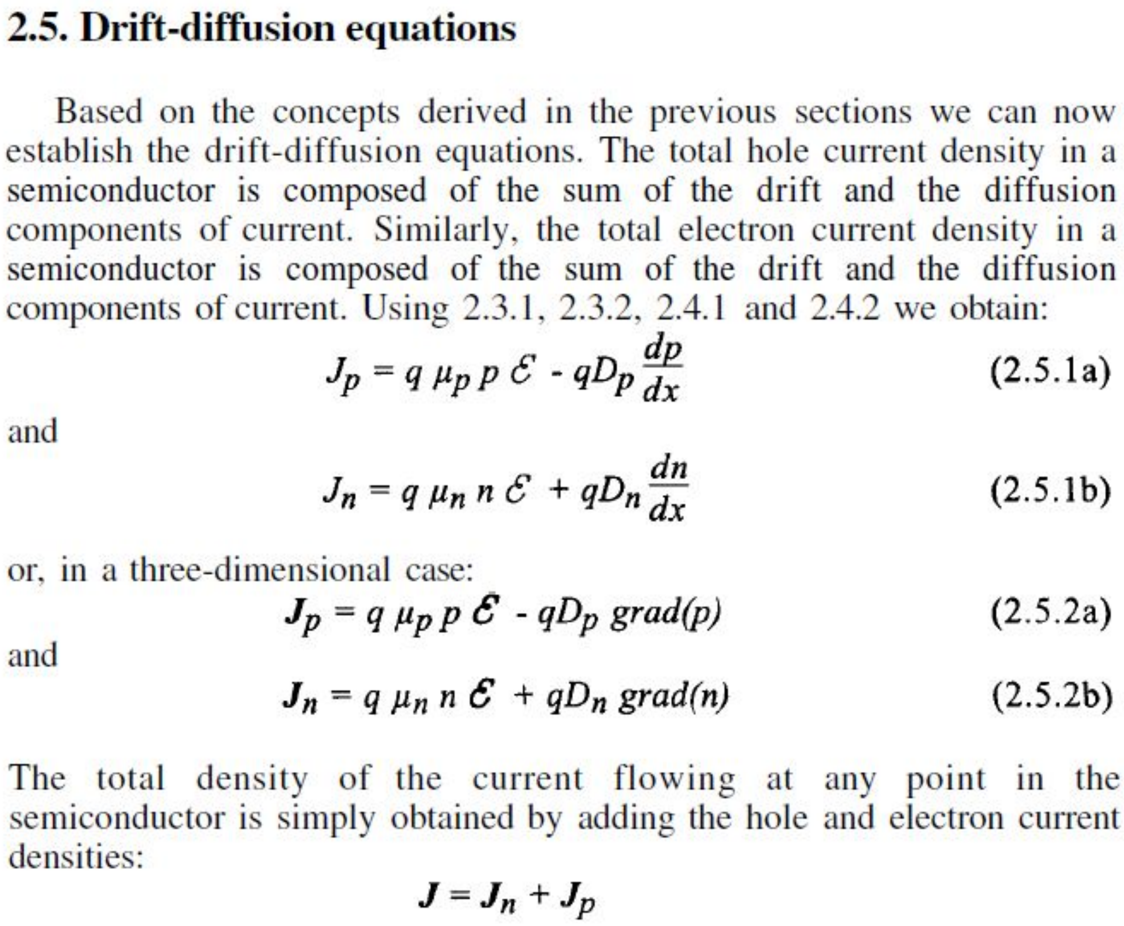
# Metal-Insulator-Semiconductor and Metal-Insulator-Metal Structures. Part II. Heterogeneous Structures. Diodes

*Alexander Gabovich*, **KPI**,

Lecture 9

# Transport properties of semiconductors



## 2.5. Drift-diffusion equations

Based on the concepts derived in the previous sections we can now establish the drift-diffusion equations. The total hole current density in a semiconductor is composed of the sum of the drift and the diffusion components of current. Similarly, the total electron current density in a semiconductor is composed of the sum of the drift and the diffusion components of current. Using 2.3.1, 2.3.2, 2.4.1 and 2.4.2 we obtain:

$$J_p = q\,\mu_p\,p\,\mathcal{E} - qD_p\frac{dp}{dx} \qquad (2.5.1a)$$

and

$$J_n = q\,\mu_n\,n\,\mathcal{E} + qD_n\frac{dn}{dx} \qquad (2.5.1b)$$

or, in a three-dimensional case:

$$\boldsymbol{J_p} = q\,\mu_p\,p\,\boldsymbol{\mathcal{E}} - qD_p\,grad(p) \qquad (2.5.2a)$$

and

$$\boldsymbol{J_n} = q\,\mu_n\,n\,\boldsymbol{\mathcal{E}} + qD_n\,grad(n) \qquad (2.5.2b)$$

The total density of the current flowing at any point in the semiconductor is simply obtained by adding the hole and electron current densities:

$$\boldsymbol{J} = \boldsymbol{J_n} + \boldsymbol{J_p}$$

# Transport properties of semiconductors

## 2.5.1. Einstein relationships

The mobility and diffusion coefficient in a semiconductor are related to each other. This relationship is derived in the following section. Consider a piece of semiconductor material with a non-uniform doping concentration. Let the doping atoms be arsenic in silicon and for the sake of simplicity we will consider a one-dimensional case. The doping impurities are N-type and their concentration is $N_d(x)$, as shown in Figure 2.6. Assuming all doping impurities are ionized, we have that $n(x) = N_d(x)$. The presence of an electron concentration gradient gives rise to an electron diffusion current. The electrons diffusing to the left "leave behind" positively charged arsenic atoms. These atoms occupy substitutional sites in the crystal lattice, and unlike electrons, cannot move. Because of the increased number of electrons in the left-hand part of the sample and the presence of positive charges in the right-hand part an internal electric field develops locally. This electric field tends to "recall" the electrons towards their place of origin. This electric field and the associated potential drop are noted $\mathcal{E}_o(x) = -d\Phi_o(x)/dx$, where the subscript zero implies an internal or "built-in" field under thermal equilibrium.

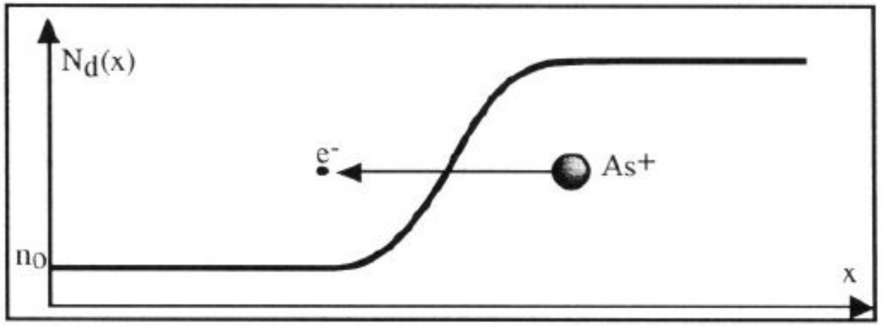
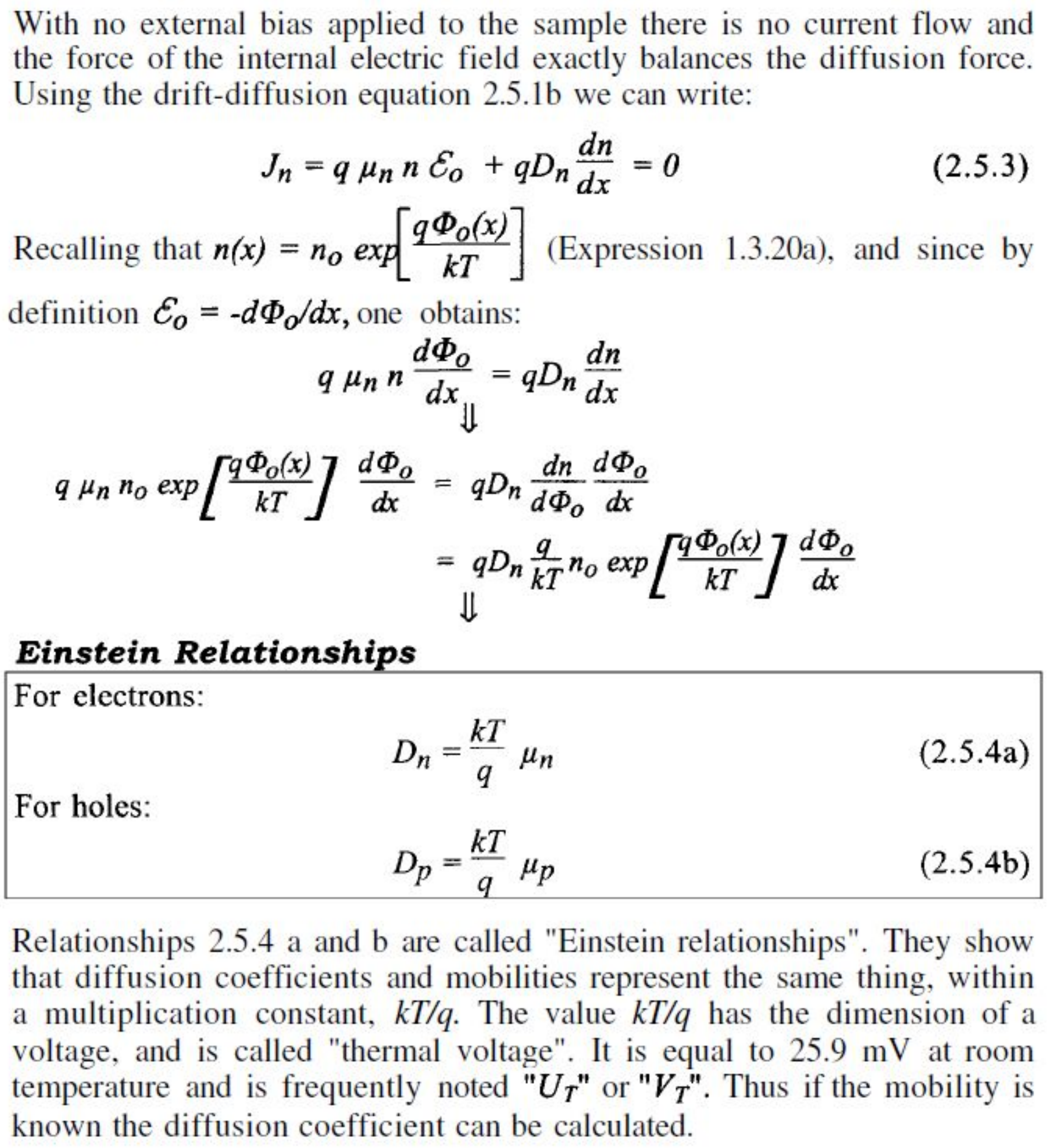


**Figure 2.6**: Non-uniform doping profile in an N-type silicon sample.

# Transport properties of semiconductors



With no external bias applied to the sample there is no current flow and the force of the internal electric field exactly balances the diffusion force. Using the drift-diffusion equation 2.5.1b we can write:

$$J_n = q\,\mu_n\,n\,\mathcal{E}_o + qD_n\frac{dn}{dx} = 0 \qquad (2.5.3)$$

Recalling that $n(x) = n_o \exp\left[\frac{q\Phi_o(x)}{kT}\right]$ (Expression 1.3.20a), and since by definition $\mathcal{E}_o = -d\Phi_o/dx$, one obtains:

$$q\,\mu_n\,n\frac{d\Phi_o}{dx} = qD_n\frac{dn}{dx}$$

$$\Downarrow$$

$$q\,\mu_n\,n_o \exp\left[\frac{q\Phi_o(x)}{kT}\right]\frac{d\Phi_o}{dx} = qD_n\frac{dn}{d\Phi_o}\frac{d\Phi_o}{dx}$$

$$= qD_n\frac{q}{kT}n_o \exp\left[\frac{q\Phi_o(x)}{kT}\right]\frac{d\Phi_o}{dx}$$

$$\Downarrow$$

***Einstein Relationships***

For electrons:

