Fundamentals of Semiconductor physics

Noël SERVAGENT

08.03.2023

Table des matières

I. Presentation *3*

II. Course *4*

1. Energy Bands.....*4*
2. Insulator, semiconductor, conductor.....*5*
3. Intrinsic semiconductors.....*6*
4. Doped (extrinsic) semiconductors.....*8*
   - *4.1. P-type semiconductors.....8*
   - *4.2. N-type semiconductors.....9*

III. Case study *11*

1. Abrupt PN junction at thermal equilibrium.....*11*
2. Abrupt junction under external bias.....*14*
   - *2.1. Current density.....14*
   - *2.2. PN junction under reverse bias.....17*
   - *2.3. Junction under DC forward bias.....18*
   - *2.4. Current-Voltage characteristic.....18*
   - *2.5. Junction under small-signal AC voltage , diffusion capacitance.....19*

IV. Exercises *21*

1. Exercise n°1.....*21*
2. Exercise n°2.....*21*

Ressources annexes *23*

Solution des exercices *24*

I.Presentation

Module:

Semiconductor sensors and applications

Author(s):

Noël SERVAGENT - École des Mines de Nantes - Laboratoire Subatech

Abstract:

In the first part of this course, simple models for intrinsic and doped semiconductors are presented. These models describe the evolution of electronic populations, and serve as a basis for understanding semiconductor-based electronic devices. In a second part, the PN junction, made of two pieces of the same semiconductor with two different doping types (P and N), is studied in detail. This simple structure represents the building block for all other semiconductor devices, such as diodes, transistors, photodiodes, light-emitting diodes, or laser diodes.

Keywords:

Semiconductors, PN Junction,

Prerequisites:

-

Learning outcomes:

-

Course overview:

- Introduction
- Energy Bands
- Insulator, semiconductor, conductor
- Intrinsic semiconductors
- Doped (extrinsic) semiconductors

Conception & production :

Le Mans Université

License:

1 - http://www.gnu.org/licenses/fdl.txt

II.Course

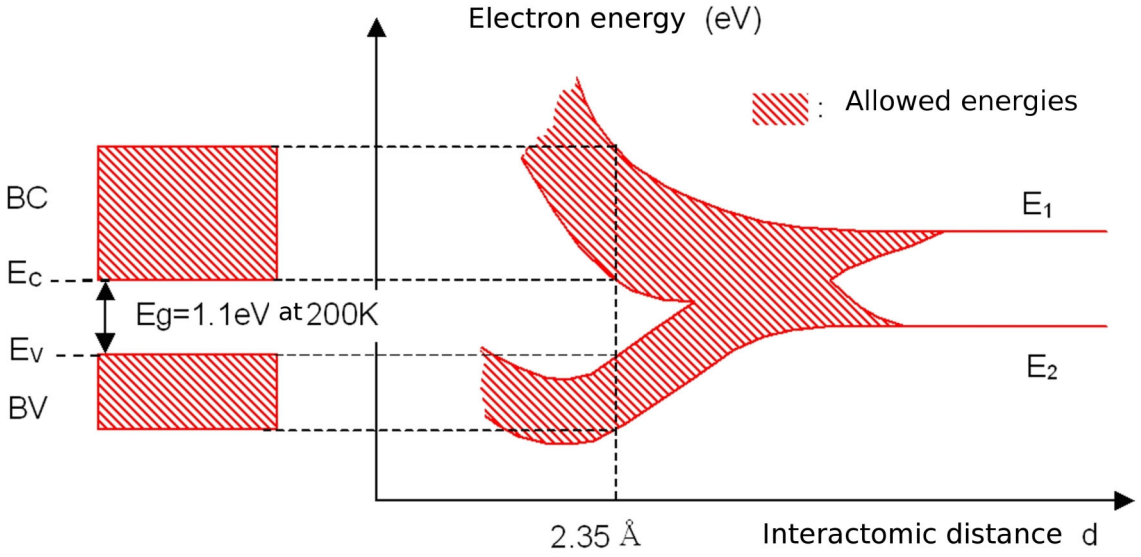
Research on semiconducting materials started in the early nineteenth century. Since then, many semiconductors have been investigated. Certainly the most well-known materials are Silicon **Si** and germanium **Ge**, which both belong to column IV of the periodic table of elements. While a bulk crystal of Si or Ge consists of the periodic arrangement of a single atom, other semiconductors like gallium arsenide **GaAs** (III-V) are built with two different elements: **Ga** (III) and **As** (V). Such composite semiconductors have electrical or optical properties that are not achievable with pure semiconductors made of only one type of atoms.

Before the bipolar transistor was invented in 1947, semiconductors were used in basically two types of electrical devices: photodiodes and rectifiers. In the nineteen fifties, germanium used to be the most frequently employed material. However, it could not be used in applications requiring weak current consumption and/or operation under high temperatures. Silicon has started overcome all other semiconductors since 1960, because it was both significantly cheaper and less power-consuming.

1. Energy Bands

Consider an isolated silicon atom; its energy levels are quantized (see the Bohr model for Hydrogen). When two identical atoms are brought closer together, the quantized energy levels hybridize and split into two different levels because of the mutual interaction of the two atoms. More generally, when N atoms are moved closer, until they reach the equilibrium inter-atomic distance d, the energy levels split into N levels. These N levels are very close to each other if N is large (which is the case in a crystal) so that they eventually form a continuous energy band.

Let's now consider silicon atoms arranged in a periodic lattice, but with a very large lattice parameter (or inter-atomic distance), in order to first consider each atom as isolated. The two levels with the highest energy are labeled E_1 and E_2. Now let's shrink homothetically the atom lattice: energy levels split and form two continuous bands known as the **conduction band** CB and the **valence band** VB, Figure 1 shows the formation of these bands as a function of the inter-atomic distance.

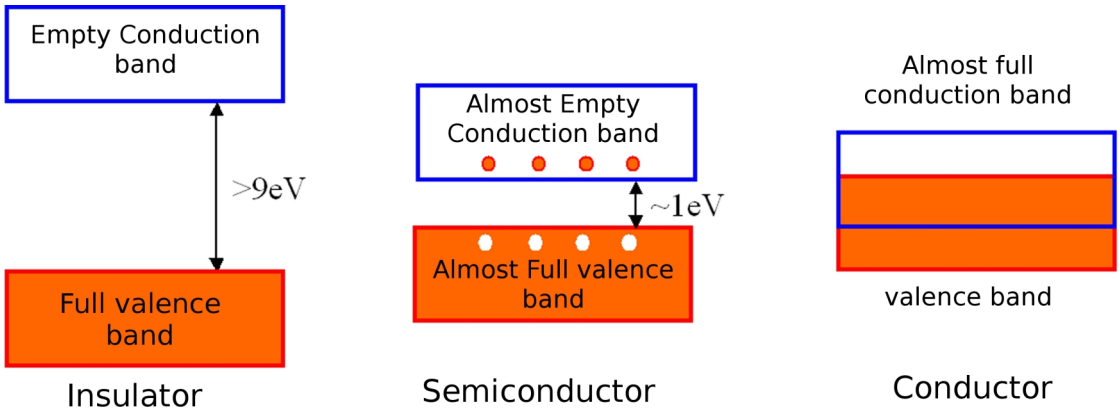


In a silicon crystal ($d_0 = 2.35\ \mathring{A}$), two continuous energy bands exist (CB and VB), separated by a forbidden band, which is not accessible for electrons. This forbidden region is called the « gap » and its width E_g is a characteristic of the material. The lowest energy level of the conduction band is denoted E_C and the highest energy level of the valence band is called E_V so that we have the relationship $E_g = E_C - E_V$. The conduction and valence bands CB and VB represent the energies accessible to electrons, or the energies of the states *potentially* occupied by electrons: they do not provide any information about the *effective* occupation of the energy states by electrons.

2. Insulator, semiconductor, conductor

Solid-state materials can be classified into three groups: insulators, semiconductors and conductors. Insulators are materials having an electrical conductivity $\sigma < 10^{-8} S/cm$ (like diamond: $10^{-14} S/cm$); semiconductors have a conductivity $10^{-8} S/cm < \sigma < 10^{3} S/cm$ (for silicon it can range from $10^{-5} S/cm$ to $10^{-3} S/cm$); at last conductors are materials with high conductivities : $10^{3} S/cm < \sigma$ (like silver: $10^{6} S/cm$.)

The electrical properties of a given material depend on the electronic populations of the different allowed bands. Electrical conduction is the result of electron motion within each band. When an electric field is applied to the material, electrons start to move in the direction opposed to the direction of the electric field. An empty energy band (in which there is no free electron) does not of course participate in the formation of an electric current. It is also the case for a *fully* occupied band. Indeed, an electron can move provided that, whenever it leaves its site, it can find some free space elsewhere (another available site within its energy band, called a "hole"), where it can go. A material with fully occupied or empty energy bands is then an **insulator**. This is the case when the gap energy exceeds ~ $9eV$, because for such gaps, the thermal energy at $300K$ (~ $25meV$) is clearly insufficient to allow electrons from the valence band to be promoted to the conduction band. In this case the valence band (and all bands of lower energy) is fully occupied, and the conduction band is empty.



A **semiconductor** is primarily an insulator at $0K$. However, since the energy gap is lower compared to insulators (~ $1eV$), the valence band is slightly thermally populated at room temperature, whereas the conduction band is slightly depopulated. Since electrical conduction is directly connected to the number of electrons in the "almost empty" conduction band and to the number of holes in the "almost fully occupied" valence band, it can be expected that the electrical conductivity of such an intrinsic semiconductor will be very small.

For a **conductor**, conduction bands and valence bands are not separated and there is therefore no energy gap. The conduction band is then partially occupied (even at low temperatures), resulting in a "high" electrical conductivity.

3. Intrinsic semiconductors

An intrinsic semiconductor is an undoped semiconductor. This means that holes in the valence band are vacancies created by electrons that have been thermally excited to the conduction band, as opposed to doped semiconductors where holes or electrons are supplied by a "foreign" atom acting as an impurity.

