11. NON-EQUILIBRIUM CONDITIONS: TRANSPORT

Up till now, we have considered non-equilibrium behavior in totally homogeneous cases. We now address non-equilibrium cases, which have spatial variation and possibly also net fluxes of particles (current!).

The conventional method to test this case is to take a closer look at the distribution function \( f(E, E_r) \) by resolving it into both space and \( k - \)dependence.

That is, we define a six-dimensional “phase space” \( k, r \)

\[
f(r, k, t) \, d^3k \, d^3r = \text{# of particles within } d^3k \text{ about } k \text{ in volume } d^3r \text{ around } r \text{ at time } t.
\]

The kinematics of particles causes an orderly and predictable variation of \( f(r, k, t) \) as follows.

In time \( \delta t \), particles at \( r, k \) go to \( r + v \delta t, k + \frac{E}{\hbar} \delta t, t + \delta t \)

Therefore, the particles that enter \( d^3r \, d^3k \) at \( r, k, t \) were located at

\[
(r - v \delta t, k - \frac{E}{\hbar} \delta t, t - \delta t).
\]

We define \( \delta f_{\text{kin}} \)

\[
\delta f_{\text{kin}} = \left[ f(r - v \delta t, k - \frac{E}{\hbar} \delta t, t - \delta t) - f(r, k, t) \right] \delta t
\]

In addition to kinematic effects that can change \( f \), we have collisions or scattering events that lead to randomize the occupation of microstates, hence serve to restore internal equilibrium within a set of states described by a quasi-Fermi level, and we have transitions that serve to restore equilibrium between sets of states having different quasi-Fermi levels.

We write, therefore
\( \delta f_{\text{total}} = \delta f_{\text{kin}} + \delta f_{\text{coll}} + \delta f_{\text{transitions}} \)

and since \( \delta f_{\text{total}} \) now take account of all mechanisms for charging \( f \), and since particles are accounted for, we can assert that \( \delta f_{\text{total}} \) is small and can say \( \delta f_{\text{total}} = 0 \).

The assumption behind this are subtle. If we assume \( \delta f_{\text{total}} = 0 \), this means that the randomizing effects of collision and transitions are in dynamic balance with the ordering effects of \( \delta f_{\text{kin}} \), and any large changes in \( f \) with time or position must arise from a mismatch in the average dynamic balance between ordering and randomizing.

In order for this to be true, we must be in the so-called hydrodynamic regime, in which, like the kinematic theory of gases, our time interval \( \delta t \) covers many collision time, and our spatial variation \( \delta r \) or \( v \delta t \) covers many mean free paths.

There are experiments in GaAs FETs in which the hydrodynamic assumption fails, and in which the transport is termed ballistic, that is it, takes place in times and over distances in which only one or two collisions can be expected to happen. In general, ballistic conditions are found in highly perfect materials at high electric fields in structures of small dimensions. The derivations that follow assume non-ballistic conditions, i.e., that on a time scale \( \delta t \), \( \delta f_{\text{total}} = 0 \), and changes occur only in trading kinematic ordering off against random scattering.

With that, we can now deal with the

**BOLTZMAN TRANSPORT EQUATION**

By Taylor’s theorem

\[ \delta f_{\text{kin}} = - (v \cdot \nabla_r f + \frac{F}{\hbar} \cdot \nabla_k f + \frac{\partial f}{\partial t}) \delta t \]

Therefore, we can write, since \( \delta f_{\text{total}} = 0 \)

\[ - \delta f_{\text{kin}} = \delta f_{\text{coll}} + \delta f_{\text{transitions}} \]

\[ \frac{\partial f}{\partial t} + v \cdot \nabla_r f + \frac{F}{\hbar} \cdot \nabla_k f = \left( \frac{\partial f}{\partial t} \right)_{\text{coll}} + \left( \frac{\partial f}{\partial t} \right)_{\text{transitions}} \]

is the BTE
Where \( f \) is the distribution function for one particular set of strongly scattering state, e.g. electrons, or holes.

**Properties of \( f\left( \vec{r}, \vec{k}, t \right) \)**

Since \( \int_{\text{Vol}} d^3 \vec{r} \int_{\text{FBZ}} d^3 \vec{k} f\left( \vec{r}, \vec{k}, t \right) = \text{total # of electrons (or holes, or etc.)} \)

**DENSITY:** \( n\left( \vec{r}, t \right) = \int d^3 \vec{k} f_n\left( \vec{r}, \vec{k}, t \right) \)

With a similar expression for holes

**CURRENT DENSITY:** To find the current density, we evaluate the average velocity:

\[
\langle \vec{v} \rangle = \frac{\int d^3 \vec{k} f\left( \vec{r}, \vec{k}, t \right) \vec{v}}{\int d^3 \vec{k} f\left( \vec{r}, \vec{k}, t \right)}
\]

\[
\Rightarrow \vec{J}_n = -qn \langle \vec{v}_n \rangle = -q \int d^3 \vec{k} f_n\left( \vec{r}, \vec{k}, t \right) \vec{v}_n
\]

\[
\vec{J}_p = +pq \langle \vec{v}_p \rangle = +q \int d^3 \vec{k} f_p\left( \vec{r}, \vec{k}, \hbar \right) \vec{v}_p
\]

where \( \vec{v}_n = \frac{1}{\hbar} \nabla_{\vec{k}} E(\vec{k}) \) for electrons in the conduction band

where \( \vec{v}_n = \frac{1}{\hbar} \nabla_{\vec{k}} E(\vec{k}) \) for holes in the valence band

We now look at Boltzman transport equations by performing various integrals. We do the electron case.

Zeroth integral: \( \int d^3 \vec{k} \)
\[
\int \frac{\partial f}{\partial t} \, d^3 k + \int \nabla \cdot \mathbf{f} \, d^3 k + \int \frac{F}{\hbar} \cdot \nabla f \, d^3 k = \int \left( \frac{\partial f}{\partial t} \right)_{\text{coll}} \, d^3 k + \int \left( \frac{\partial f}{\partial t} \right)_{\text{transitions}} \, d^3 k
\]

The terms can be taken one by one

\[
\int \frac{\partial f}{\partial t} \, d^3 k = \frac{\partial}{\partial t} \int d^3 k \, f = \frac{\partial n}{\partial t}
\]

\[
\int \nabla \cdot \mathbf{f} \, d^3 k = \nabla_{r} \int \nabla \cdot f \, d^3 k = \nabla_{r} \cdot n \langle \mathbf{v} \rangle = -\frac{1}{q} \nabla \cdot J_n
\]

\[
\int d^3 k \, \frac{F}{\hbar} \cdot \nabla f = 0 \text{ if } F \text{ not a function of } k \text{ or if }
\]

\[
F \perp \mathbf{v} \quad (\text{such as } q \mathbf{v} \times \mathbf{B})
\]

(In semiconductors
\[
\nabla_{k} f = \frac{\partial f}{\partial E} \nabla_{k} E = \hbar \mathbf{v} \cdot \frac{\partial f}{\partial E}
\]

\[
\int \left( \frac{\partial f}{\partial t} \right)_{\text{coll}} \, d^3 k = 0 \text{ if collision preserve particles count}
\]

\[
\int \left( \frac{\partial f}{\partial t} \right)_{\text{transitions}} \, d^3 k = G - R = -U \quad \text{Net generation from transitions}
\]

Thus the zeroth integral leads to continuity equations for electrons and holes.

\[
\frac{\partial n}{\partial t} - \frac{1}{q} \nabla \cdot J_n = (G - R)_n \quad \text{CONTINUITY}
\]

\[
\frac{\partial p}{\partial t} + \frac{1}{q} \nabla \cdot J_p = (G - R)_p
\]

First integral  \[ \int \vec{v} \, d^3 k \quad \text{Momentum Balance (Particle Flow)} \]

Second integral  \[ \int \mathbf{v} \cdot d^3 \mathbf{k} \quad \text{Energy Balance (Heat Flow)} \]
We first address only the momentum balance term, i.e. the equivalent of the second hydrodynamic equation. The energy balance term must be considered where the specific heat of the free particles is an important part of whole, which can occur in metals at low temperature, and in thermoelectric properties of semiconductors where the heat flux of the particles is balanced by an electrostatic potential.

