Semiconductor Device Physics

Module 1

SensL Technologies Ltd.
River View Business Park
Blackrock
Cork, Ireland

www.SensL.com

Contents

1 Objectives

This module explains semiconductor device physics and the p-n junction diode. The bandstructure and band diagrams of a semiconductor together with absorption mechanisms in a semiconductor is presented. The p-n junction diode is described and its performance under forward and reverse bias explained. The two breakdown mechanisms of a p-n junction namely, zener and avalanche, are also detailed. After reading this module the reader will:

- Understand the bandstructure of a semiconductor, be familiar with band diagrams, direct and indirect bandgaps and absorption in semiconductors.
- Appreciate the properties of p-n junctions such as the depletion region and depletion capacitance.
- Understand the concept of breakdown in a p-n junction and be aware that two types of breakdown (zener and avalanche) exist.

---

2 Semiconductor Device Physics

The electrical and optical properties of a semiconductor are largely determined by its bandstructure. The periodic potential created by the relative position and type of atoms that make up the semiconductor crystal establishes the bandstructure. The band theory of solids arises from the complex quantum-mechanical interaction of an electron with the periodic potential of the crystal lattice. One important outcome of this theory is the formation of allowed energy bands in which electrons can exist, according to the Pauli exclusion principle. The allowed energy bands are separated by forbidden energy regions where charge carriers (electrons and holes) are not allowed to exist. The band theory of solids also gives rise to the concepts of holes and carrier effective mass, leading ultimately to an understanding of the conduction processes present in a semiconductor.

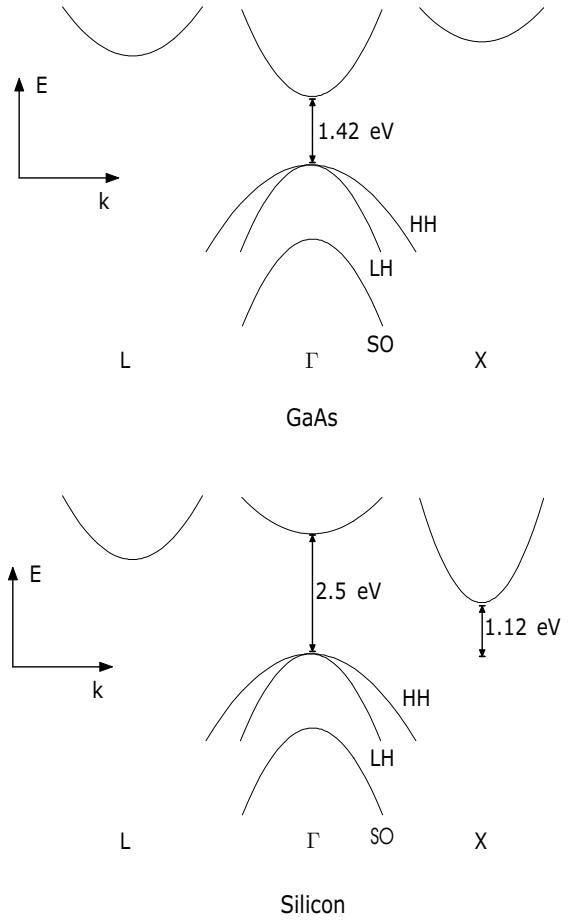

Figure 1: Simplified bandstructure representation of GaAs (direct bandgap) and Silicon (indirect bandgap).

2.1 Bandstructure

In a semiconductor the electron does not behave as a free electron. Its motion is restricted to certain allowed energy bands and its mass is effectively reduced compared to the free electron mass. The electron 'effective' mass is given by:

$$m_e^* = \hbar^2 \left(\frac{d^2 E}{dk^2} \right)^{-1} \tag{1}$$

where $\hbar$ is Plank's constant divided by 2π, E is the electric field and k is the wave-vector.

