VIII.2. A Semiconductor Device Primer

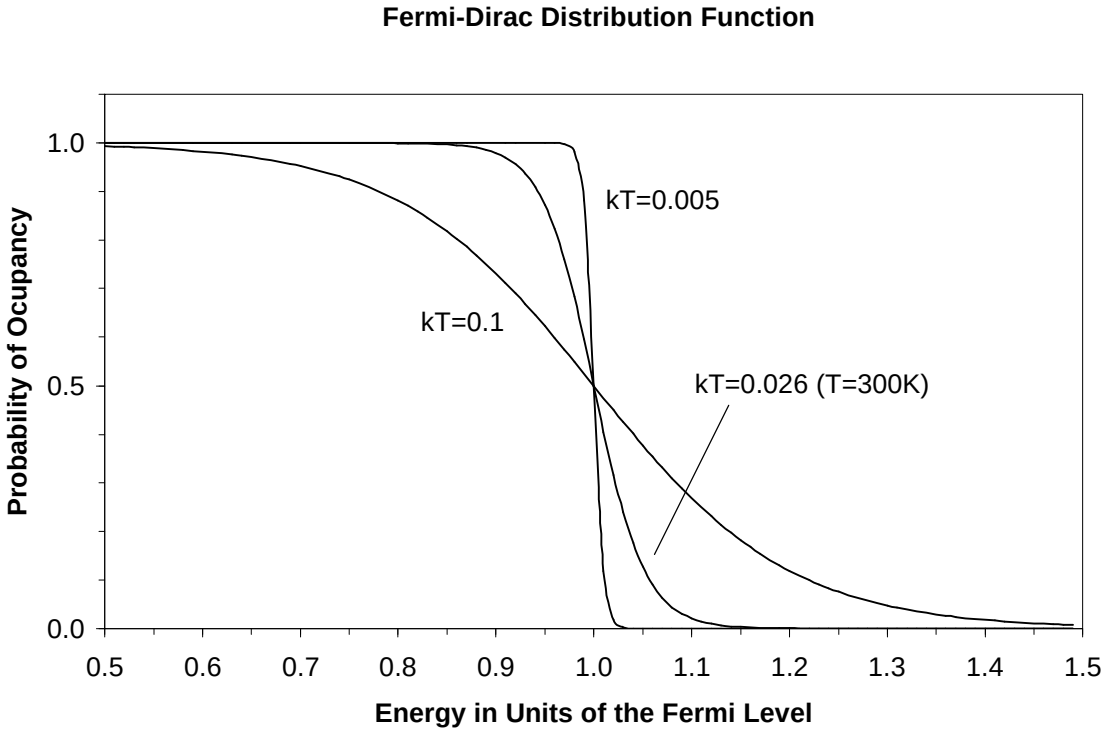
Bibliography:

1. Grove, A.S., *Physics and Technology of Semiconductor Devices* (John Wiley & Sons, New York, 1967)
2. Sze, S.M., *Physics of Semiconductor Devices* (John Wiley & Sons, New York, 1981) TK 7871.85.S988, ISBN 0-471-05661-8
3. Kittel, C., *Introduction to Solid State Physics* (John Wiley & Sons, New York, 1996) QC176.K5, ISBN 0-471-11181-3

1. Carrier Concentrations

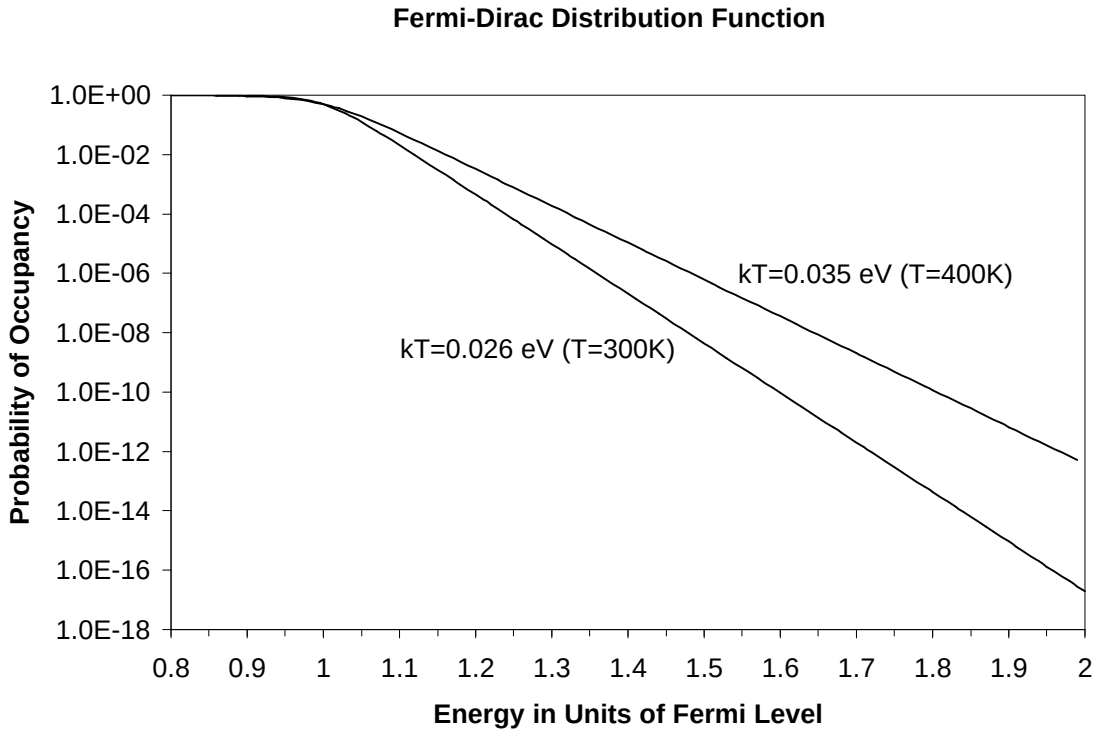
The probability that an electron state in the conduction band is filled is given by the Fermi-Dirac distribution

$$f_e(E) = \frac{1}{e^{(E-E_F)/k_B T} + 1}$$



The density of atoms in a Si or Ge crystal is about $4 \cdot 10^{22}$ atoms/cm^3.

Since the minimum carrier density of interest in practical devices is of order 10^{10} to 10^{11} cm^{-3}, very small ocupancy probabilities are quite important.



In silicon the band gap is 1.12 eV. If the Fermi level is at midgap, the band-edges will be 0.56 eV above and below E_F.

As is apparent from the plot, relatively large deviations from the Fermi level, i.e. extremely small occupancies, will still yield significant carrier densities

The number of occupied electron states N_e is determined by summing over all available states multiplied by the occupation probability for each individual state

$$N_e = \sum_i m_i f(E_i)$$

Since the density of states near the band edge tends to be quite high, this can be written as an integral

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

where $g(E)$ is the density of states.

Solution of this integral requires knowledge of the density of states.