We will treat the particle flow equation in what is called the relaxation time approximation. We assume

\[ f = f_o + f_1 \quad \text{where} \quad f_1 \ll f_o \]

\( f_o \) is the same form as the equilibrium distribution function, but has a quasi-Fermi energy (and temperature) appropriate to the particular set of states, which may be a function of position. Note that \( f_o \) is an even function of \( k \) hence, it creates no current. \( f_1 \) is a small perturbation, which is not an even function of \( k \). If is \( f_1 \) that accounts for currents

We now assume

\[
\left( \frac{\partial f_i}{\partial t} \right)_{\text{coll}} = -\frac{f_i}{\tau}
\]

\( \tau = \)relaxation time, which can be estimated from quantum mechanics

With this assumption

\[
f_1 = -\tau \left[ \frac{\partial f}{\partial t} + \vec{v} \cdot \nabla f + \frac{\vec{F}}{\hbar} \nabla_k f - \left( \frac{\partial f}{\partial t} \right)_{\text{transitions}} \right]
\]

Forces other than electric (built into \( E_{fn} \) and \( \nabla E_{fn} \))
and \( n\langle \vec{v} \rangle = \int d^3 k \vec{v} f_i \)

because \( \int d^3 k \vec{v} f_o = 0 \)

Unless \( \left( \frac{\partial f}{\partial t} \right)_{\text{transition}} \) is a function of \( \vec{v} \), \( \int d^3 k \vec{v} \left( \frac{\partial f}{\partial t} \right)_{\text{trans}} = 0 \)

And since \( f_i \ll f_o \) we can neglect all \( \nabla f_i \) terms, compared to \( \nabla f_o \) terms

We also have to assume that \( \int \frac{\partial f_i}{\partial t} \vec{v} d^3 k \) is small, that is, on the time scale of the relaxation time, the current density is very slowly varying. This is not unreasonable, since the net effect of many collisions is required to set up a change in \( \langle \vec{v} \rangle \).

Thus

\[
n\langle \vec{v} \rangle = -\int d^3 k \left[ \left( \vec{v} \vec{v} \right) \nabla, f_o \right] \tau
\]

where the \( \vec{F} \) has been dropped, because except for electric fields, which get built into \( E_{Fn} \), and magnetic fields, for which \( F_v = 0 \), there are no important forces.

Now \( f_o = \frac{1}{1 + e^{\frac{E(\vec{k}) - E_{Fn}}{kT}}} \)

\[
\nabla, f_o = \left[ -f_o (1 - f_o) \right] \nabla, \left[ \frac{E(\vec{k})}{kT} - \frac{E_{Fn}}{kT} \right]
\]

A simple algebraic proof can be written out

\[
f_o (1 - f_o)
\]

Several \( kT \)
If the temperature is a function of position, then the $\nabla \frac{E(k)}{kT}$ term and the $\frac{1}{kT}$ part of the $\nabla \frac{E_{Fn}}{kT}$ term must be kept. However, for isothermal conditions, which is approximate to the vast majority of semiconductor device problems; we can write

$$n\langle \mathbf{v} \rangle = \left[ \int d^3k \mathbf{v} \mathbf{v} \tau f_o(1-f_o) \right] \left[ -\frac{\nabla E_{Fn}}{kT} \right]$$

Flow is proportional to the gradient of electrochemical potential (quasi-Fermi level)!!

We define a conductivity effective density

$$n_c = \int d^3k f_o(1-f_o)$$

[ Note: in Boltzman limit, $f_o \ll 1$ at energies of interest $\Rightarrow n_c \rightarrow n$ ]

and we define a conductivity average:

$$\langle x \rangle_c = \frac{\int d^3k f_o(1-f_o) x}{n_c}$$

[ In Boltzman limit $\langle x \rangle_c \rightarrow \langle x \rangle$]

If $\tau$ is not a function of direction, i.e. if $\tau$ is a function of $E$ only (often a reasonable approximation), then

$$\int d^3k \mathbf{v} \mathbf{v} \tau f_o(1-f_o)$$ has only diagonal components of the form

$$\int d^3k \mathbf{v} \mathbf{v} \tau f_o(1-f_o)$$ and all these components are equal. Hence for $\tau$ isotropic,

$$\int d^3k \mathbf{v} \mathbf{v} \tau f_o(1-f_o) = n_c \langle \frac{1}{3} v^2 \tau \rangle_c$$. This quantity has another name. It is called the diffusion coefficient for the particle. That is

$$D_s = \langle \frac{1}{3} v^2 \tau \rangle_c$$ for electron

Therefore, we find
\[ n \langle \nu \rangle = D_n \ n_c \left[ -\frac{\nabla E_{Fn}}{kT} \right] \]

On, since

\[ \vec{J}_n = -q \ n \langle \nu \rangle \]

\[ \vec{J}_n = \left( \frac{q D_n}{kT} \right) n_c \nabla E_{Fn} \]

**ELECTRON FLOW**

Let us look at \( \nabla E_{Fn} \) more closely

\[ E_{Fn} = \left( E_{Fn} - E_c \right) + \frac{E_c}{\nabla} \]

- Determines \( n \)
- Follows Band

Contains Chemical Potential

Contains Electrostatic Potential

We consider a simple example to illustrate how to interpret this:

For a parabolic bond

\[ n = \frac{2}{\sqrt{\pi}} \ N_c \ F_{\frac{1}{2}} \left( \frac{E_{Fn} - E_c}{kT} \right) \]

\[ \Rightarrow \nabla n = \frac{2}{\sqrt{\pi}} \ N_c \left[ \frac{d}{dz} F_{\frac{1}{2}}(z) \right] \frac{\nabla \left( E_{Fn} - E_c \right)}{kT} \]

\[ n_c = \int d^3 k \ f_o (1 - f_o) \]

Or, by simple manipulation

\[ \vec{J}_n = q \ \mu_n \ n_c \ \vec{E} + q \ D_n \ \nabla n \]

where \[ n_c = \int d^3 k \ f_o (1 - f_o) \]

\[ \mu_n = \frac{q}{kT} D_n \]

**Einstein relation**

**Mobility**
Note that the Einstein relation comes right out of the relaxation time solution to the Boltzman transport equation.