The allowed energy bands in the semiconductor are termed the conduction

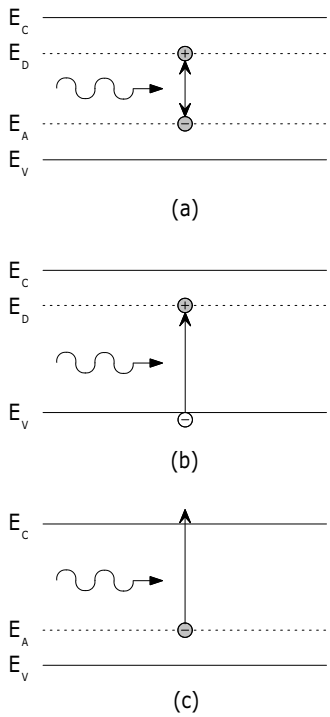

Figure 2: Photon absorption due to (a) donor-acceptor transition, (b) donor-band transition, and (c) acceptor-band transition.

and valence bands (CB and VB respectively). The separation in energy between the maximum energy in the valence band (VB) and the minimum energy in the conduction band (CB) is called the bandgap energy, E_g. The bandgap represents the extent of the forbidden energies in the bulk semiconductor. In the valence band a 'hole' is regarded as the absence of an electron and it is conceptually a positively charged particle. Under an applied external electric field holes will move (drift) in a direction opposite to that of electrons. At absolute zero $(0\ K)$ the conduction band (CB) is empty and the valence band (VB) is full of electrons (i.e. no holes). As the temperature is raised, some electrons gain sufficient energy to cross the forbidden energy gap into the conduction band, leaving behind holes in the valence band. This generation of electron-hole (e-h) pairs can also be achieved by photoexcitation of the semiconductor.

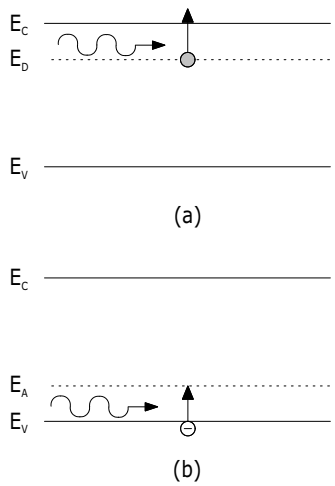

Figure 3: Low energy (a) donor-band and (b) acceptor-band absorption transitions.

2.1.1 Band Diagrams

Energy band diagrams for semiconductors are usually presented as E-k (Energy-momentum) diagrams. The bandstructure of semiconductors is extremely complicated but it can be theoretically calculated using techniques such as the 'pseudopotential' method or it can be measured experimentally by a variety of techniques. Generally it is useful to work with a simplified band diagram, concentrating on the local minima in the conduction band and maxima in the valence band. Near the band edges the band diagram can reasonably be approximated by parabolae. For semiconductors we can generalise two broad categories of bandstructure. The first category is the direct band-gap semiconductor, where the minimum energy in the CB coincides in momentum with the maximum energy in the VB. This is characteristic of many compound semiconductors such as GaAs and InP. The direct band nature of these semiconductors makes them efficient emitters of light, as no momentum change is required to support recombination of an electron from the CB with a hole in the VB, thus all the energy lost in the recombination can be dissipated as a photon of light. The second category of semicondcutor is the indirect bandgap type, characteristic of semiconductors such as silicon or germanium. For these semiconductors the minimum energy in

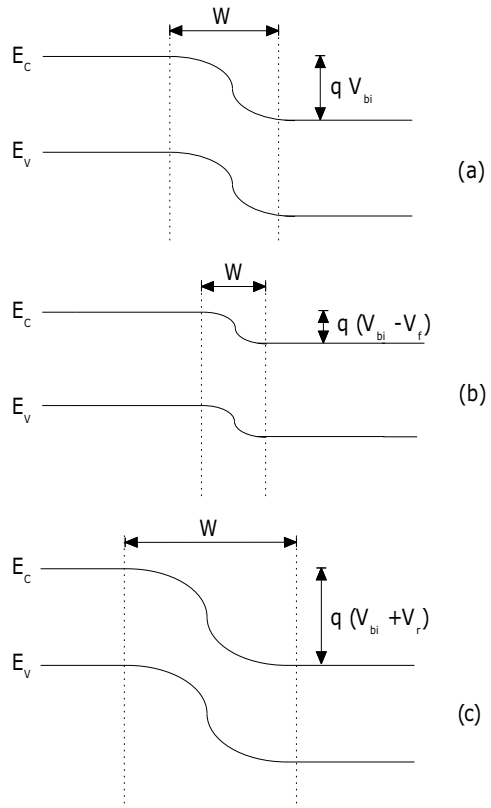

Figure 4: Energy band diagrams for a p-n junction in (a) equilibrium, (b) under forward bias and (c) under reverse bias.

the CB is located at a different k-value on the E-k diagram compared to the maximum energy in the VB, thus not offering any direct recombinational routes for an electron to recombine with a hole, unless a phonon (lattice vibration quantum) is simultaneously involved to bring about the required change in momentum. This is the main reason why Si and Ge are poor emitters of light. Simplified band diagrams for Si and GaAs are shown in Fig. 1.