In order to better understand the behavior of semiconductors, we need to investigate more deeply what are the electron and hole densities in the conduction and valence bands, respectively. We first need to introduce the notion of energy **density ofstates** N(E). This parameter gives the number of states (per unit volume and per unit energy) between E and $E + dE$: $N_c(E)$ (respectively, $N_v(E)$) physically represents the "room" available for electrons (resp. holes) in the conduction band (resp. valence band). For energies that are close to the extrema of these two bands, the density of states has a quadratic dependence with E:

$$N_c(E) = \frac{1}{2\pi^2}\left(\frac{2m_c}{\hbar^2}\right)^{3/2}\sqrt{E - E_c}$$
$$N_v(E) = \frac{1}{2\pi^2}\left(\frac{2m_v}{h^2}\right)^{3/2}\sqrt{E_v - E} \quad [cm^{-3}/eV]$$

Where $\hbar = h/2\pi$ is the normalized Planck constant ($h = 6,626.10^{-34} Js$) and m_c (resp. m_v) is the average effective mass of the conduction band (resp. of the valence band). For a direct gap semiconductor, m_c (resp. m_v) is the effective mass of an electron m_e (resp. a hole m_h) in the crystal.

The above-mentioned concept of effective mass allows considering electrons (and holes) inside the crystal as almost free particles, like free quasi-particles. The semiconductor then becomes an electron (and hole) gas, but in which electrons and holes have an effective mass which may be very different of the mass of the particle moving in free space. For example, for GaAs $m_c/m_0 = 0,066$ avec $m_0 = 0,911.10^{-30} kg$ is the free electron mass.

In order to know what is the number of electrons and holes present in each band, the density of states is not the only information that we need. We also have to know the probability for an electron to occupy a level with a given energy E. This probability is given by the **Fermi-Dirac distribution function** :

$$f(E) = \frac{1}{1 + exp[(E - E_F)/kT]}$$

Where $k = 1,38.10^{-23} JK^{-1}$ is the Boltzmann constant, T the temperature and E_F the Fermi energy, which is the chemical potential for semiconductors.

The probability for a hole to occupy a level of energy E is simply given by $1 - f(E)$ since a hole is by definition the absence of an electron.

The **electron density** n $[cm^{-3}]$ in the conduction band is obtained by integrating, over the range of energies accessible by electrons in the band, the number of states that may be occupied by electrons of energy E, weighted by the probability to "find" an electron having this given energy :

$$n = \int_{E_c}^{+\infty} N_c(E) \cdot f(E) dE$$

Similarly, the **hole density** p $[cm^{-3}]$ in the valence band writes :

$$p = \int_{-\infty}^{E_v} N_v(E) \cdot (1 - f(E)) dE$$

For a semiconductor whose Fermi level E_F is located more than $3kT$ away from the extrema, the Fermi distribution function can be written under the form of a simple exponential, so that the expressions for the charge carriers densities become :

$$n = N_c exp\left[-\frac{E_c - E_F}{kT}\right] \quad \text{avec} \quad N_c = \int_{E_c}^{+\infty} N_c(E) \cdot exp\left[-\frac{E - E_c}{kT}\right] dE$$

$$p = N_v exp\left[\frac{E_v - E_F}{kT}\right] \quad N_v = \int_{-\infty}^{E_v} N_v(E) \cdot exp\left[\frac{E - E_v}{kT}\right] dE$$

Where N_c et N_v are called the *effective densities of states*. They represent the density of "useful" states, at a given temperature T, in their respective energy band.

Let's note that the product of the two densities turns out to be independent on the position of the Fermi level. This is still true even for extrinsic semiconductors (see below). It is known as the *law of mass action* since it resembles the law of mass action encountered in chemistry (for instance the auto-ionization equilibrium of water is governed by $[H^+][OH^-]=K_e$).

$$np = n_i^2 \text{ avec } n_i = \sqrt{N_c N_v} exp\left[-\frac{E_c - E_v}{2} kT\right]$$

Where n_i is the density of intrinsic carriers (for Si at $300K$, $n_i = 10^{10} cm^{-3}$).

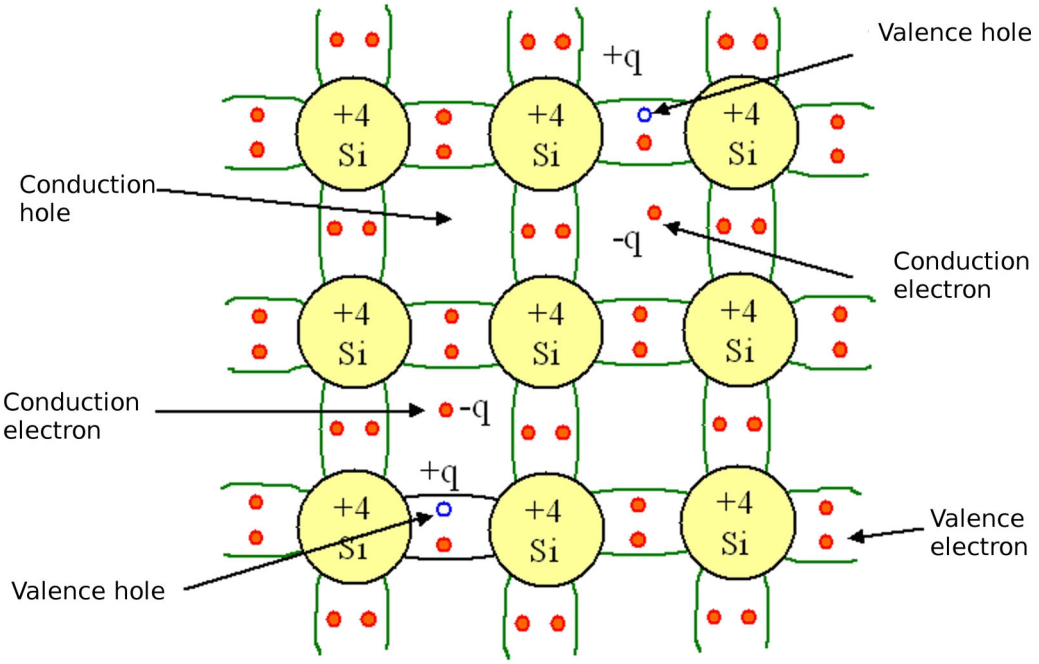


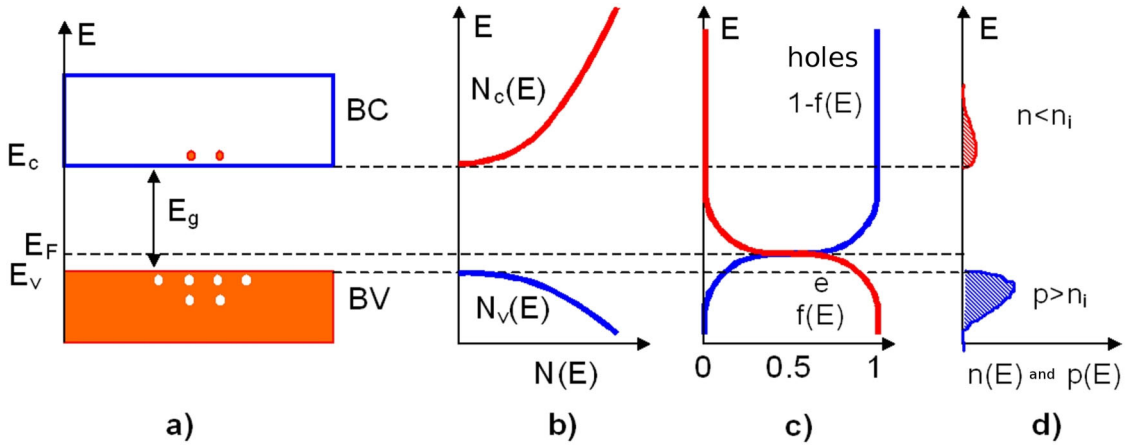
Figure 3 shows that for an intrinsic semiconductor (without impurities), each electron in the conduction band is associated with a hole in the valence band. We conclude that the electron and hole densities are equal :

$$n = p = n_i$$

By substituting the carrier densities by their respective expressions, this relationship allows defining **the Fermi level for an intrinsic semiconductor** E_{Fi}. Since at room temperature kT is significantly lower than the energy gap, this level is located near the middle of the forbidden band :

$$E_{Fi} = \frac{E_c + E_v}{2} + \frac{kT}{2} \ln \frac{N_v}{N_c} \cong \frac{E_c + E_v}{2}$$

In figure 4 are graphically summarized the important points raised above for intrinsic semiconductors.

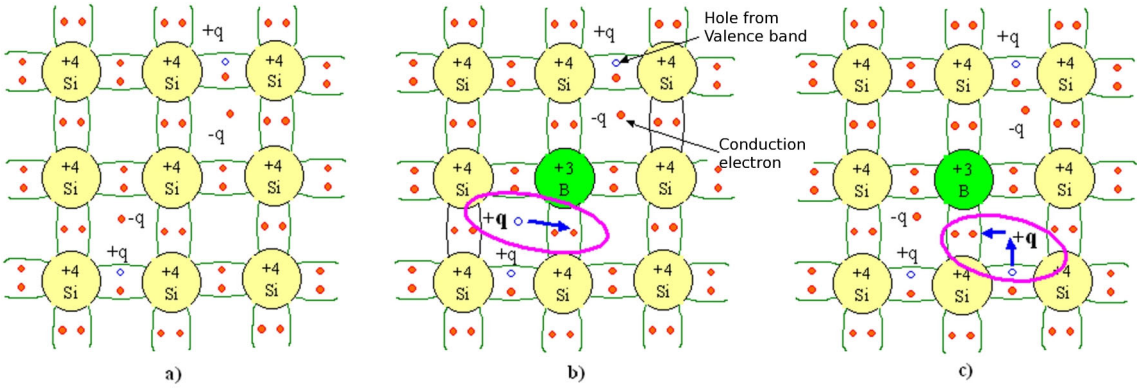


4. Doped (extrinsic) semiconductors

An extrinsic semiconductor is a semiconductor doped by a specific impurity which is able to deeply modify its electrical properties, making it suitable for electronic applications (diodes, transistors, etc.) or optoelectronic applications (light emitters and detectors).