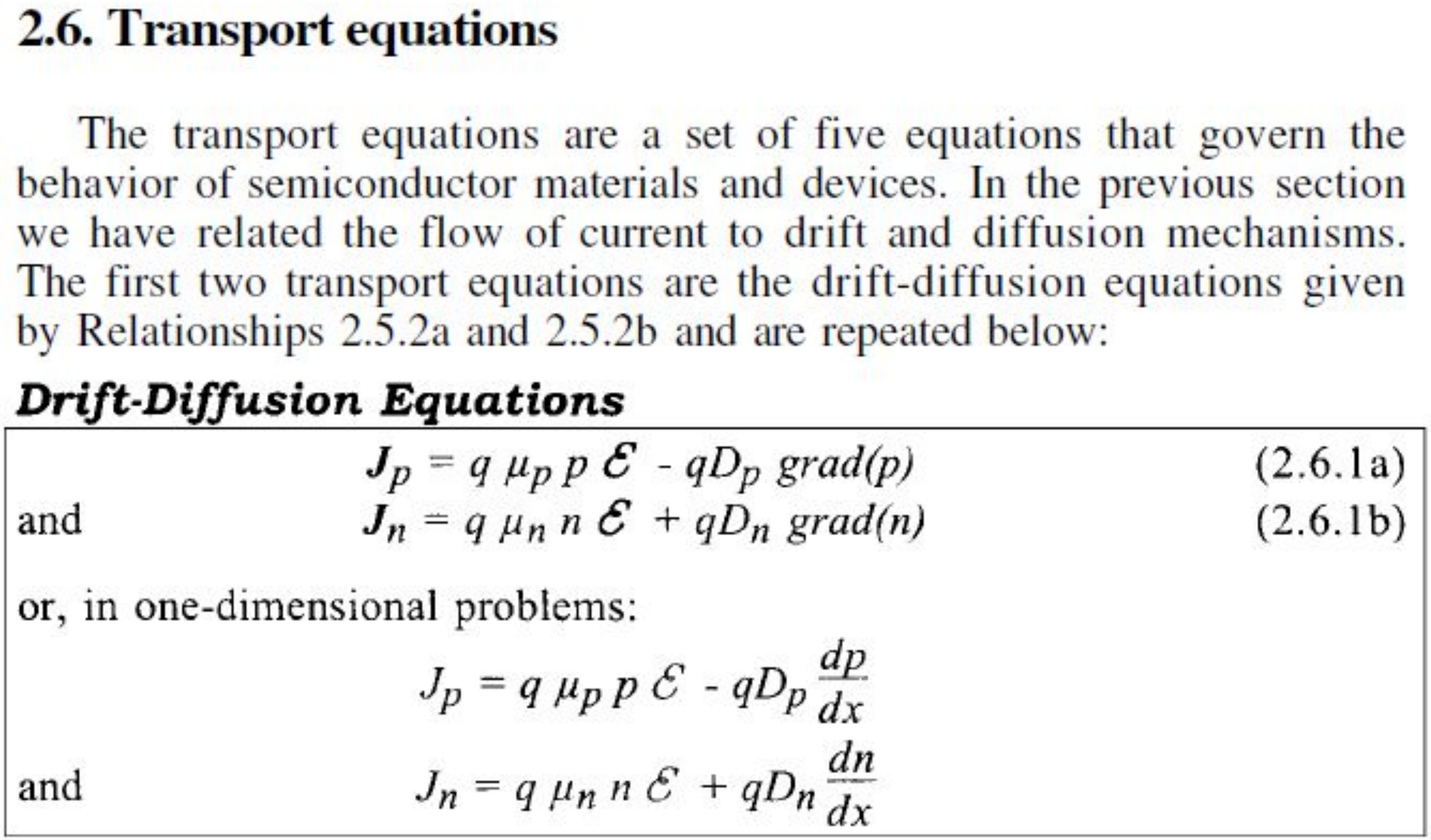
$$D_n = \frac{kT}{q}\mu_n \qquad (2.5.4a)$$

For holes:

$$D_p = \frac{kT}{q}\mu_p \qquad (2.5.4b)$$

Relationships 2.5.4 a and b are called "Einstein relationships". They show that diffusion coefficients and mobilities represent the same thing, within a multiplication constant, *kT/q*. The value *kT/q* has the dimension of a voltage, and is called "thermal voltage". It is equal to 25.9 mV at room temperature and is frequently noted "$U_T$" or "$V_T$". Thus if the mobility is known the diffusion coefficient can be calculated.

# Transport properties of semiconductors



## 2.6. Transport equations

The transport equations are a set of five equations that govern the behavior of semiconductor materials and devices. In the previous section we have related the flow of current to drift and diffusion mechanisms. The first two transport equations are the drift-diffusion equations given by Relationships 2.5.2a and 2.5.2b and are repeated below:

***Drift-Diffusion Equations***

$$\boldsymbol{J}_p = q\,\mu_p\,p\,\mathcal{E} - qD_p\,grad(p) \qquad (2.6.1a)$$

and

$$\boldsymbol{J}_n = q\,\mu_n\,n\,\mathcal{E} + qD_n\,grad(n) \qquad (2.6.1b)$$

or, in one-dimensional problems:

$$J_p = q\,\mu_p\,p\,\mathcal{E} - qD_p\frac{dp}{dx}$$

and

$$J_n = q\,\mu_n\,n\,\mathcal{E} + qD_n\frac{dn}{dx}$$

# Transport properties of semiconductors

Using the Maxwell equations $\nabla \mathcal{D} = \rho$ and $\mathcal{D} = \varepsilon\, \mathcal{E}$, where $\mathcal{D}$ is the displacement field, and using the relationship between electric field and potential $\mathcal{E}(x) = -d\Phi(x)/dx$ one readily obtains the Poisson equation:

$$\nabla^2 \Phi(x,y,z) = div(grad(\Phi(x,y,z))) = -\, div\, \mathcal{E}$$

$$= -\frac{\rho(x,y,z)}{\varepsilon_s} = -\frac{q}{\varepsilon_s}(p - n + N_d^+ - N_a^-) \qquad (2.6.2)$$

where $\varepsilon_s$ is the permittivity of the semiconductor and $\rho$ is the local charge density (C/cm$^3$) in the semiconductor. If all the doping atoms are ionized, which is the case at room temperature, one obtains:

**Poisson's Equation**

$$\nabla^2 \Phi(x,y,z) = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2}$$

$$= -\frac{q}{\varepsilon_s}\left[(p(x,y,z) - n(x,y,z) + N_d(x,y,z) - N_a(x,y,z))\right] \qquad (2.6.3a)$$

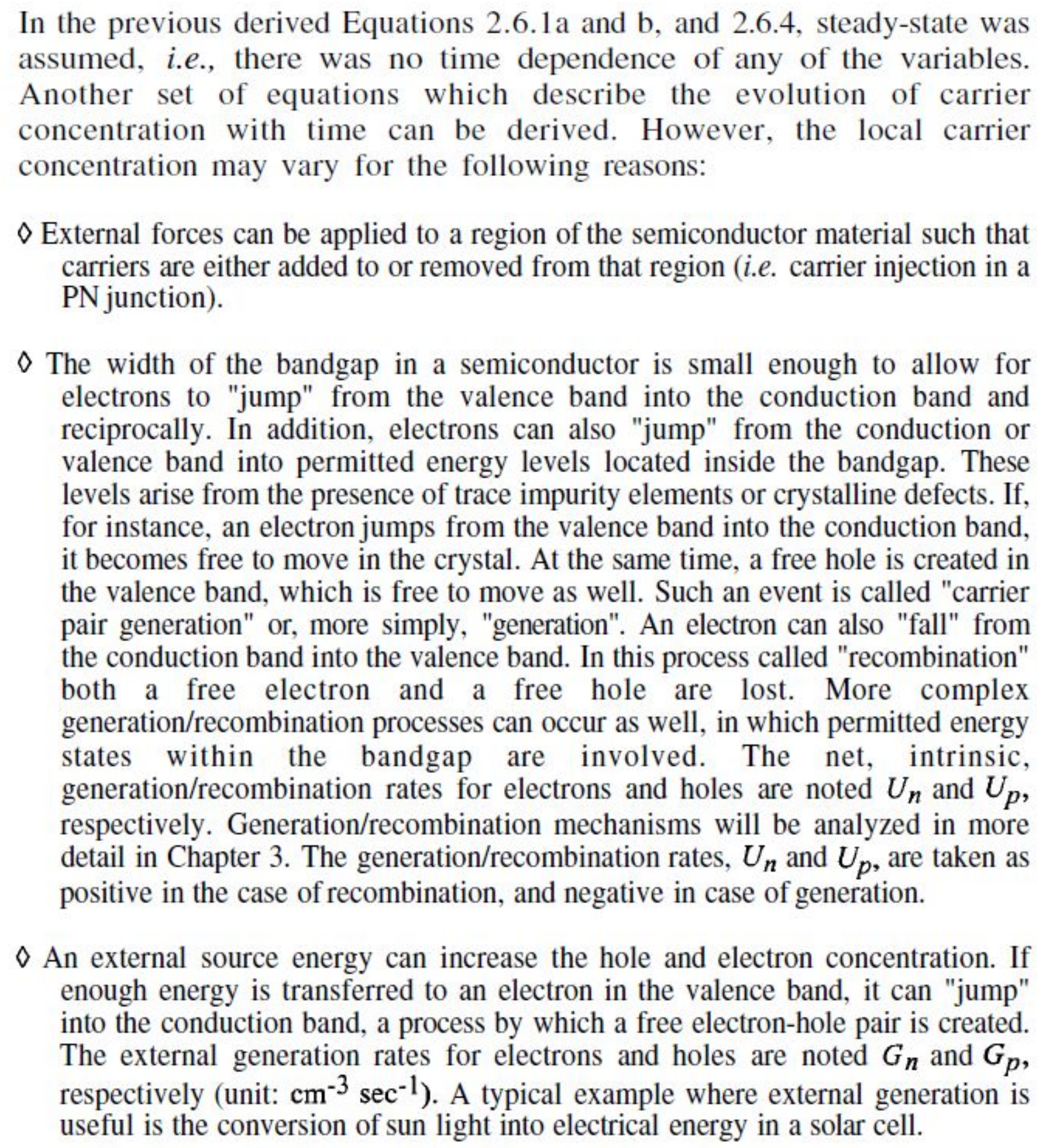
or, in short:

$$\nabla^2 \Phi = -\frac{\rho}{\varepsilon_s} \qquad (2.6.3b)$$

and for one-dimensional problems:

$$\frac{d^2 \Phi}{dx^2} = -\frac{\rho}{\varepsilon_s} \qquad (2.6.4)$$

# Transport properties of semiconductors



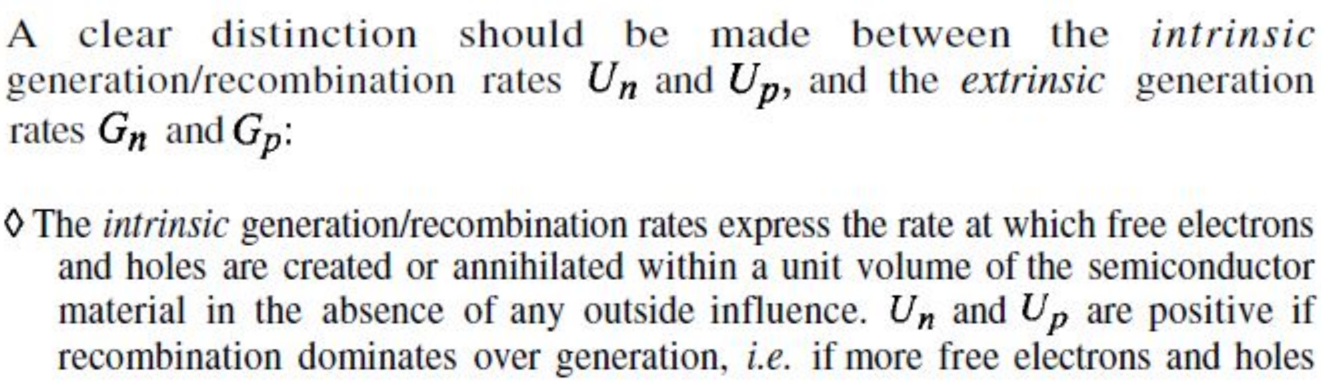
In the previous derived Equations 2.6.1a and b, and 2.6.4, steady-state was assumed, *i.e.*, there was no time dependence of any of the variables. Another set of equations which describe the evolution of carrier concentration with time can be derived. However, the local carrier concentration may vary for the following reasons:

◊ External forces can be applied to a region of the semiconductor material such that carriers are either added to or removed from that region (*i.e.* carrier injection in a PN junction).