Note also that while $\nabla n$ drives diffusion

$$n_c \vec{E} \text{ determines drift,}$$

and

$$n_c \to n \text{ only in Boltman limit.}$$

For holes: $f_o = \frac{1}{1 + e^{\left(E_{Fp} - E(k)\right)/kT}}$, $\nabla f_o = -f_o (1 - f_o) \nabla \frac{E_{Fp}}{kT}$

$$p\langle v \rangle = p_c D_p \nabla \frac{E_{Fp}}{kT}$$

$$\vec{J}_p = \left( \frac{q D_p}{kT} \right) p_c \nabla E_{Fp} \quad (T = \text{const})$$

But $\frac{d}{dz} F(z) = \int dx \frac{e^{x-z}}{(1+e^{x-z})^2} x^{1/2}$

$$= \int dx \left( \frac{1}{x^{1/2}} \right) \left[ \frac{1}{1+e^{x-z}} \right] \left[ 1 - \frac{1}{1+e^{x-z}} \right]$$

Thus $\frac{2}{\sqrt{\pi}} N_c \frac{d}{dz} F_{1/2}(z) = n_c$

Therefore, we conclude

$$\nabla \left( E_{Fn} - E_c \right) = \frac{1}{kT} \frac{1}{n_c} \nabla n$$

Furthermore, $\nabla E_c = +q \vec{E}$ where $\vec{E}$ is field
Therefore, the flow equation becomes

\[ \vec{J}_n = \left( \frac{qD_n}{kT} \right) n_c \left[ \nabla (E_{F_n} - E_c) + \nabla E_c \right] \]

or

\[ \vec{J}_n = \left( \frac{q_n}{kT} \right) q n_c \begin{pmatrix} \nabla E \text{ (Drift)} \\ kT \nabla n \text{ (Diffusion)} \end{pmatrix} \]

\[ E_{F_p} = (E_{F_p} - E_v) + E_v \]

But \( p = \frac{2}{\sqrt{\pi}} N_v \frac{E}{F_p} \left( \frac{E_{F_p} - E_v}{kT} \right) \)

\[ \Rightarrow \nabla p = -p \frac{\nabla (E_{F_p} - E_v)}{kT} \]

But \( \nabla E_v \) still is \( q \vec{E} \)

Thus \( \vec{J}_p = \left( \frac{qD_p}{kT} \right) q p_c \begin{pmatrix} \vec{E} - kT \nabla p \end{pmatrix} \]

Or \( \vec{J}_p = q \mu_p p_c (\vec{E} - qD_p \nabla p) \)

where

\[ \mu_p = \frac{q D_p}{kT} \quad D_p = \frac{1}{3} \left\langle \frac{\tau}{3} \frac{v^2}{c} \right\rangle \]

for holes

Finally, we define quasi-Fermi potentials
\[
\phi_p = -\frac{1}{q}(E_{Fp} - E_{ref}) \\
\phi_n = -\frac{1}{q}(E_{Fn} - E_{ref}) \\
\psi = -\frac{1}{q}(E_i - E_{ref})
\]

\[
n = n_i e^{kT(\psi - \phi_n)} , \quad p = n_i e^{kT(\phi_p - \psi)}
\]

in Boltzmann case \((E_{ref} \text{ cancels out})\)

**Summary of Device Equations in Boltzmann Limit**

**Poisson Equation:**
\[\nabla^2 \psi = -\frac{D}{\varepsilon} \]

**Charge Density Equation:**
\[p = q \left( N_D^+ - N_A^- + n_i e^{q(\phi_p - \psi)/kT} - n_i e^{q(\psi - \phi_n)/kT} \right) \]

**Electron Continuity Equation:**
\[\frac{\partial n}{\partial t} - \frac{1}{q} \nabla \cdot \vec{J}_n = (G - R)_n \]

**Hole Continuity Equation:**
\[\frac{\partial p}{\partial t} + \frac{1}{q} \nabla \cdot \vec{J}_p = (G - R)_p \]

**Flow Equations:**
\[\vec{J}_n = -q\mu_n n\nabla \psi + qD_n \nabla n \]
\[\vec{J}_p = -q\mu_p p\nabla \psi - qD_p \nabla p \]

**Einstein Equation:**
\[\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{q} \]

**G-R Terms:**
Determined by rate equation models, but typically of the form
\[(G - R)_n = -\frac{n}{\tau}\] in neutral material
\[(G - R)_p = -\frac{p}{\tau}\]

where \(\tau\) is minority carrier lifetime \(n = p =\) excess carriers in quasi-neutral regimes

In non-neutral regions \(G - R\) terms are more complex.

NOW WE CAN GET PRACTICAL!
we will first consider the ideal MIS diode, which will serve as a basis for understanding of nonideal MIS diode characteristics. Section 3 presents the nonideal situations caused by surface states, surface charges, and space charges in the insulator. The ion transport as discovered by Snow et al. the conductance method as proposed by Nicollian and Goetzberger, the time constant dispersion, and 1/f noise will also be considered in Section 3. Section 4 is concerned with the effects on MIS characteristics due to metal work function, crystal orientation, lattice temperature, illumination, and γ-ray radiation. Section 5 considers some of the applications of MIS diodes as electronic and optical devices, and as a tool to study fundamental physical processes. Section 6 presents a brief discussion of the carrier transport and maximum dielectric strength of thin insulating films.

2. Solution of Poisson's Equation with Mobile Carriers

2.1 Ideal Metal-Insulator-Semiconductor (MIS) Diode

The MIS structure is shown in Fig. 1 where \( d \) is the thickness of the insulator and \( V \) is the applied voltage on the metal field plate. This defines that the voltage, \( V \), is positive when the metal plate is positively biased with respect to the ohmic contact, and \( V \) is negative when the metal plate is negatively biased with respect to the ohmic contact. This convention will be used throughout this chapter.

The energy band diagram of an ideal MIS structure for \( V = 0 \) is shown in Fig. 2, where Figs. 2(a) and 2(b) are for n-type and p-type semiconductors respectively. An ideal MIS diode is defined as follows: (1) at zero applied bias there is no energy difference between the metal work function \( \phi_m \) and the semiconductor work function, or the work function difference \( \phi_{ms} \) is zero:

\[
\phi_{ms} = \phi_m - \left( x + \frac{E_g}{2q} \right) - \Psi_B = 0 \quad \text{for n-type} \\
\phi_{ms} = \phi_m - \left( x + \frac{E_g}{2q} + \Psi_B \right) = 0 \quad \text{for p-type}
\]

where \( \phi_m \) is the metal work function, \( x \) the semiconductor electron affinity, \( \chi \), the insulator electron affinity, \( E_g \) the band gap, \( \Psi_B \) the potential barrier.

---

Fig. 1 Metal-insulator-semiconductor (MIS) structure.

Fig. 2 Energy band diagrams for ideal MIS structures at \( V = 0 \).
between the metal and the insulator, and $\psi$, the potential difference between the Fermi level $E_F$ and the intrinsic Fermi level $E_i$; in other words, the band is flat (flat-band condition) when there is no applied voltage; (2) the only charges which can exist in the structure under any biasing conditions are those in the semiconductor and those with the equal but opposite sign on the metal surface adjacent to the insulator; and (3) there is no carrier transport through the insulator under dc biasing conditions, or the resistivity of the insulator is infinity. The ideal MIS diode theory to be considered in this section serves as a foundation to understand practical MIS structures and to explore the physics of semiconductor surfaces.

When an ideal MIS diode is biased with positive or negative voltages, there are basically three cases which may exist at the semiconductor surface. These cases are illustrated in Fig. 3. Consider the $p$-type semiconductor first. When

![Energy band diagrams for ideal MIS structures](image)

**Fig. 3** Energy band diagrams for ideal MIS structures when $V \neq 0$ for both $p$-type and $n$-type semiconductors.

...a negative voltage ($V < 0$) is applied to the metal plate, Fig. 3(a), the top of the valence band bends upward and is closer to the Fermi level. For an ideal MIS diode there is no current flow in the structure [or $\delta (\ln \rho_m)/\delta x = 0$], so the Fermi level remains constant in the semiconductor. Since the carrier density depends exponentially on the energy difference ($E_F - E_v$), this band bending causes an accumulation of majority carriers (holes) near the semiconductor surface. This is the case of “accumulation.” When a small positive voltage ($V > 0$) is applied, Fig. 3(b), the bands bend downward, and the majority carriers are depleted. This is the case of “depletion.” When a larger positive voltage is applied, Fig. 3(c), the bands bend even more downward such that the intrinsic level $E_i$ at the surface crosses over the Fermi level $E_F$. At this point the number of electrons (minority carriers) at the surface is larger than that of the holes, the surface is thus inverted, and this is the case of “inversion.” Similar results can be obtained for the $n$-type semiconductor. The polarity of the voltage, however, should be changed for the $n$-type semiconductor.

