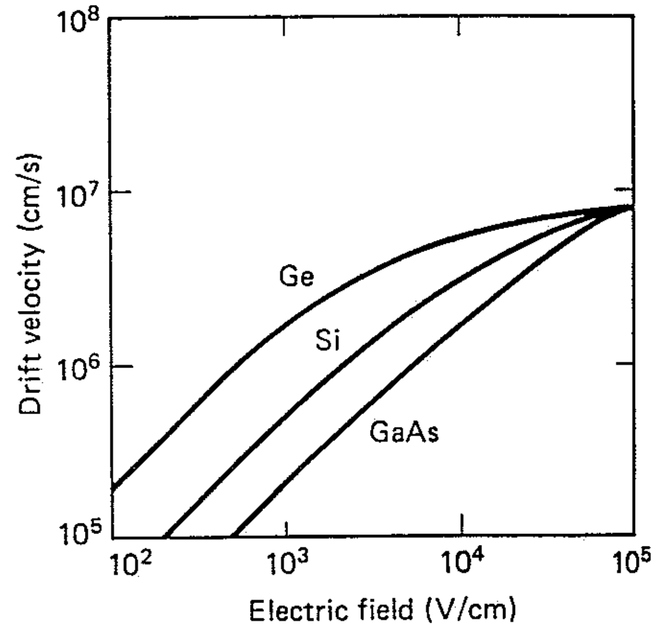

For high electric fields, carriers gain energies greater than their thermal energies. This reduces the rate of ionized impurity scattering as the carriers spend less time in the neighborhood of the scattering centers. However, scattering due to phonons is increased as the probability of the carrier having enough energy to emit a phonon increases.

Since scattering increases, mobility drops for high fields and the carrier velocity saturates rather than continuing its linear increase with field. We will address this issue in more depth later when we talk about hot electron effects.




(a)

electrons




(b)

holes

The diffusivity behaves similarly to mobility with

$$D_n = \frac{kT}{q}\mu_n. \tag{40}$$

at low and moderate doping levels. This is known as the Einstein Relation. However, at high doping levels (Fermi level approaches conduction band minima), a more general equation is required:

$$D_n = 2\left(\frac{kT}{q}\right)\mu_n\left[\frac{F_{1/2}\left(\frac{E_f - E_c}{kT}\right)}{F_{-1/2}\left(\frac{E_f - E_c}{kT}\right)}\right], \tag{41}$$