Chapter 3

Semiconductor Physics (Draft)

DAAN LENSTRA, WITH CONTRIBUTIONS OF MEINT SMIT

3.1 Introduction

It is the aim of this chapter to provide the physical background for understanding the operation of active waveguide devices, like optical amplifiers, detectors, modulators and saturable absorbers. The explanation is focused on the properties of a short active waveguide section built from semiconductor material, as depicted in Figure 3.1, which we consider as a basic cell in longer building blocks which can be analysed as concatenations of basic cells. The most important building block which can be considered as a concatenation of forward biased basic cells is the Semiconductor Optical Amplifier. Detectors, Phase modulators, and (Saturable) Absorbers can be considered as concatenations of reverse biased basic cells. The description of these building blocks is given in parts II and III of the book. In this chapter the properties of semiconductors and the basic cells in which they are applied are discussed: their band structure, the effect of doping on the semiconductor properties, pn- and pin-junctions, Double Heterostructures, Quantum Wells, light absorption, emission and amplification.

3.2 Semiconductor crystal structure

3.2.1 Crystal structure.

Modern photonic and micro-electronic components and circuits are based on semiconductor materials. The most well-known semiconductor is silicon, it is the basic material for microelectronics. Silicon is tetravalent, it has four electrons in the outer electron shell. As the outer shell has room for eight electrons, it can share electrons

tetravalent

This chapter is heavily inspired by and based on Semiconductor Physics and Devices, Basic Principles, by Donald A. Neamen; Master Course Photonics (Ghent University) by Roel Baets and Günther Roelkens; Fundamentals of Semiconductor Lasers by Takahiro Numai; Electronic and Optoelectronic Properties of Semiconductor Structures, by Jasprit Singh; SOLID STATE PHYSICS PART II: Optical Properties of Solids by S. Dresselhaus; Fundamentals of Semiconductors, by Yu and Cardona; Physics of Semiconductor Devices, by Sze, Li and Ng; Thermal Physics, by C. Kittel and H. Kroemer

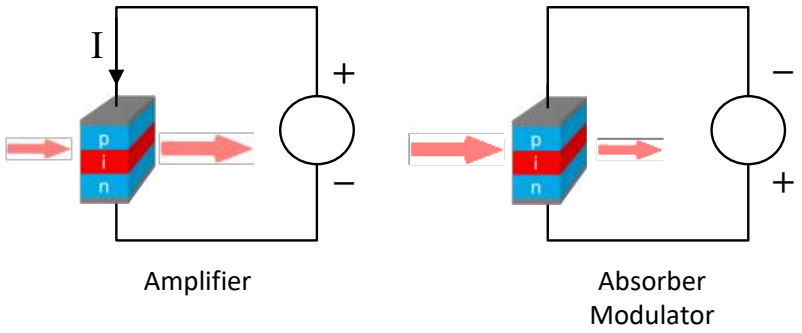

Figure 3.1: Basic Semiconductor Optical Amplifier cell, which works as an absorber, an optical detector or a modulator when reversely biased

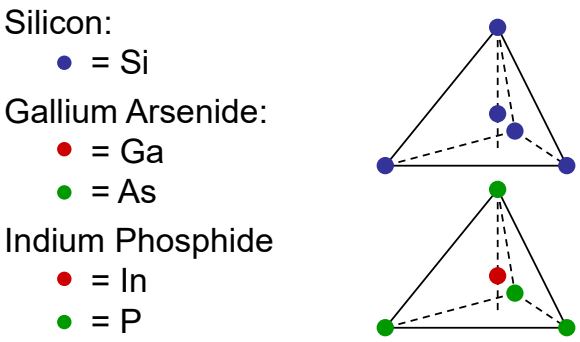

Figure 3.2: Basic structure of elemental and compound semiconductors.

tetrahedron

with four other silicon atoms in order to fill the outer shell with eight electrons. The basic structure of silicon is, therefore, a tetrahedral structure as shown in Fig. 3.2, in which each silicon atom is in the center of a tetrahedron with four other silicon atoms at the corners. The crystal structure is such, that each other silicon atom is also in the center of another tetrahedron.

compound semiconductor
group-IV material
III-V semiconductor

Silicon has excellent properties for the fabrication of electronic components and circuits. It also supports some opto-electronic functions, such as light detection and modulation, but it is not suitable for efficient generation of light, e.g. in lasers and optical amplifiers. For these functions so-called III-V compound semiconductors, such as GaAs, InP and their related materials, are much better suited. Their structure is very similar to the structure of silicon. Whereas silicon is a so-called group-IV material (it is in the IV-th column of the periodic system), III-V compound semiconductors are composed of elements from the group-III column (In or Ga) and the group-V column (As and P), in such a way that each group-III atom is surrounded by four group-V atoms, and the other way around, as depicted in Fig. 3.2. In this structure each group-III atom shares its three electrons with the five electrons of the adjacent group-V atoms, so that together they fill the outer shell with eight electrons, just like silicon. The electronic properties of the compound semiconductors are, therefore, similar to those of silicon, although most of them support higher operation speeds than silicon. The main difference is in the optical properties, where they are used for fabrication of high-quality lasers, modulators and optical amplifiers.

unit cell

The tetrahedric structure shown in Fig. 3.2 is part of the unit cell of semiconductor

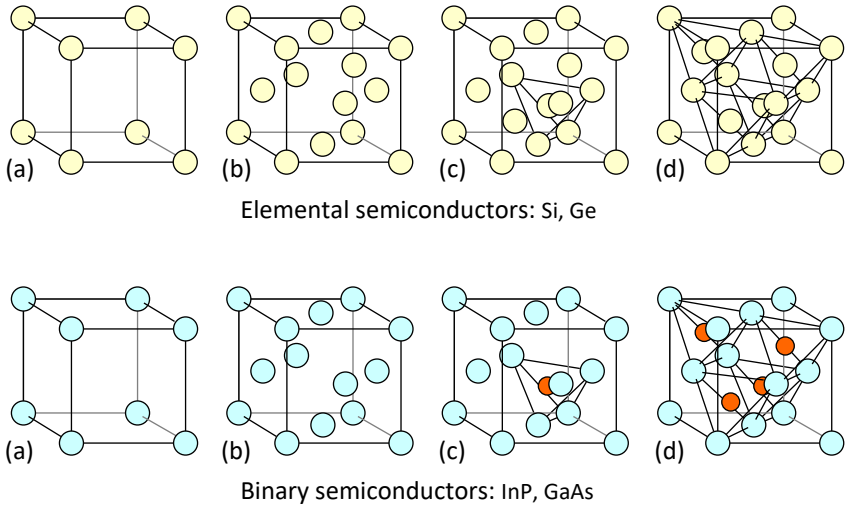

Figure 3.3: Structure of a the cubic unit cel in a Zinc Blende crystal structure, for elemental (Si, Ge) and Binary semiconductors (InP, GaAs). The pictures labeled (d) show the full structure, the pictures labeled (a), (b) and (c) clarify how it is composed, as explained in the text.

crystals, the unit cell is more complicated, however. The unit cell of a crystal is the smallest cell from which the crystal can be built by periodic repetition in all three directions.

cubic crystal

Figure 3.3(d) shows the unit cells of silicon and compound semiconductor crystals. The cell is cubic, which means that its dimension, which is called the lattice constant, is the same in all three directions. And further it means that the faces of the unit cell are orthogonal to each other. It is a fairly complex picture, which can be understood better if we build it step by step. In Fig. 3.3(a) we see that the cell has eight atoms at the corners of the cube. Further, it has six atoms at the center of each of the six faces, as shown in Fig. 3.3(b). Fig. 3.3(c) shows how three of the face-centered atoms, together with one corner atom, form a tetrahedron, which has one atom in its center. For a binary semiconductor the latter is from the other group. Figure (d) shows that the unit cell contains four of these tetrahedrons, each of them with one atom in its center. This can best be seen for the binary semiconductors where the central atom has a different color. For the elemental semiconductors Si and Ge the structure is the same, however.

binary semiconductor

From the lower Fig. 3.3(d) it may seem that the unit cell contains much more atoms from one group than from the other. This is not correct, however. As we see, there are four atoms of one group fully inside the cell. For the other group there are six atoms in the center of the faces, each of which counts half because it is shared with the adjacent cell, and eight atoms at the corners, which count only for 1/8, because they are shared with seven other cells. So, together also four atoms.

Zinc Blende

The cubic crystal structure with four tetrahedrons inside is known as the Zinc Blende structure. The crystal structure is very important for the electronic, the optical, the chemical and the mechanical properties. Cleaving planes, for example, follow crystal planes, and wet selective etchants produce side walls, the slope of which is determined by the crystal orientation. For the electronic and photonic properties the crystal structure is even more important as will be described in the following sections.