4.1. P-type semiconductors

A **P-type semiconductor** is an intrinsic semiconductor (like Si) in which an impurity acting as an *acceptor* (like e.g. boron B in Si) has been intentionally added. These impurities are called acceptors since once they are inserted in the crystalline lattice, they lack one or several electrons to realize a full bonding with the rest of the crystal.



From figure 5, we see that a p-type semiconductor has a lower electron density n and a higher hole density p than the same intrinsic semiconductor. Electrons are said to be the **minority carriers** whereas holes are the **majority carriers**.

For extrinsic semiconductors, the dopant density is always far higher than the intrinsic carrier density : $N_A >> n_i$. In the case of a p-type material, the hole density is then close to the dopant density N_A. Since the law of mass action is always true, we obtain the following expressions for the carrier densities

$$n = \frac{n_i^2}{N_A}$$

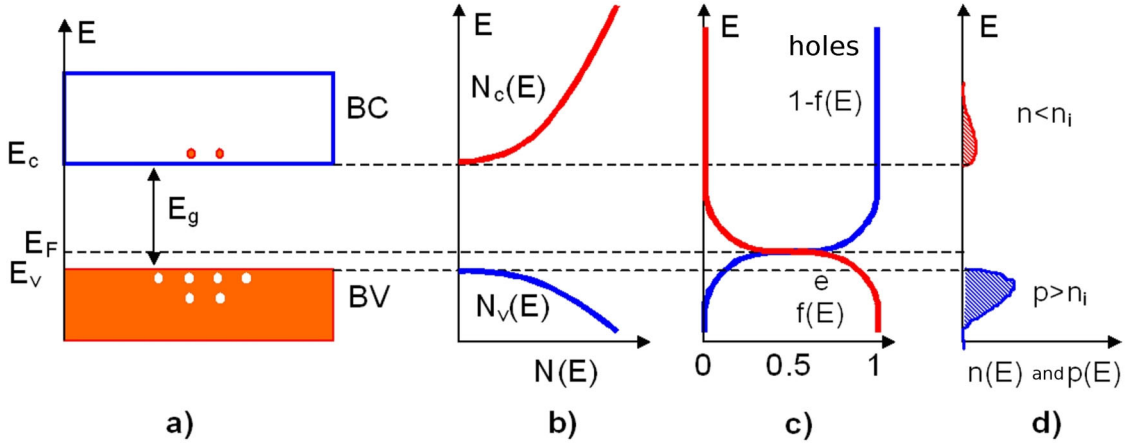
$$p = N_A$$

The **Fermi level for a p-type semiconductor** or chemical potential is then :

$$E_{Fp} = E_v + kT \ln \frac{N_v}{N_A}$$

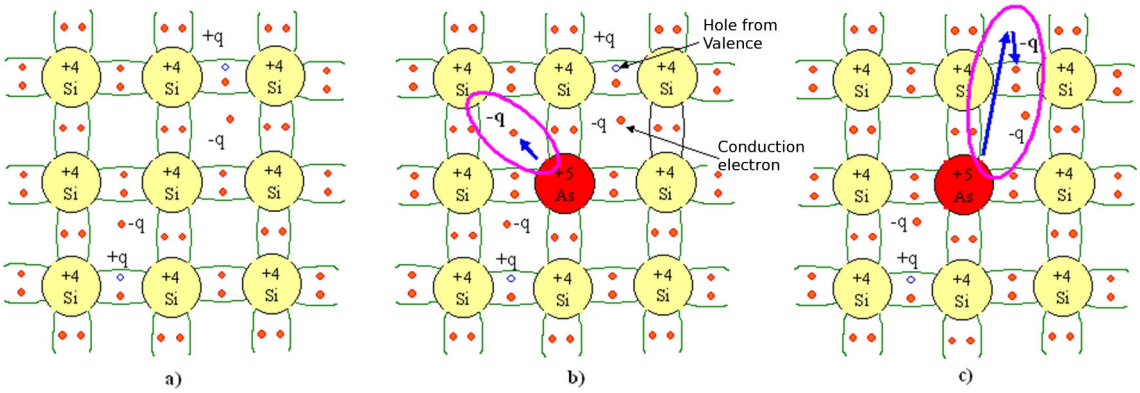
When the acceptor density is increased, the Fermi level moves closer to the edge of the valence band. If $N_A = N_v$ the Fermi level enters the valence band, the semiconductor is then said to be degenerate.

The important points regarding p-type semiconductors are summarized graphically in figure 6.



4.2. N-type semiconductors

A **N-type semiconductor** is an intrinsic semiconductor (e.g. silicon **Si**) in which a donor impurity (e.g. arsenic **As** in Si, or Si in GaAs) has been intentionally introduced. The impurities are called donor impurities since they have to give an extra electron to the conduction band in order to make all the bonds with neighboring atoms (As is pentavalent while Si is tetravalent).



From figure 7, we see that a n-type semiconductor has a higher electron density n and a lower hole density p than the same intrinsic semiconductor. Holes are said to be the **minority carriers** whereas electrons are the **majority carriers**.

Like in p-type semiconductors, we can write the following relationships, where N_D is the donor density :

$$n = N_D$$

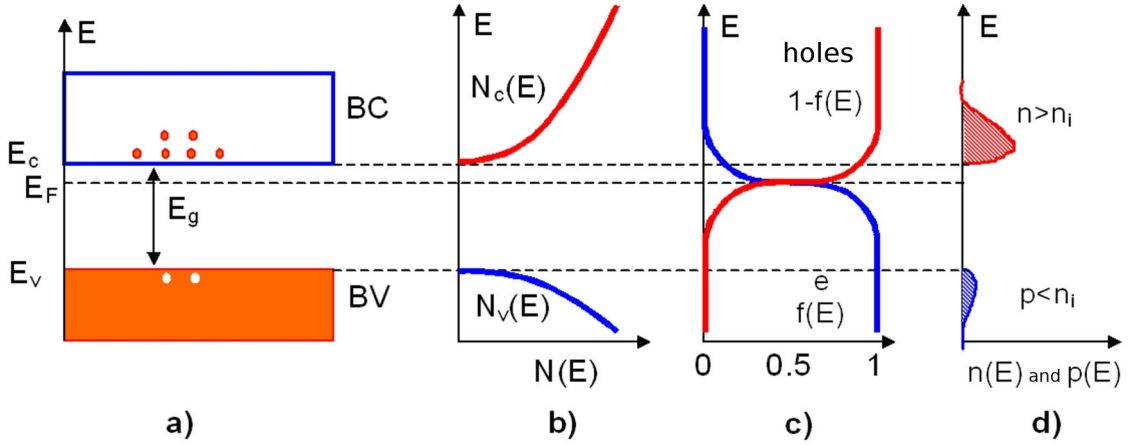
$$p = \frac{n_i^2}{N_D}$$

The **Fermi level for a n-type semiconductor** is then :

$$E_{Fn} = E_c - kT \ln \frac{N_c}{N_D}$$

When the donor density is increased, the Fermi level moves closer to the edge of the conduction band. If $N_D = N_c$ the Fermi level enters the conduction band, the semiconductor is then said to be degenerate.

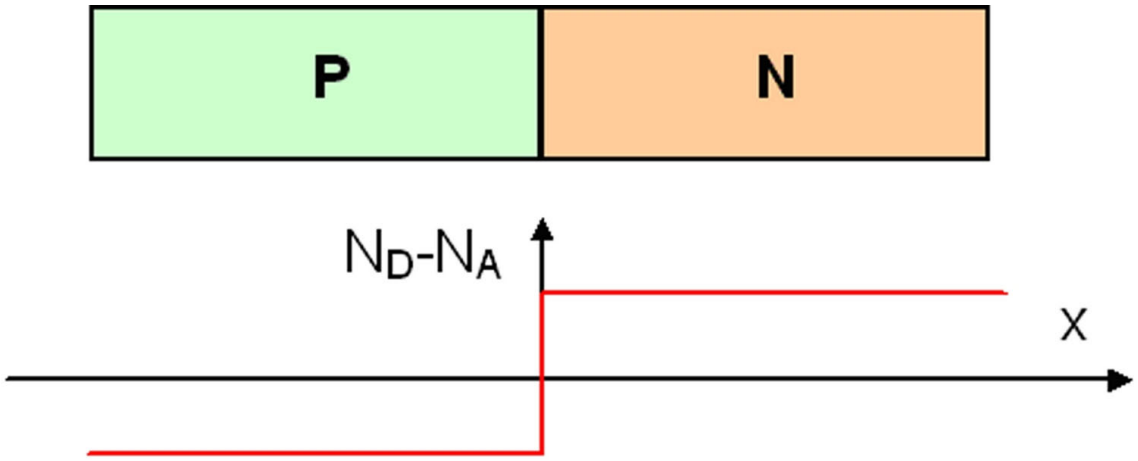
The important points regarding n-type semiconductors are summarized graphically in figure 8.



III.Case study

1. Abrupt PN junction at thermal equilibrium

A **PN junction** is the juxtaposition of a n-type and a p-type piece of semiconductor, taken originally from the same block of crystal. The difference between the densities of donors and acceptors $N_D - N_A$ undergoes a very sharp variation from a negative value in the P region to a positive value in the N region. An abrupt junction is by definition a junction in which the doping type changes over a very small distance compared to the spatial extent of the depletion region (see below.)



There are some other types of junctions (exponential, linear junctions...). However we will only focus here on the abrupt junction, because results are much simpler and can be easily generalized to more complex cases.