### (1) Surface Space-Charge Region

We shall derive the relations between the surface potential, space charge, and electric field in this subsection. These relations will then be used to derive the capacitance-voltage characteristics of the ideal MIS structure in the next subsection.

Figure 4 shows a more detailed band diagram at the surface of a $p$-type semiconductor. The potential $\psi$ is defined as zero in the bulk of the semiconductor, and is measured with respect to the intrinsic Fermi level $E_i$ as shown. At the semiconductor surface $\psi = \psi_s$, and $\psi_s$ is called the surface potential. The electron and hole concentrations as a function of $\psi$ are given by the following relations:

$$n_p = n_{po} \exp(q\psi/kT) = n_{po} \exp(\beta\psi) \quad (2)$$
$$\rho_p = \rho_{po} \exp(-q\psi/kT) = \rho_{po} \exp(-\beta\psi) \quad (3)$$

where $\psi$ is positive when the band is bent downward (as shown in Fig. 4), $n_{po}$ and $p_{po}$ are the equilibrium densities of electrons and holes respectively in the bulk of the semiconductor, and $\beta = q/kT$. At the surface the densities are

$$n_s = n_{po} \exp(\beta\psi_s)$$
$$\rho_s = \rho_{po} \exp(-\beta\psi_s) \quad (4)$$

It is obvious from the previous discussions and with the help of Eq. (4) that the following regions of surface potential can be distinguished:
1. Ideal Metal-Insulator-Semiconductor (MIS) Diode

where \( N_D^+ \) and \( N_A^- \) are the densities of the ionized donors and acceptors respectively. Now, in the bulk of the semiconductor, far from the surface, charge neutrality must exist. Therefore \( \rho(x) = 0 \) and \( \psi = 0 \), and we have

\[
N_D^+ - N_A^- = n_p - p_p. \tag{7}
\]

In general for any value of \( \psi \), we have from Eqs. (2) and (3)

\[
p_p - n_p = p_p \exp(-\beta \psi) - n_p \exp(\beta \psi). \tag{8}
\]

The resultant Poisson's equation to be solved is therefore

\[
\frac{\partial^2 \psi}{\partial x^2} = \frac{q}{\varepsilon_i} \left[ p_p(e^{-\beta \psi} - 1) - n_p(e^{\beta \psi} - 1) \right]. \tag{9}
\]

Integration of Eq. (9) from the bulk toward the surface gives

\[
\int_0^{\psi(x)} \frac{\partial \psi}{\partial x} \, dx = \frac{q}{\varepsilon_i} \int_0^x \left[ p_p(e^{-\beta \psi} - 1) - n_p(e^{\beta \psi} - 1) \right] \, dx \tag{10}
\]

gives the relation between the electric field \( \varepsilon \equiv -\partial \psi/\partial x \) and the potential \( \psi \):

\[
\varepsilon^2 = \left( \frac{2kT}{q} \right)^2 \left( \frac{q p_p \beta}{2 \varepsilon_i} \right)^2 \left[ e^{-\beta \psi} + \beta \psi - 1 \right] + \frac{n_p}{p_p} \left( e^{\beta \psi} - \beta \psi - 1 \right) \tag{11}
\]

We shall introduce the following abbreviations:

\[
L_D \equiv \frac{2kT \varepsilon_i}{\sqrt{q p_p q_i^2}} = \sqrt{\frac{2 \varepsilon_i}{q p_p \beta}}, \tag{12}
\]

and

\[
F \left( \beta \psi, \frac{n_p}{p_p} \right) \equiv \left[ e^{-\beta \psi} + \beta \psi - 1 \right] + \frac{n_p}{p_p} \left( e^{\beta \psi} - \beta \psi - 1 \right)^{1/2} \geq 0 \tag{13}
\]

where \( L_D \) is called the extrinsic Debye length for holes. Thus the electric field becomes

\[
\varepsilon = -\frac{\partial \psi}{\partial x} = \pm \frac{2kT}{q L_D} F \left( \beta \psi, \frac{n_p}{p_p} \right) \tag{14}
\]

with positive sign for \( \psi > 0 \) and negative sign for \( \psi < 0 \). To determine the electric field at the surface, we let \( \psi = \psi_s \):

\[
\varepsilon_s = \pm \frac{2kT}{q L_D} F \left( \beta \psi_s, \frac{n_p}{p_p} \right). \tag{15}
\]

Similarly, by Gauss' law the space charge per unit area required to produce this field is

\[
Q_s = \varepsilon_s \varepsilon \varepsilon_s = \pm \frac{2 \varepsilon_i kT}{q L_D} F \left( \beta \psi_s, \frac{n_p}{p_p} \right). \tag{16}
\]
To determine the change in hole density, $\Delta p$, and electron density, $\Delta n$, per unit area when the $\psi$ at the surface is shifted from zero to a final value $\psi_s$, it is necessary to evaluate the following expressions:

$$\Delta p = p_m \int_0^\infty (e^{-\beta \psi} - 1) \, dx$$

$$= \frac{q_p n_m L_D}{2kT} \int_0^\infty \frac{(e^{-\beta \psi} - 1)}{F(\beta \psi, n_m, p_m)} \, d\psi \quad (\text{cm}^{-2}). \quad (17)$$

$$\Delta n = n_m \int_0^\infty (e^{\beta \psi} - 1) \, dx$$

$$= \frac{q_n n_m L_D}{2kT} \int_0^\infty \frac{(e^{\beta \psi} - 1)}{F(\beta \psi, n_m, p_m)} \, d\psi \quad (\text{cm}^{-2}). \quad (18)$$

A typical variation of the space-charge density $Q_s$ as a function of the surface potential $\psi_s$ is shown in Fig. 5 for a $p$-type silicon with $N_A = 4 \times 10^{15}$ cm$^{-3}$ at room temperature. We note that for negative $\psi_s$, $Q_s$ is positive and corresponds to the accumulation region. The function $F$ is dominated by the first term in Eq. (13), i.e., $Q_s \sim \exp(q/2kT)$ for $\psi_s = 0$. For $\psi_s > 0$, $Q_s$ is negative and we have the depletion case. The function $F$ is now dominated by the second term, i.e., $Q_s \sim \sqrt{\psi_s}$. For $\psi_s \gg \psi_B$, we have the inversion case with the function $F$ dominated by the fourth term, i.e., $Q_s \sim -\exp(q/2kT)$. We also note that for the depletion case, $\psi_{(\text{inv})} \approx 2\psi_B \approx \frac{2kT}{q} \ln \left( \frac{N_A}{n_i} \right)$.

(19)

The differential capacitance of the semiconductor space-charge region is given by

$$C_D = \frac{Q_s}{\psi_s} = \frac{e}{L_D} \left[ \frac{1 - e^{-\beta \psi_s}}{F(\beta \psi_s, n_m, p_m)} \right] \quad \text{farad/cm}^2. \quad (20)$$

At flat-band condition, i.e., $\psi_s = 0$, $C_D$ can be obtained by expanding the exponential terms into series, and we obtain

$$C_D(\text{flat-band}) = \sqrt{2e_i/L_D} \quad \text{farad/cm}^2. \quad (21)$$

Figure 5 shows a variation of the space-charge density $Q_s$ as a function of the surface potential $\psi_s$ for a $p$-type silicon with $N_A = 4 \times 10^{15}$ cm$^{-3}$ at room temperature. $\psi_s$ is the potential difference between the Fermi level and the intrinsic level of the bulk. (After Garrett and Brattain, Ref. 10.)