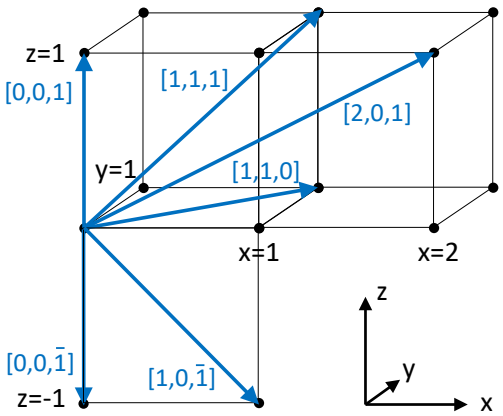

Figure 3.4: Miller indices for crystal directions.

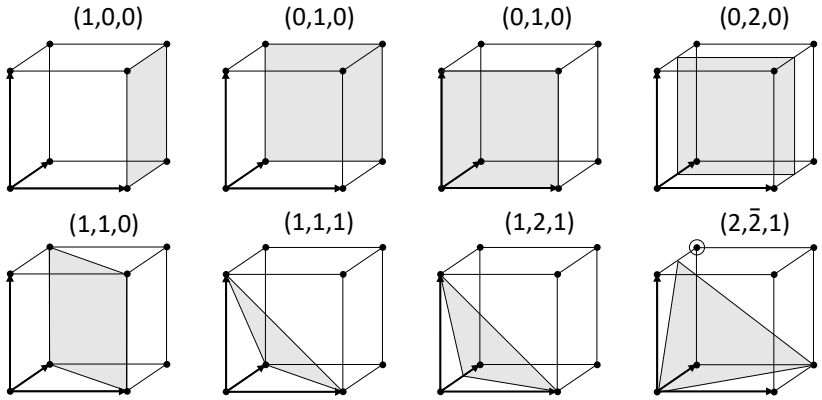

Figure 3.5: Miller indices for crystal planes.

3.2.2 Miller Indices

Miller indices

Because many properties of the the semiconductor are dependent on the crystal orientation it is important to have a good description of crystal directions and crystal planes. Such a description is provided by the so-called Miller indices. We will briefly describe them here for materials with a cubic unit cell, such as silicon and the III-V compound semiconductors.

lattice constant

Crystal directions. The length of the ribs of the unit cell is called the lattice constant of the crystal. For cubic crystals the lattice constant is equal in the x, y and z-direction. For InP the lattice constant is 5.8687 A. For the definition of the Miller indices the ribs of the unit cube are normalized. The Miller indices of a crystal direction are the coordinates of the direction vectors in the normalized crystal space, as indicated in Fig. 3.4. The Miller indices of crystal directions are enclosed between square brackets [], as opposed to the coordinates of crystal planes, which are enclosed between curved brackets (). In the literature we see that the brackets are often omitted, so the (1,1,1)-plane is simply denoted as 111-plane.

Fig. 3.4 shows examples for some crystal directions. The Miller indices of the x-axis are $[1,0,0]$ and of the z-axis $[0,0,1]$. $[1,1,0]$ is the diagonal of the two-dimensional xy-

Problem: Show that the shaded plane in the lower right unit cell of Figure 3.5 is the $(2,\bar{2},1)$plane.

Solution: The shaded plane intersects the x-axis and the y-axis at x=y=0 and the z-axis at z=2. If we put a second unit cell on top of the first one, as shown in Fig. 3.6 we can extend the plane into that cell by extending the left side of the shaded triangle to the point (0,1,2) and going from there back to the point (1,1,0). We see that in this cell the plane intersects the x, y and z-axis at non-zero values. If we take the point (0,1,1) as the origin of a new coordinate system, we see that in this system the plane intersects the x-axis at x=1/2, the y-axis at y=-1/2 and the z-axis at z=1. So the reciprocal coordinates of the plane are $(2,\bar{2},1)$.

Problem 3.1: Calculation of the Miller index.

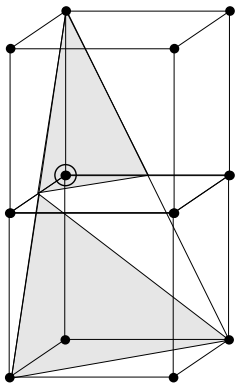

Figure 3.6: Configuration of the $(2,\bar{2},1)$-plane

plane. [1,1,1] is the spatial diagonal of the three-dimensional unit cell. $[2,0,1]$ is the direction in the x,z plane going through the point (x=2, z=1). Negative directions are indicated with a small bar on top of the coordinate. So $[1,0,\bar{1}]$ is the planar diagonal of the lower (south-east) quadrant of the x-z plane. In a cubic crystal it is perpendicular to the [1,0,1] direction.

Crystal planes Crystal planes are indicated with curved brackets () around the reciprocal coordinates of their intersections with the x, y and z-axis. A few examples are given in Fig. 3.5. The x, y and z-axis are indicated with bold arrows. The (1,0,0) plane is the shaded plane, intersecting the x, y and z-axis at $x = 1, y = \infty, z = \infty$, so the reciprocal coordinates are (1,0,0). In the same way the (0,1,0) plane is the x,z plane intersecting the y-axis at y=1. Because of the periodicity of the crystal structure, the (0,1,0) back face (y=1) and the (0,1,0) front face (y=0) of the unit cell are identical. A plane halfway the (0,1,0) faces of the cube is the (0,2,0) plane. It intersects the y-axis at y=1/2, so its reciprocal Miller index is 2.

Planes orthogonal to the axes have only one non-zero Miller index. Rotated or tilted planes have more non-zero indices. The (1,1,0) plane intersects the x and y-axis at x=y=1 and it is parallel to the z-axis. This plane is the natural cleaving plane of InP crystals. The (1,1,1) plane intersects the axes at x=y=z=1. The (1,2,1) plane intersects the x and z-axis at x=z=1 and the y-axis at y=1/2. If the plane intersects an axis in its negative range, the reciprocal Miller index is provided with a bar, e.g. $\bar{2}$.

cleaving plane

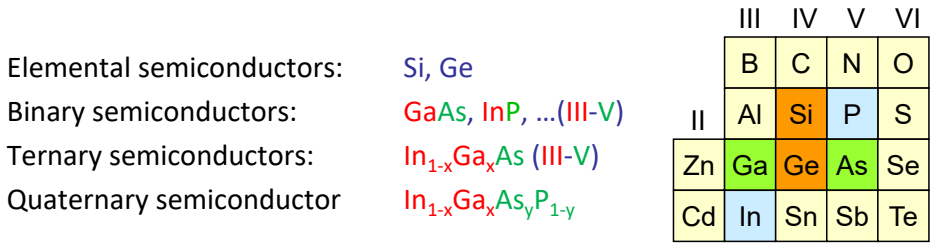

Figure 3.7: Relevant part of the periodic system for the formation of elemental, III-V and II-VI semiconductors.

For the definition of the Miller indices of a plane it is important that the plane does not go through the origin of the cell, as shown in Figure 3.5, because this would yield a reciprocal Miller index ∞. For such planes we have to find which corner of the unit cell can be used as origin of a coordinate system in which the plane intersects all three axes at non-zero values between -1 and 1. For the plane in the lower right figure this is the encircled corner point. In Problem 3.1 it is shown how the Miller indices of this plane can be found.

3.2.3 Compound semiconductors

Silicon and Germanium are so-called elemental semiconductors. Their elemental crystals consist for 100% of silicon or germanium atoms and, as shown in Fig. 3.3 they are organized in a so-called Zinc Blende structure. A similar structure can also be formed by a combination of group-III and group-V materials, or group-II and group-VI materials, as shown in the picture. Figure 3.7 shows the most commonly used materials in III-V and II-VI semiconductors. In its simplest form such a III-V semiconductor consists of two materials in a 50/50 ratio, for example In and P. We can make more complex materials with different properties, however, by replacing part of the In atoms by Ga atoms. This can be done during the epitaxial growth by controlling the flow of the precursor gases, e.g. $In(CH_3)_3$ and $Ga(CH_3)$, as described in Sec. 4.7 of Chapter 4. In this way a so-called ternary semiconductor is formed: InGaAs.

ternary semiconductor

Lattice match. Changing the composition of the semiconductor will change its properties, e.g., the bandgap and the corresponding emission and absorption wavelength. But also properties like the refractive index. We are not free, however, to change the ratio of In and Ga at will, because the Ga atom is significantly smaller than the In atom, so that the lattice constant will shrink when we add more Ga. This is called lattice mismatch. It will cause strain in the epitaxially grown layer, as shown in Fig. 3.8(b). The left figure shows the ideal case in which the lattice constant of the grown material matches that of the substrate. If it is different, however, the epitaxially grown layer will be strained and if the strain becomes to large the grown material will adapt to its own intrisic lattice constant, which will cause defects at the interface layer.

lattice mismatch *strain* *defect*

The strain caused by replacing In with Ga can be reduced, however, if we replace at the same time part of the P atoms with As atoms. The latter are larger than the P atoms, so they increase the lattice constant and counter the effect of replacing In atoms by Ga atoms. If we replace the In atoms and the P atoms in a proper ratio with Ga and As, respectively, the lattice constant remains constant, which allows us to grow thick defect-free crystalline layers.