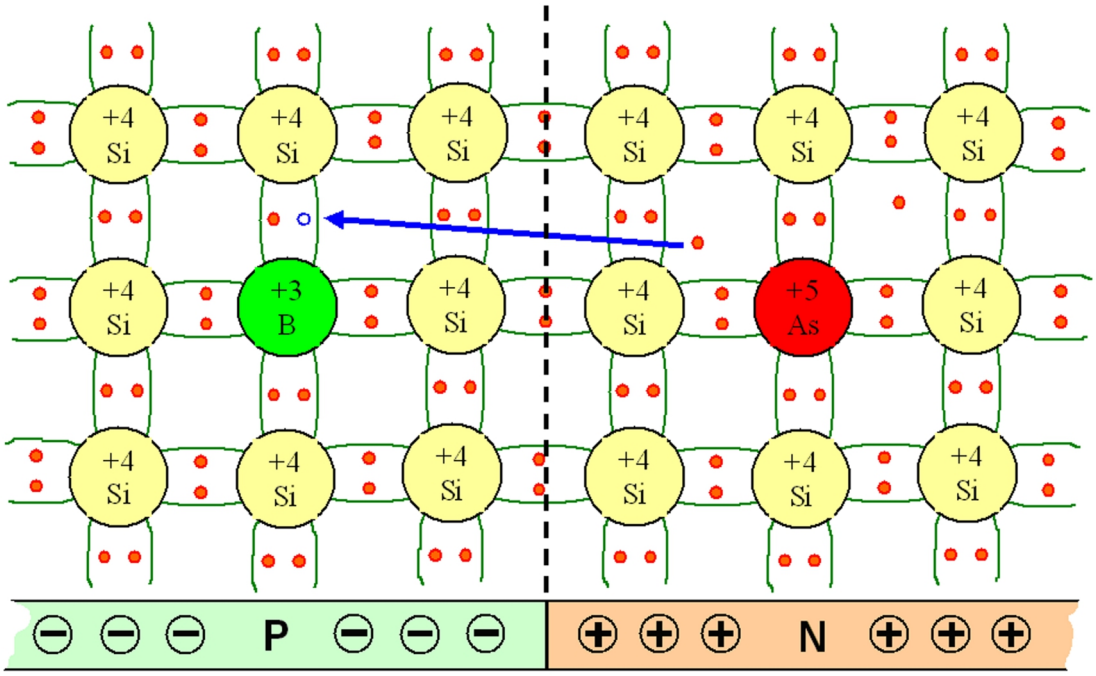
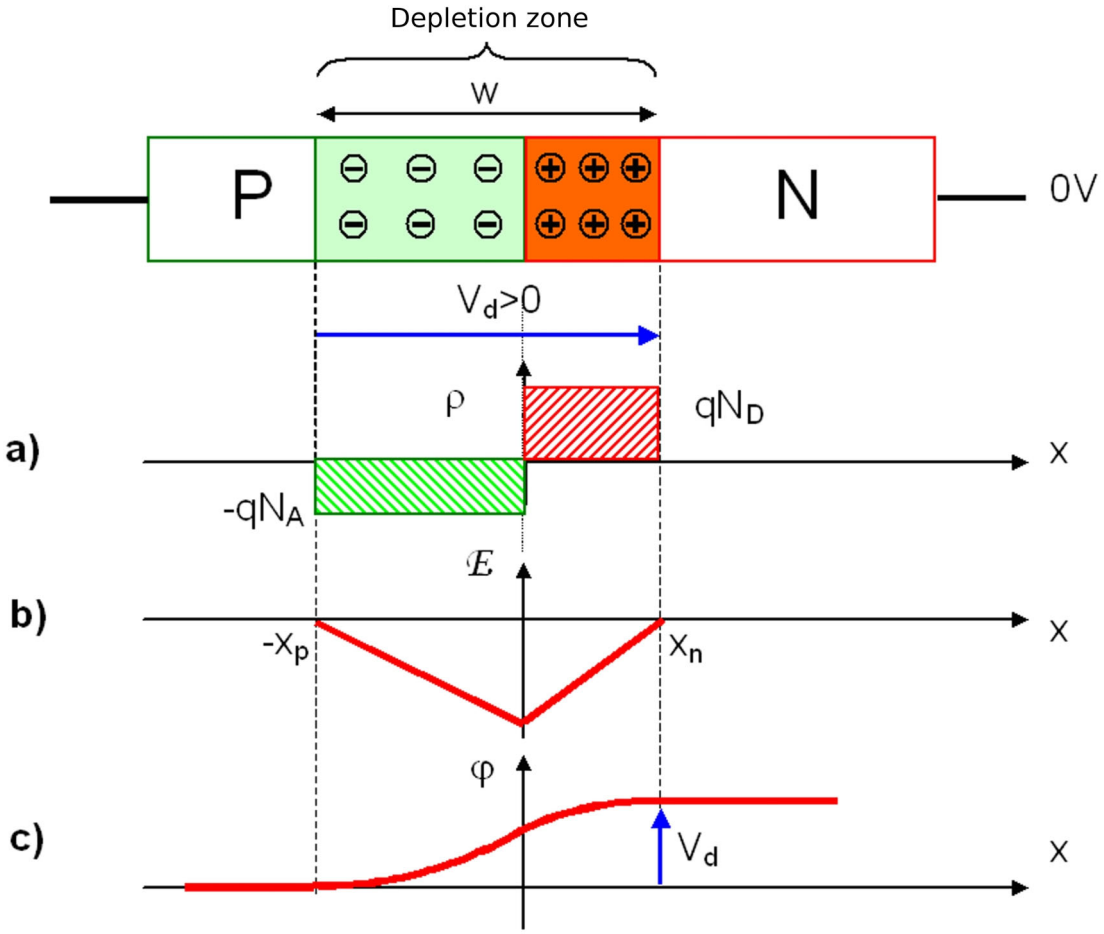


Figure EC2 enables understanding what happens when a N-type semiconductor is in contact with a P-type semiconductor. The electrons, abundant in the N-type region, will cross the

junction in order to recombine with holes present in the P-type region. Therefore, a negative static space charge builds up in the P region (the trivalent boron atoms have a static negative charge since they release a positively-charged hole in the valence band) whereas a positive space charge builds up in the N-type zone for the same reasons. The small volume in which this space charge is created is called **the spacecharge zone** or **depletion zone**. Because there is a strong electric field in this area, the density of free carriers is negligible at thermal equilibrium. In addition, the frontier between the depletion zone and the neutral parts of the junction is very sharp.



Just after the P- and N-type semiconductors have been drawn closer, a potential barrier appears for both holes an electrons. Indeed, the static space charges accumulated at the borders of the junction (positive charges in the N zone, negative charges in the P zone) create an electric field pointing from N towards P, which prevents diffusion and further recombination of electrons and holes: diffusion is then rapidly stopped by the buildup of this internal electric field. In addition, because of the existence of this double layer of charges on both sides of the junction, the electrostatic potential varies sharply inside the depletion zone and the potential difference V_d, called the diffusion potential (or built-in potential) reaches non-negligible values (e.g. $0.8V$ for Si). However, if we attempt to measure a voltage by connecting a multimeter at both sides of the PN diode, we will find a zero voltage drop, because a multimeter measures a difference in electrochemical potentials, not just electrostatic potentials. Indeed, here the electrochemical potential is constant throughout the crystal, including in the space charge zone, because this potential takes into account not only the electric field but also the concentration gradient of charge carriers. The diffusion potential due to this gradient compensates exactly for the electrostatic potential.

Rappel

The relationship between the space $\rho[C \times cm^{-3}]$ charge density , the electric field E and the electrostatic potential φ is :

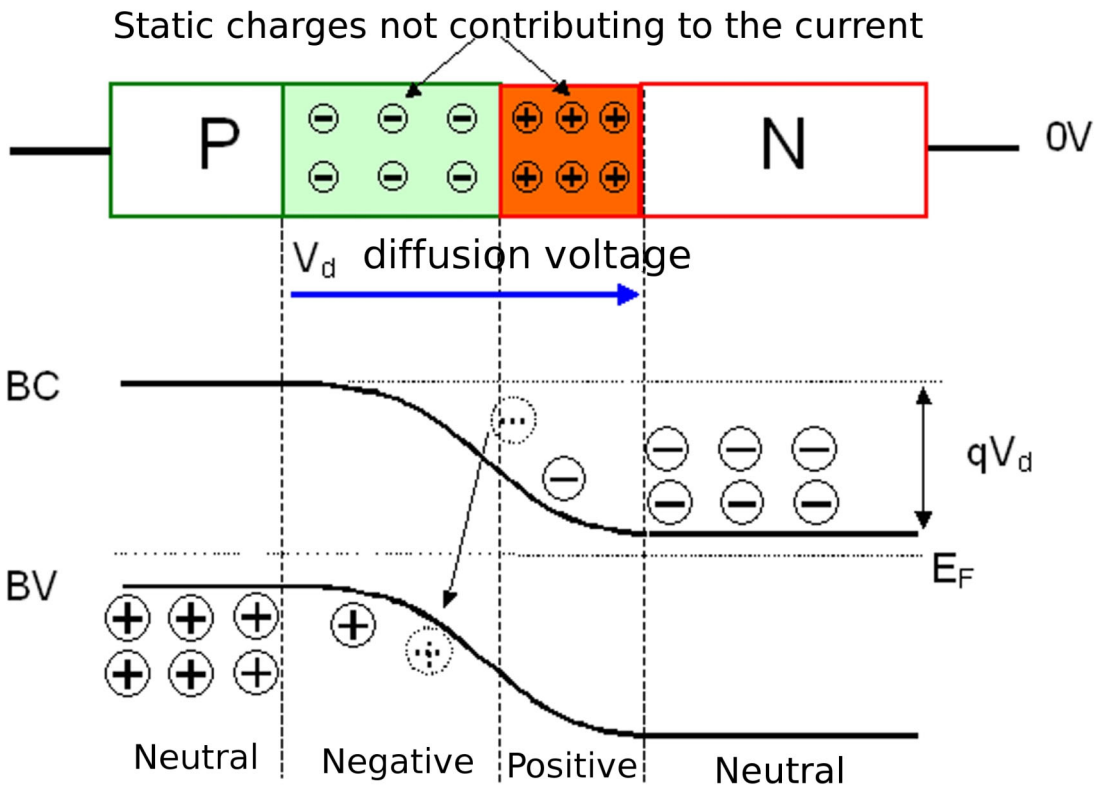
$$\frac{d^2\varphi}{dx^2} = \frac{-dE}{dx} = \frac{-\rho}{\epsilon_s}$$

Where ϵ_s is the permittivity of the medium ($10^{-10} F/m$ for silicon).