### (2) Ideal MIS Curves

Figure 6(a) shows the band diagram of an ideal MIS structure with the band bending of the semiconductor identical to that shown in Fig. 4. The charge distribution is shown in Fig. 6(b). For charge neutrality of the system, it is required that

$$Q_M = Q_e + qN_A W = Q_s \quad (22)$$

where $Q_M$ is charges per unit area on the metal, $Q_e$ is the electrons per unit area in the inversion region, $qN_A W$ is the ionized acceptors per unit area in the space-charge region with space-charge width $W$, and $Q_s$ is the total charge per unit area in the semiconductor. The electric field and the potential as obtained by first and second integrations of Poisson's equation are shown in Fig. 6(c) and 6(d) respectively.
\[ \mathcal{E}(x) = \pm \left[ \frac{2kT}{e} \int_0^{\psi(x)} \rho(\psi') \, d\psi' \right]^{1/2}. \]

Here, \( \Psi = \frac{\Psi}{\text{class}} \)

where \( \Psi = \beta \psi \)

\( \beta = \frac{qV}{kT} \) and \( \psi \) is band bending

This is the central relation in the theory of both accumulation and depletion layers at semiconductor junctions with uniform doping, applicable whenever the space charge density \( \rho \) depends on the position \( x \) only implicitly, through its dependence on the band bending parameter \( \psi \), with no explicit \( x \)-dependence. If we insert our particular form of \( \rho(\psi) \),

\[ \rho(\psi) = qN_D [1 - \exp(-\psi)] , \quad (\ast 3-18) \]

we find easily

\[ \mathcal{E}(x) = \pm \left[ \frac{2kT N_D}{e} \right]^{1/2} \left[ \psi(x) + \exp[-\psi(x)] - 1 \right]^{1/2}. \quad (\ast 3-19) \]

The sign of \( \mathcal{E} \) is not specified by (\ast 3-17) and (\ast 3-19). One finds easily that the following rule holds: If the semiconductor extends to the right of the interface, the sign of \( \mathcal{E} \) is opposite to that of \( \psi \), while for a semiconductor extending to the left, \( \mathcal{E} \) and \( \psi \) have the same sign.

The relation (\ast 3-19) is the central relation in the theory of accumulation and depletion layers for the simplest case of common interest: that all impurity levels are ionized, that the semiconductor is non-degenerate, and that minority carriers may be ignored. In the remainder of this Section we treat the limiting case of strongly depleted depletion layers, deferring the more general case, including accumulation layers until later.

### 3.2) Strong-Depletion Limit:
The Depletion Approximation Recovered

Inside the depletion layer of a p-n junction, the amount of band bending usually becomes sufficiently large compared to \( kT \) that the exponential term in (\ast 3-19) may be neglected, in which case (\ast 3-19) reduces to

\[ \left| \mathcal{E}(x) \right| = \left[ \frac{2qN_D}{e} \cdot \frac{kT}{q} \cdot [\psi(x) - 1] \right]^{1/2}. \quad (\ast 3-20) \]
Suppose now that we apply this result to the p-n interface plane $x = 0$, of an abrupt p-n junction, or to the metal-semiconductor interface of a Schottky barrier. In that case, $\psi(0) - kT/q$ is simply that part $V_n$ of the built-in potential that occurs within the n-type semiconductor, and we may write

$$E_{\text{max}} = \left( \frac{2qN_D}{\varepsilon} \left[ V_n - kT \right] \right)^{1/2}.$$  

(3.21)

If we had calculated $|E_{\text{max}}|$ using the depletion approximation, we would have obtained a result that would have differed from (3.21) by the absence of the term $-kT/q$. This term is the Gummel end correction on the n-type side, and our procedure represents a derivation of the earlier claim that the magnitude of this correction is just $kT/q$. Note that this correction has this simple form only if the band bending is sufficiently large compared to $kT$ that the exponential term in (3.19) is negligible.

The field $|E_{\text{max}}|$ is the same as if the electron associated with the donors had been depleted over a distance

$$w_n = \frac{eE_{\text{max}}}{qN_D} = \left( \frac{2e}{qN_D} \left[ V_n - kT \right] \right)^{1/2}.$$  

(3.22)

with no electron depletion outside this range. This, then, is the effective depletion depth in the sense of the depletion approximation. Note again the occurrence of the Gummel correction term.

### 3.3) The p-type Case

An essentially identical treatment can be given for the band bending inside a p-type semiconductor. A similar band bending parameter $\psi_p$ is introduced on the p-type side, defined according to

$$\psi_p(x) = -[E_V(x) - E_V(\infty)]/kT,$$  

(3.23)

which differs from (3.6) by a sign reversal. This sign convention is chosen so that a positive value of $\psi_p$ corresponds to depletion of holes, just as a positive value of $\psi_n$ corresponds to the depletion of electrons. With this choice, the p-side equivalent of the central relation (3.19) assumes the form completely analogous to (3.19)

$$E(x) = \pm \left( \frac{2kTN_A}{\varepsilon} \right)^{1/2} \left\{ \psi_p(x) + \exp[-\psi_p(x)] - 1 \right\}^{1/2}.$$  

(3.24)
without a change of signs inside this relation. Experience has shown that this sign
convention simplifies the treatment. Note, however, that this choice also means that,
whereas \( \psi_n > 0 \) corresponds to upward band bending, \( \psi_p > 0 \) corresponds to
downward band bending.

The equivalents of \((3-21)\) and \((3-22)\) for the case of deep depletion of holes now
become

\[
E_{\text{max}} = \left( \frac{2qN_A}{\varepsilon} \left[ V_p - kT \right] \right)^{1/2}. \tag{3-25}
\]

\[
w_p = \frac{\varepsilon E_{\text{max}}}{qN_A} = \left( \frac{2\varepsilon}{qN_A} \left[ V_p - kT \right] \right)^{1/2}. \tag{3-26}
\]

### 3.4 Application to p-n Homojunctions and Heterojunctions

We now combine \((3-21)\) and \((3-25)\) to treat the case of an abrupt p-n junction,
including that of a p-n heterojunction. At the interface we must have

\[
D = \varepsilon E = \text{continuous.} \tag{3-27}
\]

The dielectric permittivities of the two semiconductors need not be the same, hence
we cannot simply assume continuity of the electric field \(E\). We write \((3-27)\) in the
form

\[
(\varepsilon E_{\text{max}})_n^2 = (\varepsilon E_{\text{max}})_p^2 = D_{\text{max}}^2, \tag{3-28}
\]

which implies, from \((3-21)\) and \((3-25)\),

\[
\varepsilon_n N_D [V_n - kT/q] = \varepsilon_p N_A [V_p - kT/q]. \tag{3-29}
\]

One sees easily that

\[
V_n - kT/q = (V_p + V_n - 2kT/q) \cdot \frac{\varepsilon_p N_A / (\varepsilon_p N_A + \varepsilon_n N_D)}{\varepsilon_n N_D / (\varepsilon_p N_A + \varepsilon_n N_D)}, \tag{3-30a}
\]

\[
V_p - kT/q = (V_p + V_n - 2kT/q) \cdot \frac{\varepsilon_p N_A / (\varepsilon_p N_A + \varepsilon_n N_D)}{\varepsilon_n N_D / (\varepsilon_p N_A + \varepsilon_n N_D)}. \tag{3-30b}
\]

These results may be inserted into the earlier relations \((3-21)\), \((3-22)\), \((3-25)\), and
\((3-26)\) for the fields and the effective depletion depths on both sides of the junction.
Evidently, all of these quantities have voltage dependences of the form \((V_p + V_n -
2kT/q)^{1/2}\). This dependence suggests that we generalize the definition of the built-in
current \(V_{bi}\) to heterojunctions by defining it as the sum of the two band bendings on
3.1) Abrupt P-N Junctions

Charge and Field Distributions

Our point of departure is a plot of space charge density and electric field through the space charge layer of an abrupt p-n junction, shown on the next page. At the top of the figure we show the various contributions to the space charge, and crosshatched the net space charge. Overall electrical neutrality requires that the two oppositely crosshatched areas are equal. Below the space charge plot we show the resulting electric field, obtained from the Poisson Equation.