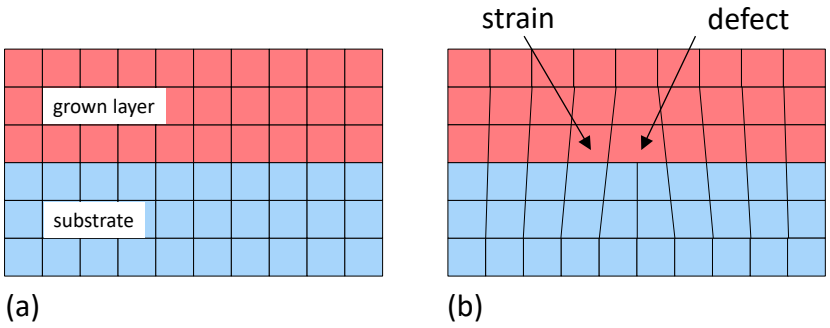

Figure 3.8: Illustration of lattice mismatch leading to strain and defects.

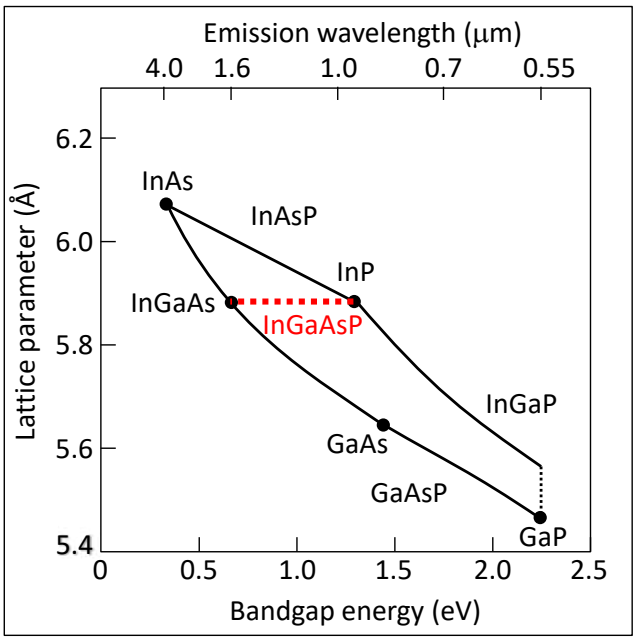

Figure 3.9: Bandgap vs. lattice constant diagram for III-V semiconductors.

By doing this we get a so-called quaternary semiconductor, InGaAsP, and we describe its composition with the indices x and y, which are the fractions of In atoms replaced by Ga and the fraction of P atoms replaced by As, respectively, i.e., $In_{1-x}Ga_xAs_yP_{1-y}$.

quaternary semiconductor

Bandgap Engineering The bandgap is a characteristic feature of every semiconductor. It will be introduced and explained in the next section. The bandgap energy is denoted by E_g and has values in the order of an eV. In a quaternary semiconductor like $In_{1-x}Ga_xAs_yP_{1-y}$ we have two degrees of freedom for varying the material properties, we can chose both x and y. In order to keep the lattice constant equal to that of the InP substrate, we have to sacrifice one degree of freedom because this requires a fixed ratio of x and y. Then we are stil left with one degree of freedom, with which we can vary the emission wavelength of the semiconductor and also other properties.

Fig. 3.9 shows how the composition is related with the band gap of the compound semiconductor. The dots in the graph indicate the combination of bandgap and lattice

constant for a few binary semiconductors. The lines between the dots indicate how the bandgap and the lattice constant change if we mix the two binary materials. For example, if we replace a fraction y of the P atoms in InP by As atoms, we move from the dot InP (y=0) to the dot InAs (y=1) via the line describing the bandgap and lattice constant of ternary $InAs_yP_{1-y}$ material. From InAs (x=0) we can move to GaAs (x=1) by replacing a fraction x of the In by Ga, which will give us ternary $In_{1-x}Ga_xAs$ material. Now we see that by controlling the ratio of x and y we can move from InP (y=0) to InGaAs (y=1) along a horizontal line, which means that the lattice constant is the same everywhere along this line which describes the quaternary semiconductor $In_{1-x}Ga_xAs_yP_{1-y}$. From the graph we see that by varying y between 0 and 1 we can change the emission wavelength (the wavelength corresponding to the band gap) from 0.92 eV to 1.65 eV, which covers the full C-band and O-band windows for optical communication.

bandgap engineering

We call this approach to vary the properties of ternary and quaternary semiconductors by controlling the composition during epitaxial growth bandgap engineering. It is widely used for InGaAsP, but it can also be used for other compound semiconductors. More information about bandgap engineering is provided in Section 4.7 of Chapter 4.

3.3 Band structure, Schrödinger equation and Fermi-Dirac distribution

Emission and absorption of light as well as electrical properties of semiconductors have a close relation to the electron energies, which in semiconductors are characterized by band structures. In a semiconductor crystal each electron is represented by

wave function

a quantum mechanical wave function of the form

$$\psi_{n,\boldsymbol{k}}(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\, u_{n,\boldsymbol{k}}(\boldsymbol{r}), \tag{3.1}$$

where $u_{n,\boldsymbol{k}}$ is a function with the same periodicity R as the crystal, or

$$u_{n,\boldsymbol{k}}(\boldsymbol{r}) = u_{n,\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{R}), \tag{3.2}$$

crystal periodicity band index

with $\boldsymbol{R}$ a vector indicating the periodicity of the crystal. In Eqs. (3.1) and (3.2), n is the band index, it labels the orbitals of the electrons around the atoms or molecules in the crystal. Because in the crystal the outer orbits are strongly coupled they split up in a large number of energy levels or states. Each quantum mechanical state is

quantum state Pauli principle

labeled by the combination $(n,\boldsymbol{k})$ and can be occupied by two electrons, one with spin up, the other with spin down (Pauli principle). Because of the exclusion principle, the next two electrons must have a different wave vector $\boldsymbol{k}$. In fact, the wave vector $\boldsymbol{k}$ is quantized, with one vector per $(2\pi)^3/V_c$ volume in $\boldsymbol{k}$-space (reciprocal space), where V_c is the crystal volume. Hence, the density of states in $\boldsymbol{k}$-space is $V_c/(2\pi)^3$, but for a "macroscopically" large crystal of 1 mm^3, say, the wave vector $\boldsymbol{k}$ can be considered for all practical purposes as continuous.

Schrödinger equation

The wave function (3.1) satisfies the Schrödinger equation

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + U_c(\boldsymbol{r})\right]\psi_{n,\boldsymbol{k}}(\boldsymbol{r}) = E_n(\boldsymbol{k})\psi_{n,\boldsymbol{k}}(\boldsymbol{r}) \tag{3.3}$$

crystal potential energy spectrum

where $U_c(\boldsymbol{r})$ is the (periodic) crystal potential with periodicity vector $\boldsymbol{R}$, i.e., $U_c(\boldsymbol{r}) = U_c(\boldsymbol{r}+\boldsymbol{R})$, and the corresponding energy spectrum is denoted by $E_n(\boldsymbol{k})$, m is the mass of the particle and the reduced Plank's constant $\hbar = h/2\pi$ equals Plank's constant h

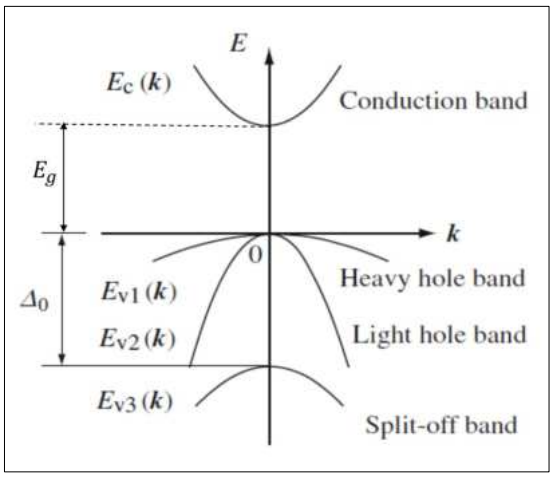

Figure 3.10: Typical band structure for a direct semiconductor, near $\boldsymbol{k} = 0$. There is one conduction band, $E_c(\boldsymbol{k})$, and three valence bands $E_{vi}(\boldsymbol{k}), (i = 1,2,3)$. The energy gap E_g of forbidden energy is indicated. The top of the split-off band $E_{v3}(\boldsymbol{k})$ lies Δ_0 below the top of the other two valence bands. This split-off energy is related to the interaction of spin and angular momentum.