The (chemical) potential of a semiconductor being equal to its Fermi energy, the built-in potential or **diffusion potential** is proportional to the difference of the Fermi energies of the two unbounded semiconductors :

$$V_d = \frac{1}{q}(E_{Fp} - E_{Fn}) = \frac{kT}{q} \ln\left[\frac{N_A N_D}{n_i^2}\right]$$

When the junction meets thermal equilibrium, the Fermi energy has a constant value throughout the whole device. The energies of conduction and valence bands are therefore shifted up or down, and exhibit a smooth variation across the depletion region. As a consequence, there is an electrostatic potential energy difference appearing between the P and N region, equal to qV_d.



2. Abrupt junction under external bias

2.1. Current density

In order to describe the behavior of a semiconductor out of thermal equilibrium (upon application of an external voltage), we have to estimate the electrical currents resulting from the motion of charge carriers. The latter, electrons and holes, will move because either an electric field is applied, or because a concentration gradient of charge carriers exists. In the first case, the current is called a conduction current, while in the second case it is called a diffusion current. Moreover, it is more relevant to consider a *current density* J (which is a current, *i.e*. an amount of charges per unit time, per unit area) instead of the current itself.

When electrons and holes are submitted to an electric field created by an applied voltage, their flow generates a **conduction current** :

$$J_{n.c.} = n \cdot q \cdot \mu_n E$$
$$J_{p.c.} = p \cdot q \cdot \mu_p E$$

Where $n(p)$ is the electron (hole) density, $q = 1,602.10^{-19} C$ is the electron elementary charge, E is the electric field of the polarized junction, and μ_n et μ_p are the mobilities of electrons and holes, respectively.

Besides, when electron and hole densities do not have constant values throughout the semiconductor, there is a diffusion flow that tends to flatten the density profile. This **diffusioncurrent** is proportional to the concentration gradient :

$$J_{n.d.} = q \cdot D_n \cdot \frac{\partial n}{\partial x}$$
$$J_{n.d.} = -q \cdot D_p \cdot \frac{\partial p}{\partial x}$$

Where D_n and D_p are the diffusion constants of both types of carriers.

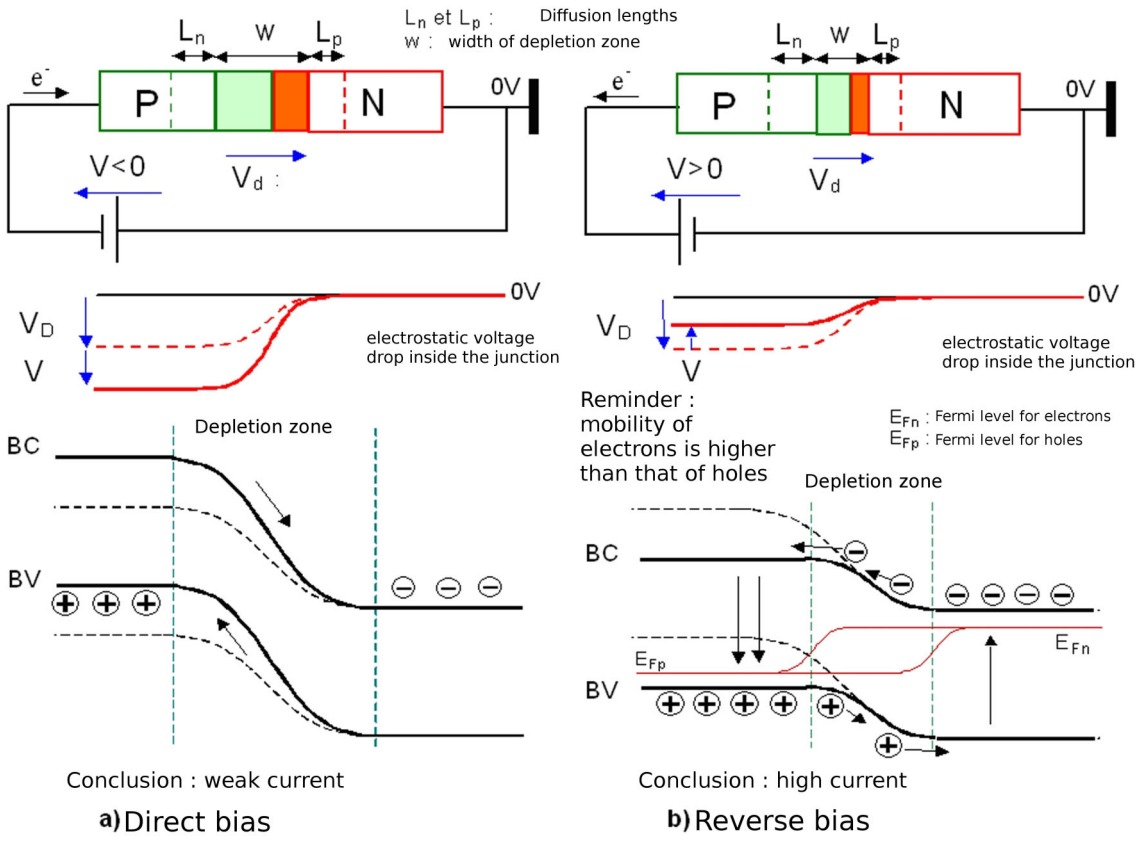
The electron mobility being higher than that of holes, the Einstein relationship shows that, for a given concentration gradient, the diffusion current for electrons is higher than the diffusion current for holes.

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{q}$$

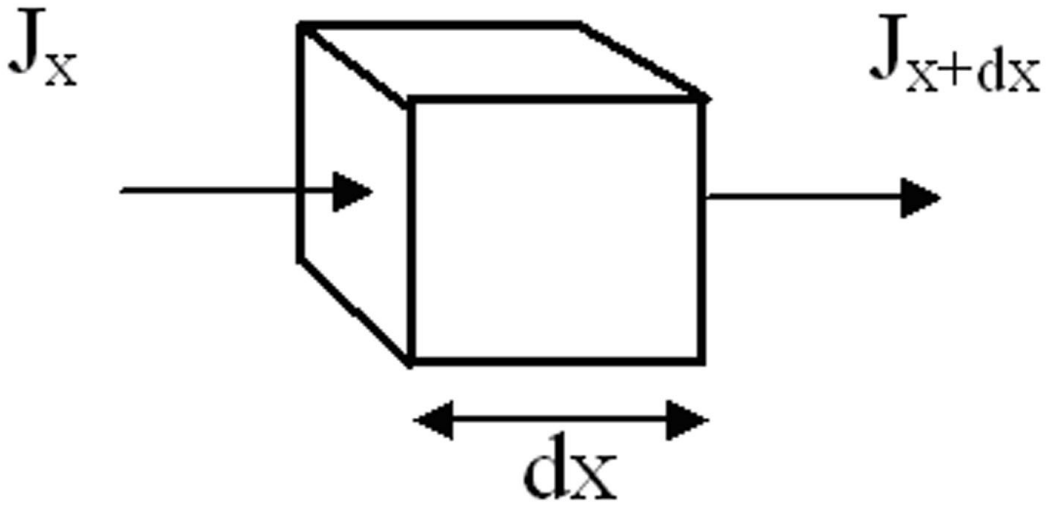
It is obvious that the total current is constant throughout the whole device. Then, in an attempt to estimate this current, we may choose a region in which calculations will be as simple as possible. This region corresponds to the neutral parts of the N and P sections. Indeed, the electric field E, in the low injection regime, is negligible out of the depletion zone, in virtue of the non-zero conductivity of semiconductors (if the field was not zero, free charges would escape and the region would not be neutral any more.) The total current is therefore only the combination of the hole and electron diffusion currents. But, in neutral zones (out of the depletion zone), the spatial distribution of majority carriers (electrons in the N part, holes in the P part) is constant. Therefore, since diffusion currents are proportional to carrier concentration gradients, the total current turns out to be generated by minority carriers, i.e. electrons in the P side and holes in the N side. The expression of the **total current density** is then :

$$J = J_{n.d.}(x_p) + J_{p.d.}(x_n)$$

Where x_p (resp. x_n) is the limit of the depletion zone in the P side (resp. N side.)



In order to calculate J as a function of the external voltage V, we have to estimate the minority carriers densities n_p ($= n(x_p)$) and p_n ($= p(x_n)$) in the neutral zones. To do so, we have to write the continuity equations giving the evolution of the carrier densities versus time.



Let's consider an infinitesimally small volume of semiconductor of unit section area and length dx. Let n_p (resp. p_n) be the electron density in the P side (resp. the hole density in the N side.) The number of carriers in the small volume can change because the current density is not homogeneous in space (there are more or less charges exiting the volume than entering the same volume: the variation per unit time is equal to $\frac{1}{q}\frac{\partial J_x}{\partial x}$), or because carriers have recombined (which is described by a lifetime τ_n or τ_p). We do not allow carriers to be generated here. We then obtain the following **charge conservation equations** :

$$\frac{dn_p}{dt} = \frac{1}{q}\frac{\partial J_{n.d.}}{\partial x} - \frac{n_p - n_{p0}}{\tau_n}$$

$$\frac{dp_n}{dt} = \frac{1}{q}\frac{\partial J_{n.d.}}{\partial x} - \frac{n_p - p_{n0}}{\tau_p}$$

Where $n_{p0} = n(x_p)$ (resp. $p_{n0} = p(x_n)$) is the electron density (resp. hole density) in the P side (resp. N) for the passive junction, and τ_n,τ_p are the lifetimes of the carriers in neutral regions.