◊ The width of the bandgap in a semiconductor is small enough to allow for electrons to "jump" from the valence band into the conduction band and reciprocally. In addition, electrons can also "jump" from the conduction or valence band into permitted energy levels located inside the bandgap. These levels arise from the presence of trace impurity elements or crystalline defects. If, for instance, an electron jumps from the valence band into the conduction band, it becomes free to move in the crystal. At the same time, a free hole is created in the valence band, which is free to move as well. Such an event is called "carrier pair generation" or, more simply, "generation". An electron can also "fall" from the conduction band into the valence band. In this process called "recombination" both a free electron and a free hole are lost. More complex generation/recombination processes can occur as well, in which permitted energy states within the bandgap are involved. The net, intrinsic, generation/recombination rates for electrons and holes are noted $U_n$ and $U_p$, respectively. Generation/recombination mechanisms will be analyzed in more detail in Chapter 3. The generation/recombination rates, $U_n$ and $U_p$, are taken as positive in the case of recombination, and negative in case of generation.

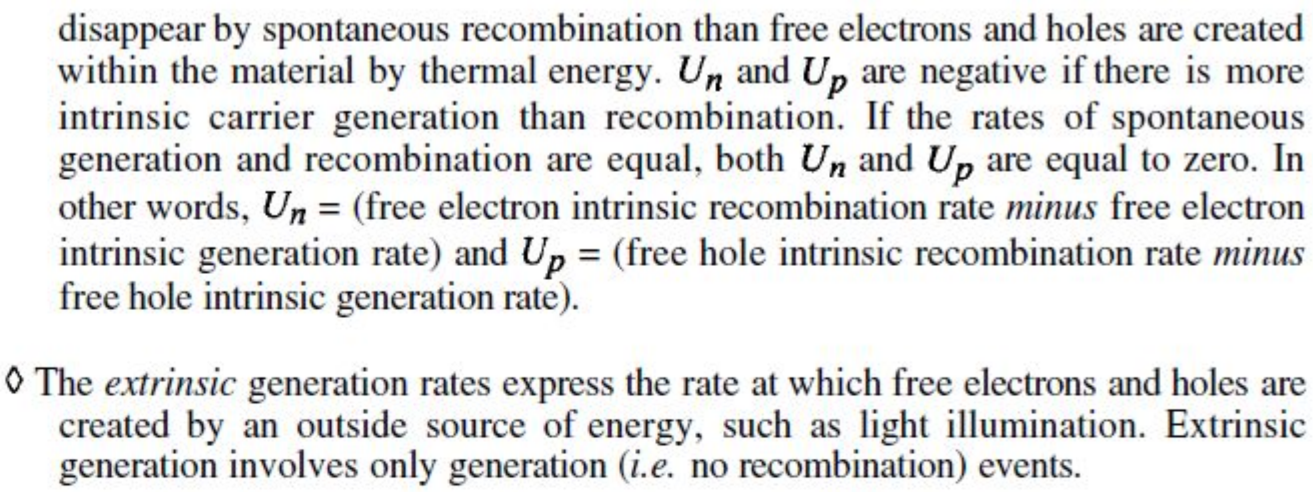
◊ An external source energy can increase the hole and electron concentration. If enough energy is transferred to an electron in the valence band, it can "jump" into the conduction band, a process by which a free electron-hole pair is created. The external generation rates for electrons and holes are noted $G_n$ and $G_p$, respectively (unit: $\text{cm}^{-3}\ \text{sec}^{-1}$). A typical example where external generation is useful is the conversion of sun light into electrical energy in a solar cell.

# Transport properties of semiconductors



A clear distinction should be made between the *intrinsic* generation/recombination rates $U_n$ and $U_p$, and the *extrinsic* generation rates $G_n$ and $G_p$:

◊ The *intrinsic* generation/recombination rates express the rate at which free electrons and holes are created or annihilated within a unit volume of the semiconductor material in the absence of any outside influence. $U_n$ and $U_p$ are positive if recombination dominates over generation, *i.e.* if more free electrons and holes disappear by spontaneous recombination than free electrons and holes are created within the material by thermal energy. $U_n$ and $U_p$ are negative if there is more intrinsic carrier generation than recombination. If the rates of spontaneous generation and recombination are equal, both $U_n$ and $U_p$ are equal to zero. In other words, $U_n$ = (free electron intrinsic recombination rate *minus* free electron intrinsic generation rate) and $U_p$ = (free hole intrinsic recombination rate *minus* free hole intrinsic generation rate).



◊ The *extrinsic* generation rates express the rate at which free electrons and holes are created by an outside source of energy, such as light illumination. Extrinsic generation involves only generation (*i.e.* no recombination) events.

# Transport properties of semiconductors

To derive the equations describing the variation of the number of carriers due to generation/recombination events we will consider a differential volume of semiconductor material (Figure 2.7). The cross-sectional area of the volume under consideration is $A$ with length $dx$. An electron current density $J_n(x)$ (unit: Amps/cm$^2$) enters the volume and a current density $J_n(x+dx)$ flows out of it.

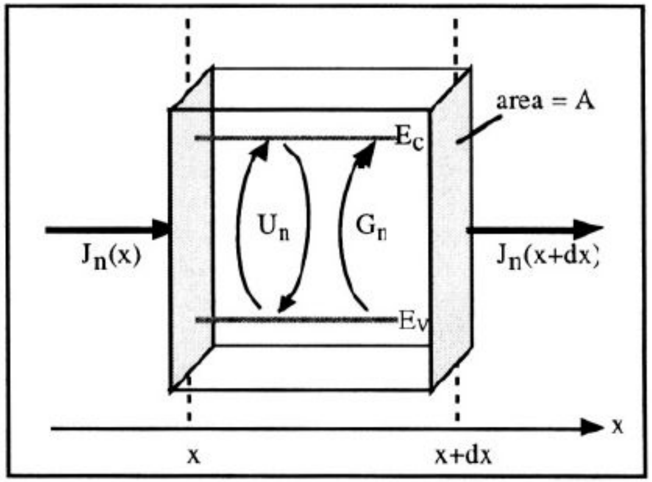


**Figure 2.7**: Elementary volume used for the derivation of the continuity equations. [6]

For one-dimensional current flow in the $x$-direction the variation of the number of free electrons in the volume $Adx$ as a function of time is given by the number of electrons entering the volume, minus the number of electrons flowing out of the volume, plus the number of electrons generated minus the number of electrons recombined:

$$A \frac{\partial n}{\partial t} dx = A \left( \frac{J_n(x)}{-q} - \frac{J_n(x+dx)}{-q} \right) + A\,(G_n - U_n)\,dx \qquad (2.6.5)$$

# Transport properties of semiconductors

$J_n(x+dx)$ can be developed in series, which yields: $J_n(x+dx) = J_n(x) + \frac{dJ_n(x)}{dx} dx + \ldots$

Using the latter result Equation 2.6.5 can then be rewritten to obtain the continuity equation for electrons:

$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n(x)}{\partial x} + (G_n - U_n) \qquad (2.6.6a)$$

A similar calculation, made for holes would yield:

$$\frac{\partial p}{\partial t} = - \frac{1}{q} \frac{\partial J_p(x)}{\partial x} + (G_p - U_p) \qquad (2.6.6b)$$

Extending Expressions 2.6.6a and 2.6.6b to three dimensions one obtains the continuity equations:

**Continuity Equations**

$$\frac{\partial n}{\partial t} = \frac{1}{q} \operatorname{div} \mathbf{J}_n + (G_n - U_n) \qquad (2.6.7a)$$

and

$$\frac{\partial p}{\partial t} = - \frac{1}{q} \operatorname{div} \mathbf{J}_p + (G_p - U_p) \qquad (2.6.7b)$$

or, in one-dimensional problems:

$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} + (G_n - U_n) \qquad (2.6.6a)$$

and

$$\frac{\partial p}{\partial t} = - \frac{1}{q} \frac{\partial J_p}{\partial x} + (G_p - U_p) \qquad (2.6.6b)$$

The set of equations composed of the *drift-diffusion equations*, the *Poisson equation*, and the *continuity equations* is called the "*transport equations*". The transport equations allows one to derive most properties of semiconductor devices.

# Transport properties of semiconductors

## 2.7. Quasi-Fermi levels

At thermodynamic equilibrium, and in the absence of applied external forces, the equilibrium carrier concentrations are a function of the internal potential $\Phi_o(x,y,z)$ in the semiconductor. The carrier concentrations are related to the internal potential by the Boltzmann relationships 1.3.20a and 1.3.20b. These can be rewritten in the following form:

$$n(x,y,z) = n_i \exp\left[\frac{E_F - E_{io}}{kT}\right] \exp\left[\frac{q\Phi_o(x,y,z)}{kT}\right] \qquad (2.7.1)$$

$$p(x,y,z) = n_i \exp\left[-\frac{E_F - E_{io}}{kT}\right] \exp\left[\frac{-q\Phi_o(x,y,z)}{kT}\right] \qquad (2.7.2)$$

and the *pn* product is given by:

$$p(x,y,z)\, n(x,y,z) = n_i^2 \qquad (2.7.3)$$

Under thermodynamic equilibrium conditions the Fermi level, $E_F$, is unique for both electrons and holes.

# Transport properties of semiconductors

Under non-equilibrium conditions, however, this is no longer the case. For instance when excess carriers are continuously injected into the semiconductor material or if light is continuously shone on it, the relationship between the internal potential $\Phi(x,y,z)$ and the electron and hole concentrations, $n(x,y,z)$ and $p(x,y,z)$ becomes more complicated. The Boltzmann relationships, however, are still valid if one introduces the notion of "quasi-Fermi levels". Quasi-Fermi levels are also called "imref", which means "*imaginary reference*", and quite conveniently, corresponds to the word "Fermi" spelled backwards. Instead of a single Fermi level common to both types of carriers let us define an electron quasi-Fermi level, $E_{Fn}(x,y,z)$, and a hole quasi-Fermi level, $E_{Fp}(x,y,z)$. The Boltzmann relationships can be rewritten in the following form:

$$n(x,y,z) = n_i \, exp\left[\frac{E_{Fn}(x,y,z)-E_{io}}{kT}\right] exp\left[\frac{q\Phi(x,y,z)}{kT}\right] \qquad (2.7.4)$$

$$p(x,y,z) = n_i \, exp\left[-\frac{E_{Fp}(x,y,z)-E_{io}}{kT}\right] exp\left[\frac{-q\Phi(x,y,z)}{kT}\right] \qquad (2.7.5)$$

and the *pn* product is equal to:

$$p(x,y,z)\; n(x,y,z) = n_i^2 \; exp\left[\frac{E_{Fn}(x,y,z) - E_{Fp}(x,y,z)}{kT}\right] \qquad (2.7.6)$$

# Transport properties of semiconductors

From Equation 2.6.1b we know that the electron current density is given by:

$$J_n = q\, \mu_n\, n\, \mathcal{E} + qD_n\, grad(n) \qquad (2.7.7)$$

Taking the derivative of Expression 2.7.4 we can write:

$$grad(n) = n_i\, exp\left[\frac{E_{Fn}-E_{io}}{kT}\right] exp\left[\frac{q\Phi}{kT}\right] \frac{1}{kT}\left(grad(E_{Fn}) + q\, grad(\Phi)\right)$$

$$= \frac{n}{kT}\, [grad(E_{Fn}) + q\, grad(\Phi)] \qquad (2.7.8)$$

Introducing the result of Equation 2.7.8 into Relationship 2.7.7 one obtains:

$$J_n = q\, \mu_n\, n\, \mathcal{E} + qD_n\, \frac{n}{kT}\, [grad(E_{Fn}) + q\, grad(\Phi)] \qquad (2.7.9)$$

Using the Einstein Relationship $D_n = \frac{kT}{q}\mu_n$ we finally obtain:

$$J_n = n\, \mu_n\, grad(E_{Fn}) \qquad (2.7.10a)$$

A similar calculation, made for holes, would yield:

$$J_p = p\, \mu_p\, grad(E_{Fp}) \qquad (2.7.10b)$$

The two last relationships show that, *in the most general case, the current is not linked to the gradient of the internal potential,* $\Phi_o$*, but to the gradient of the quasi-Fermi levels.* Under thermodynamic equilibrium conditions and in the absence of external forces, however, $E_{Fn} = E_{Fp} = E_F$ = a constant, and therefore, $J_n = J_p = 0$, and $pn = n_i^2$.