divided by 2π. A typical semiconductor band structure in one dimension is sketched in Fig. 3.10. As indicated, the bands are separated by a gap E_g of forbidden energy. No electron states are available in the gap. The occurrence of a bandgap is typical for semiconductors; its value is characteristic for the specific semiconductor, whether compound or pure. The three uppermost bands below the energy gap are the valence bands. The lowest band above the gap is the conduction band. Figure 3.10 shows only one band above the gap; in principle there can be more with higher energy, but they are less relevant generally. In case the minimum of the conduction band and the maximum of the valence band occur for the same $\boldsymbol{k}$-vector, one speaks of a *direct* semiconductor. When this is not the case the semiconductor is referred to as *indirect*. Transitions between the conduction band and the valence band (e.g., in case of absorption or emission of a photon) are much more probable for direct semiconductors than for indirect semiconductors, where a phonon is required to bridge the difference in $\boldsymbol{k}$-vector. Direct bandgap semiconductors (like most III-V-semiconductors) have, therefore, higher absorption or spontaneous emission than indirect bandgap semiconductors (like silicon and germanium), and they support stimulated emission, as opposed to indirect semiconductors.

band structure
energy gap
valence band
conduction band
direct semiconductor
indirect semiconductor

In an ideal semiconductor at temperature $T = 0\,\mathrm{K}$, all states up to the valence bands are occupied by electrons and the conduction band is empty. At $T > 0$ a fraction of electrons from the valence band will be thermally excited to the conduction band, leaving behind holes, or missing electrons, in the valence band. There may be other mechanisms leading to electrons in the conduction band, such as by doping the material with special atoms (see Section 3.4) or by absorption of light (Section 3.8.1).

The band structure and the band gap in particular are crucial in understanding electrical and optical properties of a semiconductor. For many applications this is especially true in the vicinity of $\boldsymbol{k} = 0$. This point in the reciprocal space (the $\boldsymbol{k}$-space) is called the Γ-point. It can be derived from the Schrödinger equation that important physical quantities are related to the band structure.

Gamma point

Analogy between optical physics and quantum mechanics. As the Schrodinger equation is a wave equation, there is a far reaching analogy between quantum mechanics and optical waveguide physics. If we substitute the description of a waveguide mode $\vec{U} = U_{xy}e^{-j\beta}$ in the Helmholtz equation $\nabla^2\vec{U} + k^2\vec{U} = 0$ we obtain

$$[\nabla_{xy}^2 + n^2(x,y)k_0^2 - \beta_n^2]U_{xy} = 0$$

where $n(x,y)$ is the refractive index profile, k_0 is the wavenumber in vacuum and β_n is the propagation constant of the n-th mode. As we see, this formula is very similar to the Schrödinger equation

$$\left[\nabla^2 - \frac{2m}{\hbar^2}[U_c(\boldsymbol{r}) - E_n(\boldsymbol{k})]\right]\psi_{n,\boldsymbol{k}}(\boldsymbol{r}) = 0$$

where the potential function $\frac{2m}{\hbar^2}U_c(\boldsymbol{r})$ has the same role as the index profile $n^2(x,y)k_0^2$ and the energy $E_n(\boldsymbol{k})$ compares to β^2. In two dimensions a quantum-mechanical wave can be compared with a waveguide mode, and its energy with the (squared) propagation constant, which is also quantized; it can assume a large number of discrete values if the waveguide is strongly multimoded. So the wave functions of electrons and holes in a quantum well look very similar to the modes in a dielectric waveguide, with a cosine-like profile in the well and exponential tails in the barriers. For photonic crystals the analogy with quantum mechanics is even stronger, including forbidden bandgaps.

electron velocity

The velocity of an electron in state $(n, \boldsymbol{k})$, for example, is given by

$$v_n(\boldsymbol{k}) = \frac{1}{\hbar}\nabla_k E_n(\boldsymbol{k}) \tag{3.4}$$

group velocity

where ∇_k denotes the divergence in $\boldsymbol{k}$-space, with components $\partial k_x, \partial k_y, \partial k_z$. Note that Equation 3.4 is the quantum mechanical analogue of the group velocity of electromagnetic waves related to the dispersion relation.

effective mass
light hole band
heavy hole band

The effective mass m_n of the electrons and holes is related to the curvature $\partial^2 E_n(k)/\partial k^2$ of the bands. A strong band curvature corresponds to a light mass (e.g., the light hole band in Figure 3.10), a weak curvature to a heavy mass (e.g., the heavy hole band). The effective mass is dependent on the magnitude and the direction of the $\boldsymbol{k}$-vector and should for accurate calculations be described with a tensor.

Velocity and effective mass play important roles for describing the properties of a semiconductor. For example, if $f_n(\boldsymbol{k})$ denotes the probability that a state $(n, \boldsymbol{k})$ is occupied by an electron, the total current density can be expressed as

$$\boldsymbol{j} = \frac{2e}{V_c}\sum_{n,\boldsymbol{k}} f_n(\boldsymbol{k})\,\boldsymbol{v}_n(\boldsymbol{k}) \tag{3.5}$$

where the factor 2 in front is due to the spin degeneracy. By recalling the density of wave vectors in $\boldsymbol{k}$-space $V_c/(2\pi)^3$, we can approximate the summation over $\boldsymbol{k}$ by an integration:

$$\boldsymbol{j} = \frac{e}{4\pi^3}\sum_n \int d^3k\, f_n(\boldsymbol{k})\, v_n(k). \tag{3.6}$$

In the absence of applied fields, the energy bands and their occupation probabilities are inversion symmetric, i.e., invariant for $\boldsymbol{k} \to -\boldsymbol{k}$. This implies that a semiconductor normally exhibits no net current, but an applied electric field can lead to a skew occupation probability and thus to a net non-zero electrical current (see Figure 3.11).

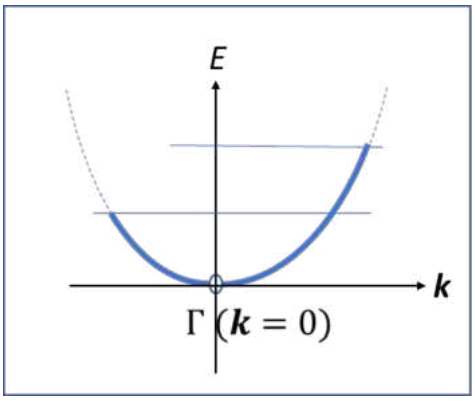

Figure 3.11: Example of skew band occupation representing a situation with non-zero current.

The effective mass is relevant when considering electron mobility and conductance. In thermal equilibrium, the occupation probability function, i.e. the probability that an electron occupies a quantum mechanical state with energy E, is called the Fermi-Dirac distribution, and can be derived from statistical considerations (See Kittel and Kroemer [101]). We will sketch a simplified picture for the situation at $T = 0$ K, when the crystal is in its ground state, i.e., the state with lowest total electron energy. Each state $(n, \boldsymbol{k})$ can only be occupied by two electrons at most. The lowest total energy is achieved when the lowest energy bands are occupied up to a certain maximum energy E_F, also referred to as the Fermi energy. In this case, the Fermi-Dirac distribution function f_F is simply given by

electron mobility
occupation probability function
Fermi-Dirac distribution
Fermi energy

$$f_F(E) = \left\{ \begin{array}{ll} 1 & E < E_F \\ 0 & E > E_F \end{array} \right\}, \tag{3.7}$$

where the value of E_F follows from the total number of electrons (see Equation 3.9). Note that Eq. 3.7 is valid only at $T = 0$ K. For $T > 0$, some electrons have jumped to higher energy levels, leaving behind empty states below E_F. In that case, there is a non-zero probability that some energy states above E_F become occupied by electrons and some states below E_F will be empty.

The Fermi-Dirac distribution function for a system in equilibrium at temperature T is given by (see Kittel and Kroemer [101])

$$f_F(E) = \frac{1}{1 + exp(\frac{E - E_F}{k_B T})}, \tag{3.8}$$

where k_B is the Boltzmann constant. It should be noted that the Fermi energy E_F depends in principle on T because of the requirement that the total number of electrons N_e satisfies

$$N_e = 2 \sum_{n,\boldsymbol{k}} f_F(E_n(\boldsymbol{k}) \tag{3.9}$$

A plot of the Fermi-Dirac function for various temperatures is given in Figure 3.12. Note that $f_F(E_F) = 1/2$.