In the parabolic band approach the CB is given by:

$$E(k) = \frac{\hbar^2 k^2}{2m_e^*} \tag{2}$$

where k is the wavevector (momentum space) and m_e^* is the electron effective mass. The value of m_e^* depends on which conduction band the electron occupies (i.e. Γ, X, or L). The complexities of these issues are beyond the scope of the

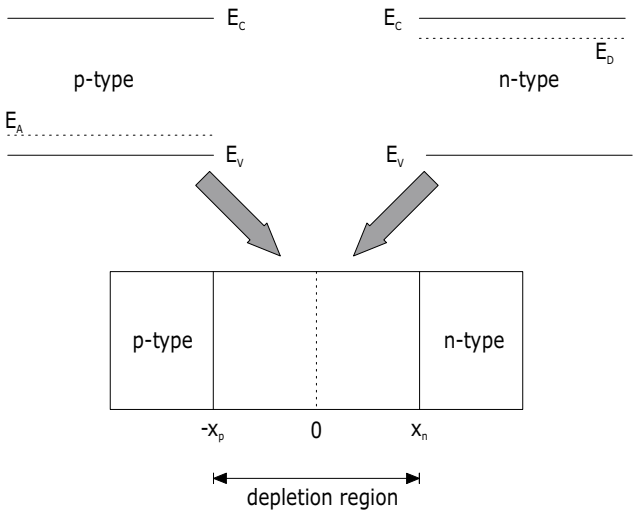

Figure 5: Formation of a p-n junction, illustrating the depletion region, when p-type and n-type materials are brought together.

present discussion, but the interested reader is referred to [1, 2].

There are three main valence bands to consider:

- heavy-hole band (HH)
- light-hole band (LH)
- split-off band (SO)

The effective mass of the hole is different in each. As a general rule the magnitude of the effective mass is directly related to the curvature of the parabola that describes the band. Thus in the heavy-hole band the hole effective mass is larger than in the light-hole band.

Changes in temperature and pressure have a profound effect on the bandstructure of the semiconductor. We have already noted that the bandstructure is determined by the relative positions of the atoms that make up the semiconductor crystal. Variations in temperature will cause the crystal lattice to either expand or contract, thereby altering the relative positions of the atoms in the crystal and hence the periodic lattice potential. The net result of this is a change in the fundamental bandgap energy, as well as a shift in the relative positions of

the various energy bands. These changes in turn fundamentally alter the optical and electrical properties of the semiconductor.

2.2 Absorption in Semiconductors

The process by which electron-hole (e-h) pairs are *created* or generated in a semiconductor is largely by absorption of a photon of energy $\hbar\omega > E_g$ where E_g is the bandgap energy and ω is the angular frequency of the light. The process of absorption and carrier generation is fundamental to the understanding of the operation of all photodetectors.

For direct bandgap semiconductors the absorption of a photon generally promotes a VB electron to the CB without any appreciable change in k (momentum or wavevector). In an indirect bandgap semiconductor, such as silicon, the absorption leads to an indirect carrier transition which involves a significant change in momentum. This momentum change may be provided by single or multiple phonons.

In addition to VB-CB transitions by absorption, in a doped semiconductor absorption can occur between donor and acceptor energy levels. These transitions usually require photon energies less than the bandgap energy. It is also possible to have absorption via VB-donor band states and acceptor band states-CB. These are illustrated schematically in Fig. 2.

Low energy absorption can occur via donor band-CB or VB-acceptor band states as shown in Fig. 3. These are termed low-energy impurity band transitions.

Intraband transitions can occur in p-type semiconductors where absorption of photons can produce transitions from LH to HH, SO to HH and SO to LH bands, depending also on the doping level and temperature of the semiconductor. Finally, free carrier absorption results in promoting a carrier within a band to a higher energy in the same band.