The **steady-state continuity equations write** :

$$\frac{\partial^2 n_p}{\partial x^2} - \frac{n_p - n_{p0}}{L_n^2} = 0$$

$$\frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{n0}}{L_p^2} = 0$$

Where $N_{n/p} = \sqrt{D_{n/p} \cdot \tau_{n/p}}$ is the diffusion length of electrons/holes.

The expression of the diffusion potential (see above) can be written as a relationship between the densities of minority carriers (n_{p0} and p_{n0}) and majority carriers n_{n0} (= N_D) and p_{p0} (= N_A), for the junction at thermal equilibrium :

$$n_{n0} = n_{p0} \cdot exp\left[\frac{qV_d}{kT}\right]$$

$$p_{p0} = p_{n0} \cdot exp\left[\frac{qV_d}{kT}\right]$$

When the junction is connected to an external forward voltage source V, the electrostatic potential difference becomes V_d–V. Like in the junction at thermal equilibrium, the majority and minority carriers densities are linked by :

$$n_n = n_p \cdot exp\left[\frac{q(V_d - V)}{kT}\right]$$

$$p_p = p_n \cdot exp\left[\frac{q(V_d - V)}{kT}\right]$$

In the low injection regime, the densities of majority carriers remain almost constant whatever the forward voltage applied to the diode; however it is not the case for minority carriers :

$$n_p - n_{p0} = n_{p0} \cdot \left(exp\left[\frac{qV}{kT}\right] - 1\right)$$

$$p_n - p_{n0} = p_{n0} \cdot \left(exp\left[\frac{qV}{kT}\right] - 1\right)$$

We are now able to solve the steady-state differential continuity equations; for this purpose boundary conditions have to be determined. Applying an external bias on the device cannot modify the density of minority carriers far from the junction, which can be written as : $n_p(x = -\infty) = n_{p0}$ and $p_n(x = +\infty) = p_{n0}$. The variations in minority carriers density can be written as a function of the position x inside the junction, out of the depletion zone :

$$n_p - n_{p0} = n_{p0} \cdot (exp\left[\frac{qV}{kT}\right] - 1) exp\left[\frac{(x - x_p)}{L_n}\right]$$

$$p_n - p_{n0} = p_{n0} \cdot (exp\left[\frac{qV}{kT}\right]-1)exp\left[\frac{(-(x-x_n)}{L_p}\right]$$

Since the current is constant through the whole device, it can be calculated for instance at the boundaries of the depletion zone. The total **current density across the diode is** :

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

With

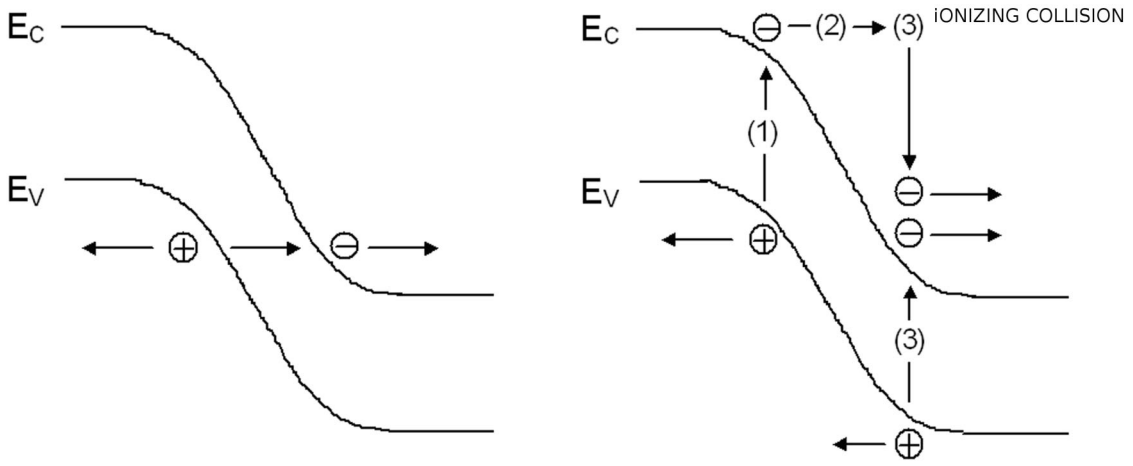
$$J_s = \frac{qn_{p0}D_n}{L_n} + \frac{qp_{n0}D_p}{L_p}$$

2.2. PN junction under reverse bias

Under **reverse bias** ($V < 0$), whatever the weak voltage applied to the junction, the total current is fixed to $-J_s$. This current is called the saturation current. However, for high reverse voltages, the reverse current can suddenly increase. It is then possible to reach the **breakdown voltage** of the junction, denoted V_c. Indeed, when the reverse voltage is increased, the electric field is increased accordingly inside the junction. But the electric field cannot exceed a given value E_0. This is because as the electric field increases, so does the electrostatic force $\overrightarrow{F} = -q \cdot \overrightarrow{E}$ acting on the electrons linked to the crystalline network; this force can at some point become so high that it exceeds the binding force linking the valence electrons to their respective nuclei. As these electrons are released, the crystal becomes conductive and neither the electric field nor the voltage can increase any more. This means that the maximum electric field that can be applied to a semiconductor crystal is the field that would lead to the direct excitation of an electron from the valence band to the conduction band, or stated differently the field leading to the material ionization.

This breakdown effect can be related to two different physical phenomena. The first one is a tunneling effect called the **Zener effect**. The presence of a high electric field ($\sim 10^6 V/cm$ for silicon) creates electron-hole pairs. The electrons associated to these pairs are emitted through the depletion zone, from the valence band to the conduction band, without loss or gain in energy, hence the term of tunneling used for this effect. In practice, one may observe the Zener effect only for highly-doped PN junctions, in which the space charge zone is narrow $(w =\sim 500\mathring{A})$ so that the "tunnel" is short enough to be crossed.

When the depletion region is not so narrow, typically for $w > 1000\mathring{A}$, another phenomenon called the **avalanche breakdown effect** arises and provokes the junction breakdown well before the Zener effect can be observed. For electric fields of the order of $10^5 V/cm$, which is about 10 times smaller than the threshold for the Zener effect, the acceleration of some carriers is sufficient to enable the generation of electron-hole pairs by collisions with the atoms of the crystal network. These pairs are then accelerated and provoke the creation of other pairs. The result is a chain reaction creating an avalanche effect. This process is illustrated in figure EC7. It can be described as follows: phase (1) corresponds to the thermal generation of an electron-hole pair ; during phase (2) the electron is accelerated by the electic field and is subsequently promoted to a higher lying level within the conduction band, where it is referred to as a "hot carrier"; in phase (3) the electron kinetic energy is high enough to create another electron-hole pair thanks to collisions: after the impact, called the ionization impact, the electron having lost its energy is brought back to the bottom of the conduction band while a second electron-hole pair is created. If the thickness of the depletion band is large enough the process can repeat itself. The process described here for electrons can also occur with holes.



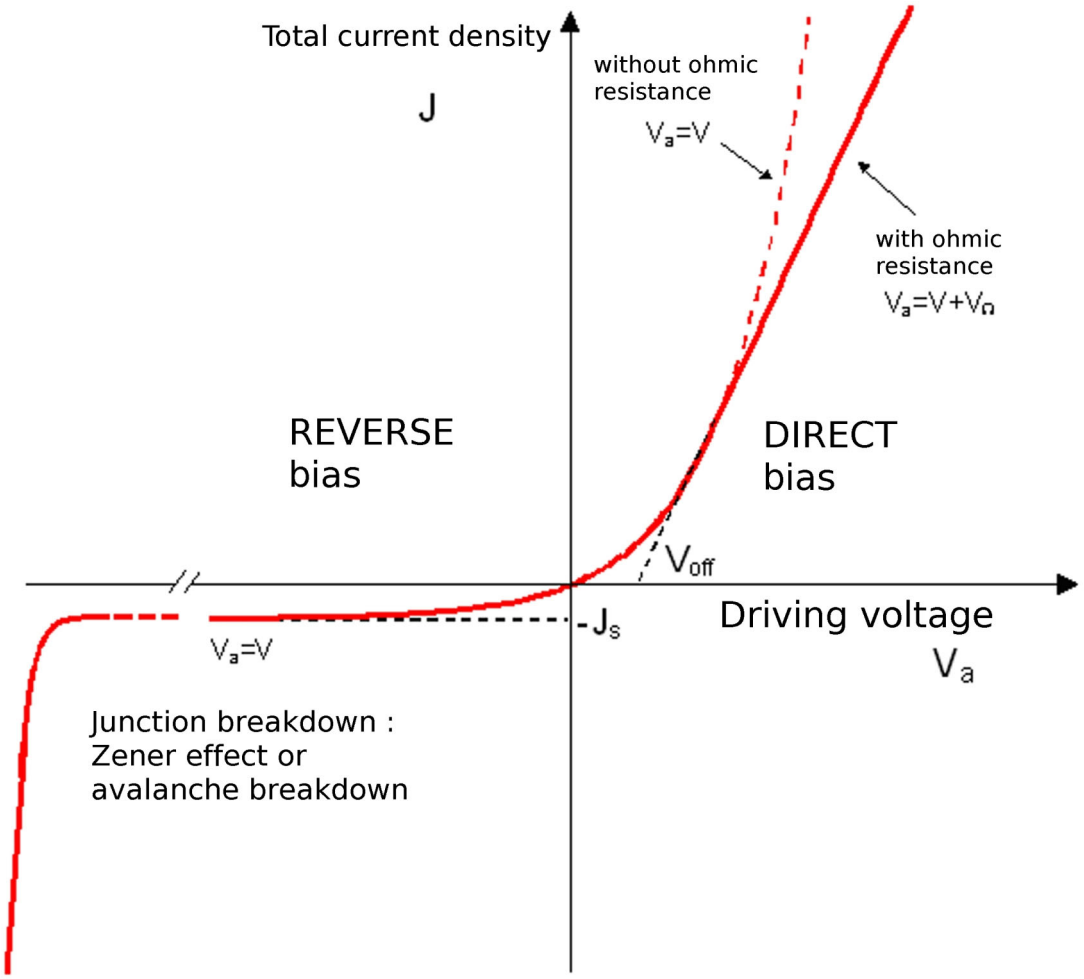
a) Tunneling b) Avalanche effect

2.3. Junction under DC forward bias

Under a DC **forward bias** ($V > 0$) and within the framework of the weak injection model, the total current flowing through the diode is an exponential function of the applied bias V. However, if the forward bias is very high, the ohmic resistances of the two $N-$ and $P-$ doped parts are not negligible any more. It is no longer possible to assume that the voltage applied to the device (noted V_a) does not induce a voltage drop in these doped regions; instead it must be written as $V + V_\Omega$ (cf. p 23)where V is the voltage required to lower the junction potential barrier to a value around kT/q while $V_\Omega = R \cdot J$ represents the voltage drop due to the ohmic resistance of the N and P regions.