In many cases, the band structure of the conduction band can be approximated near $\boldsymbol{k} = 0$ as

$$E(\boldsymbol{k}) = E_c + \frac{\hbar^2 k^2}{2 m_c}, \tag{3.10}$$

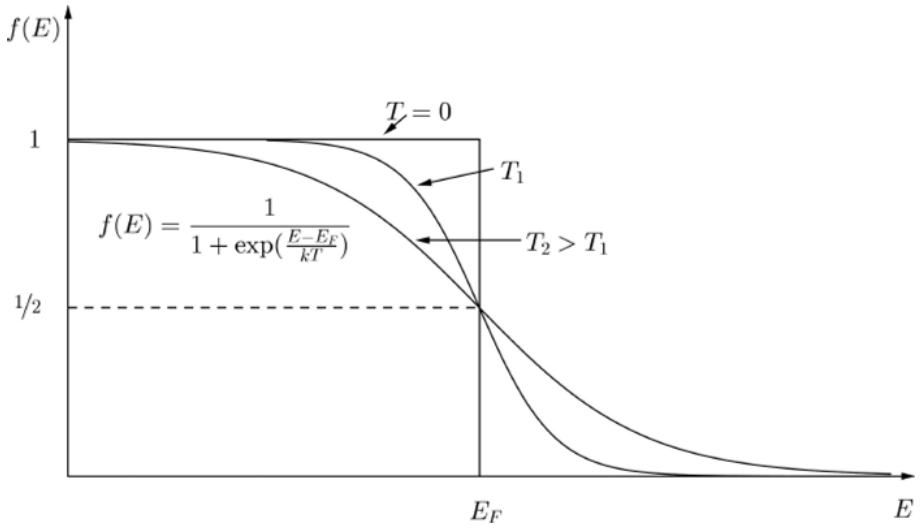

Figure 3.12: The Fermi-Dirac distribution function versus energy for different temperatures.

where E_c is the energy of the minimum of the conduction band and m_c is the conduction band effective mass. Similarly, the valence band energy can be approximated as

$$E(\boldsymbol{k}) = E_v - \frac{\hbar^2 k^2}{2m_v}, \tag{3.11}$$

with E_v the energy of the top of the valence band and m_v the valence band effective mass. Eqns. 3.10 and 3.11 are referred to as the parabolic band approximations near $\boldsymbol{k} = 0$. In the parabolic approximation, a surface of equal energy $E \geq E_c$ in $\boldsymbol{k}$-space is a sphere with radius $k = \sqrt{2m_c(E-E_c)}/\hbar$ and the volume in $\boldsymbol{k}$-space of a shell with thickness dE equals $4\pi k^2 \frac{dk}{dE} dE = 2\pi(\frac{2m_c}{\hbar^2})^{3/2}(E-E_c)^{1/2}dE$. When multiplied with the density of states in $\boldsymbol{k}$-space, $V_c/(2\pi)^3$, we arrive at the number of conduction-band states with energy between E and $E+dE$:

parabolic band approximation

$$N_c(E)dE = \frac{2V_c}{(2\pi)^2}(\frac{2m_c}{\hbar^2})^{3/2}(E-E_c)^{1/2}dE, \tag{3.12}$$

where the factor 2 in front of V_c is due to the electron spin. Then, the total density of electrons in the conduction band in equilibrium at temperature T is given by multiplying Equation 3.12 with the Fermi-Dirac function and integrating over the conduction band energy:

$$n_e = \frac{1}{V_c}\int_{E_c}^{\infty} dE N_c(E) f_F(E) = \frac{1}{2\pi^2}(\frac{2m_c}{\hbar^2})^{3/2}\int_{E_c}^{\infty} dE \frac{(E-E_c)^{1/2}}{1+exp(\frac{E-E_F}{k_B T})} \tag{3.13}$$

Here we replaced the upper value of the conduction band by ∞, which is a justified approximation because of the rapid decline of the Fermi-Dirac distribution. Moreover, if we assume that the Fermi energy lays within the gap and T is small enough that for $E \geq E_c, E-E_F \gg k_B T$, we can approximate $\frac{1}{1+exp(\frac{E-E_F}{k_B T})}$ by $exp[-\frac{E-E_F}{k_B T}]$, which is known as the *Boltzmann approximation*. With this approximation Eq. 3.13 can be simplified

Boltzmann approximation

Problem: Calculate at room temperature the carrier and hole concentrations in intrinsic InP and lattice-matched InGaAs.

Solution: ??

Problem 3.2:

to (see Appendix A for the derivation)

$$n_e = N_c exp^{-\frac{E_c - E_F}{k_B T}}, \tag{3.14}$$

where N_c is called the effective density of states in the conduction band,

$$N_c \cong \frac{1}{4}(\frac{2m_c k_B T}{\pi\hbar^2})^{3/2}. \tag{3.15}$$

For the holes in the valence band we can follow a similar procedure. Since a hole is a non-occupied electron state, the hole equilibrium distribution is not $f_F(E)$ but rather

$$1 - f_F(E) = \frac{1}{1 + exp(\frac{E_F - E_)}{k_B T}} \approx e^{-\frac{E_F - E}{k_B T}}, \tag{3.16}$$

where the last step is again the above-introduced Boltzmann approximation, valid when $E_F - E_v \gg k_B T$ and $E \leq E_v$.

In the same way as for n_e in Appendix A, we find for the thermal hole concentration

$$n_h = N_v exp^{-\frac{E_F - E_v}{k_B T}}, \tag{3.17}$$

where N_v is called the effective density of states in the valence band,

$$N_v \cong \frac{1}{4}(\frac{2m_v k_B T}{\pi\hbar^2})^{3/2}. \tag{3.18}$$

In an intrinsic semiconductor we have $n_e = n_h$ (charge neutrality). Hence Eqs. 3.14 and 3.17 are identical in this case.

It is noted that in many practical cases the effective electron mass in the conduction band is of the same order as the electron mass $m_c \approx m_e$, whereas the valence band effective mass can be larger than the electron mass, $m_v > m_e$, by a factor 10 or more; "holes are heavier than electrons". Also note that we have been dealing so far with intrinsic, or pure, semiconductors, for which E_F is in the center of the energy gap.

3.4 Doped semiconductors

A semiconductor consisting of a pure crystal without impurities is called an intrinsic semiconductor. The intrinsic semiconductor may be an interesting material, but the real power of semiconductors is realized by adding small controlled amounts of specific dopant, or impurity, atoms. The technique of adding impurity atoms to a semiconductor material to change its conductivity is called *doping*. Doping can significantly alter the electrical characteristics of the semiconductor. The doped material, called an extrinsic semiconductor, is essential for fabrication of the various semiconductor devices that we will treat in later chapters. Consider an intrinsic perfect semiconductor crystal from group IV, such as silicon, for which the valence band is filled

intrinsic semiconductor

doping

extrinsic semiconductor

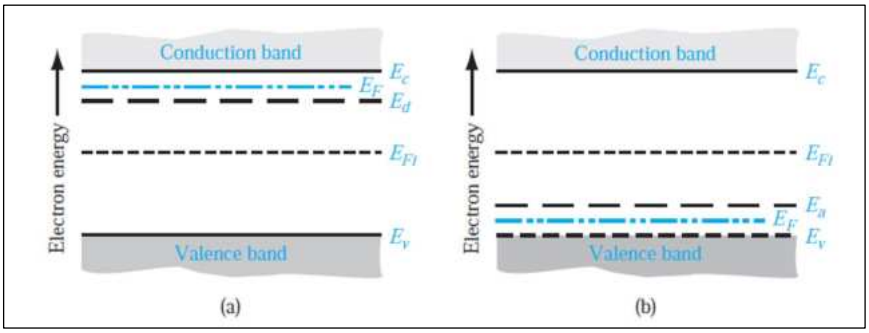

Figure 3.13: Simplified energy-band diagrams for (a) n-type and (b) p-type semiconductor at T=0. E_{Fi} is the intrinsic Fermi level, E_d the donor level and E_a the acceptor level. $E_c - E_d$ is the binding energy of the outermost electron of the type-V atom; $E_a - E_v$ is the binding energy of the hole generated by the missing electron of the type-III atom.

with electrons. Adding a small fraction of group V atoms to the crystal will add excess electrons without changing the band structure. For these extra electrons, there is room only in the conduction band. Indeed, the energy needed to separate these electrons from their atom is much smaller than the thermal energy $k_B T$ at room temperature.

donor impurity

Hence, these so-called donor impurity atoms add electrons to the conduction band without creating holes in the valence band. The resulting semiconductor is referred

n-type

to as n-type (n for negatively charged electron). On the other hand, adding a small fraction of group III atoms will contribute one valency electron less and, by a similar mechanism as for the donor electrons, the missing electron gives rise to a hole in the

acceptor impurity

valence band. These group III atoms are called acceptor impurity atoms. They generate holes in the valence band without adding electrons to the conduction band. The

p-type

resulting semiconductor is referred to as p-type (p for positively charged hole). The Fermi energy will change if dopant atoms are added.

For III-V semiconductors group VI atoms may be added as donor atoms, making the material n-type, and group II atoms to make it p-type. Adding group IV atoms can make the material n-type or p-type, dependent on whether they replace a group III or a group V atom in the crystal.