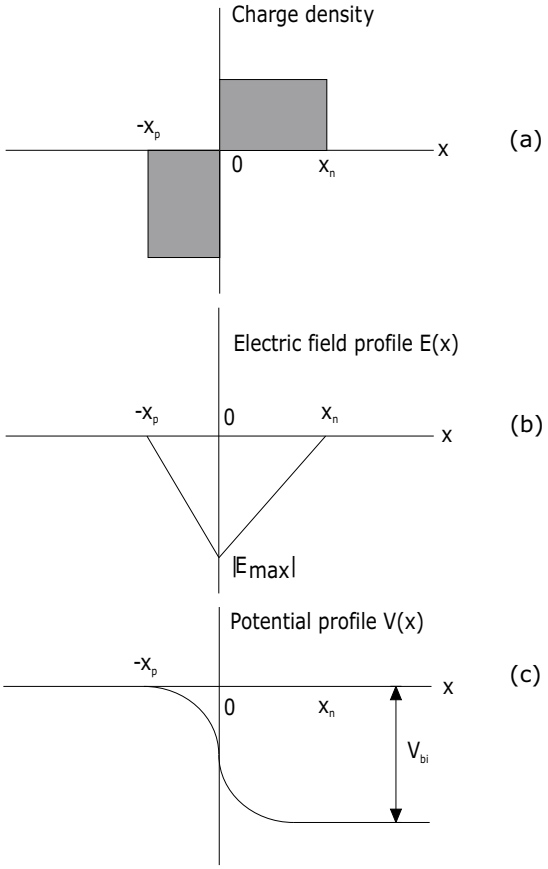

Figure 6: (a) Doping concentration, (b) space charge density, (c) electric field profile and (d) potential profile for a p^+-n junction.

3 The p-n Junction Diode

The simplest type of semiconductor diode is formed when p-type and n-type semiconductors are brought into contact. The p-n junction diode is essential for the operation of most optoelectronic devices, both in forward and reverse bias modes of operation. This chapter will outline the principle electro-static operation of the p-n junction and explains the notions of *built-in potential*, *depletion region* and *junction capacitance*. The current-voltage relationship for the diode is developed for both the forward and reverse bias cases. The different types of p-n junctions are described along with phenomena such as junction breakdown. It is important to note that the special class of p-n heterojunctions is not considered

here as the primary device of interest is the silicon based photodiode, which is typically formed using a p-n homojunction.

3.1 The Equilibrium State of the p-n Junction

The p-n junction is shown in Fig. 4. The p- and n-type semiconductors have large concentrations of free holes and electrons respectively. When they are brought together to form a junction there is initially a large concentration gradient for both electrons and holes on either side of the junction. This gradient initiates diffusion of free carriers across the junction, with electrons exiting the n-region leaving behind positively ionised donor atoms (holes), while holes exiting the p-region leave behind negatively ionised acceptor atoms, thus forming two regions of immobile charge carriers on either side of the junction. As the diffusion of carriers continues the width of these two regions expands and an electric field is created as these regions become depleted of mobile charges. The diffusion of carriers continues until the electric field is sufficient to oppose further diffusion of carriers across the junction. At this stage thermal equilibrium is reached and there is no net flow of current across the junction in the absence of an external bias. The region of immobile charge carriers on either side of the junction is called the *depletion* or *space-charge* region (see Fig. 5).

The width of the depletion region, $W = x_n + x_p$, where x_n and x_p are the extents of the space charge layer into each of the two regions. The extent of the depletion region on each side of the junction is inversely proportional to the doping concentration. In other words the depletion region spreads further into the more lightly doped semiconductor.

The built-in electric field created in the space charge region reaches a maximum at the junction and falls to zero at the edges of the space charge region. The presence of the electric field, E, results in a potential gradient, V_{bi} causing the bands (CB and VB) to bend in the depletion region by an energy qV_{bi}. This is known as the *built-in potential* of the junction. If we assume that the junction doping is *abrupt* i.e. there is a step change in doping concentration from p-type

to n-type at the junction, it is evident from Fig. 6 that:

$$qN_Ax_p = qN_Dx_n = \epsilon_s\epsilon_0 E \tag{3}$$

where ϵ_s is the relative permittivity of the semiconductor and N_A and N_D are the p- and n-type doping densities respectively. We are assuming here that all the dopants are ionised. This equation is known as the depletion approximation and it shows that the total ionised charge density on either side of the junction is equal. From Poisson's equation we can derive the electric field profile on either side of the junction:

$$E(x) = \frac{qN_D}{\epsilon_s\epsilon_0}(x - x_n) \tag{4}$$

for the n-side of the junction and:

$$E(x) = \frac{qN_A}{\epsilon_s\epsilon_0}(x + x_p) \tag{5}$$

on the p-side. From these equations it can be shown that the built-in potential is given by:

$$V_{bi} = \frac{q}{2\epsilon_s\epsilon_0}\left[N_Ax_p^2 + N_Dx_n^2\right] \tag{6}$$