# The PN junction diode

## 4.1. Introduction

A PN junction is formed when a P-type and an N-type semiconductor are in contact. If the N-and P-type regions are made out of the same semiconductor material (e.g. N-type silicon and P-type silicon), the junction is a homojunction. If the semiconductor materials are different (e.g. N-type silicon and P-type germanium), the junction is a heterojunction. Heterojunctions are dealt with in Chapter 9.

A diode is a semiconductor device consisting of a single PN junction (Figure 4.1). Unlike a resistor, it has a highly non-linear current-voltage characteristic and is often used as a rectifying element. Some diodes can emit light (light-emitting diodes), and others can emit laser light (laser diodes). The proper combination of two PN junctions produces a bipolar transistor, a device capable of amplifying electric signals.

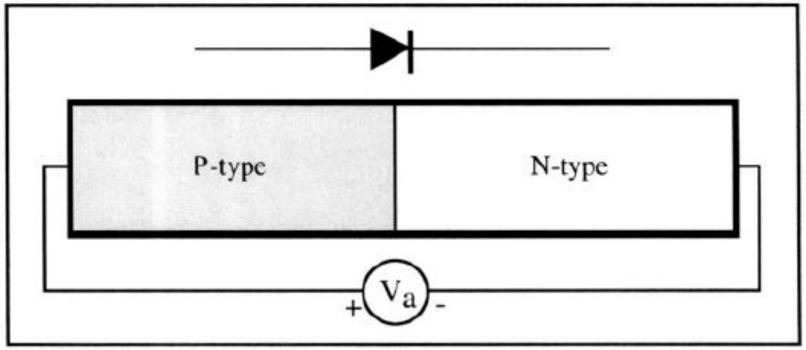


**Figure 4.1**: PN junction and symbol representing a diode.

The PN junction presents the following property: It allows current flow in one bias direction, but not in the other bias direction. Hence it rectifies the current. The sign convention used in this chapter is shown in Figure

# The PN junction diode

4.1. The applied voltage, $V_a$, is positive if the potential applied to the P-side is higher than that on the N-side. As illustrated in Figure 4.2 current flows through the diode if $V_a$ is positive, and does not if $V_a$ is negative. If $V_a > 0$ the junction is said to be forward biased, and if $V_a < 0$ it is reverse biased.

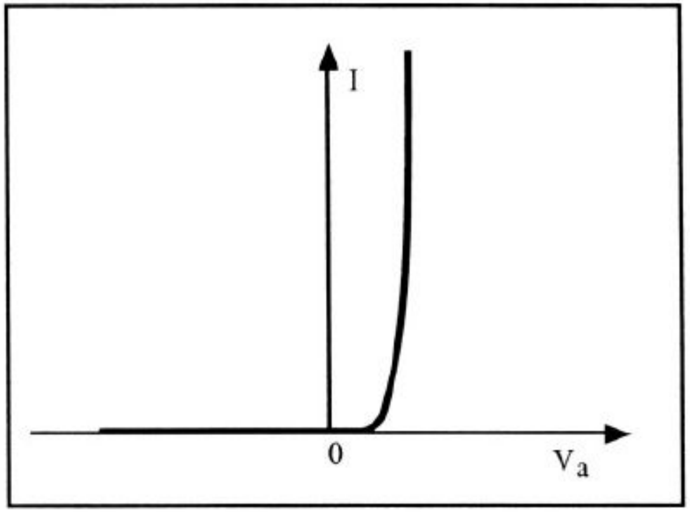


**Figure 4.2**: Current-voltage characteristics of a PN junction.

Experimental measurements show that the current in a PN junction, I, obeys the following equation:

$$I = I_s\left(exp\left[\frac{qV_a}{kT}\right] - 1\right) \qquad (4.1.1)$$

where $I_s$ is a constant and $V_a$ is the voltage applied to the diode.

# The PN junction diode

An analogy of the diode is a valve which controls liquid flow (Figure 4.3). When a pressure differential is applied in the forward direction, the valve opens and allows the liquid flow. If the pressure differential is applied in the reverse direction, the valve closes, and no liquid flows, except for a few drops if the valve is imperfect and somewhat "leaky".

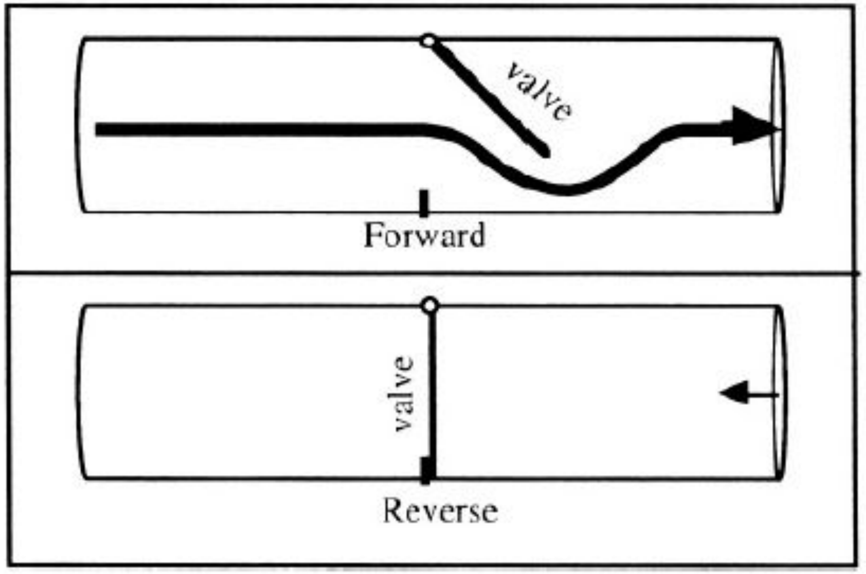
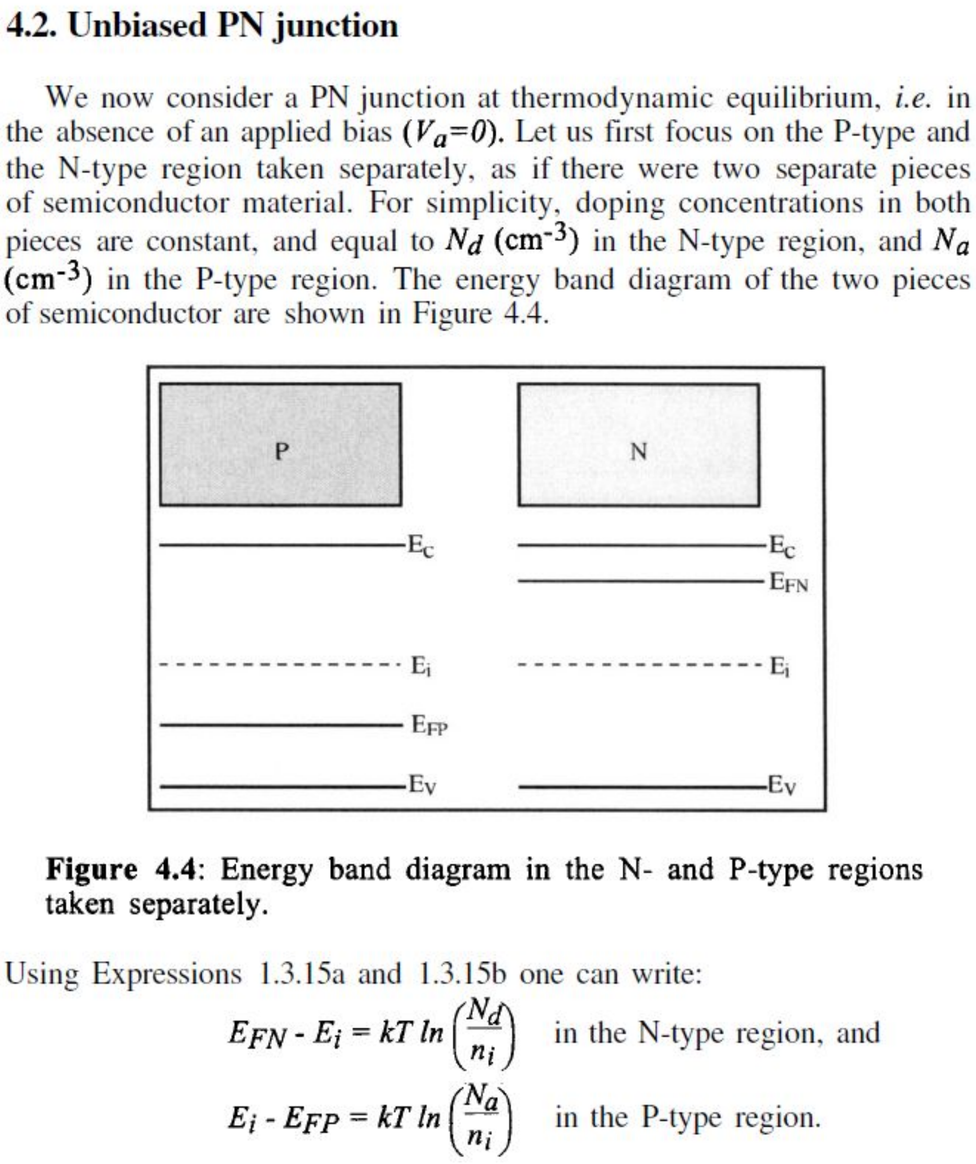


**Figure 4.3**: Fluid mechanics analogy of a pn junction to a valve.

# The PN junction diode



## 4.2. Unbiased PN junction

We now consider a PN junction at thermodynamic equilibrium, *i.e.* in the absence of an applied bias ($V_a=0$). Let us first focus on the P-type and the N-type region taken separately, as if there were two separate pieces of semiconductor material. For simplicity, doping concentrations in both pieces are constant, and equal to $N_d$ (cm$^{-3}$) in the N-type region, and $N_a$ (cm$^{-3}$) in the P-type region. The energy band diagram of the two pieces of semiconductor are shown in Figure 4.4.

**Figure 4.4**: Energy band diagram in the N- and P-type regions taken separately.

Using Expressions 1.3.15a and 1.3.15b one can write:

$$E_{FN} - E_i = kT \ln\left(\frac{N_d}{n_i}\right) \quad \text{in the N-type region, and}$$

$$E_i - E_{FP} = kT \ln\left(\frac{N_a}{n_i}\right) \quad \text{in the P-type region.}$$

# The PN junction diode

Let us now build the PN junction by connecting the P-type region to the N-type region. The surface where the contact is made is called the "metallurgical junction". A junction where the doping concentration "abruptly" switches from P-type to N-type (at the metallurgical junction) is called a step junction. We already know from Section 1.4 that the Fermi level is unique and constant in a structure under equilibrium: electrons instantly diffuse from the electron-rich N-type region into the electron-poor P-type region, and holes from the P-type material diffuse into the N-type region. As a result of the charge displacement an internal built-in potential called junction potential, $\Phi_o$, is formed at the junction, as shown in Figure 4.5.