Fortuitously, to a good approximation the density of states near the band edge has a parabolic distribution

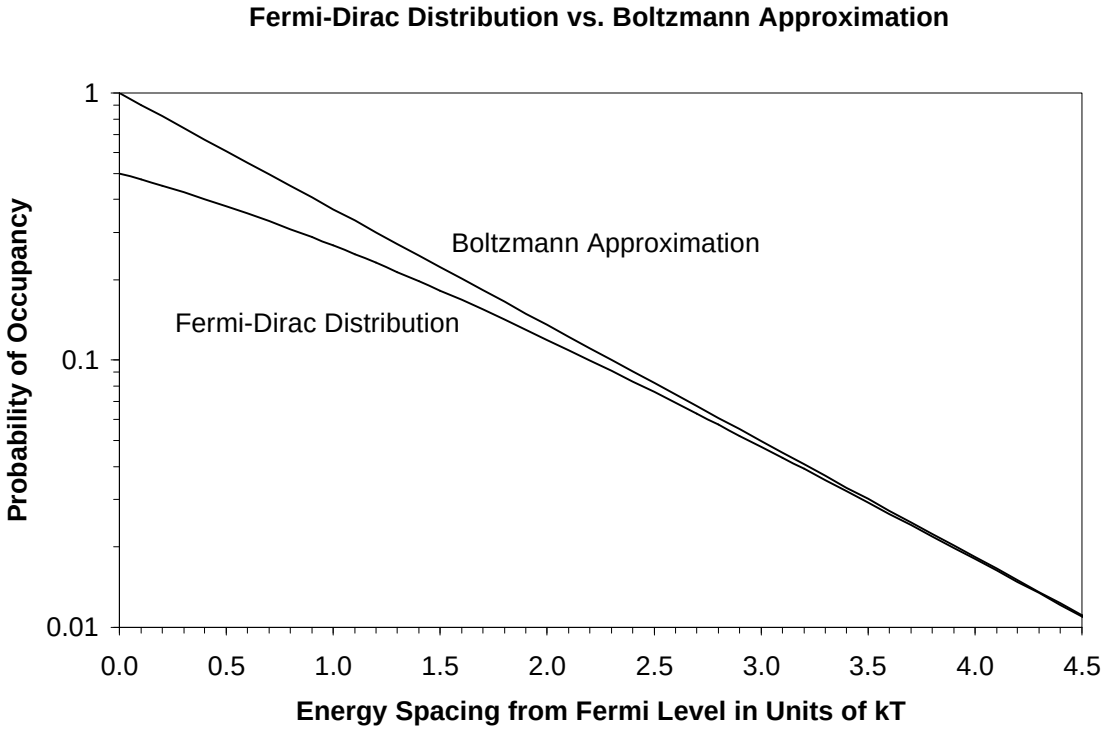
$$g(E)dE \propto (E - E_c)^{1/2}$$

As the energy increases beyond the band edge, the distribution will deviate from the simple parabolic form, but since the probability function decreases very rapidly, the integral will hardly be affected.

The second obstacle to a simple analytical solution of the integral is the intractability of integrating over the Fermi distribution.

Fortunately, if E-E_F is at least several times k_BT, the Fermi distribution can be approximated by a Boltzmann distribution

$$1+e^{(E-E_F)/k_BT} \approx e^{(E-E_F)/k_BT} \quad \Rightarrow \quad f(E) \approx e^{-(E-E_F)/k_BT}$$



At energies beyond 2.3 k_BT of the Fermi level the difference between the Boltzmann approximation and the Fermi Distribution is <10%, for energies >4.5 k_BT it is less than 1%.

Applying the approximation to the occupancy of hole states, the probability of a hole state being occupied, i.e. a valence state being empty is

$$f_h(E) = 1 - f_e(E) = \frac{1}{e^{(E_F-E)/k_BT}+1} \approx e^{-(E_F-E)/k_BT}$$

Since the band gap is of order 1 eV and k_BT at room temperature is 0.026 eV, the conditions for the Boltzmann approximation are fulfilled for excitation across the band gap.

With these simplifications the number of electrons in the conduction band in thermal equilibrium is

$$n_e \propto (k_B T)^{3/2} e^{-(E_c - E_F)/k_B T}$$

or

$$n_e = N_c e^{-(E_c - E_F)/k_B T}$$

where N_c is the effective density of states at the band edge.

Correspondingly, the hole concentration

$$p = N_v e^{-(E_F - E_v)/k_B T}$$

In a pure semiconductor

$$n = p = n_i$$

where n_i is the number of electrons or holes intrinsic to a pure semiconductor, i.e. where only source of mobile carriers is thermal excitation across the band gap without any additional impurity atoms or crystal imperfections that would allow other excitation mechanisms.

Silicon (E_g = 1.12 eV):	n_i= 1.45·10^{10} cm^{-3}	at 300K
Germanium (E_g = 0.66 eV):	n_i= 2.4·10^{13} cm^{-3}	at 300K

For comparison:

> The purest semiconductor material that has been fabricated is Ge with active impurity levels of about 3·10^{10} cm^{-3}.

Using the above results

$$n_i = N_c\, e^{-(E_c - E_F)/k_B T} = N_v\, e^{-(E_F - E_v)/k_B T}$$

which one can solve to obtain E_F

$$E_F = \frac{E_c + E_v}{2} - \frac{k_B T}{2} \log\left(N_c / N_v\right)$$

If the band structure is symmetrical ($N_c = N_v$), the intrinsic Fermi level lies in the middle of the band gap.

Even rather substantial deviations from a symmetrical band structure will not affect this result significantly, as N_c / N_v enters logarithmically and $k_B T$ is much smaller than the band gap.

A remarkable result is that the product of the electron and hole concentrations

$$np = n_i^2 = N_c N_v e^{-(E_c - E_v)/k_B T} = N_c N_v e^{-E_g / k_B T}$$

depends only on the band gap E_g and not on the Fermi level.

This result, the law of mass action, is very useful in semiconductor device analysis. It requires only that the Boltzmann approximation holds.

Qualitatively, it says that if one carrier type exceeds this equilibrium concentration, recombination will decrease the concentrations of both electrons and holes to maintain $np = n_i^2$.

2. Carrier Concentrations in Doped Crystals

The equality

$$n_e = n_h$$

only holds for pure crystals, where all of the electrons in the conduction band have been thermally excited from the valence band.

In practical semiconductors the presence of impurities tips the balance towards either the electrons or holes.

Impurities are an unavoidable byproduct of the crystal growth process, although special techniques can achieve astounding results. For example, in the purest semiconductor crystals – "ultrapure" Ge – the net impurity concentration is about $3 \cdot 10^{10}$ cm^{-3}.

In semiconductor device technology impurities are introduced intentionally to control the conductivity of the semiconductor.

Let N_d^+ be the concentration of ionized donors and N_a^- the concentration of ionized acceptors.

Overall charge neutrality is preserved, as each ionized dopant introduces a charged carrier and an oppositely charged atom, but the net carrier concentration is now

$$\Delta n = n - p = N_d^+ - N_a^-$$

or

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

Assume that the activation energy of the donors and acceptors is sufficiently small so that they are fully ionized

$$N_D^+ \approx N_D \quad \text{and} \quad N_A^- \approx N_A$$

Then

$$p + N_D = n + N_A \,,$$

which, using $np = n_i^2$, becomes

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

If the acceptor concentration $N_A >> N_D$ and $N_A >> n_i$

$$\frac{p}{N_A} + \frac{N_D}{N_A} = \frac{n_i}{p}\frac{n_i}{N_A} + 1 \quad \Rightarrow \quad p \approx N_A\,,\; n \approx \frac{n_i^2}{N_A} << N_A$$

i.e. the conductivity is dominated by holes.

Conversely, if the donor concentration $N_D >> N_A$ and $N_D >> n_i$ the conductivity is dominated by electrons.