2.4. Current-Voltage characteristic

We can now plot the current density-versus-voltage (J-V) curve for a PN junction, taking into account the special features of reverse and direct bias that we have discussed before (fig. EC8). This curve shows the existence of an "offset" (or "threshold") voltage V_{off} in the case where the ohmic resistance of the semiconductors is weak.



2.5. Junction under small-signal AC voltage , diffusion capacitance

Let's consider the case where the bias applied to the diode consists of a **DC direct voltage** $V_0 > 0$ superimposed with a small-signal AC component with a weak amplitude ΔV , at frequency f. The ohmic resistance of the doped semi-conductors is neglected (regime of weak injection.) This voltage creates a current which can be written as a combination of a DC current J_0 and a AC current of amplitude ΔJ (at the same frequency), provided that a linear approximation of the J-V characteristic can be done around the (V_0 , J_0) point.

$V = V_0 + \Delta V \cdot exp(j\omega t)$

And

$J = J_0 + \Delta J \cdot exp(j\omega t)$

Where $\omega = 2\pi f$ is the modulation angular frequency.

For low modulation frequencies such as $\omega\tau << 1$ where $\tau = \tau_n$ or τ_p is the carrier lifetime, the complex admittance $y = \frac{\Delta J}{\Delta V}$ is :

$y = g_d + j \cdot C_d \cdot \omega$ with :

$$g_d = \frac{q^2}{kT} exp\left[\frac{qV_0}{kT}\right] \cdot \left(\frac{p_{n0} \cdot D_p}{L_p} + \frac{n_{p0} \cdot D_n}{L_n}\right)$$

$$C_d = \frac{q^2}{2} kT exp\left[\frac{qV_0}{kT}\right] \cdot (p_{n0} \cdot L_p + n_{p0} \cdot L_n)$$

Where g_d is the **diffusion conductance** , C_d is the **diffusion capacitance**.

This expression shows that the amplitude of the AC component varies exponentially with the DC component. Furthermore, there is a phase shift between the AC voltage and the AC current, modelled by a diffusion capacitor. This capacitance is due to the high finite mobility of carriers. Indeed, electrons and holes have a nonzero effective mass (which depends on the nature of the crystal: for instance in GaAs $m_e = 0.06.10^{-30} kg$); under the AC voltage, due to their quite high mobility they can reach reasonably high velocities. When the voltage is reversed rapidly the carriers cannot respond instantaneously to the driving voltage because of their mass or inertia. Therefore the total generated current is phase shifted with respect to the voltage.

For higher modulation frequencies such as $\omega\tau << 1$, the complex admittance becomes :

$y = g_d + j(\omega) + j \cdot C_d(\omega) \times \omega$ with :

$$g_d(\omega) = \frac{q^2}{kT\sqrt{2}} e^{\frac{qV_0}{kT}} \cdot (p_{n0} \cdot \sqrt{D_p} + n_{p0} \cdot \sqrt{D_n}) \sqrt{\omega}$$

$$C_d(\omega) = g_d / \omega$$

IV.Exercises

1. Exercise n°1

Intrinsic semiconductor

Consider an intrinsic semiconductor in which the effective density of states is N_C in the conduction band and N_v in the valence band.

Question 1

[Solution n°1 p 24]

Give the expressions of the electron density n (hole density p) in the conduction band (valence band, resp.).

Question 2

[Solution n°2 p 24]

Use previous result to derive the intrinsic density n_i and the position of the intrinsic Fermi level E_{Fi}.

The semiconductor considered here is Silicon, with an energy gap $E_g = 1,1eV$ and with $N_C = 2,7.1019cm^{-3}$ and $N_V = 1,1.10^{19}cm^{-3}$.

Question 3

[Solution n°3 p 24]

Calculate the intrinsic density and the Fermi level position for the following temperatures: $27°C$, $127°C$ and $227°C$. Hint: at $300K$, $kT = 0.026eV$; the energy reference will be taken at the highest occupied level of the valence band ($EV = 0eV$).

Extrinsic semiconductor

Silicon is doped with phosphorus atoms (column V of Mendeleev table) with a concentration of $10^{18}cm^{-3}$.

Question 4

[Solution n°4 p 24]

What is, at $27°C$, the electron density in doped Si ? Use this result to derive the hole density. Which type of semiconductor is obtained ?

Question 5

[Solution n°5 p 25]

Calculate, at $27°C$, the position of the Fermi level E_F and plot the band diagram.

2. Exercise n°2

PN junction

The junction is made from Silicon.

Question 1

[Solution n°6 p 25]

Show that the depletion zone is wider on the lower-doped side.

Question 2

[*Solution n°7 p 26*]

Calculate the diffusion voltage V_d at $T = 300K$ as a function of the dopant densities $N_A = 10^{19}cm^{-3}$, $N_D = 10^{16}cm^{-3}$ and the intrinsic carriers density $n_i \approx 10^{10}cm^{-3}$ $(300K)$. Data : the Boltzamnn constant $k = 1,38.10^{-23}JK^{-1}$ and the electron charge : $q = 1,602.10^{-19}C$.

Question 3

[*Solution n°8 p 27*]

Give the expression of the depletion zone width w versus N_A, N_D, V_d and the permittivity $\varepsilon_s = 10^{-10}F/m$. Estimate this width at $300K$ with the numerical values given in question 2.

Ressources annexes

-

$$V + V_{\Omega}$$

Solution des exercices

>Solution n°1 *(exercice p. 21)*

The expressions for n and p are :

$$n = N_c exp\left[-\frac{E_c - E_F}{kT}\right]$$

$$p = N_v exp\left[-\frac{E_v - E_F}{kT}\right]$$

>Solution n°2 *(exercice p. 21)*

For an intrinsic semiconductor $n = p = n_i$. Then the intrinsic carrier density is :

$$n_i = N_c exp\left[-\frac{E_c - E_{Fi}}{kT}\right] = N_v exp\left[-\frac{E_v - E_{Fi}}{kT}\right] \Leftrightarrow n_i = \sqrt{N_c N_v} exp\left[-\frac{E_c - E_v}{2kT}\right]$$

The Fermi level can be obtained by writing:

$$\frac{n}{p} = 1 \Leftrightarrow \frac{N_c}{N_v} exp\left[-\frac{E_c - E_v - 2E_{Fi}}{kT}\right] = 1$$

$$\Leftrightarrow \frac{-E_c + E_v + 2E_{Fi}}{kT} = \ln\frac{N_v}{N_c}$$

$$\Leftrightarrow E_{Fi} = \frac{E_c - E_v}{2} + \frac{kT}{2}\ln\frac{N_v}{N_c}$$

>Solution n°3 *(exercice p. 21)*

Calculation of the intrinsic carrier density :

$$n_i = \sqrt{N_c N_v} exp\left[-\frac{E_c - E_v}{2kT}\right]$$

$$n_i = \sqrt{2.7\times10^{19}\times1.1\times10^{19}} exp\left[-\frac{1.1}{2}\times0.026\right] \Leftrightarrow n_i = 1.12\times10^{10} cm^{-3}$$

Calculation of the Fermi level (taking the valence energy being equal to zero): :

$$E_{Fi} = \frac{E_c - E_v}{2} + \frac{kT}{2}\ln\frac{N_v}{N_c}$$

$$E_{Fi} = \frac{1.1}{2} + \frac{0.026}{2}\ln\frac{1.1\times10^{19}}{2.7\times10^{19}} \Leftrightarrow E_{Fi} = 0.538eV$$

>Solution n°4 *(exercice p. 21)*

Phosphorus is like arsenic, a donor-type impurity : $N_D = 10^{18} cm^{-3}$.