Within a multiplication factor *-q* the junction potential is equal to the curvature of the energy bands:

$$E_{FN} - E_{FP} = q\Phi_o = kT\, ln\left(\frac{N_d}{n_i}\right) + kT\, ln\left(\frac{N_a}{n_i}\right) = kT\, ln\left(\frac{N_a N_d}{n_i^2}\right) \qquad (4.2.1)$$

and thus:

$$\Phi_o = \frac{kT}{q}\, ln\left(\frac{N_a N_d}{n_i^2}\right) \qquad (4.2.2)$$

# The PN junction diode

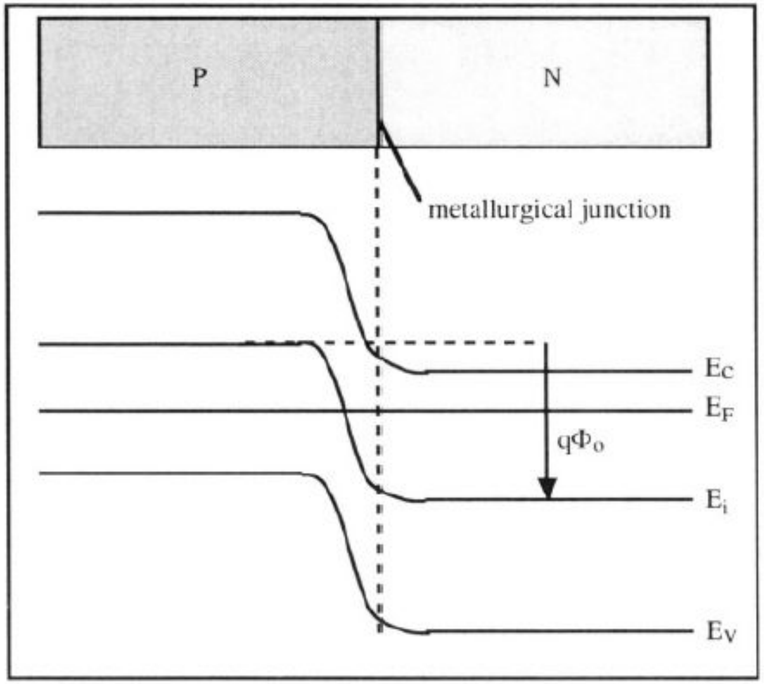


**Figure 4.5**: PN junction and corresponding energy band diagram.[1]

When electrons diffuse from the N-type region into the P-type material, they "leave behind" the ionized donor atoms they originated from. These atoms occupy substitutional sites in the crystal lattice and cannot move within the crystal. The region where these positively charged ions are located constitutes a space-charge region called a "depletion region" because it is depleted of electrons (Figure 4.6).

# The PN junction diode

The positive charge in the depletion region attracts electrons such that at equilibrium, the force of diffusion pushing electrons into the P-type region is exactly balanced by the force of the built-in electric field that "recalls" the electrons back into the N-type region. Similarly, the diffusion of holes from the P-type into the N-type region gives rise to a depletion region in the P-type material. This region is depleted of holes and bears a negative charge because of the presence of negatively charged acceptor ionized atoms. There are several names for the depletion region

located around the metallurgical junction; it can be called the "depletion region", the "space-charge region" or the "transition region".

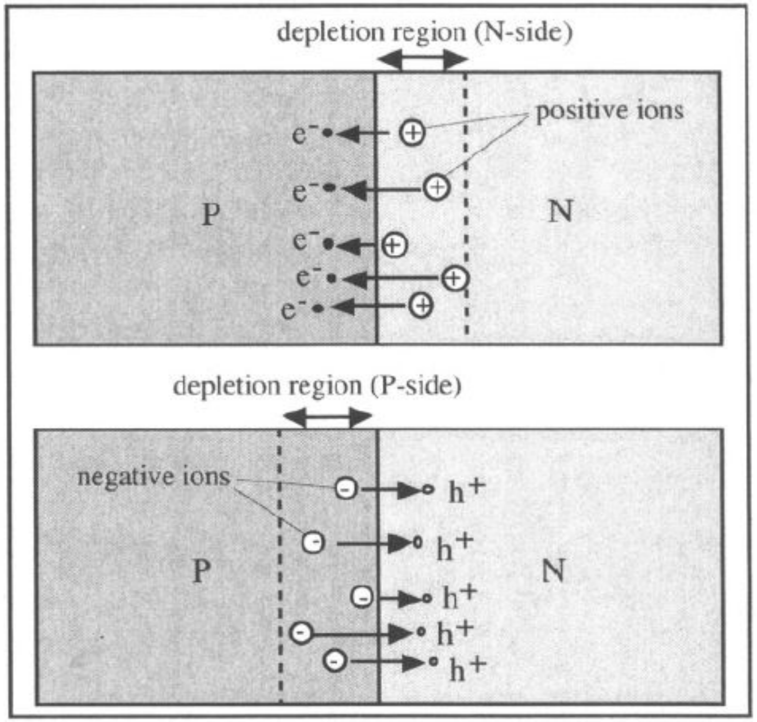


**Figure 4.6**: Creation of depletion regions by the diffusion of electrons and holes.

# The PN junction diode

The electric field and the potential variation in the space-charge region can be calculated using the Poisson equation (Expression 2.6.2). For a one-dimensional junction the problem simplifies to:

$$\frac{d^2\Phi(x)}{dx^2} = -\frac{q}{\varepsilon_s}(p - n + N_d^+ - N_a^-) \qquad (4.2.3a)$$

Using the Boltzmann Relationships 1.3.20a and 1.3.20b we obtain:

$$\frac{d^2\Phi(x)}{dx^2} = -\frac{q}{\varepsilon_s}\left\{p_o\, exp\left[\frac{-q\Phi(x)}{kT}\right] - n_o\, exp\left[\frac{q\Phi(x)}{kT}\right] + N_d^+ - N_a^-\right\} \qquad (4.2.3b)$$

with $N_d^+ = N_d$ and $N_a^+ = N_a$.

Equation 4.2.3b cannot be solved analytically and a close-form solution for the potential cannot be found. It can, however, be simplified by using the "depletion approximation". The depletion approximation assumes that the space charge is composed only of ionized doping impurities, and that the contribution of free carriers to the local charge is negligible.

Furthermore, the carrier depletion in the space-charge regions is assumed to be complete. In other words, there are no free electrons in the depletion region on the N-type side, and no free holes in the depletion region on the P-type side. As a result, the charge densities in the depletion regions are equal to $qN_d$ in the N-type material, and $-qN_a$ in the P-type material. The depletion regions extent to a distance $l_{no}$ on the N-type side, and a distance $-l_{po}$ on the P-type side, where the metallurgical junction is taken as the origin (Figure 4.7). Additionally, the electric field and potential are shown in Figure 4.7, which can also be derived from Poisson's equation with the appropriate boundary conditions.

# The PN junction diode

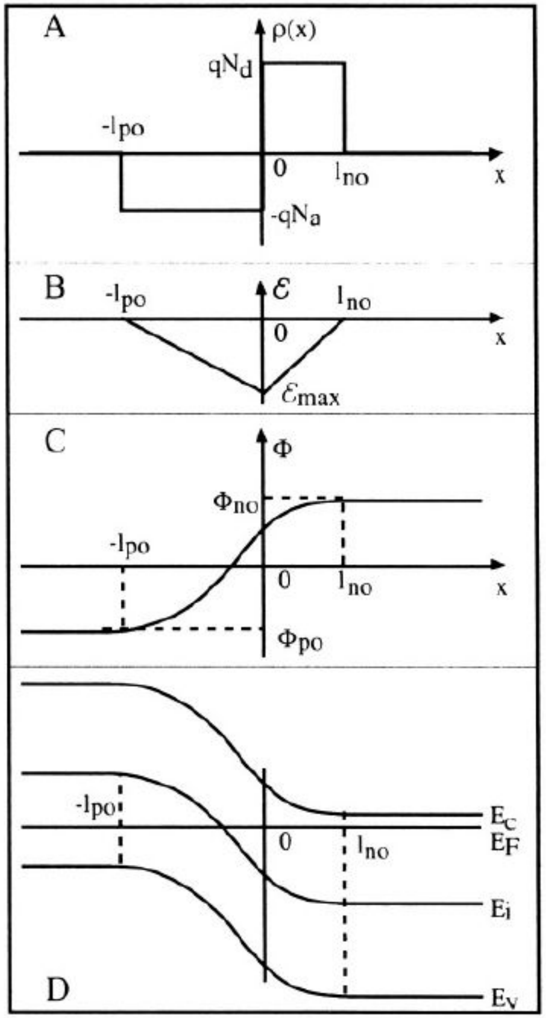


**Figure 4.7**: Charges (A), electric field (B) potential (C) and energy bands (D) in a PN junction.[2]

# The PN junction diode

With the depletion approximation, a closed-form analytical expression can be found for the electric field $\mathcal{E}(x)$, the potential $\Phi(x)$, as well as for $l_{po}$ and $l_{no}$ by utilizing Poisson's equation and Gauss' law. The value of the charge density $\rho(x)$ can be expressed for four separate regions and are given by:

$$\rho(x) = \begin{array}{llll} 0 & \text{for} & -\infty < x < -l_{po} & \text{(quasi-neutral region)} \\ -qN_a & \text{for} & -l_{po} < x < 0 & \text{(space-charge region)} \\ qN_d & \text{for} & 0 < x < l_{no} & \text{(space-charge region)} \\ 0 & \text{for} & l_{no} < x < \infty & \text{(quasi-neutral region)} \end{array}$$

We will assume that charge neutrality exists in the quasi-neutral regions. Therefore, the electric field is zero in these regions. Using all the above assumptions the Poisson equation can be integrated a first time to yield the electric field:

for $-\infty < x < -l_{po}$: $\mathcal{E}(x)=0$

for $-l_{po} < x < 0$: $$\frac{d^2\Phi(x)}{dx^2} = -\frac{d\mathcal{E}(x)}{dx} = \frac{q}{\varepsilon_s} N_a \quad \text{with} \quad \mathcal{E}(-l_{po})=0$$

$$\Downarrow$$

$$\mathcal{E}(x) = -\frac{qN_a}{\varepsilon_s}(x + l_{po}) \qquad (4.2.4)$$

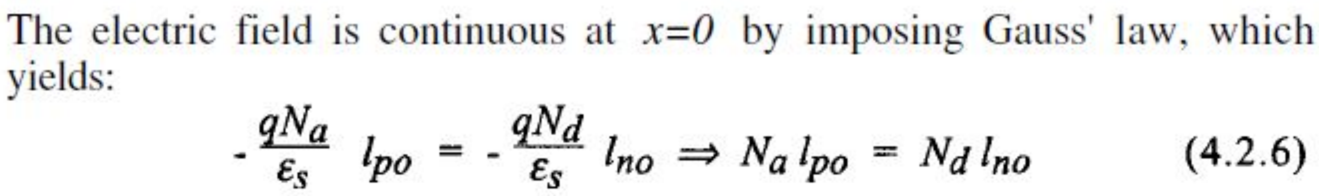
for $0 < x < l_{no}$: $$\frac{d^2\Phi(x)}{dx^2} = -\frac{d\mathcal{E}(x)}{dx} = -\frac{q}{\varepsilon_s} N_d \quad \text{with} \quad \mathcal{E}(l_{no})=0$$