If the conductivity is dominated by only one type of carrier, the Fermi level is easy to determine. If, for example, $n >> p$

$$p + N_D = n + N_A$$

can be written

$$n = N_D - N_A$$

$$N_c e^{-(E_c - E_F)/k_B T} = N_D - N_A$$

yielding

$$\frac{E_c - E_F}{k_B T} = \log\left(\frac{N_c}{N_D - N_A}\right)$$

If $N_D >> N_A$, then E_c-E_F must be small, i.e. the Fermi level lies close to the conduction band edge.

In reality the impurity levels of common dopants are not close enough to the band edge for the Boltzmann approximation to hold, so the calculation must use the Fermi distribution and solve numerically for E_F. Nevertheless, the qualitative conclusions derived here still apply.

It is often convenient to refer all of these quantities to the intrinsic level E_i , as it accounts for both E_c *and* E_v. Then

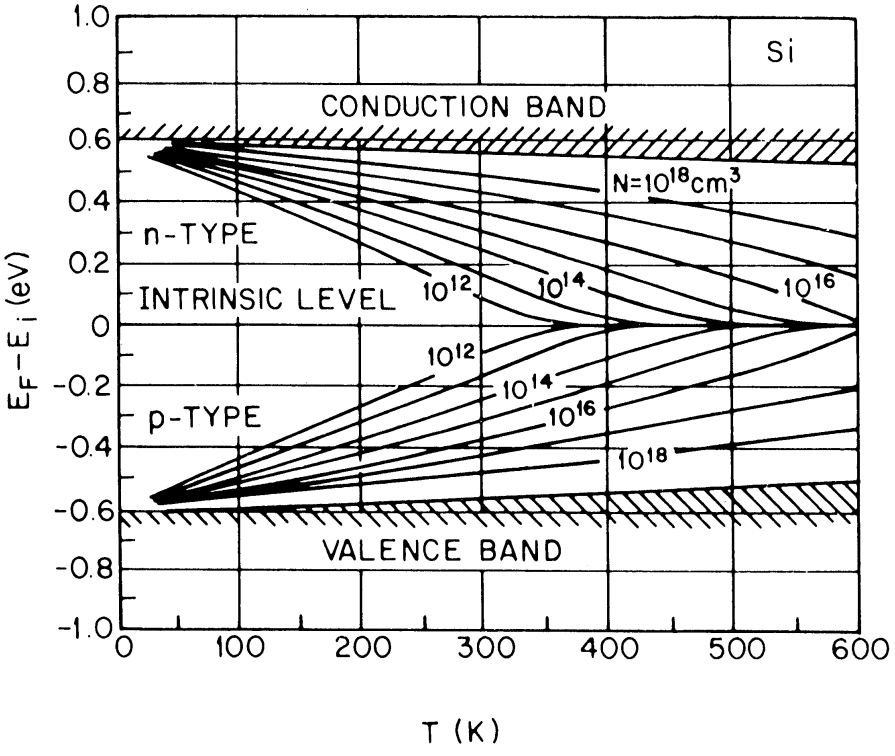
$$n = N_c e^{-(E_c - E_F)/k_B T} = n_i e^{-(E_F - E_i)/k_B T}$$

$$p = N_v e^{-(E_F - E_v)/k_B T} = n_i e^{-(E_i - E_F)/k_B T}$$

and the Fermi level

$$E_F - E_i = -k_B T \log \frac{N_A - N_D}{n_i}$$

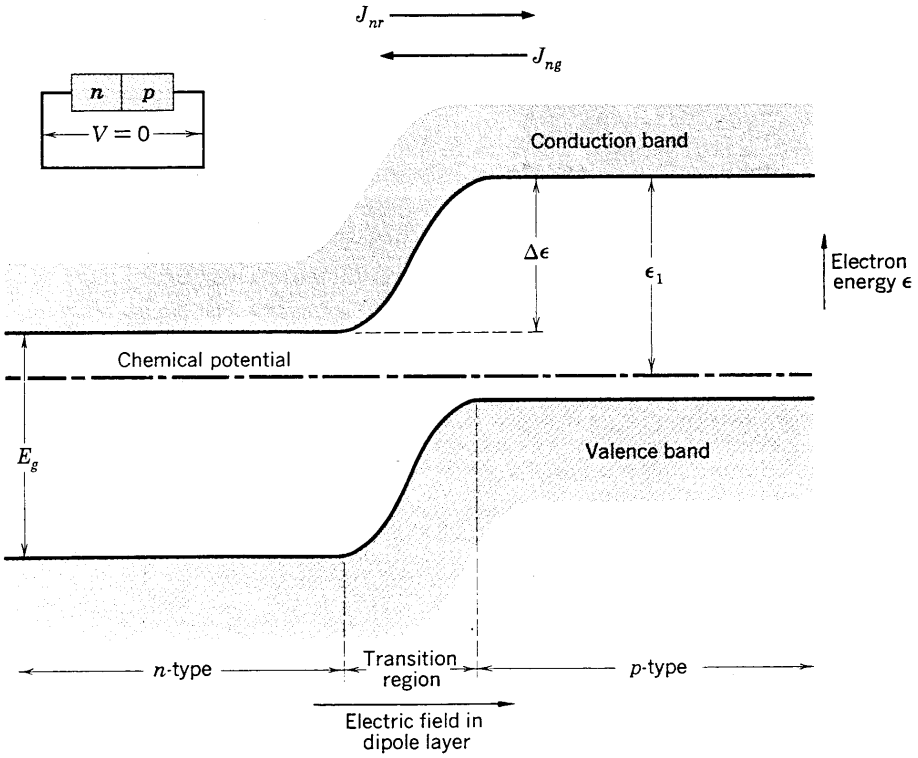
Variation of Fermi level with doping and temperature, including narrowing of the band gap with temperature:



(from Sze)

3. *p*-*n* Junctions

A *p*-*n* junction is formed at the interface of a *p*- and an *n*-type region.



(from Kittel)

Since the electron concentration in the *n*-region is greater than in the *p*-region, electrons will diffuse into the *p*-region.

Correspondingly, holes will diffuse into the *n*-region.

As electrons and holes diffuse across the junction, a space charge due to the ionized donor and acceptor atoms builds up. The field due to this space charge is directed to impede the flow of electrons and holes.

The situation is dynamic:

The concentration gradient causes a continuous diffusion current to flow.

The field due to the space charge drives a drift current in the opposite direction.

Equilibrium is attained when the two currents are equal, i.e. the sum of the diffusion and drift currents is zero.

The net hole current density is

$$J_p = -q_e D_p \frac{dp}{dx} + q_e p \mu_p E_p$$

where D_p is the diffusion constant for holes and E_p is the electric field in the p-region.

To solve this equation we make use of the following relationships:

The hole concentration is

$$p = n_i e^{(E_i - E_F)/k_B T} ,$$

so its derivative

$$\frac{dp}{dx} = \frac{p}{kT}\left(\frac{dE_i}{dx} - \frac{dE_F}{dx}\right)$$

Since the force on a charge q_e due to an electric field E is equal to the negative gradient of the potential energy,

$$q_e E = -\frac{dE_c}{dx} = -\frac{dE_v}{dx} = -\frac{dE_i}{dx}$$

As only the gradient is of interest and E_c , E_v *and* E_i differ only by a constant offset, any of these three measures can be used. We'll use the intrinsic Fermi level E_i since it applies throughout the sample.