In the regions of full depletion, the space charge is uniform; hence the slope of the \( \mathcal{E} \)-vs-\( x \) curve is constant.

\[
ed \mathcal{E}/dx = \rho = qN_D \quad \text{(n-side);} \quad \text{or } = qN_A \quad \text{(p-side)}. \tag{2-1}
\]

If this slope is extrapolated to \( \mathcal{E} = 0 \), the extrapolation intercepts the x-axis at \( x = -w_n \) and \( x = +w_p \). We call these intercept points the effective depletion layer edges.

From Gauss' Law, the magnitude \( \mathcal{E}_{\text{max}} \) of the maximal field is then

\[
\mathcal{E}_{\text{max}} = qN_D w_n / \epsilon = qN_A w_p / \epsilon . \tag{2-2}
\]

From electrical neutrality:

\[
N_D w_n = N_A w_p . \tag{2-3}
\]

These lead to

\[
w_n = w N_A / (N_A + N_D) ; \quad w_p = w N_D / (N_A + N_D) . \tag{2-4a,b}
\]

where

\[
w = w_n + w_p \tag{2-4c}
\]

is the overall effective depletion layer width. If (2-4a,b) are inserted into (2-2):

\[
\mathcal{E}_{\text{max}} = (q/\epsilon) w N_A N_D / (N_A + N_D) ; \tag{2-5}
\]
The electrostatic potential drops on the two sides are simply the areas under the $E$-vs-$x$ curve:

$$\Delta \phi_n = E_{\text{max}} \frac{w_n}{2} + V_{\text{Gn}}$$

$$\Delta \phi_p = E_{\text{max}} \frac{w_p}{2} + V_{\text{Gp}}$$

Fig. 1: Charge and field distribution for an abrupt p-n junction

Copyright © 1985, 1988 by H. Kroemer
Here \( V_{Gn} \) and \( V_{Gp} \) are the so-called Gummel end corrections, that is the contributions of the tails in the \( \mathcal{E} \)-vs.-\( x \) distributions in Fig. 1 to the potential drops. The exact calculation of these corrections is somewhat more difficult, but the final result of that calculation is extremely simple: We shall see later that under most circumstances the corrections are simply given by

\[
V_{Gn} = V_{Gp} = kT/q ,
\]

(a small but not always negligible value.

The total electrostatic potential drop across the space charge region is clearly

\[
\Delta\phi_n + \Delta\phi_p = \frac{1}{2} \mathcal{E}_{\text{max}} \cdot w + 2kT/q = \frac{q}{2\varepsilon} \frac{N_A N_D}{N_A + N_D} \cdot w^2 + 2kT/q .
\]

In the absence of an external bias this is of course the built-in potential. If a (reverse) bias voltage \( V_r \) is present, we have

\[
\Delta\phi_n + \Delta\phi_p = V_{bi} + V_r .
\]

If we insert (2-9) into (2-8), the resulting dual equation may be solved for the total depletion layer width and for the maximum electric field, both as functions of applied bias and doping levels, leading to the relations

\[
w = \left[ \frac{2\varepsilon}{q} \frac{N_A + N_D}{N_A N_D} \left( V_r + V_{bi} - \frac{2kT}{q} \right) \right]^{1/2} .
\]

\[
\mathcal{E}_{\text{max}} = \left[ \frac{2q}{\varepsilon} \frac{N_A N_D}{N_A + N_D} \left( V_r + V_{bi} - \frac{2kT}{q} \right) \right]^{1/2} .
\]

These relations are central to the theory of p-n junctions.

**Space Charge Layer Capacitance**

The effective depletion layer width is easily measured, by measuring the capacitance per unit area, \( C \), of a reverse-biased p-n junction. Any change of the bias voltage with time causes a displacement current to flow, with the current density

\[
J_{\text{Displ}} = \frac{d\mathcal{E}_{\text{max}}}{dt} = \frac{d\mathcal{E}_{\text{max}}}{dV_r} \frac{dV_r}{dt} .
\]
But from the definition of the capacitance, we also have

\[ J_{\text{Displ}} = C \frac{dV}{dt} , \quad (2.13) \]

where \( C \) is the capacitance per unit area. Evidently there is a space charge layer capacitance associated with a p-n junction, given by

\[ C = \frac{dQ_{\text{max}}}{dV_r} . \quad (2.14) \]

By executing the differentiation of (2.11), and comparing the result with (2.10), one finds immediately

\[ C = \frac{\varepsilon_r}{w} , \quad (2.15) \]

the same as for an ordinary parallel-plate capacitor.

According to (2.10) and (2.11), plot of \( 1/C^2 \) vs. \( V_r \) follows a straight line (Fig.2). Such plots are widely used in the study of p-n junctions. If extrapolated backwards, the extrapolation intercepts the voltage axis at the point \( V_r = -V_{\text{int}} \), where the intercept voltage \( V_{\text{int}} \) is given by

\[ V_{\text{int}} = V_{\text{bi}} - 2kT/q \quad (2.16) \]

Fig. 2. Plot of \( 1/C^2 \) vs. reverse bias for an abrupt p-n junction

Furthermore, the slope of such a plot is related to the doping levels according to
\[ \frac{d}{dV_T} \frac{1}{C^2} = \frac{w^2}{\varepsilon^2} = \frac{2e}{q\varepsilon} \left( \frac{1}{N_D} + \frac{1}{N_A} \right). \]  

(2-17)

A relation that forms the basis of determining doping levels by capacitance-vs.-voltage measurements, referred to as C-V Profiling. Often, one of the two doping levels is much larger than the other. In this case, the larger of the two doping levels drops out of (2-17), permitting the determination of the lower doping level directly.

The maximum electric field inside the junction is important in its own right, because it dominates the electrical breakdown behavior of the junction.

**Typical Numbers**

For typical p-n junctions, the products \(N_D w_n\) and \(N_A w_p\) tend to be of the order 10\(^{12}\) cm\(^{-2}\), a value that would be achieved, for example, by a doping level of \(N = 10^{16}\) cm\(^{-3}\) and a depletion thickness \(w = 10^{-4}\) cm = 1\(\mu\)m, both fairly typical values. The electric field that can be supported by such a sheet charge is

\[ E = \frac{qNw}{\varepsilon}. \]  

(2-18)

Semiconductor permittivities tend to be somewhat above 10\(\varepsilon_0\). Assuming 10\(\varepsilon_0\) as a reference value, a reference sheet charge of 10\(^{12}\) cm\(^{-2}\) yields \(E = 1.81 \times 10^5\) V/cm, a useful number to remember. For Si (\(\varepsilon_r = 11.9\)) one finds 1.52 \(\times 10^5\) V/cm, for GaAs (\(\varepsilon_r = 13.1\)): 1.38 \(\times 10^5\) V/cm. A useful general numerical scaling relation is

\[ E = 1.81 \times 10^5 \text{V/cm} \cdot \frac{Nw}{10^{12} \text{cm}^{-2}} \cdot \frac{10\varepsilon_0}{\varepsilon} \]  

(2-19)

If the electric field on one side of such a layer is zero, the potential difference due to band bending is

\[ V = \frac{Ew}{2} = \frac{qNw}{2\varepsilon}. \]  

(2-20)

For the above reference values one finds 9.05 V.