$$\Downarrow$$

$$\mathcal{E}(x) = -\frac{qN_d}{\varepsilon_s}(l_{no} - x) \qquad (4.2.5)$$

and, for $l_{no} < x < \infty$ one obtains: $\mathcal{E}(x)=0$

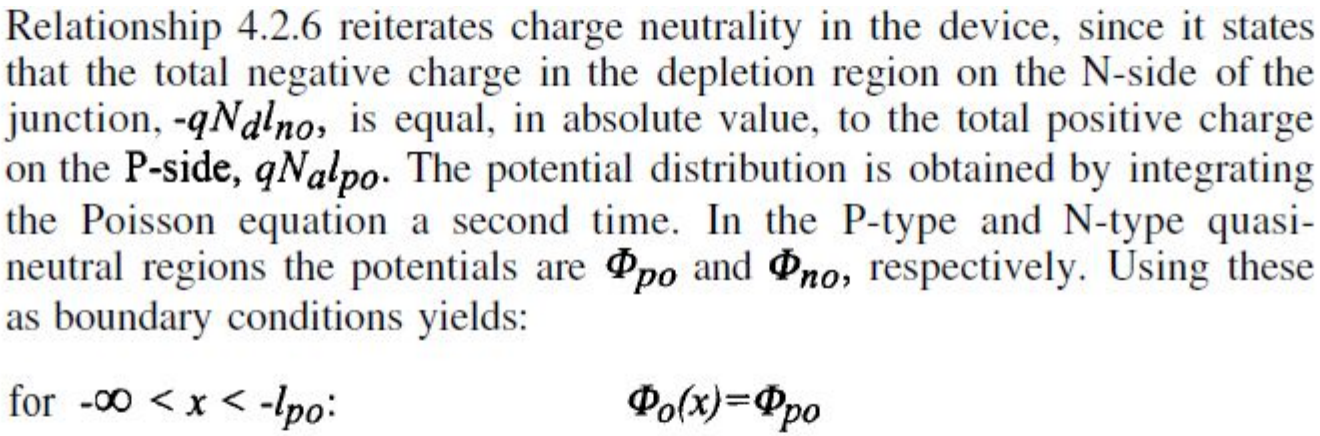
# The PN junction diode



The electric field is continuous at $x=0$ by imposing Gauss' law, which yields:

$$-\frac{qN_a}{\varepsilon_s} l_{po} = -\frac{qN_d}{\varepsilon_s} l_{no} \Rightarrow N_a l_{po} = N_d l_{no} \qquad (4.2.6)$$

Actually, the electromagnetic induction D rather than the field E should be continuous. However, here it does not matter because the dielectric constant ε is the same in the whole area.



Relationship 4.2.6 reiterates charge neutrality in the device, since it states that the total negative charge in the depletion region on the N-side of the junction, $-qN_d l_{no}$, is equal, in absolute value, to the total positive charge on the P-side, $qN_a l_{po}$. The potential distribution is obtained by integrating the Poisson equation a second time. In the P-type and N-type quasi-neutral regions the potentials are $\Phi_{po}$ and $\Phi_{no}$, respectively. Using these as boundary conditions yields:

for $-\infty < x < -l_{po}$: $\Phi_o(x) = \Phi_{po}$

# The PN junction diode

for $-l_{po} < x < 0$:

$$-\mathcal{E}(x) = \frac{d\Phi(x)}{dx} = \frac{qN_a}{\varepsilon_s}(x + l_{po})$$

$$\Downarrow$$

$$\Phi_o(x) = \frac{qN_a}{2\varepsilon_s}(x + l_{po})^2 + \Phi_{po} \qquad (4.2.7)$$

for $0 < x < l_{no}$:

$$-\mathcal{E}(x) = \frac{d\Phi(x)}{dx} = \frac{qN_d}{\varepsilon_s}(l_{no} - x)$$

$$\Downarrow$$

$$\Phi_o(x) = \Phi_{no} - \frac{qN_d}{2\varepsilon_s}(l_{no} - x)^2 \qquad (4.2.8)$$

for $l_{no} < x < \infty$:

$$\Phi_o(x) = \Phi_{no}$$

The potential is a continuous function at $x=0$. Combined with 4.2.2 this condition gives an alternate expression for the junction potential, $\boldsymbol{\Phi_o}$:

$$\frac{qN_a}{2\varepsilon_s} l_{po}^2 + \Phi_{po} = \Phi_{no} - \frac{qN_d}{2\varepsilon_s} l_{no}^2$$

$$\Downarrow$$

**Junction Potential**

$$\Phi_o = \Phi_{no} - \Phi_{po} = \frac{qN_a}{2\varepsilon_s} l_{po}^2 + \frac{qN_d}{2\varepsilon_s} l_{no}^2 = \frac{kT}{q} \ln\left(\frac{N_a N_d}{n_i^2}\right) \qquad (4.2.9)$$

# The PN junction diode

The electric field has a single maximum value at *x=0*. Its expression can be obtained using 4.2.4 or 4.2.5:

**Maximum electric field**

$$\mathcal{E}_{max} = -\frac{qN_a}{\varepsilon_s} l_{po} = -\frac{qN_d}{\varepsilon_s} l_{no} \qquad (4.2.10)$$

Using Expressions 4.2.6 and 4.2.9 the width of the depletion regions, $l_{po}$ and $l_{no}$, can be expressed as a function of the junction potential:

**Width of Depletion Regions**

$$l_{po} = \sqrt{\frac{2\varepsilon_s}{q} \frac{\Phi_o N_d}{N_a (N_a+N_d)}} \qquad (4.2.11a)$$

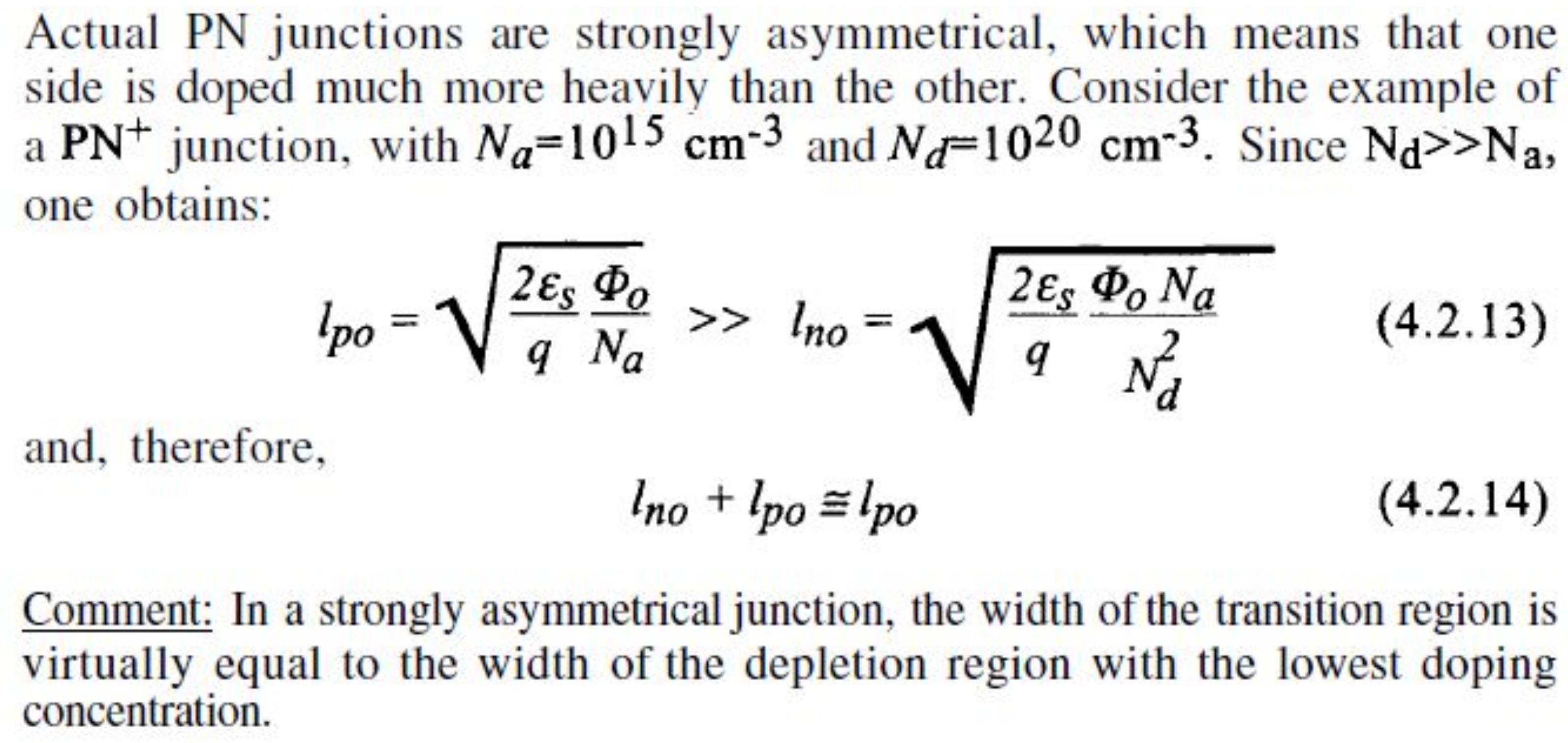
and

$$l_{no} = \sqrt{\frac{2\varepsilon_s}{q} \frac{\Phi_o N_a}{N_d (N_a+N_d)}} \qquad (4.2.11b)$$

The sum of the depletion regions is called the "transition region" which contains both ionized acceptor and donor impurities. The width of the transition region is given by:

$$l_{no} + l_{po} = \sqrt{\frac{2\varepsilon_s}{q} \frac{\Phi_o (N_a+N_d)}{N_a N_d}} \qquad (4.2.12)$$

# The PN junction diode



Actual PN junctions are strongly asymmetrical, which means that one side is doped much more heavily than the other. Consider the example of a **PN$^+$** junction, with $N_a$=$10^{15}$ cm$^{-3}$ and $N_d$=$10^{20}$ cm$^{-3}$. Since $N_d$>>$N_a$, one obtains:

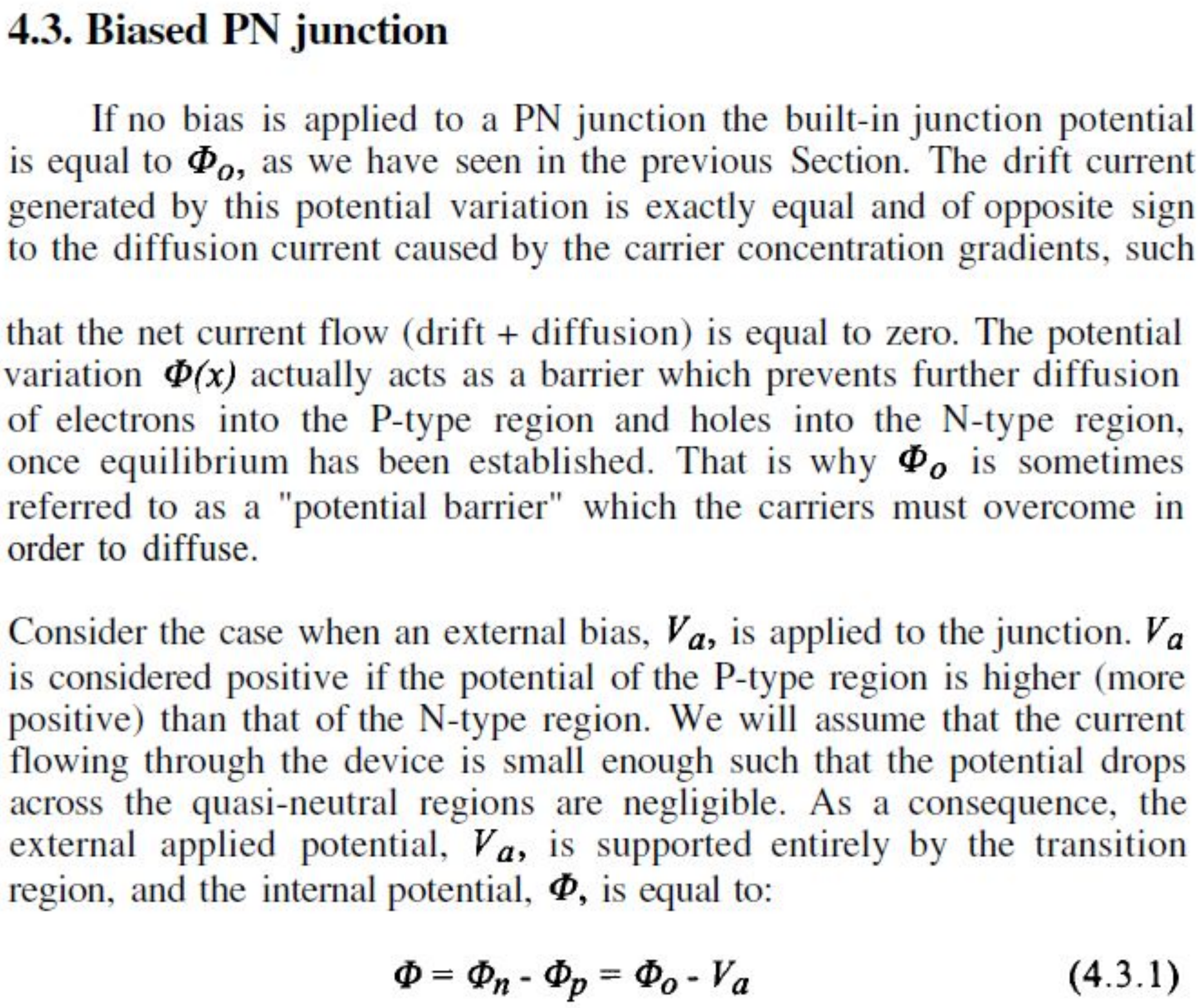
$$l_{po} = \sqrt{\frac{2\varepsilon_s}{q}\,\frac{\Phi_o}{N_a}} >> l_{no} = \sqrt{\frac{2\varepsilon_s}{q}\,\frac{\Phi_o\,N_a}{N_d^2}} \qquad (4.2.13)$$

and, therefore,

$$l_{no} + l_{po} \cong l_{po} \qquad (4.2.14)$$

Comment: In a strongly asymmetrical junction, the width of the transition region is virtually equal to the width of the depletion region with the lowest doping concentration.