The remaining ingredient is the Einstein relationship, which relates the mobility to the diffusion constant

$$\mu_p = \frac{q_e D_p}{k_B T}$$

Using these relationships the net hole current becomes

$$J_p = q_e p \frac{D_p}{k_B T} \frac{dE_F}{dx} = \mu_p p \frac{dE_F}{dx}$$

Accordingly, the net electron current

$$J_n = -q_e n \frac{D_n}{k_B T} \frac{dE_F}{dx} = -\mu_n n \frac{dE_F}{dx}$$

Since, individually, the net hole and electron currents in equilibrium must be zero, the derivative of the Fermi level

$$\frac{dE_F}{dx} = 0$$

$\Rightarrow$ in thermal equilibrium the Fermi level must be constant throughout the junction region.

For the Fermi level to be flat, the band structure must adapt, since on the *p*-side the Fermi level is near the valence band, whereas on the *n*-side it is near the conduction band.

If we assume that the dopants are exclusively donors on the *n*-side and acceptors on the *p*-side, the difference in the respective Fermi levels is

$$\Delta E_F = -k_B T \log \frac{N_A N_D}{n_i^2}$$

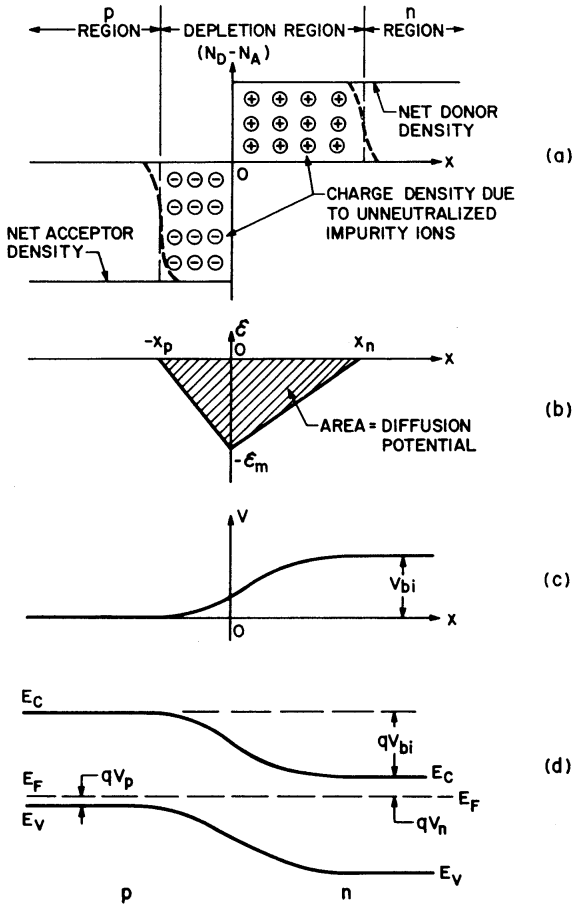
This corresponds to an electric potential

$$\Delta V_F = \frac{1}{q} \Delta E_F \equiv V_{bi}$$

often referred to as the “built-in” voltage of the junction.

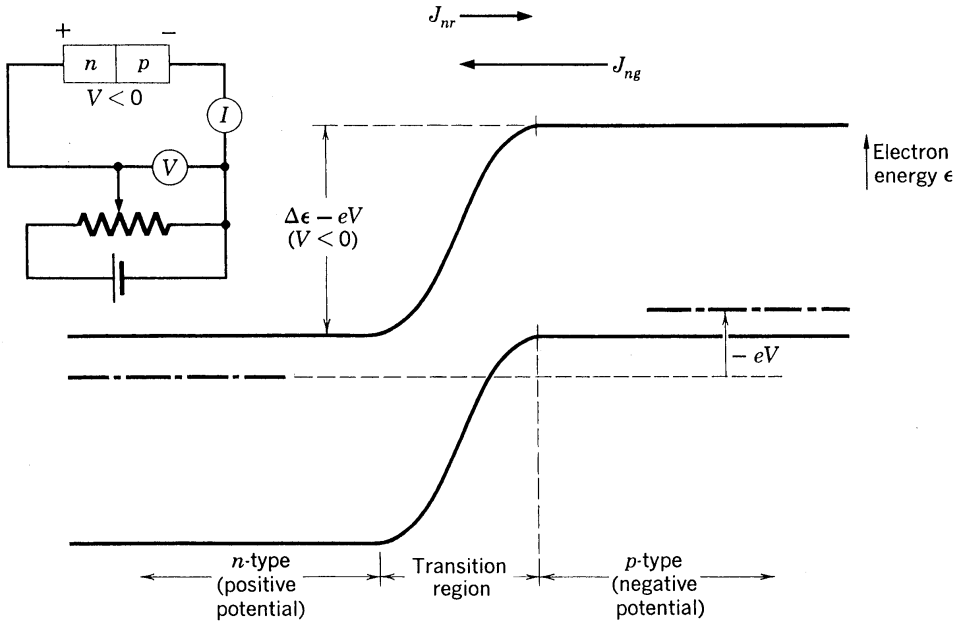
As either N_A or N_D increases relative to n_i, the respective Fermi level moves closer to the band edge, increasing the built-in voltage.

With increasing doping levels the built-in voltage approaches the equivalent potential of the band-gap E_g/q_e.



(from Sze)

The inherent potential distribution in the junction leads to a depletion region, whose width can be increased by application of an external potential, i.e. reverse biasing the junction.



(from Kittel)

This was discussed in a previous lecture.

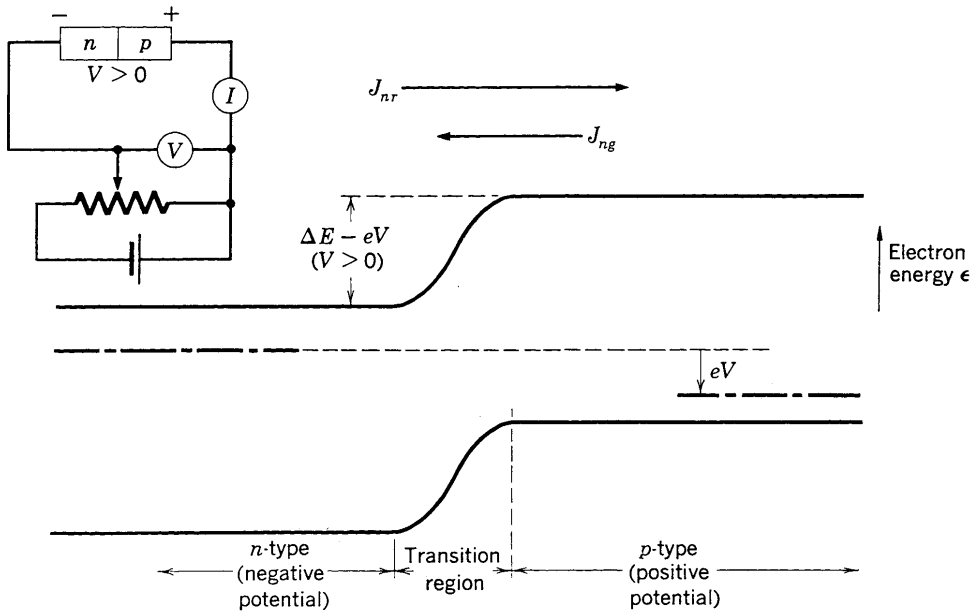
Now the forward bias mode will be treated in more detail.

Complication:

> Applying an external bias leads to a condition that deviates from thermal equilibrium, i.e. the Fermi level is no longer constant throughout the junction.

4. The Forward-Biased *p*-*n* Junction

If a positive voltage is applied to the *p*-electrode relative to the *n*-electrode, the total variation of the electric potential across the junction will decrease.