Fig. 3.13 shows schematic simplified energy-band diagrams for n-type and p-type semiconductors at $T = 0$. In (a) the excess electrons of the donor atoms occupy the donor energy levels in the bandgap just below the conduction band. The Fermi level is precisely halfway the topmost occupied energy level (the donor level E_d) and the conduction band minimum E_c. In (b) the acceptor atoms give rise to empty acceptor energy levels in the gap just above the valence band. The Fermi energy is now halfway the acceptor level E_a and the top of the valence band E_v. The position of the intrinsic Fermi level E_{Fi} at $T = 0$ is indicated.

In Fig. 3.14 the situation is sketched in case (a) the donor electrons have been thermally excited to the conduction band and (b) electrons from the valence band have been thermally activated to occupy all acceptor states, leaving behind holes in the uppermost part of the valence band. Since the activation energies are typically small fractions of the bandgap energy, the thermal activation will be achieved for temperatures much smaller than room temperature.

If a semiconductor contains both donor and acceptor impurities in the same region, it

compensated semiconductor

is referred to as compensated semiconductor. In an n-type compensated semiconduc-

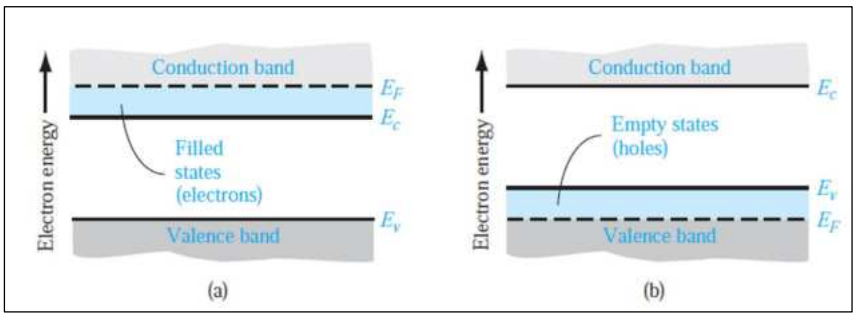

Figure 3.14: Simplified energy-band diagrams for (a) n-type and (b) p-type semiconductor at $T > 0$.

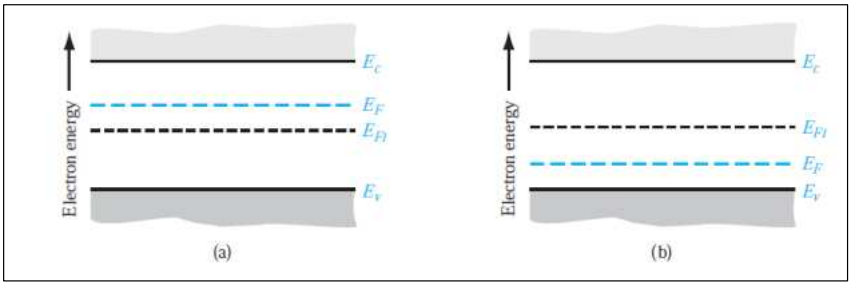

Figure 3.15: Position of the Fermi level for (a) n-type ($N_d > N_a$) and (b) p-type ($N_d < N_a$) compensated semiconductor.

tor we have $N_d > N_a$ and in a p-type $N_d < N_a$. A completely compensated semiconductor, i.e. with $N_d = N_a$, has the characteristics of an intrinsic material. The Fermi level in an n-type compensated semiconductor is in the bandgap and above the intrinsic Fermi level; likewise, in a p-type compensated semiconductor the Fermi level is in the bandgap and below the intrinsic Fermi level. This is schematically indicated in Fig. 3.15.

When we assume that the Fermi energy of the doped semiconductor is within the forbidden-energy bandgap and not too close to E_c or E_v, the theory of the previous section applies, and the density n_e of electrons in the conduction band is given by (see Eq. 3.14)

forbidden-energy bandgap

$$n_e = N_c exp^{-\frac{E_c - E_F}{k_B T}},$$

where we have assumed that $(E_c - E_F) \gg k_B T$, so that the Fermi probability function reduces to the Boltzmann approximation and N_c is the effective density of quantum states in the conduction band given by Eq. 3.15. Similarly, in the same approximation the density of holes in the valence band is given by Eq. 3.17,

$$n_h = N_v exp^{-\frac{E_F - E_v}{k_B T}},$$

with N_v the effective density of states in the valence band given by Eq. 3.18.

When the concentration of electrons in the conduction band exceeds the density of states N_c, the Fermi energy lies within the conduction band. This type of semiconductor is called a degenerate n-type semiconductor. In a similar way, the Fermi energy will lie in the valence band when the concentration of holes exceeds the density of states

degenerate semiconductor

Problem: Calculate the carrier and hole concentrations at room temperature for the following layers of the standard waveguide structure (see Appendix C, Fig. C.2):

1. n-type InP substrate with a doping level of $10^{18}/\text{cm}^3$
2. n-type Q1.25 waveguide layer with a doping level of $10^{16}/\text{cm}^3$ (d=500 nm)
3. p-type InP cladding layer with a doping level of $10^{18}/\text{cm}^3$(d=1000 nm)
4. p-type InGaAs contact layer with a doping level of $10^{19}/\text{cm}^3$(d=300 nm)

Solution: ??

Problem: Calculate the square resistance of the layers 2-4 above.

Solution: ??

Problem 3.3:

N_v. This type of semiconductor is called a degenerate p-type semiconductor. Degenerately doped p and n-type semiconductors are very important in high-speed transistors and in certain types of sensors and detectors. The increased carrier concentration allows for better control of electrical conductivity and enhanced electrical and optical performance of devices.

3.5 The pn-junction

3.5.1 Zero applied bias

We now will study situations where a p-type and an n-type semiconductor are brought into tight contact with each other such that carriers can freely move through the interface, or junction, between the two materials. This configuration is referred to as pn-junction. Most semiconductor devices have at least one such (or similar) junction; the device characteristics are intimately connected to these pn-junctions.

pn-junction

Now consider the situation sketched in Fig. 3.16(a), where the left part of a single semiconductor crystal has been doped with acceptor atoms of concentration N_a and the right part with donor atoms of concentration N_d. This is referred to as homojunction, because the semiconductor crystal is the same left and right of the junction. In this situation we initially have an abrupt change in density of electrons at the interface (Fig. 3.16(b), dashed horizontal lines). In general, carriers will then redistribute by diffusion in such a manner as to reduce the density gradient. Thus, electrons move from the n-region into the p-region, or equivalently, holes move from the p-region into the n-region. Since the structure was initially charge neutral, the consequence of the diffusive charge redistribution is the formation of a dipole layer on both sides of the interface, with negative space charge to the left and positive space charge to the right. This continues until equilibrium is reached where the diffusion force acting on the electrons will be compensated by the counteracting electrostatic Coulomb force induced by the dipole-induced electric field. This is schematically illustrated in Fig. 3.16(c).

homojunction

The indicated space-charge region is also called depletion region. If we assume that no voltage is applied to the pn-junction, then the system will be in thermal equilibrium,

depletion region

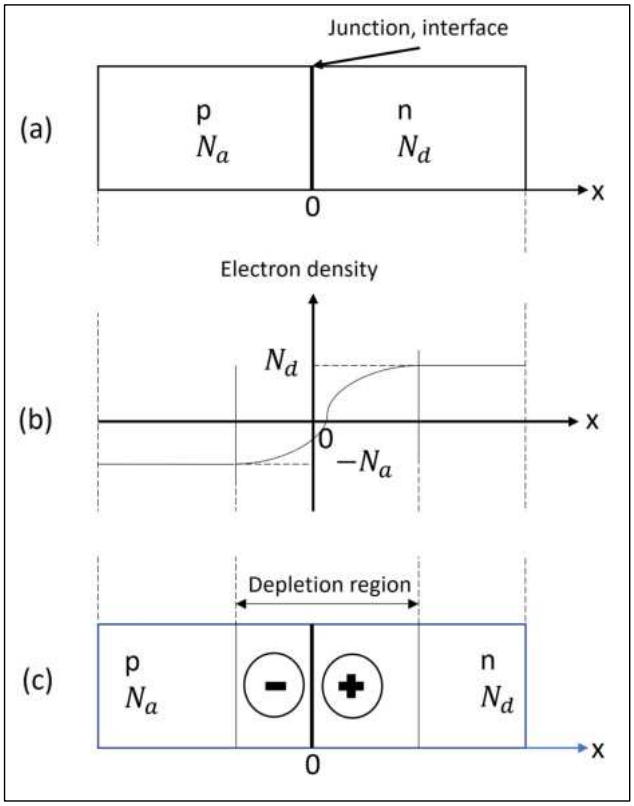

Figure 3.16: Charge redistribution when (a) two doped semiconductor crystals are put together to form a pn-junction. In (b) the large density gradient induces diffusion forces driving electrons to the left (and holes to the right), leading to (c) an equilibrium situation with a space-charge or depletion region as indicated.