At the diode junction the voltage is given by:

$$V(0) = -\frac{qN_Ax_p^2}{2\epsilon_s\epsilon_0} = -\frac{qN_Dx_n^2}{2\epsilon_s\epsilon_0} \tag{7}$$

The built-in potential can also be expressed in terms that are either known or measurable parameters of the materials used to make the junction as:

$$V_{bi} = \frac{k_BT}{q}ln\left(\frac{N_AN_D}{n_i^2}\right) \tag{8}$$

where n_i is the intrinsic carrier concentration of the material.

3.1.1 Width of the Depletion Layer

With reference to Fig. 6 the space charge density is qN_D and qN_A on the n- and p-sides of the junction respectively. The space charge density goes to zero

outside the depletion region. The maximum electric field, E_{max}, appears at the junction and for an abrupt junction is given by:

$$\epsilon_s \epsilon_0 E_{max} = qN_D x_n = qN_A x_p \tag{9}$$

The area enclosed by the electric field profile is the built-in voltage, V_{bi}, hence:

$$|-V_{bi}| = \frac{E_{max}}{2}(x_n + x_p) \tag{10}$$

Since the depletion region width, $W = x_n + x_p$, it follows that:

$$W = \sqrt{\frac{2\epsilon_s \epsilon_0}{q}\left(\frac{1}{N_D} + \frac{1}{N_A}\right)|-V_{bi}|} \tag{11}$$

For the situation where $N_A >> N_D$ (or vice-versa), as is the case in a one-sided step junction:

$$W = \sqrt{\frac{2\epsilon_s \epsilon_0}{qN_D}|-V_{bi}|} \tag{12}$$

In this case it is apparent that the depletion region extends further into the more lightly doped side of the junction. In reality the depletion approximation does not hold true as the free carrier density does not end abruptly at the junction - instead the n-side space charge density is given by $q[N_D - n(x)]$ instead of simply qN_D, with a similar equation for the p-side. Thus, it can be shown that:

$$W = \sqrt{\frac{2\epsilon_s \epsilon_0}{qN_B}\left(V_{bi} - \frac{2k_B T}{q}\right)} = L_D\sqrt{2\left(\frac{qV_{bi}}{k_B T} - 2\right)} \tag{13}$$

where N_B is either N_A or N_D and L_D is the Debye length.

3.1.2 Capacitance of the Junction

The depletion region in a p-n junction behaves like the insulator separating the conductive plates of a parallel plate capacitor whose capacitance is given by:

$$C = \frac{\epsilon_s \epsilon_0 A}{W} \tag{14}$$

where A is the junction area and W is the width of the depletion region (corresponding to the separation of the *parallel plates*). Substituting the expression for

the depletion region width gives:

$$C = \frac{\epsilon_s \epsilon_0 A}{\sqrt{\frac{2\epsilon_s \epsilon_0}{q}\left(V_{bi} - V_a\right)\frac{N_A + N_D}{N_A N_D}}} \tag{15}$$

where V_a is the applied bias. In the case of the asymmetric junction where $N_A >> N_D$ then:

$$C \cong \frac{\epsilon_s \epsilon_0 A}{\sqrt{\frac{2\epsilon_s \epsilon_0}{q N_D}\left(V_{bi} - V_a\right)}} \tag{16}$$

From these equations it is evident that there is a bias dependence of the capacitance of the p-n junction. The previous equation can be rearranged to give:

$$\frac{1}{C^2} = \frac{2}{q\epsilon_s \epsilon_0 N_D A^2}\left(V_{bi} + V_a\right) \tag{17}$$

This equation can be used to determine the donor concentration in a p^+-n junction, as well as the built-in voltage from direct measurements of the capacitance as a function of applied bias.