(from Kittel)

Since this reduces the electric field across the junction, the drift component of the junction current will decrease. Since the concentration gradient is unchanged, the diffusion current will exceed the drift current and a net current will flow.

This net current leads to an excess of electrons in the *p*-region and an excess of holes in the *n*-region. This “injection” condition lead to a local deviation from equilibrium, i.e. $pn > n_i^2$. Equilibrium will be restored by recombination.

> Note that a depletion region exists even under forward bias, although its width is decreased. The electric field due to the space charge opposes the flow of charge, but the large concentration gradient overrides the field.

Consider holes flowing into the *n*-region. They will flow through the depletion region with small losses due to recombination, as the electron concentration is small compared with the bulk.

When holes reach the *n*-side boundary of the depletion region the concentration of electrons available for recombination increases and the concentration of holes will decrease with distance, depending on the cross-section for recombination, expressed as a diffusion length.

Ultimately, all holes will have recombined with electrons. The required electrons are furnished through the external contact from the power supply.

On the *p*-side, electrons undergo a similar process. The holes required to sustain recombination are formed at the external contact to the *p*-region by electron flow toward the power supply, equal to the electron flow toward the *n*-contact.

The steady-state distribution of charge is determined by solving the diffusion equation.

$$D_n \frac{d^2 n_p}{dx^2} - \frac{n_p - n_{p0}}{\tau_n} = 0$$

Electrons flowing into the p region give rise to a local concentration n_p in excess of the equilibrium concentration n_{p0}. This excess will decay with a recombination time τ_n, corresponding to a diffusion length L_n.

The first boundary condition required for the solution of the diffusion equation is that the excess concentration of electrons vanish at large distances x,

$$n_p(\infty) = n_{p0}$$

The second boundary condition is that the carriers are injected at the origin of the space charge region $x = 0$ with a concentration $n_p(0)$.

This yields the solution

$$n_p(x) = n_{p0} + \left(n_p(0) - n_{p0}\right)e^{-x/L_n}$$

From this we obtain the electron current entering the p-region

$$J_{np} = -q_e D_n \left.\frac{dn_p}{dx}\right|_{x=0} = q_e D_n \frac{n_p(0) - n_{p0}}{L_n}$$

This says that the electron current is limited by the concentration gradient determined by the carrier density at the depletion edge $n_p(0)$ and the equilibrium minority carrier density n_{p0}.

Determining the equilibrium minority n_{p0} is easy

$$n_{p0} = n_i^2 / N_A$$

The problem is that $n_p(0)$ is established in a non-equilibrium state, where the previously employed results do not apply.

> To analyze the regions with non-equilibrium carrier concentrations a simplifying assumption is made by postulating that the product pn is constant. In this specific quasi-equilibrium state this constant will be larger than n_i^2, the pn-product in thermal equilibrium.

In analogy to thermal equilibrium, this quasi-equilibrium state is expressed in terms of a "quasi-Fermi level", which is the quantity used in place of E_F that gives the carrier concentration under non-equilibrium conditions.

The postulate *pn= const.* is equivalent to stating that the non-equilibrium carrier concentrations are given by a Boltzmann distribution, so the concentration of electrons is

$$n = n_i e^{(E_{Fn}-E_i)/k_B T}$$

where E_{Fn} is the quasi-Fermi level for electrons, and

$$p = n_i e^{(E_i-E_{Fp})/k_B T}$$

where E_{Fp} is the quasi-Fermi level for holes.

The product of the two carrier concentration in non-equilibrium is

$$pn = n_i^2 e^{(E_{Fn}-E_{Fp})/k_B T}$$

If *pn* is constant throughout the space-charge region, then E_{Fn}-E_{Fp} must also remain constant.

Using the quasi-Fermi level and the Einstein relationship, the electron current entering the *p*-region becomes

$$J_{np} = -q_e D_n \left.\frac{dn_p}{dx}\right|_{x=0} = --q_e D_n \frac{d}{dx}(n_i e^{(E_{Fn}-E_i)/k_B T}) = -\mu_n n \frac{dE_{Fn}}{dx}$$

These relationships describe the behavior of the quasi-Fermi level in the depletion region. How does this connect to the neutral region?

In the neutral regions the *majority* carrier motion is dominated by drift (in contrast to the injected *minority* carrier current that is determined by difusion). Consider the n-type region. Here the bulk electron current that provides the junction current

$$J_{nn} = -\mu_n n \frac{dE_i}{dx}$$

Since the two electron currents must be equal

$$J_{nn} = J_{np}$$

it follows that

$$\frac{dE_{Fn}}{dx} = \frac{dE_i}{dx}$$

i.e. the quasi-Fermi level follows the energy band variation.

$\Rightarrow$ in a neutral region, the quasi-Fermi level for the majority carriers is the same as the Fermi level in equilibrium.

At current densities small enough not to cause significant voltage drops in the neutral regions, the band diagram is flat, and hence the quasi-Fermi level is flat.

In the space charge region, *pn* is constant, so the quasi-Fermi levels for holes and electrons must be parallel, i.e. both will remain constant at their respective majority carrier equilibrium levels in the neutral regions.

If an external bias V is applied, the equilibrium Fermi levels are offset by V, so it follows that the quasi-Fermi levels are also offset by V.

$$E_{Fn} - E_{Fp} = q_e V$$

Consequently, the *pn*-product in non-equilibrium

$$pn = n_i^2 e^{(E_{Fn} - E_{Fp})/k_B T} = n_i^2 e^{q_e V / k_B T}$$

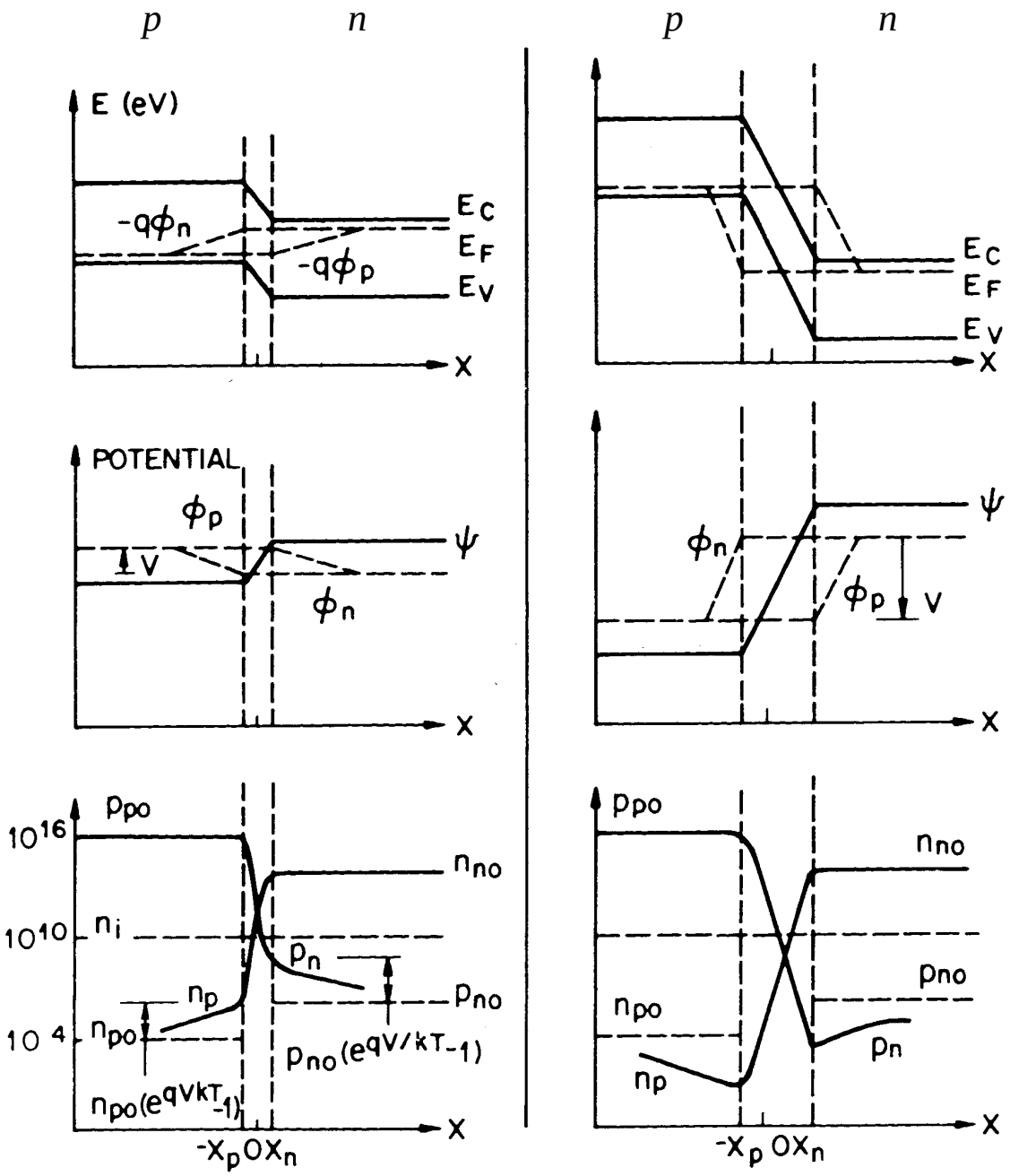
If the majority carrier concentration is much greater than the concentration due to minority injection ("low-level injection"), the hole concentration at the edge of the *p*-region remains essentially at the equilibrium value. Consequently, the enhanced *pn*-product increases the electron concentration.