# The PN junction diode. Biased junction



## 4.3. Biased PN junction

If no bias is applied to a PN junction the built-in junction potential is equal to $\Phi_o$, as we have seen in the previous Section. The drift current generated by this potential variation is exactly equal and of opposite sign to the diffusion current caused by the carrier concentration gradients, such that the net current flow (drift + diffusion) is equal to zero. The potential variation $\Phi(x)$ actually acts as a barrier which prevents further diffusion of electrons into the P-type region and holes into the N-type region, once equilibrium has been established. That is why $\Phi_o$ is sometimes referred to as a "potential barrier" which the carriers must overcome in order to diffuse.

Consider the case when an external bias, $V_a$, is applied to the junction. $V_a$ is considered positive if the potential of the P-type region is higher (more positive) than that of the N-type region. We will assume that the current flowing through the device is small enough such that the potential drops across the quasi-neutral regions are negligible. As a consequence, the external applied potential, $V_a$, is supported entirely by the transition region, and the internal potential, $\Phi$, is equal to:

$$\Phi = \Phi_n - \Phi_p = \Phi_o - V_a \qquad (4.3.1)$$

# The PN junction diode. Biased junction

Noting that $-l_p$ and $l_n$ are the edges of the transition region (Figure 4.8), the distribution of charges in the structure are:

$$\rho(x) = \begin{array}{llll} 0 & \text{for} & -\infty < x < -l_p & \text{(quasi-neutral region)} \\ -qN_a & \text{for} & -l_p < x < 0 & \text{(space-charge region)} \\ qN_d & \text{for} & 0 < x < l_n & \text{(space-charge region)} \\ 0 & \text{for} & l_n < x < \infty & \text{(quasi-neutral region)} \end{array}$$

The Poisson equation can be solved just as it was in Equations 4.2.4 to 4.2.12, by replacing $l_{no}$, $l_{po}$ and $\Phi_o$ by $l_n$, $l_p$ and $(\Phi_o - V_a)$, respectively. The result is:

$$l_p = \sqrt{\frac{2\varepsilon_s}{q} \frac{(\Phi_o - V_a)\, N_d}{N_a\,(N_a+N_d)}} \qquad (4.3.2)$$

and

$$l_n = \sqrt{\frac{2\varepsilon_s}{q} \frac{(\Phi_o - V_a)\, N_a}{N_d\,(N_a+N_d)}} \qquad (4.3.3)$$

The total width of the transition region is equal to:

$$l_n + l_p = \sqrt{\frac{2\varepsilon_s}{q} \frac{(\Phi_o - V_a)\,(N_a+N_d)}{N_a N_d}} \qquad (4.3.4)$$

It is worth noting that the width of the transition region increases when a reverse bias is applied ($V_a < 0$) and that it decreases when a forward bias ($V_a > 0$) is applied (Figure 4.8).

# The PN junction diode. Biased junction

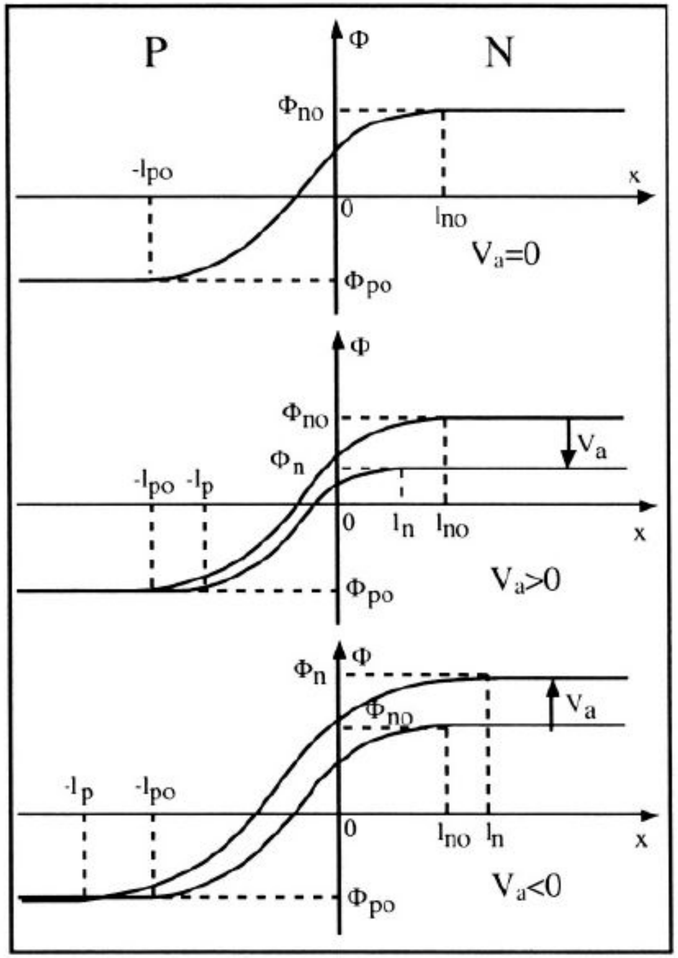
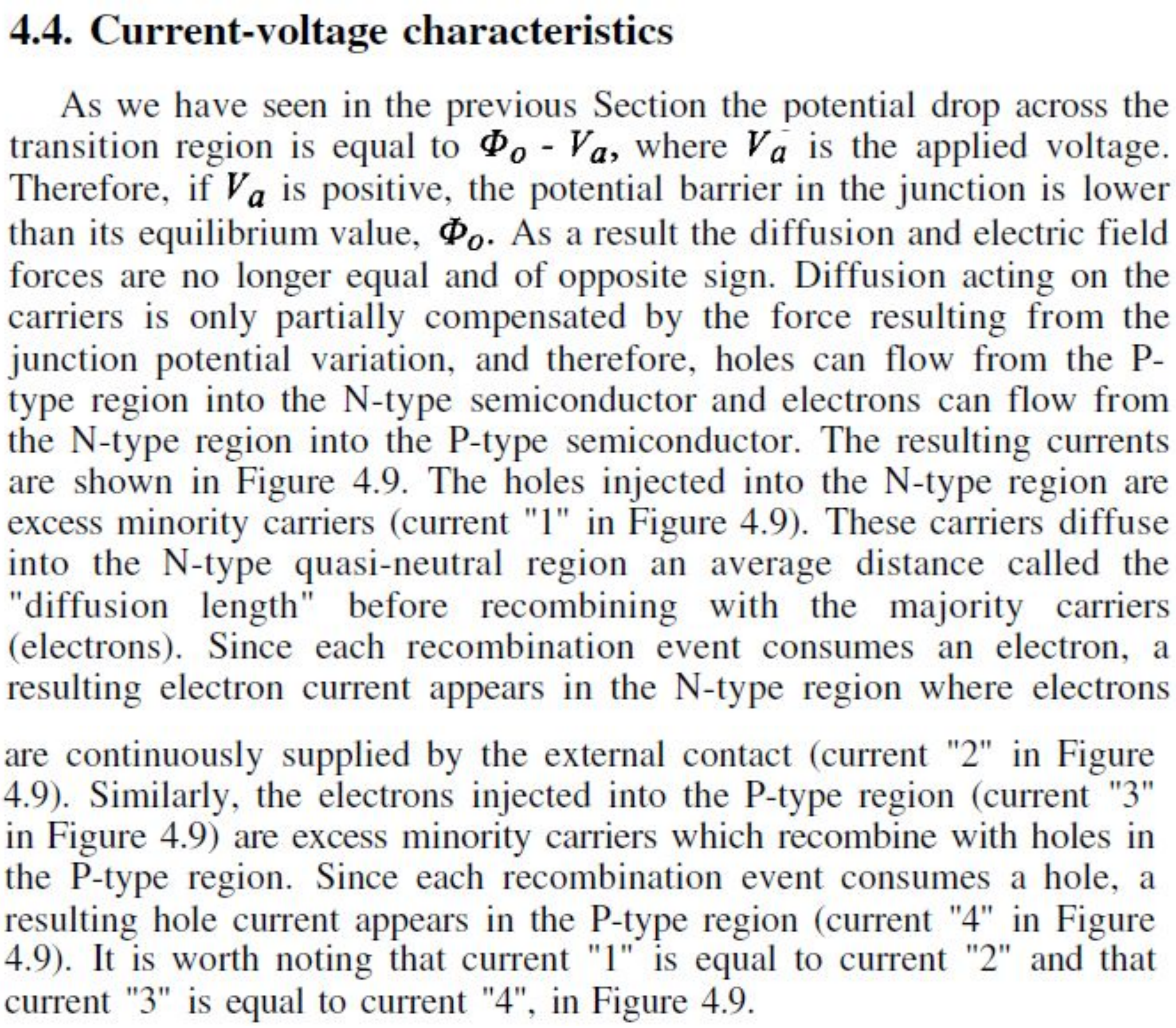


**Figure 4.8**: Potential in a PN junction for $V_a$=0, $V_a$>0 and $V_a$<0.

# The PN junction diode. Biased junction



## 4.4. Current-voltage characteristics

As we have seen in the previous Section the potential drop across the transition region is equal to $\Phi_o$ - $V_a$, where $V_a$ is the applied voltage. Therefore, if $V_a$ is positive, the potential barrier in the junction is lower than its equilibrium value, $\Phi_o$. As a result the diffusion and electric field forces are no longer equal and of opposite sign. Diffusion acting on the carriers is only partially compensated by the force resulting from the junction potential variation, and therefore, holes can flow from the P-type region into the N-type semiconductor and electrons can flow from the N-type region into the P-type semiconductor. The resulting currents are shown in Figure 4.9. The holes injected into the N-type region are excess minority carriers (current "1" in Figure 4.9). These carriers diffuse into the N-type quasi-neutral region an average distance called the "diffusion length" before recombining with the majority carriers (electrons). Since each recombination event consumes an electron, a resulting electron current appears in the N-type region where electrons are continuously supplied by the external contact (current "2" in Figure 4.9). Similarly, the electrons injected into the P-type region (current "3" in Figure 4.9) are excess minority carriers which recombine with holes in the P-type region. Since each recombination event consumes a hole, a resulting hole current appears in the P-type region (current "4" in Figure 4.9). It is worth noting that current "1" is equal to current "2" and that current "3" is equal to current "4", in Figure 4.9.