3.2 The p-n Junction Under Forward Bias

At zero bias the built-in potential barrier due to the formation of the depletion region restricts the motion of majority carriers. For any current to flow the built-in voltage, V_{bi}, must be reduced. This situation occurs when a forward bias voltage (positive voltage to p-side of the diode) is applied. Most of the applied bias is dropped across the high resistance depletion region. If the applied forward bias is V_f then the effective bias across the depletion region is $(V_{bi} - V_f)$. The forward bias voltage also causes narrowing of the depletion width. Fig. 4(b) shows the band-diagram for a forward biased p-n junction.

A current flows across the junction under forward bias because the drift and diffusion components of the current are no longer balanced as they were in equilibrium. Majority carriers crossing the junction become minority carriers on the other side of the junction. This results in an excess minority carrier density on both sides of the junction, denoted as $\delta n(x)$ and $\delta p(x)$ on the p- and n-sides respectively.

The excess minority carrier density at the edge of the depletion region on the n-side is

$$\delta p(0) = p_{N0}\left(e^{\frac{qV_f}{k_BT}} - 1\right) \tag{18}$$

and it's distribution with distance into the n-side of the diode is

$$\delta p(x) = p_{N0}\left(e^{\frac{qV_f}{k_BT}} - 1\right)e^{\frac{-x}{L_h}} \tag{19}$$

where p_{N0} is the equilibrium minority carrier density on the n-side and $L_h = \sqrt{\tau_h D_h}$ is the diffusion length for holes, where τ_h is the hole excess carrier lifetime and D_h is the hole diffusion constant. The diffusion current resulting from these excess carriers is

$$J_h(x) = -qD_h\frac{d}{dx}\delta p(x) \tag{20}$$

At the edge of the depletion region

$$J_h(0) = \frac{qD_h p_{N0}}{L_h}\left(e^{\frac{qV_f}{k_BT}} - 1\right) \tag{21}$$

Similarly the current resulting from diffusion of electrons on the p-side is

$$J_e(0) = \frac{qD_e n_{P0}}{L_e}\left(e^{\frac{qV_f}{k_BT}} - 1\right) \tag{22}$$

The total current under forward bias is then

$$J = J_h(0) + J_e(0) = q\left(\frac{qD_h p_{N0}}{L_h} + \frac{qD_e n_{P0}}{L_e}\right)\left(e^{\frac{qV_f}{k_BT}} - 1\right) \tag{23}$$

3.3 The p-n Junction Under Reverse Bias

The band diagram for the reverse biased p-n junction is given in Fig. 4 (c). Here the reverse bias voltage is dropped across the depletion region, but this time adds to the built-in potential resulting in an increased barrier $(V_{bi} + V_r)$. Under these circumstances the depletion width is also increased. Eq. 23 also applies for the reverse bias case if V_f is replaced by $-V_r$. Furthermore, if $|V_r|$ is large then:

$$J \cong -q\left(\frac{qD_h p_{N0}}{L_h} + \frac{qD_e n_{P0}}{L_e}\right) = J_s \tag{24}$$

where J_s is the reverse saturation current of the diode.

3.4 The Diode Current-Voltage Relation

The current voltage relation for the diode can be given as:

$$J = J_s \left(e^{\frac{qV_a}{k_B T}} - 1 \right) \tag{25}$$

where V_a is the applied bias. For values of $|V_r|$ larger than a few millivolts $J = -J_s$. Under forward bias greater than a few millivolts (typically $> 3k_BT$)

$$J \cong J_s e^{\frac{qV_a}{k_B T}} \tag{26}$$

An ideality factor, η, is usually incorporated as:

$$J \cong J_s e^{\frac{qV_a}{\eta k_B T}} \tag{27}$$

In an ideal diode η should be close to one. In real diodes $\eta > 1$ because of generation and recombination processes in the depletion region. There are also ohmic losses due to the small resistances of the neutral n- and p-regions.

3.5 Graded p-n Junctions

Diode junctions fabricated using a diffusion process for doping are normally graded in concentration, unlike the abrupt junctions that are achievable using ion-implantation or epitaxial growth techniques. Here we present the properties of a linearly graded junction as detailed in Fig. 7.