$N_D >> n_i$ therefore the electron density is equal to the donor density :

$n = N_D = 10^{18} cm^{-3}$

At $T = 27°C = 300K$, the hole density is given by :

$$p = \frac{n_i^2}{n} = \frac{n_i^2}{N_D} = \frac{(1.12\times10^{10})^2}{10^{18}}$$

$p = 125 cm^{-3}$

The obtained semiconductor is N-doped

>Solution n°5 *(exercice p. 21)*

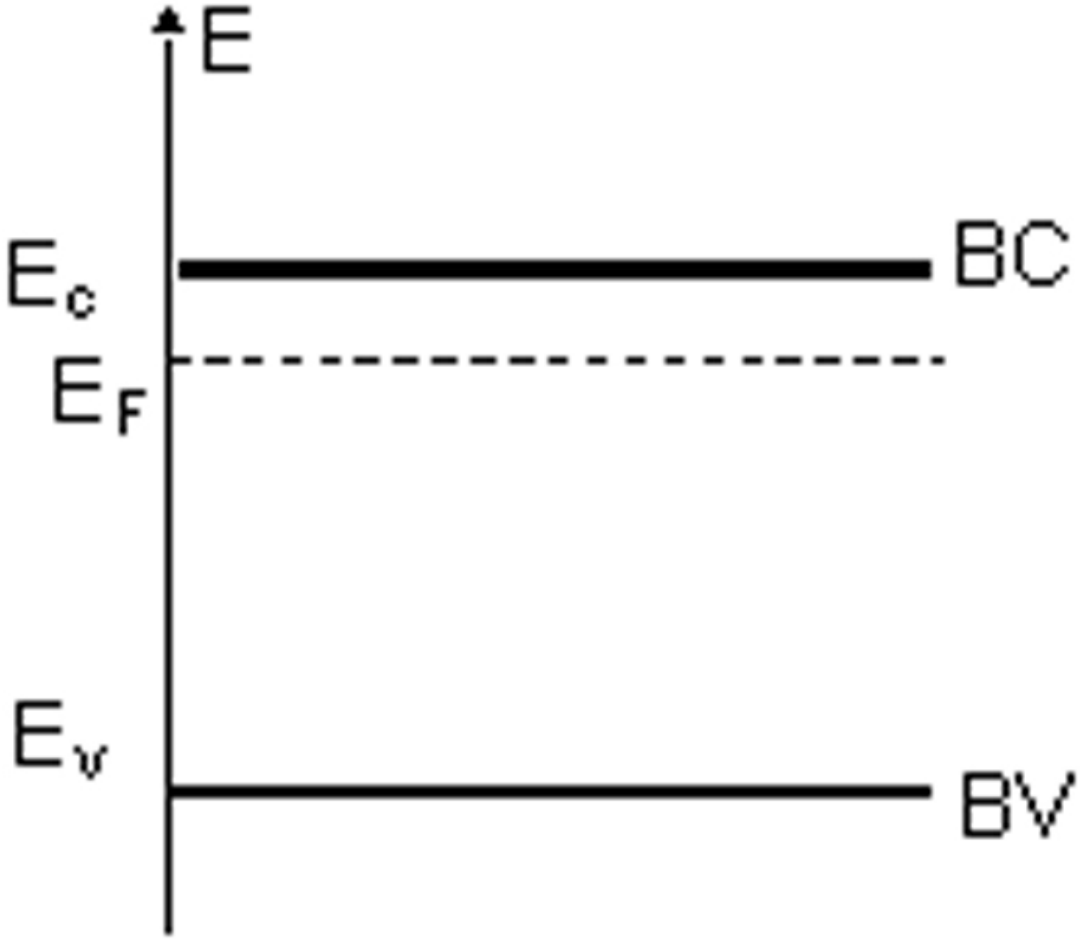
The Fermi energy can be inferred from the electron density as follows:

$$n = N_D = N_c exp\left[-\frac{E_c - E_F}{kT}\right] \Leftrightarrow N_D = N_c exp\left[-\frac{E_c - E_{Fi} + E_{Fi} - E_F}{kT}\right]$$

$$\Leftrightarrow N_D = n_i \, exp\left[-\frac{E_{Fi} - E_F}{kT}\right]$$

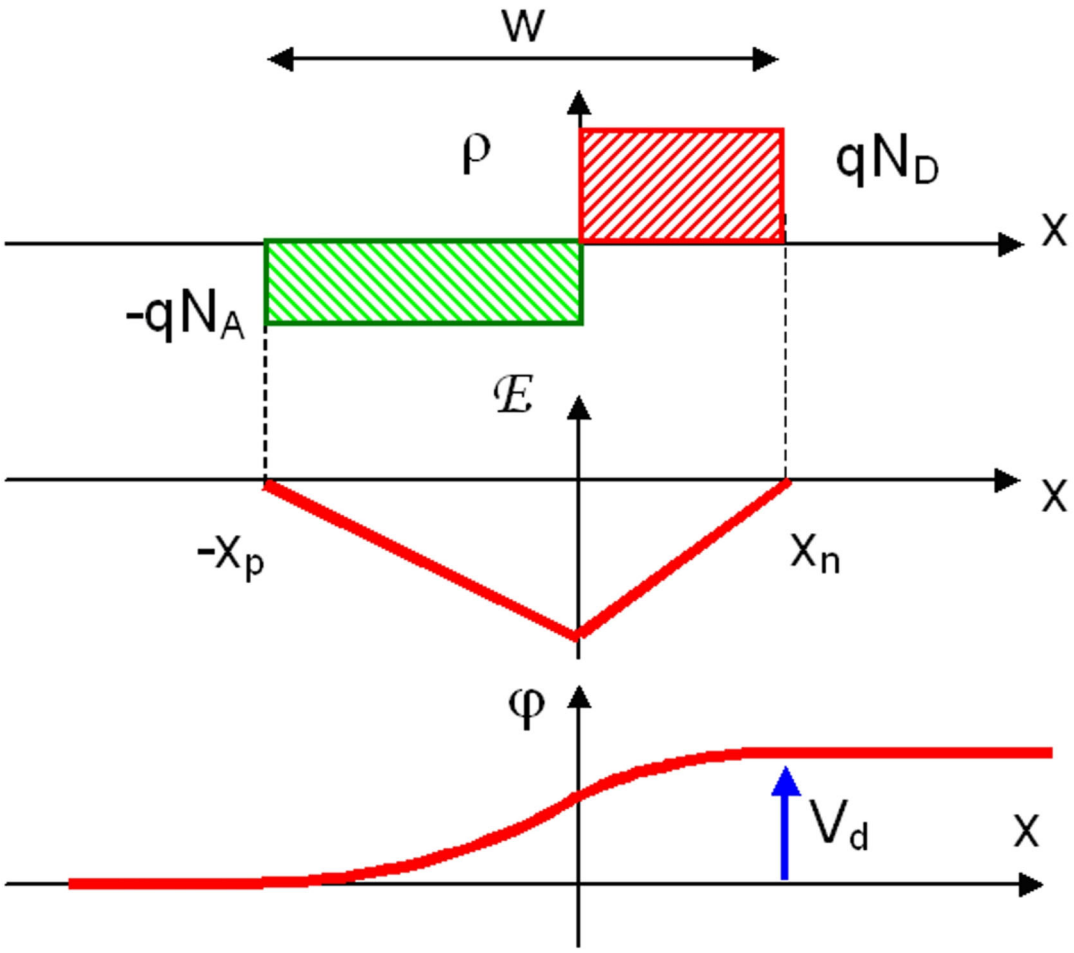
$$\Leftrightarrow E_F - E_{Fi} + kT \times \ln(\frac{N_d}{n_i})$$

$$E_F = 0.538 + 0.026 \times \ln(\frac{10^{18}}{1.12 \times 10^{10}}) \Leftrightarrow E_F = 1.014eV$$



>Solution n°6 *(exercice p. 21)*

For an abrupt PN junction, the dopant concentration varies suddenly from N_A to N_D. The following figure gives the charge, electric field and electrical potential distributions inside the depletion zone.



The electric field is given by :

$$\frac{dE}{dx} = \frac{\rho}{\varepsilon_s} \Leftrightarrow E = \frac{1}{\varepsilon_s}\int_{-x_p}^{x_n} \rho \times dx = 0$$

$$\Leftrightarrow \int_{-x_p}^{0} -qN_A dx + \int_{0}^{x_n} qN_D dx = 0$$

$$\Leftrightarrow N_A x_p = N_D x_n$$

This relationship shows that the depletion zone is more extended in the lowest-doped side.

>Solution n°7 *(exercice p. 22)*

The diffusion voltage is given by:

$$V_d = \frac{kT}{q}\ln\left[\frac{N_A N_D}{n_i^2}\right] \Leftrightarrow V_d = \frac{1.38\times10^{-23}\times300}{1.602\times10^{-19}} \times \ln\left[\frac{10^{19}\times10^{16}}{10^{20}}\right] \Leftrightarrow V_d = 0.892V$$

>Solution n°8 *(exercice p. 22)*

The width of the depletion zone is :
$w = x_n + x_p$ but we have shown in a) the following relationship : $N_A x_p = N_D x_n$
Hence :

$$w = \frac{N_A + N_D}{N_D} x_p = \frac{N_A + N_D}{N_A} x_n$$

so that :

$$x_n = \frac{N_A}{N_A + N_D} w$$

$$x_p = \frac{N_D}{N_A + N_D} w$$

Furthermore, the electrostatic potential is related to the charge distribution by :

$$\frac{d^2\varphi}{dx^2} = \frac{-\rho}{\varepsilon_s}$$

It is then possible to derive the expression for the diffusion voltage :

$$V_d = [\varphi]_{-x_p}^{x_n} = -\frac{1}{\varepsilon_s} \int_{-x_p}^{x_n} \int_{-x_p}^{x_n} \rho dx$$

$$\Leftrightarrow V_d = -\frac{1}{\varepsilon_s} \left(-qN_A \int_{x_p}^{x_n} (x + x_p) dx + qN_D \int_{x_p}^{x_n} (x - x_n) dx \right)$$

$$\Leftrightarrow V_d - \frac{1}{\varepsilon_s} \left(-qN_A \left[\frac{x^2}{2} + x_p x \right]_{-x_p}^{0} + qN_D \left[\frac{x^2}{2} + x_n x \right]_{0}^{x_n} \right)$$

$$\Leftrightarrow V_d = -\frac{q}{2\varepsilon_s} (N_A x_p^2 + N_D x_n^2)$$

Since :

$$x_n = \frac{N_D}{N_A + N_D} w$$

$$x_p = \frac{N_A}{N_A + N_D} w$$

we get :

$$V_d = \frac{q}{2\varepsilon_s} \frac{N_A N_D}{N_A + N_D} w^2$$

The depletion zone width is then :

$$w = \sqrt{\frac{2\varepsilon_s V_d}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right)}$$

Numerical application :

$$w = \sqrt{\frac{2 \times 10^{-12} \times 0.892}{1.602 \times 10^{-19}} \left(\frac{1}{10^{19}} + \frac{1}{10^{16}} \right)} \Leftrightarrow w = 3.34 \times 10^{-5} cm = 0.334 \mu m$$