# The PN junction diode. Biased junction

If the junction is reverse-biased ($V_a<0$) the amplitude of the potential barrier is increased beyond its equilibrium value, $\Phi_o$. Diffusion of holes in the N-type region and diffusion of electrons in the P-type region are reduced and net current, resulting from the drift of holes from the N-type region into the P-type region and the drift of electrons from the P-type region into the N-type region, is observed. The magnitude of this current, however, is extremely small since it involves only *minority* carriers in the vicinity of the edges of the transition region.

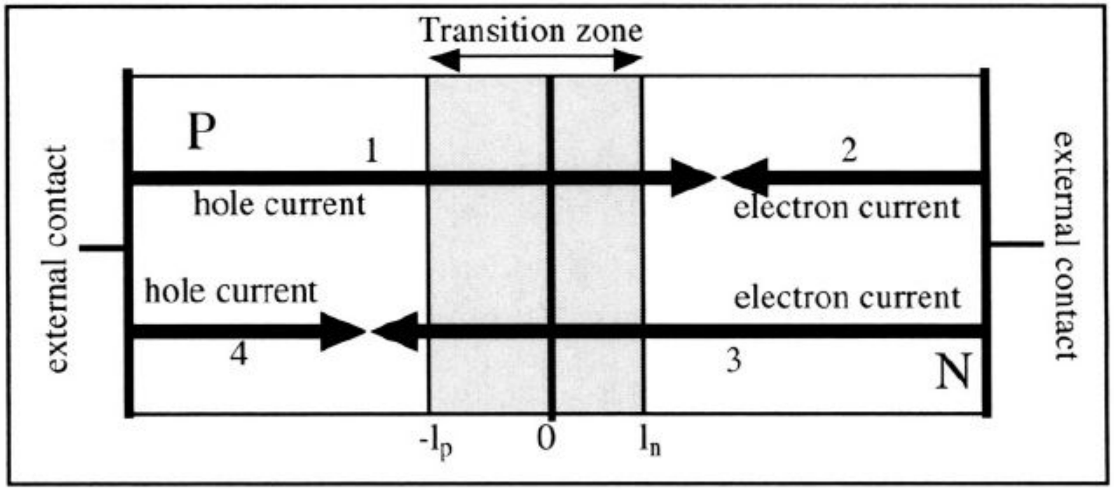
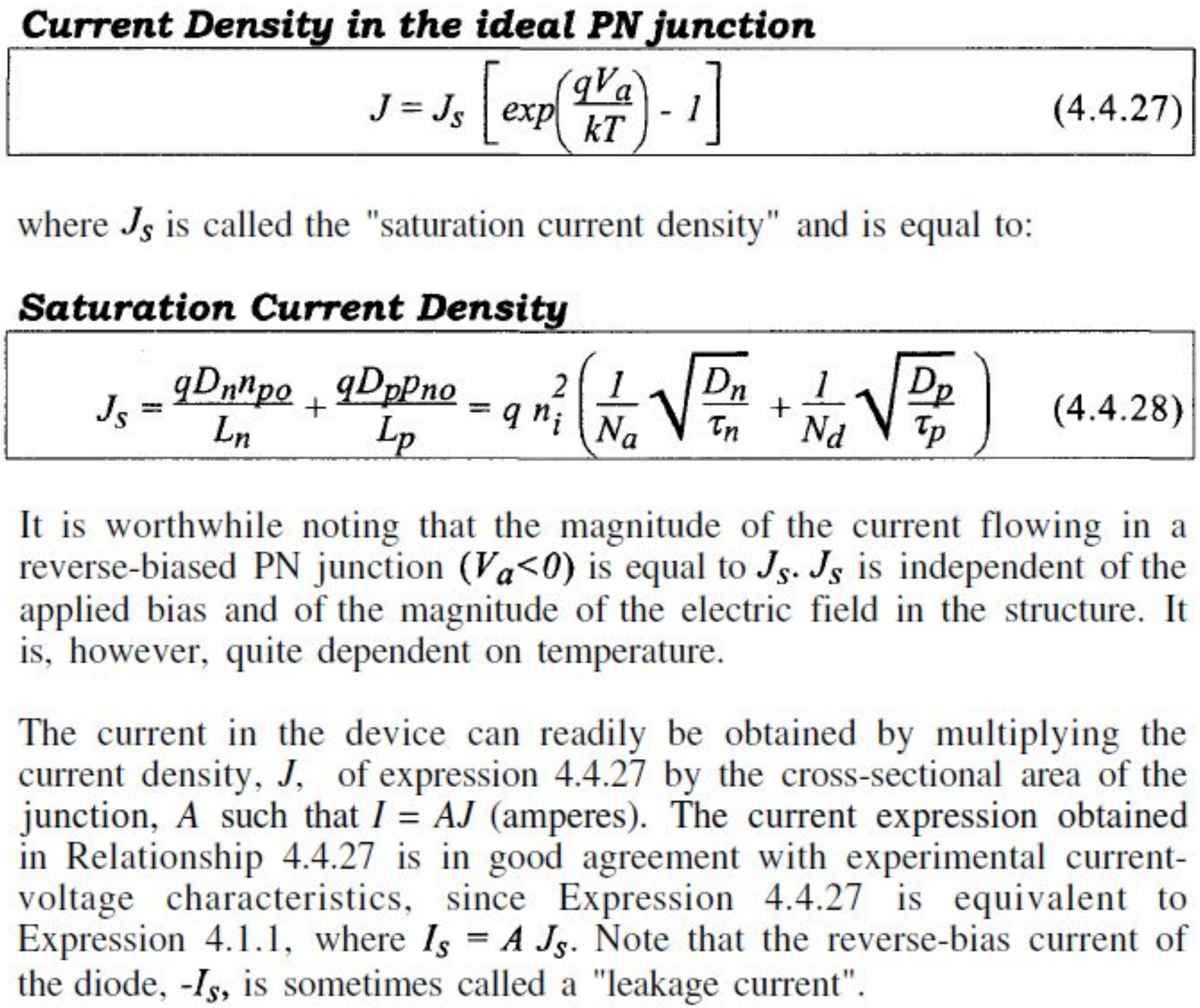


**Figure 4.9**: Forward-biased PN junction; 1: holes injected from the P-type region into the N-type region; 2: electrons recombining with the holes injected in the N-type region; 3: electrons injected from the N-type region into the P-type region; 4: holes recombining with the electrons injected in the N-type region.

# The PN junction diode. Biased junction



**Current Density in the ideal PN junction**

$$J = J_s \left[ exp\left(\frac{qV_a}{kT}\right) - 1 \right] \qquad (4.4.27)$$

where $J_s$ is called the "saturation current density" and is equal to:

**Saturation Current Density**

$$J_s = \frac{qD_n n_{po}}{L_n} + \frac{qD_p p_{no}}{L_p} = q\, n_i^2 \left( \frac{1}{N_a}\sqrt{\frac{D_n}{\tau_n}} + \frac{1}{N_d}\sqrt{\frac{D_p}{\tau_p}} \right) \qquad (4.4.28)$$

It is worthwhile noting that the magnitude of the current flowing in a reverse-biased PN junction ($V_a<0$) is equal to $J_s$. $J_s$ is independent of the applied bias and of the magnitude of the electric field in the structure. It is, however, quite dependent on temperature.

The current in the device can readily be obtained by multiplying the current density, $J$, of expression 4.4.27 by the cross-sectional area of the junction, $A$ such that $I = AJ$ (amperes). The current expression obtained in Relationship 4.4.27 is in good agreement with experimental current-voltage characteristics, since Expression 4.4.27 is equivalent to Expression 4.1.1, where $I_s = A\, J_s$. Note that the reverse-bias current of the diode, $-I_s$, is sometimes called a "leakage current".

# The PN junction diode. Solar cell

## 4.7. Solar cell

A solar cell is a PN junction in which the generation of carriers by an external source of energy, usually sunlight, is utilized to generate electrical power. In other words a solar cell directly converts solar energy into electrical power. The design of most solar cells is quite elaborate, such that the efficiency of energy conversion is maximized. In this Section, however, we will exemplify the operation of a solar cell using a simple PN junction structure. Solar cell operation is based on the generation of electron-hole pairs in the transition region, and the separation of both types of carriers by the junction electric field. Let's take the example of the **$P^+N$** junction shown in Figure 4.23. We will assume that illumination by sunlight uniformly generates G electron-hole pairs per cubic centimeter and per second, at any location in the semiconductor material. Using the same notations as before, the transition region extends from $-l_p$ to $l_n$. The bias applied to the device is $V_a$.

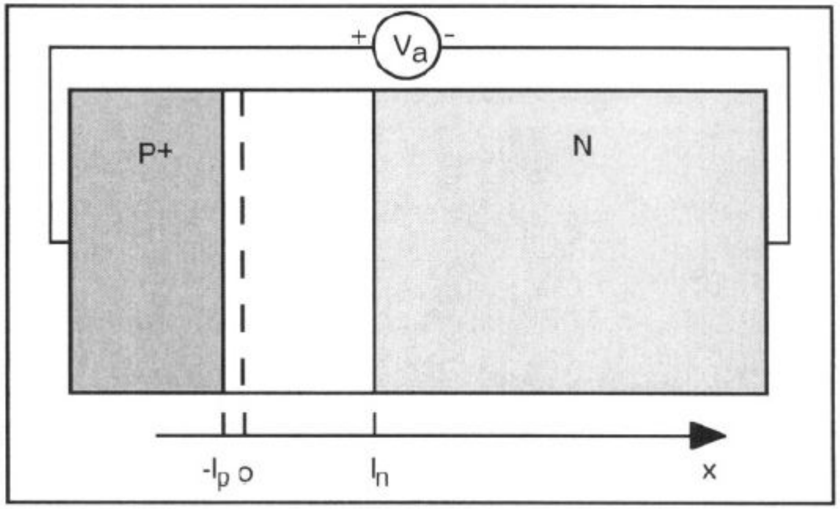


**Figure 4.23**: Geometry of a simple solar cell.

# The PN junction diode. Solar cell

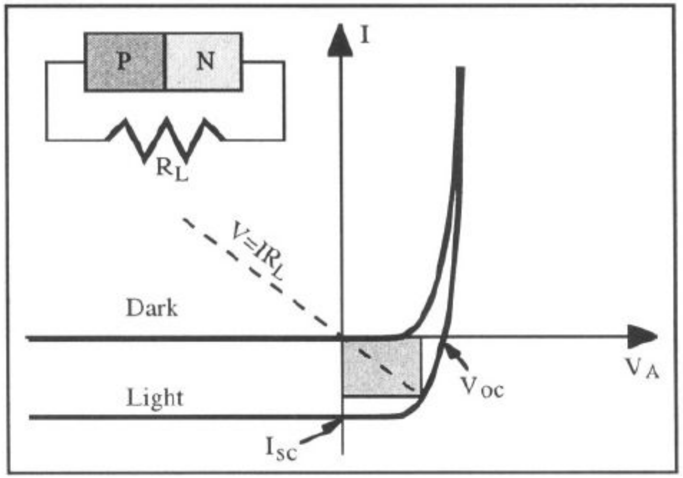


**Figure 4.24**: Current-voltage characteristics of a solar cell in the dark and under illumination. The area gray rectangle represents the power supplied by the cell to a load having a resistance $R_L$.

Figure 4.24 shows the current-voltage characteristics of a solar cell in the dark and under illumination. The insert shows a simple circuit where the solar cell under illumination delivers electrical power to a load resistor, $R_L$. The operation point of the circuit is given by the intersection of the *I-V* characteristics of the illuminated cell with the load line $V=IR_L$. The area of the gray rectangle represents the power supplied by the solar cell to the load. Optimization of solar cell performance involves the use of anti-reflection coatings, which increases light absorption, and therefore, the generation rate, *G*. The use of high-quality semiconductor material with a high minority carrier lifetime, and the choice of a load resistance value, $R_L$, maximizes the power transferred to the load (*i.e.:* which maximizes the area of the gray rectangle in Figure 4.24).