**Abrupt vs. Augmented Depletion Approximation**

In the elementary treatments of the p-n junction found in many papers and textbooks (including almost all undergraduate texts) the Gummel end corrections are usually neglected. Those treatments are usually based on what is called the (abrupt) depletion approximation: The actual electron and hole distributions are approximated by abrupt distributions, in which electrons and holes are assumed to be fully depleted from the inside the space charge layer, and not depleted at all on the outside, with an
abrupt transition at the (effective) depletion layer edges. Such an approximation has zero end corrections. Inasmuch as the end corrections are indeed small, the depletion approximation is a good approximation for large reverse biases. On the other hand, the Gummel end corrections are so simple that it is just as easy to include them as to omit them. There is hence no justification for even a small loss in accuracy, and under many conditions even the small correction is by no means negligible, especially in the treatment of forward-biased junctions, and of heterojunctions. We shall therefore routinely include the Gummel end corrections without specifically drawing attention to them, and we shall in the future understand the term depletion approximation always as including the Gummel correction, unless specifically stated otherwise. When it is necessary to distinguish our treatment from that of the more conventional (unmodified) depletion approximation, we shall refer to our treatment including the Gummel end corrections as to the augmented depletion approximation. The results of the conventional treatment found in the literature are easily converted by recalling that the actual bias voltage is always larger by a 1kT/q contribution from each end, and by subtracting 2kT/q from the actual bias voltage (or from the built-in potential) before applying the conventional depletion approximation.

2.2) P-I-N Junctions

The simplest quasi-abrupt junction is one where a thin layer (thickness w₁) of undoped ("intrinsic") semiconductor has been inserted between the otherwise uniformly doped n- and p-regions. Usually, there will be some residual doping left, but we ignore this here.

2.3) Moment Theorem for General "Pseudo-Abrupt" Junctions

One must often deal with p-n junctions in which the transition from n-type to p-type doping is not abrupt, but with the two doping levels becoming eventually becoming constant far enough away from the interface, and where the two depletion layer edges fall within the regions of constant doping, at least above a certain minimum voltage. Above that voltage, such pseudo-abrupt junctions behave very similarly to truly abrupt junctions.

Let

\[ N(x) = N_D^+(x) - N_A^+(x) \]  \hspace{1cm} (2-21)

be the non-abrupt net ionized doping concentration (donors minus acceptors) of the pseudo-abrupt p-n junction, and assume that N(x) has already reached its asymptotic
limits at the two edges of the space charge layer. We compare this distribution with an associated abrupt distribution that has the same asymptotic limits,

\[ N_0(x) = \begin{cases} \text{N}(+\infty) & \text{for } x > 0 \\ \text{N}(-\infty) & \text{for } x < 0 \end{cases}, \tag{2.22} \]

where we choose the origin of the x-axis such that

\[ \int_{-\infty}^{+\infty} [N(x) - N_0(x)] \, dx = 0, \tag{2.23} \]

which means that the net charge difference between the true distribution and the associated abrupt distribution is zero.

If now \( x_n \) and \( x_p \) are the two depletion layer edges for a given voltage \( V_0 \) applied to the abrupt p-n junction, the same depletion layer edges will occur for the original junction, but at a different voltage. By integrating the Poisson equation from \( x_n \) to \( x_p \), one finds, after some manipulation, that the electrostatic potential difference is given by

\[ \Delta \phi = [\phi(x_p) - \phi(x_n)] - [\phi_0(x_p) - \phi_0(x_n)] = \frac{q}{\varepsilon} \int_{x_n}^{x_p} [N(x) - N_0(x)] \, dx. \tag{2.24} \]

Evidently, the pseudo-abrupt junction requires a different applied bias voltage to have the same depletion layer edges. Care is in order with respect to the signs. As defined in (2.24), \( \Delta \phi \) is a positive quantity if the pseudo-abrupt junction is graded relative to the abrupt case, regardless of whether the n-type side is on the left or on the right (Note that the integration runs from the right to the left in the latter case). One finds easily that a positive value of \( \Delta \phi \) corresponds to a bias voltage shift in the forward direction, that is, to a reduction of the reverse bias. Hence, a positive \( \Delta \phi \) corresponds to a positive shift \( \Delta V_{\text{int}} \) of the intercept voltage in a \( 1/C^2 \)- vs. \( V_r \) plot, as in Fig. 2, namely,

\[ \Delta V_{\text{int}} = \Delta \phi = \frac{q}{\varepsilon} \int_{x_n}^{x_p} [N(x) - N_0(x)] \, dx. \tag{2.25} \]
Note that (2-25) simply represents the electrostatic potential difference across the electrostatic dipole formed between the actual donor distribution and the associated abrupt distribution. Hence, (2-25) is often referred to as the Moment Theorem.

Note that for junctions that are more gradual than an abrupt junction, the intercept voltage (always considered a positive voltage) always increases. P-n junctions for which the doping distribution is such as to lead to a negative value of $\Delta \phi$ are called hyper-abrupt junctions.

2.4) Linearly Graded Junction

For many p-n junctions the doping levels at the edges of the depletion layer are not constant. The depletion approximation remains a good approximation in most cases, as discussed in many textbooks for the simplest case, a linearly graded junction. The most important complication arises from the fact that the diffusion potential is no longer a constant, but depends on the width of the depletion region, and hence on the bias voltage. This leads to transcendental equations relating the various quantities of interest, which can no longer be solved in closed form. However, very accurate numerical values are still obtainable by numerical iteration, which tends to converge very rapidly.
Shockley Read Hall Statistics

Semiconductor behavior is determined primarily by controlled impurities. Shallow \( \Rightarrow \) dopants, Deep \( \Rightarrow \) Traps. In either case the occupancy of all states is determined by the occupancy function whether the states are in the bands or in the gap.

Exchanges with the conduction band are dealt as electron capture & emission, whereas exchanges with the VB are dealt as hole capture and emission. In the figure above the arrows denote electron transitions.

Let \( N_t \) = Concentration of trap centers
\( f \cdot N_t \) = Concentration of occupied traps
\( (1-f) \cdot N_t \) = Concentration of empty traps
where \( f \) = Occupancy function

In equilibrium
\[
f = \frac{1}{1 + e^{(E_t - E_f)/kT}}
\]
where \( E_t \) = Trap energy
\( E_f \) = Fermi energy

In non-equilibrium a quasi-fermi level should be used.

Each of these processes has a rate, \( r \).

\[
r_a \propto n \cdot N_t (1-f)
\]

Conc. of available electrons \( \text{(Cm}^{-3}\text{)} \)
Conc. of empty traps \( \text{(Cm}^{-3}\text{)} \)
What is the proportionality constant? To calculate it, recognize that electrons must be in the vicinity of the trap to be captured. Call that region $\sigma_n \left( \text{cm}^2 \right)$ a capture cross section.

The numbers of electrons that sweep past the trap per second are contained in the volume shown below:

$$V = \sigma_n \text{cm}^2 \cdot v_{th} \left( \text{cm/s} \right) = \text{cm}^3 / \text{s}$$

where $v_{th}$ is the thermal velocity of the electron. Consider an electron $v_{th} \text{cms}$ away from the trap position $x_o$. After 1 sec the electron will be at $x$, and therefore in the capture cross section of the trap. Any electron $v_{th} + \Delta L_2$ away will after 1 sec still be $\Delta L_2$ away (case 2) from $x_o$ and hence not be captured. All electrons closer than $v_{th}$
away (as for case 3 of the electron \( \Delta L_3 \) closers) would have intersected the capture cross-section and be captured. Hence all electrons in the volume

\[
V = \sigma_n \cdot v_{th}
\]

will be captured /sec by available empty traps

OR all # of electrons available to be captured /sec = \( n \cdot \frac{\sigma_n \cdot v_{th}}{\text{cm}^3/\text{s}} \)

The # of available empty traps \( \text{cm}^{-3} \)

is \( N_t (1 - f) \)

\[
\therefore r_a = v_{th} \cdot \sigma_n \cdot n \cdot \frac{N_t (1 - f)}{\text{sec}^{-1}} \text{ cm}^{-3}/\text{s}
\]

# of \( \text{cm}^3/\text{s} \)

OR the proportionality constant is \( \sigma v_{th} \) (for the rate of electron capture)

\[
v_{th} = \sqrt{\frac{2v_{th}^2}{m^*}} = \sqrt{2 \cdot \frac{3kT}{2m^*}}
\]

\[
v_{th} = \sqrt{\frac{3kT}{m^*}} = 10^7 \text{ cm/s}
\]

For the emission process, b,

\[
r_b = e_n \cdot \frac{N_t \cdot f}{\text{cm}^3} \]

Emission rate for electrons

Available trapped electrons cm\(^3\).

the capture rate for holes through c, will be analogous to a,

\[
r_c = v_{th} \cdot \sigma \cdot p \cdot \frac{N_t \cdot f}{\text{cm}^3} \]

Concentration of traps that can capture holes ⇒ concentration of filled traps
Finally, the emission of holes has a rate

\[ r_d = e_p \cdot \frac{N_t (1 - f)}{} \]

Emission rate of holes
Conc. of traps with no electrons empty.