$$n_p(0) = n_{p0} e^{q_e V / k_B T}$$

Correspondingly, the hole concentration in the *n*-region at the edge of the depletion zone becomes

$$p_n(0) = p_{n0} e^{q_e V / k_B T}$$

Energy band diagrams showing the intrinsic Fermi level Ψ, the quasi-Fermi levels for electrons Φ_n and holes Φ_p, and the carrier distributions for forward (a) and reverse bias conditions (b).

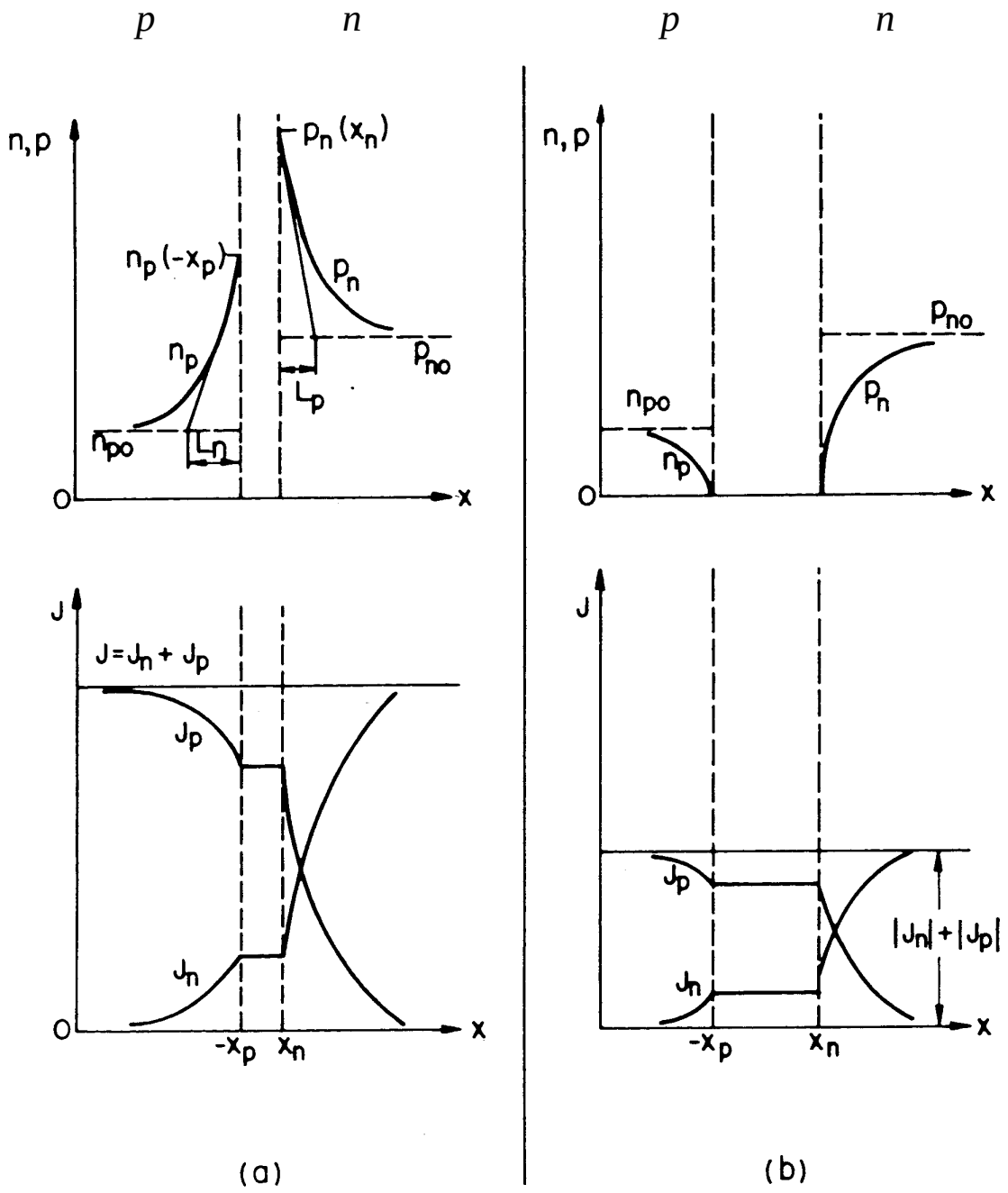


(a) forward bias

(b) reverse bias

(from Sze)

Carrier distributions and current densities for forward (a) and reverse bias conditions (b)



forward bias

reverse bias

(from Sze)