which means that the Fermi energy level is constant everywhere in the system. In terms of the band diagram for this situation this means in case of compensated semiconductors, that the conduction and valence bands must bend as we go through the space charge region, as sketched in Fig. 3.17. Far away from the junction the band diagrams are as shown in Fig. 3.15. Here, we will use the index i to indicate quantities that pertain to the intrinsic, i.e., non-doped semiconductor. The electron concentration in the conduction band of the n-region is (see Eq. 3.14)

$$n_e = N_d = N_c e^{-\frac{E_c - E_F}{k_B T}} = n_i e^{\frac{E_F - E_{Fn}}{k_B T}} = n_i e^{-\frac{e\phi_{Fn}}{k_B T}}, \tag{3.19}$$

where N_d is the donor concentration, $n_i \equiv N_c e^{-\frac{E_c - E_{Fn}}{k_B T}}$ is the intrinsic carrier concentration with $E_F n$ the intrinsic Fermi energy in the n-region, and ϕ_{Fn} is defined as (see Fig. 3.17)

$$e\phi_{Fn} \equiv E_{Fn} - E_F. \tag{3.20}$$

Similarly, in the p-region, the hole concentration is (27)

$$n_h = N_a = n_i e^{\frac{E_{Fp} - E_F}{k_B T}} = n_i e^{\frac{e\phi_{Fp}}{k_B T}}, \tag{3.21}$$

with N_a the acceptor concentration in the p-region and the potential ϕ_{Fp} is defined (see Fig. 3.17) by

$$e\phi_{Fp} = E_{Fp} - E_F. \tag{3.22}$$

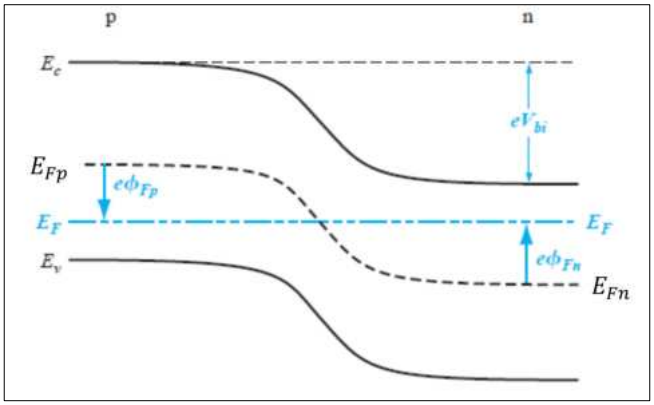

Figure 3.17: Energy-band diagram of a pn-junction of compensated semiconductors in thermal equilibrium; no electrical current flows through the structure. Far away from the junction the band diagrams are as shown in Fig. 3.15.

built-in potential barrier V_b

Electrons in the conduction band of the n-region are blocked by the built-in potential barrier V_b, as indicated in Fig. 3.17 and similarly for holes in the p region. This barrier is a consequence of thermal equilibrium and maintains equilibrium between majority electrons in the n-region and minority electrons in the p-region, and similarly between majority holes in the p-region and minority holes in the n-region. V_b can be expressed as the difference between the intrinsic Fermi levels in the p and n-regions in terms of the potentials ϕ_{Fn} and ϕ_{Fp}, i.e.,

$$V_b = |\phi_{Fn}| + |\phi_{Fp}|. \tag{3.23}$$

According to Eq. 3.19 we can express the potential ϕ_{Fn} as

$$\phi_{Fn} = -\frac{k_B T}{e} \ln(\frac{N_d}{n_i}) \tag{3.24}$$

with N_d the donor concentration. Similarly, we can derive from Eq. 3.21

$$\phi_{Fp} = +\frac{k_B T}{e} \ln(\frac{N_a}{n_i}). \tag{3.25}$$

The built-in potential barrier V_b for the junction, introduced in Eq. 3.23, can be expressed, using Eqs. 3.23, 3.24 and 3.25 as

$$V_b = \frac{k_B T}{e} \ln(\frac{N_a N_d}{n_i^2}) = V_T \ln(\frac{N_a N_d}{n_i^2}) \tag{3.26}$$

thermal voltage

where $V_T \equiv \frac{k_B T}{e}$ is the thermal voltage.

3.5.2 Space charge width in the junction area

In Fig. 3.16, the depletion region around the pn-junction was introduced. This region is characterized by space charge, which is illustrated in Fig. 3.18, where we have assumed that the space-charge region abruptly ends in the n-region at $x = x_n$ and in the p-region at $x = -x_p (x_p > 0)$. We denote the electric potential by $\phi(x)$ which is related to the charge density $\rho(x)$ by the Poisson equation (in 1 dimension)

$$\frac{d^2\phi(x)}{dx^2} = -\frac{\rho(x)}{\epsilon_s}, \tag{3.27}$$

Problem: Calculate the potential barrier in the InP cladding layer on top of the standard waveguide (see Appendix C, Fig. C.2), with a p-type doping level of 10^{17} and an n-type doping level of 10^{16}

Solution:

Problem 3.4:

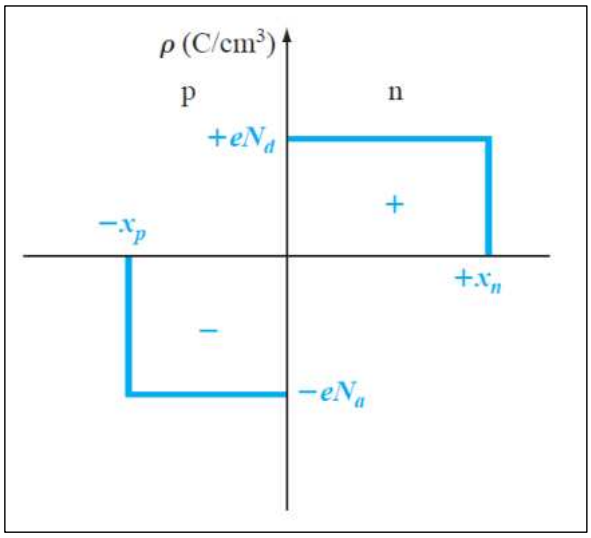

Figure 3.18: Space charge density corresponding to the abrupt junction approximation in Fig. 3.16(c).

with ϵ_s the permittivity of the semiconductor. The charge density is (see Fig. 3.18)

$$\rho(x) = \left\{ \begin{array}{ccc} 0 & , & x < -x_p \,\&\, x > x_n \\ -eN_a & , & -x_p < x < 0 \\ eN_d & , & 0 < x < x_n \end{array} \right\}. \tag{3.28}$$

Note that, because of charge neutrality, we must have

$$N_a x_p = N_d x_d, \tag{3.29}$$

and the boundary conditions (see Fig. 3.17) are (37)

$$\phi(x) = \left\{ \begin{array}{ccc} 0 & , & x < -x_p \\ V_b & , & x > x_n \end{array} \right\}. \tag{3.30}$$

It can easily be verified that the solution of Eq. 3.27 with Eq. 3.28, which is continuous, and satisfies Eq. 3.30, is given by

$$\phi(x) = \left\{ \begin{array}{ccc} \frac{eN_a}{2\epsilon_s}(x+x_p)^2 & , & -x_p < x < 0 \\ \frac{eN_d}{\epsilon_s}(x_n x - \frac{1}{2}x^2) + \frac{eN_a}{2\epsilon_s}x_p^2 & , & 0 < x < x_n \end{array} \right\}. \tag{3.31}$$

Therefore, since $V_b = \phi(x_n) - \phi(-x_p) = \phi(x_n)$, we have (39)

$$V_b = \frac{e}{2\epsilon_s}(N_d x_n^2 + N_a x_p^2). \tag{3.32}$$

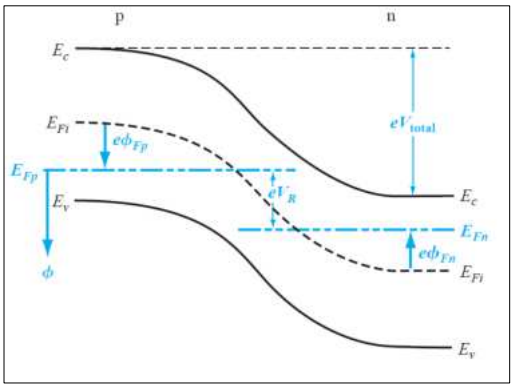

Figure 3.19: Energy-band diagram of a pn-junction under reverse bias V_R

The built-in potential is through Formula 3.31 related to the space-charge penetration distances x_n and x_p. Combined with Eq. 3.29 we can solve for x_n and x_p,

$$x_n = \sqrt{\frac{2\epsilon_s V_b}{e} \frac{N_a}{N_d(N_a+N_d)}}, \tag{3.33}$$

$$x_p = \sqrt{\frac{2\epsilon_s V_b}{e} \frac{N_d}{N_a(N_a+N_d)}} \tag{3.34}$$

The total width W of the space-charge region can then be expressed as

$$W = x_n + x_p = \sqrt{\frac{2\epsilon_s V_b}{e} \frac{(N_a+N_d)}{N_a N_d}} \tag{3.35}$$

Note that V_b is given by Eq. 3.26.