For an impurity concentration gradient, a, the electric field profile is given by

$$E(x) = \frac{-qa}{\epsilon_s \epsilon_0} \frac{\left(\frac{W}{2}\right)^2 - x^2}{2} \tag{28}$$

and the maximum electric field is at $x = 0$:

$$|E_{max}| = \frac{qaW^2}{8\epsilon_s \epsilon_0} \tag{29}$$

The built-in potential is:

$$V_{bi} = \frac{qaW^3}{12\epsilon_s \epsilon_0} \cong \frac{k_B T}{q} ln \left(\frac{aW}{2n_i} \right)^2 \tag{30}$$

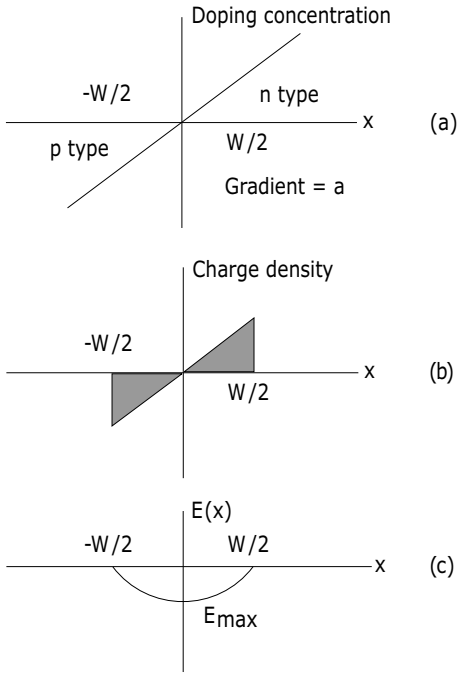

Figure 7: Linearly graded junction (a) energy band diagram, (b) space-charge distribution and (c) electric-field profile.

which gives the depletion width, W, as:

$$W = \left(\frac{12\epsilon_s \epsilon_0 V_{bi}}{qa} \right)^{1/3} \tag{31}$$

The depletion layer capacitance per unit area as a function of applied bias is

$$C = \frac{\epsilon_s \epsilon_0}{W} = \left[\frac{qa(\epsilon_s \epsilon_0)^2}{12(V_{bi} \pm V_a)} \right]^{1/3} \tag{32}$$

3.6 Breakdown Phenomena in Diodes

At sufficiently large reverse biases the diode junction breaks down allowing significant current flow, limited largely by the resistance of the external circuit. The main mechanisms that lead to junction breakdown are described here.

3.6.1 Zener Breakdown

Zener breakdown occurs when the depletion region is sufficiently thin for tunnelling to take place across the junction. This happens if both the p- and n-type doping are relatively large. The tunnelling component of the current increases with increasing reverse bias, eventually leading to complete breakdown of the junction. In a properly designed zener diode the current in the diode is independent of the voltage after breakdown. The zener breakdown process is independent of temperature.

3.6.2 Avalanche Breakdown

Avalanche breakdown occurs in moderately doped diodes with a depletion region wider than that required for zener breakdown. The large electric field present in the depletion region accelerates carriers there to near saturation velocity. These carriers can collide with atoms in the crystal lattice and if the energy of the carriers is sufficiently high these collisions can break lattice bonds thereby ionising the atoms and releasing bound carriers. This process is known as impact ionisation. After a collision the original carrier and the e-h pair created by the ionising collision move under the influence of the electric field and undergo several more ionising collisions as they cross the depletion region. This leads to a chain reaction causing the reverse current to increase and the diode eventually breaks down. Unlike zener breakdown, avalanche breakdown is temperature dependent and the breakdown voltage increases with increasing temperature.

---

4 Summary

This module has explained the bandstructure and absorption mechanisms of a semiconductor. The module has also described the p-n junction and properties of the junction such as the depletion region, the current-voltage relationship and

depletion capacitance. The breakdown methods of the junction have also been discussed, with zener and avalanche breakdown being explained.

---

5 Acknowledgements

SensL Technologies Ltd. would like to acknowledge Dr. Aoife M. Moloney of the School of Electronic and Communications Engineering in the Dublin Institute of Technology and Dr. Alan P. Morrison of the Department of Electrical and Electronic Engineering in University College Cork, for writing and coordinating the PCEdu-1 Photon Counting Fundamentals Pack.

References

[1] S. M. Sze. *Physics of Semiconductor Devices*. John Wiley and Sons, 1981.

[2] Pallab Bhattacharya. *Semiconductor Optoelectronic Devices*. Prentice Hall, 1994.