Since the equilibrium concentrations

$$n_{p0} = \frac{n_i^2}{N_A} \quad \text{and} \quad p_{n0} = \frac{n_i^2}{N_D}$$

the components of the diffusion current due to holes and electrons are

$$J_n = q_e D_n \frac{n_i^2}{N_A L_n}\left(e^{q_e V / k_B T} - 1\right)$$

$$J_p = q_e D_p \frac{n_i^2}{N_D L_p}\left(e^{q_e V / k_B T} - 1\right)$$

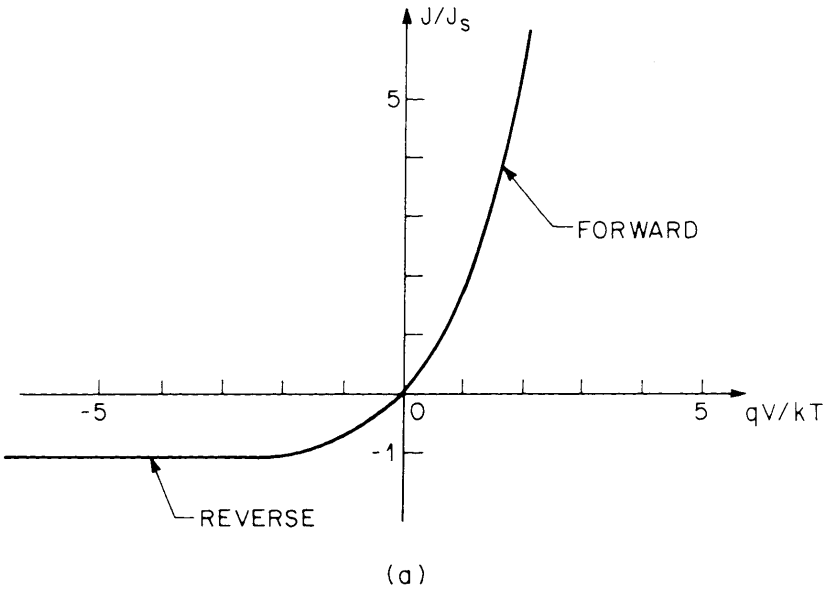
The total current is the sum of the electron and hole components

$$J = J_n + J_p = J_0(e^{q_e V / k_B T} - 1)$$

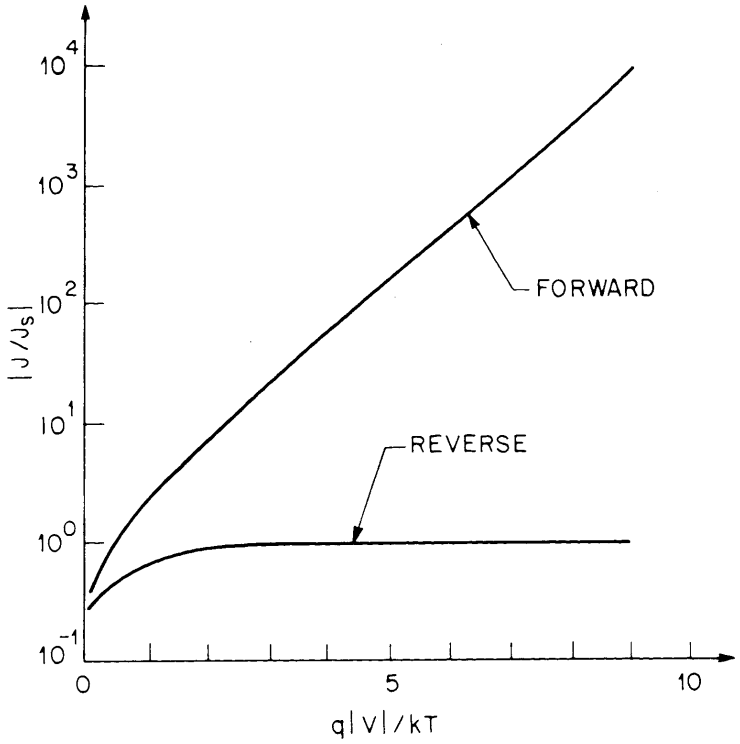
where

$$J_0 = q_e n_i^2 \left(\frac{D_n}{N_A L_n} + \frac{D_p}{N_D L_p} \right)$$

Forward and reverse characteristics of a *pn*-junction diode



(a)



(from Sze)

Note that in the diode equation

a) The band gap does not appear explicitly
   (only implicitly via n_i)

b) The total current has two distinct components, due to electrons and holes.

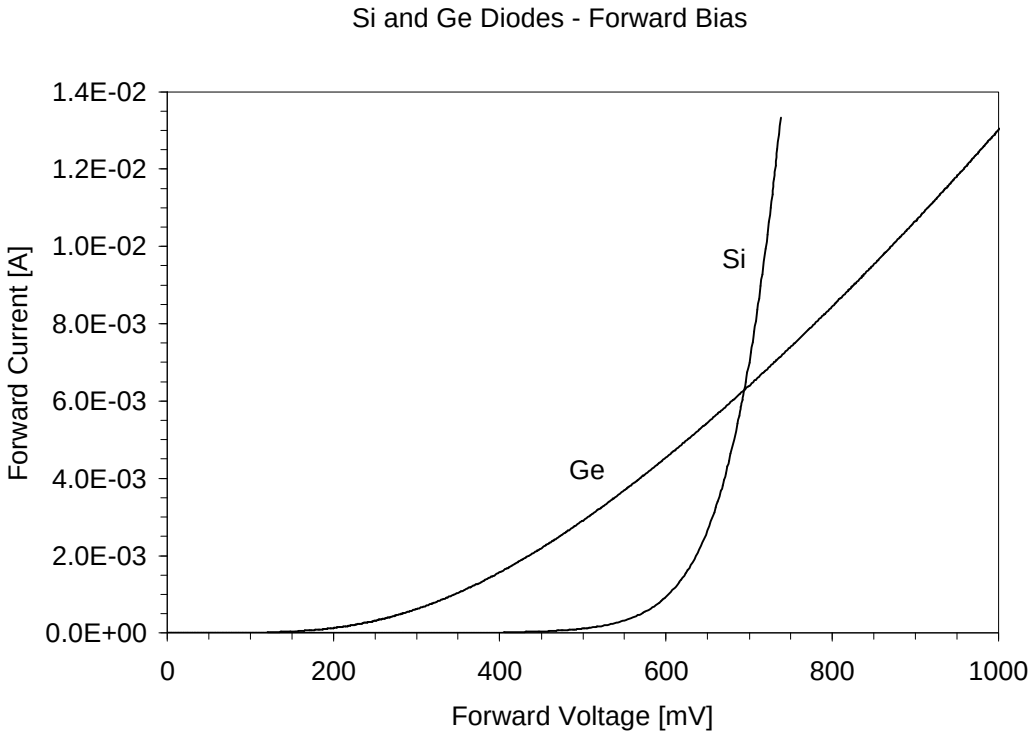
c) The electron and hole currents are not necessarily equal

$$\frac{I_n}{I_p} = \frac{N_D}{N_A} \quad \text{if} \quad \frac{D_n}{L_n} = \frac{D_p}{L_p}$$

d) Current flows for all values of V. However, when plotted on a linear scale, the exponential appears to have a knee, often referred to as the “turn-on” voltage