Next step is to determine the emission probabilities \( e_n \) and \( e_p \). In general this is a very difficult problem since, \( f \), is known only in equilibrium. So first consider equilibrium and evaluate \( e_n \) and \( e_p \).

In equilibrium transistor rates into and out of the CB must be equal or \( r_a = r_b \)

Inserting \( n = N_c e^{-(E_c - E_F)/kT} = n_i e^{(E_F - E_i)/kT} \) into \( r_a = r_b \)

\[ \Rightarrow v_{th} \sigma_n n_i \mathcal{N}_t (1 - f) = e_n \mathcal{X}_t f \]

\[ e_n = v_{th} \sigma_n n_i e^{(E_F - E_i)/kT} \cdot \frac{(1 - f (E_t))}{f(E_t)} \]

\[ f E_t = \frac{1}{1 + e^{-(E_t - E_F)/kT}} \quad ; \quad 1 - f (E_t) = \frac{e^{(E_t - E_F)/kT}}{1 + e^{(E_t - E_F)/kT}} \]

or \[ 1 - f (E_t) = e^{(E_t - E_F)/kT} \]

\[ \therefore e_n = v_{th} \sigma_n n_i \cdot e^{(E_F - E_i)/kT} \cdot e^{(E_t - E_F)/kT} \]

\[ e_n = v_{th} \sigma_n n_i e^{(E_t - E_F)/kT} \]

or

\[ e_n = v_{th} \sigma_n N_C e^{-(E_c - E_t)/kT} \]

OR emission probability of electrons into the conduction band rises exponentially as the trap gets closer to \( E_c \) (which we expect in intuitively).
From \( r_c = r_d \) and \( p = n_i e^{(E_e - E_i)/kT} \)

\[
e_p = v_{th} \sigma_p N_v e^{-\frac{(E_p - E_v)}{kT}} = v_{th} \sigma_p n_i e^{\frac{(E_l - E_i)}{kT}}
\]

In non-equilibrium (the case of most interest) \( f \) is unknown. Assume that non-equilibrium is generated by optical excitation resulting in a generation rate.

\[
\begin{align*}
G_i & \quad \uparrow \quad r_a \quad \downarrow \quad r_b \quad \uparrow \quad r_c \quad \downarrow \quad r_d \\
G_i & \quad \uparrow \quad r_a \quad \downarrow \quad r_b \quad \uparrow \quad r_c \quad \downarrow \quad r_d
\end{align*}
\]

\( r_a \) thru \( r_d \) continue to take place.

In steady state the concentration electrons \( n_n \) and holes \( p_n \) (assuming \( n - \) type s/c) is not a \( f(t) \).

\[
\frac{dn_n}{dt} = G_L - (r_a - r_b) = 0
\]

\[
\frac{dp_n}{dt} = G_L - (r_c - r_d) = 0
\]

\[
\therefore \quad r_a - r_b = r_c - r_d
\]

Since we are in non-equilibrium, \( f(E_f) \), the distribution function for the traps has to be calculated from the above equation. Substituting for \( r_a \) thru \( r_d \)

\[
f(E_f) = \frac{\sigma_n n + \sigma_p n_i e^{\frac{(E_l - E_i)}{kT}}}{\sigma_n \left[ n + n_i e^{\frac{(E_l - E_i)}{kT}} \right] + \sigma_p \left[ p + n_i e^{\frac{(E_l - E_i)}{kT}} \right]}
\]

Re-substituting to find a net rate of recombination

\[
U = r_a - r_b = r_c - r_d
\]
\[ U = \frac{\sigma_p \sigma_n v_{th} N_i (pn - n_i^2)}{\sigma_n \left[ n + n_i e^{(E_i - E_f)/kT} \right] + \sigma_p \left[ p + n_i e^{(E_f - E_i)/kT} \right]} \]

\[ U = \sigma v_{th} N_i \cdot \frac{pn - n_i^2}{n + p + 2n_i \cosh \left( \frac{E_i - E_f}{kT} \right)} \quad \text{for} \quad \sigma_n = \sigma_p = \sigma \]

\[ U = \frac{1}{\tau} \cdot \frac{pn - n_i^2}{n + p + 2n_i} \quad \text{for} \quad \sigma_n = \sigma_p = \sigma \\
\& \\
E_i = E_f \quad \text{Max. rate} \]

\( pn - n_i^2 = \) Driving force for recombination
\( n + p + 2n_i = \) Resistance to recombination minimized when \( n + p \) is minimized.

\( \Leftrightarrow \)

Low-level injection. Assume \( n_n \gg fn \)

Also \( n_n \gg n_i e^{(E_i - E_f)/kT} \)

As \( E_T \sim E_i \) for efficient recombination

\[ \therefore U = \frac{\sigma_p \sigma_n v_{th} N_i}{\sigma_n n_n} \left[ n_n - P_n - n_i^2 \right] \]

\[ = \sigma_p v_{th} N_i \left[ p_n - n_i^2 / n_n \right] \]

\[ = \sigma_p v_{th} N_i \left[ p_n - p_{no} \right] \]

\[ U = \frac{\Delta p_n}{\tau_p} = \frac{1}{\tau_p} = \sigma_p v_{th} N_i \]

Minority carrier lifetime

Rate-limiting step is the capture of the minority carrier.
This is also achieved at by recognizing the hole capture rate \( r_c \) is the dominant step.

\[ r_a, r_d \] are proportional to \( (1 - f) \). If \( f \to 1 \) \( r_a, r_d \to 0 \)

\[ r_b = e_n N_t f \quad \text{and} \quad r_c = v_{th} \sigma_p p N_t f \]

\[ f (E_T) \to 1 \] in an n-type s/c.

Typical values \( \sigma = 10^{-13} \text{cm}^2 \) Coulombic

\( 10^{-18} \text{cm}^2 \) Neutral

\( \sigma = 10^{-15} - 10^{-16} \text{cm}^2 \)

Normally

Generation occurs when \( n_i^2 \gg pn \)

\[ U = -\frac{\sigma_p \sigma_n v_{th} N_t n_i^2}{\sigma_n \left[ \mu + n_i e^{E_i - E_L} / kT \right] + \sigma_p \left[ \nu + n_i e^{E_i - E_L} / kT \right]} \]

generation

\[ = -\frac{\sigma_p \sigma_n v_{th} N_t n_i}{\sigma_n e^{(E_i - E_L)/kT} + \sigma_p n_i e^{(E_i - E_L)/kT} / kT} \]

\[ = -\frac{n_i}{2\tau_o} \]

when \( \sigma_n = \sigma_p = \sigma \)

\[ U = -\frac{\sigma v_{th} N_t n_i}{2\cosh \left[ (E_i - E_L) / kT \right]} \]

Generation peaks @ mid-gap.

\[ \therefore U = -\frac{n_i}{2\tau} \]

When \( E_i \to E_L \) the lifetime

\[ \tau = \frac{1}{\sigma v_{th} N_t} \]