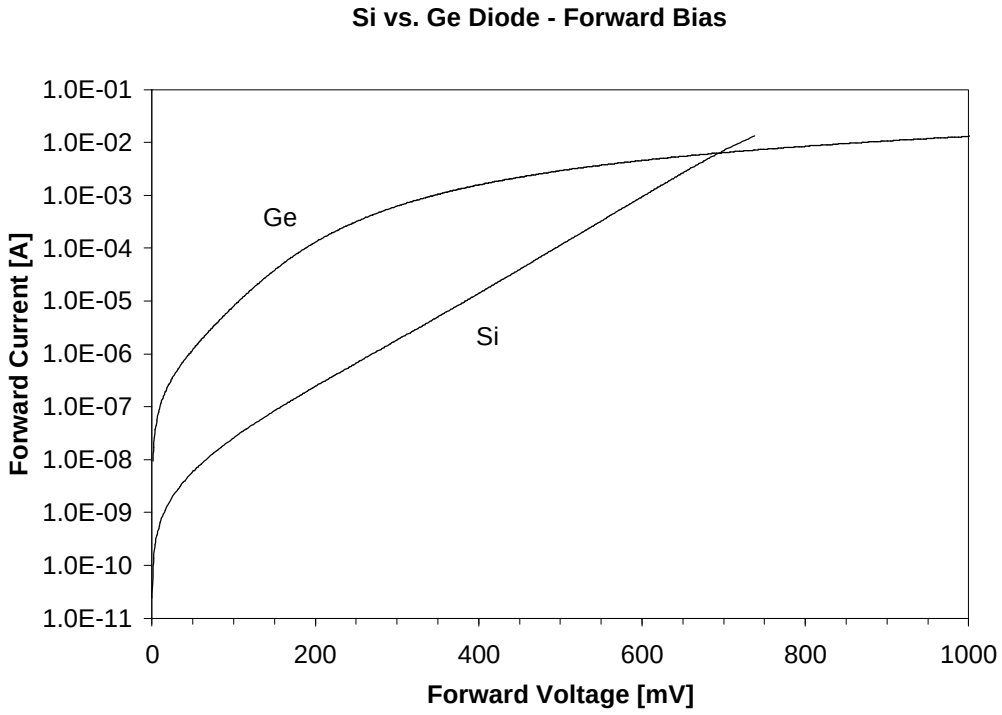
e) The magnitude of the turn-on voltage is determined by I_0. Diodes with different band-gaps will show the same behavior if I_0 is the same.

Comparision between commercial Si and Ge junction diodes (1N4148 and 1N34A)

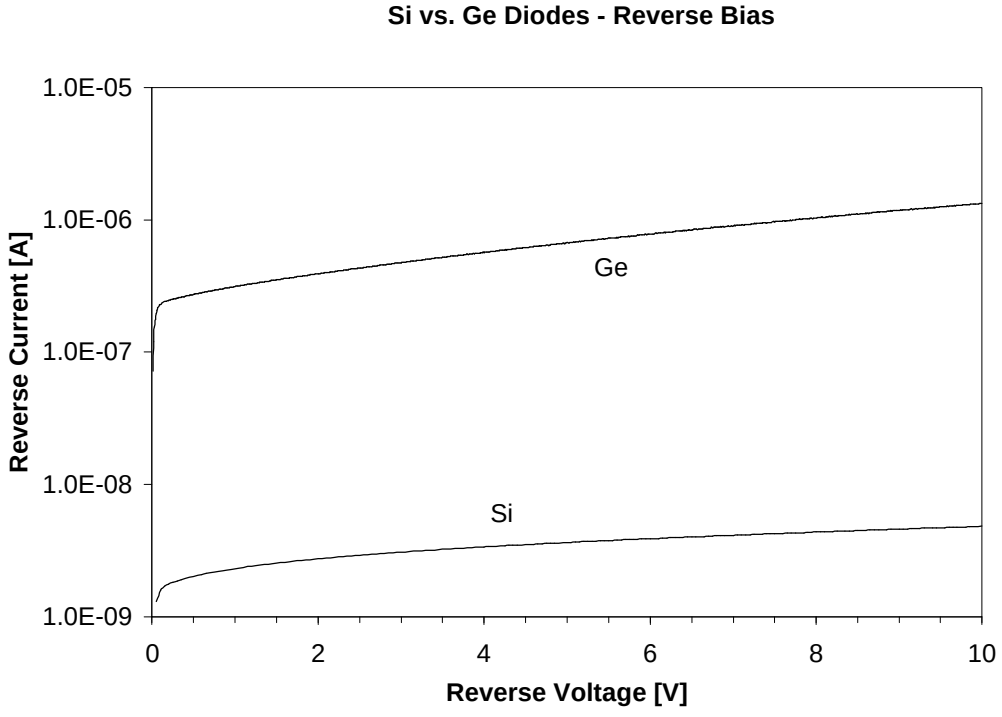


On a linear scale the Ge diode “turns on” at 200 – 300 mV, whereas the Si diode has a threshold of 500 – 600 mV

However, on a logarithmic scale it becomes apparent that both diodes pass current at all voltages >0.

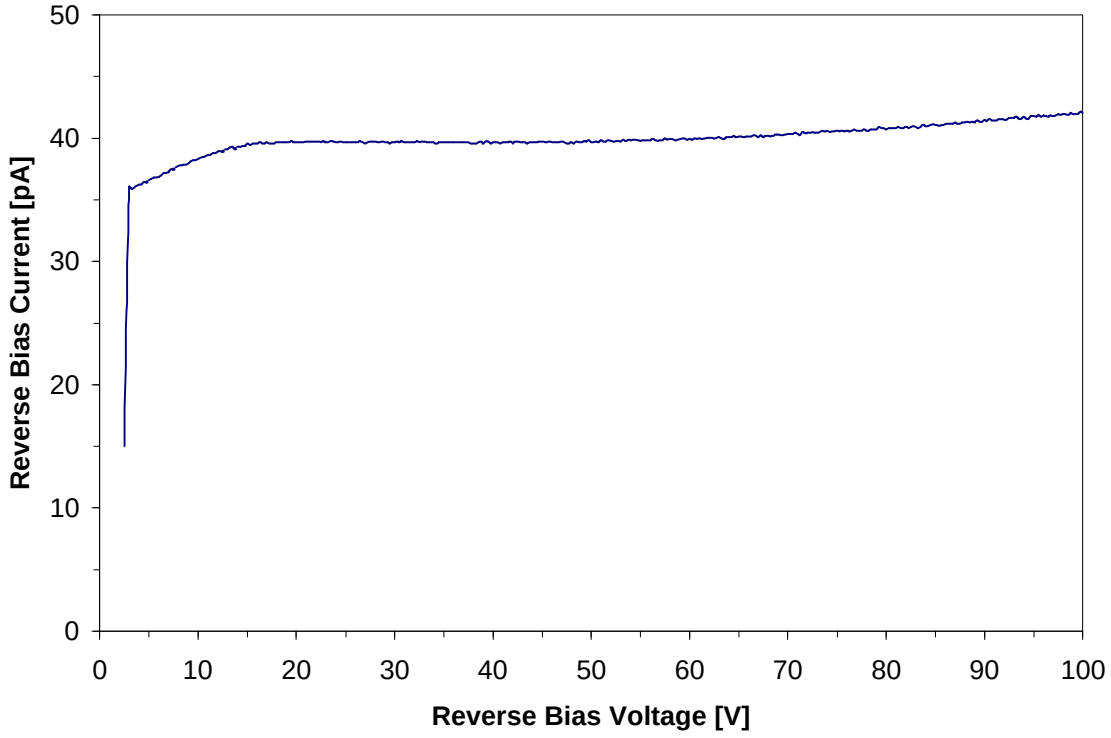


The reverse current shows why the Ge diode shows greater sensitivity at low voltages (smaller band-gap $\Rightarrow$ increased n_i).



The Si diode shows a “textbook” exponential forward characteristic at currents >10 nA, whereas the Ge diode exhibits a more complex structure.

A “state-of-the-art” reverse diode characteristic (Steve Holland)



The depletion width is 300 μm, attained at about 20 V.

The area of the diode is 9 mm^2, so the reverse leakage current of 40 pA corresponds to 450 pA/cm^2, which is about 10x the theoretical value.

The discrepancies between the measured results and the simple theory require the analysis of additional processes in the depletion zone.

One can recognize four regions in the forward current:

a) generation- recombination in the depletion region
b) diffusion current (as just calculated for the ideal diode)
c) high-injection region where the injected carrier concentration affects the potentials in the neutral regions.
d) voltage drop due to bulk series resistance

The reverse current is increased due to generation currents in the depletion zone.