3.5.3 Reverse applied bias, junction capacitance and reverse current

If a positive potential is applied between the p and n regions, the system will no longer be in thermal equilibrium. Far away from the junction each part, n or p, will be in local equilibrium with its own Fermi level. The two Fermi levels, E_{Fp} and E_{Fn} will now differ by the applied voltage eV_R, as indicated in the sketched energy band diagram under reverse bias in Fig. 3.19.

Hence, the total internal barrier height V_{total} is now given by

$$V_{total} = V_b + V_R. \tag{3.36}$$

A consequence of the applied voltage V_R is that additional charge has displaced to enlarge the space charge region, the width of which is again given by Eq. 3.35 with V_b replaced by V_{total}, i.e.,

$$W = \sqrt{\frac{2\epsilon_s(V_b+V_R)}{e} \frac{(N_a+N_d)}{N_a N_d}}. \tag{3.37}$$

The width of the space-charge region has increased due to the reverse bias. This increase of the total electric charges around the interface of the pn-junction means that a capacitance C_{pn} can be assigned to the junction. It is defined as

$$C_{pn} = \frac{dQ}{dV_R}, \tag{3.38}$$

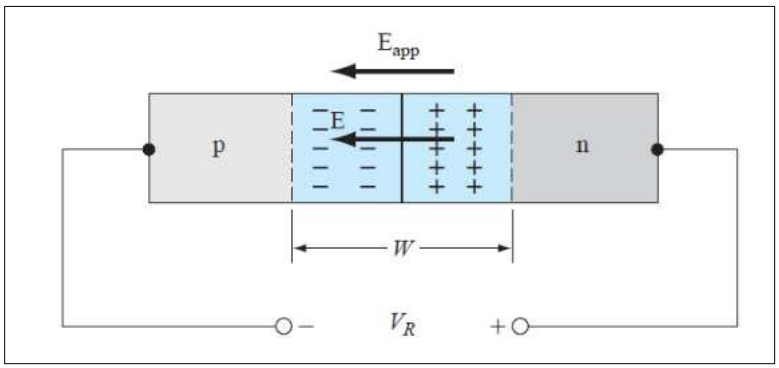

Figure 3.20: A pn-junction with applied reverse bias voltage $V_R > 0$, showing the direction of the current and the induced electric field.

where Q is the displaced charge, see Eq. 3.29

$$Q = eN_a x_p = eN_d x_n \tag{3.39}$$

For the barrier under reverse bias we can write, using Eq. 3.33,

$$x_n = \sqrt{\frac{2\epsilon_s(V_b + V_R)}{e} \frac{N_a}{N_d(N_a + N_d)}}. \tag{3.40}$$

The junction capacitance can now be written, using Eqs. 3.38-3.40 as

$$C_{pn} = eN_d \frac{dx_n}{dV_R} = \sqrt{\frac{e\epsilon_s}{2(V_b + V_R)} \frac{N_a N_d}{(N_a + N_d)}}. \tag{3.41}$$

It is also referred to as depletion-layer capacitance.

The effective built-in potential-energy barrier is $eV_{total} = e(V_b + V_R)$, with V_R the applied reverse bias voltage, see Fig. 3.19. Here, the positive pole is connected to the n-side and the negative pole to the p-side, see Fig. 3.20. The only way carriers can pass this barrier is by thermal activation, for which the probability is proportional to the Boltzmann factor $e^{-\frac{e(V_b+V_R)}{k_B T}}$. To deduce the dependence on V_R of the current that flows across the pn-junction, we must consider separately the electrical currents of electrons and holes, j_e and j_h , respectively. When $V_R = 0$, both j_e and j_h vanish, since the system is in thermal equilibrium. This does not, of course, mean that no individual carriers flow across the junction, but rather that as many electrons (or holes) flow in one direction as in the other. This balance is disrupted when $V_R \neq 0$. Consider, for example, the electrical current of holes across the depletion layer, which has two components:

1. Holes flow from the n- to the p-side of the junction leading to an electrical current known as the hole generation current. This current arises from holes that are generated just on the n-side of the depletion layer by thermal excitation of electrons out of the valence band. Although the density of such holes on the n-side (“minority carriers”) is very small compared with the density of electrons (“majority carriers”), they play an important role in carrying current across the junction. This is because any such hole that wanders into the depletion layer is immediately swept over to the p-side of the junction by the strong electric field that prevails within the layer. The resulting generation current is insensitive to the size of the potential drop across the depletion layer, since any hole, having entered the layer from the n-side, will be swept through to the p-side.

2. Holes flow from the p- to the n-side of the junction leading to an electrical current known as hole recombination current. This name stems from the fact that such holes upon arriving on the n-side of the junction will quickly recombine with one of the abundant electrons. Since these holes have to move against the potential barrier the number of successful holes is proportional to the Boltzmann factor $e^{-\frac{e(V_b+V_R)}{k_BT}}$ and the corresponding electrical current density satisfies

$$j_h^{rec} \propto e^{-\frac{e(V_b+V_R)}{k_BT}}. \tag{3.42}$$

In contrast to the generation current, the recombination current is highly sensitive to the applied (reverse) voltage V_R. Since at $V_R = 0$, there can be no net hole current across the junction, we have

$$j_h^{rec} = j_h^{gen} e^{-\frac{eV_R}{k_BT}}. \tag{3.43}$$

The total hole electrical current density flowing through the junction in the np-direction is given by the generation current minus the recombination current,

$$j_h = j_h^{gen} - j_h^{rec} = j_h^{gen}(1 - e^{-\frac{eV_R}{k_BT}}). \tag{3.44}$$

The same analysis applies to the components of the electron current, except that the generation and recombination currents for electrons flow oppositely to the corresponding hole currents. However, since electrons are oppositely charged, the electrical currents are parallel, thus leading to the total (reverse) electrical current density in the np-direction given by

$$J_R = (j_h^{gen} + j_e^{gen})(1 - e^{-\frac{eV_R}{k_BT}}) \equiv J_S^0(1 - e^{-\frac{eV_R}{k_BT}}). \tag{3.45}$$

valid for

$$V_R \geq 0 \tag{3.46}$$

and where $J_S^0 = j_h^{gen} + j_e^{gen}$ is a material-specific current, the reverse-saturation current density,

$$J_S^0 \propto e^{-\frac{eV_b}{k_BT}} \tag{3.47}$$

Note that for increasing V_R, the reverse current saturates at the value J_S^0.

3.5.4 Forward applied bias: the pn-junction diode

In case of forward bias voltage V_a, Equation 3.45 with V_R replaced by $-V_a$ yields for the forward current flowing in the pn direction (see the inset of Fig. 3.21)

$$J^0 = J_S^0(e^{\frac{eV_a}{k_BT}} - 1), \tag{3.48}$$

valid as long as the effective barrier $V_b - V_a$ exists, i.e., $V_a < V_b$. The current density is plotted in Fig. 3.21 as a function of the applied forward-bias voltage V_a, which can assume negative values as well. Equation 3.48 is known as the ideal-diode equation. In the derivation of the ideal current-voltage relationship, we neglected any effects occurring within the space-charge region. These effects cause the I-V relationship to deviate substantially from the ideal expression. The effects in question are related to the presence of trap levels within the bandgap, which provide the dominant mechanism for electron and hole generation in the junction region. With these traps as intermediate

Problem a): Estimate the width of the depletion layer in the standard waveguide layer for a reverse bias voltage of

- 0 V
- 1 V
- 5 V

Solution a):

Problem b): Estimate the capacitance of the depletion layer in the standard modulator based on the standard waveguide layer. with a width of 2 μm and a length of 500 μm, for a reverse bias voltage of

- 0 V
- 1 V
- 5 V

Solution b):

Problem c): Estimate the resistance of the path between two standard phase modulators in the arms of a Mach-Zehnder Interferometer, which are connected to an MMI coupler with waveguide sections of 100 μm long, if the 1 μm thick p-doped InP top layer

- is present
- is removed

Solution c):

Problem d): Calculate the leakage current if one electrode is grounded and the other is biased at -1 V, if the 1 μm thick p-doped InP top layer

- is present
- is removed

Solution d):

Problem e): Estimate the voltage at the second modulator, if the first one is biased at -5 V and the second one is floating.

Solution d):

Problem 3.5:

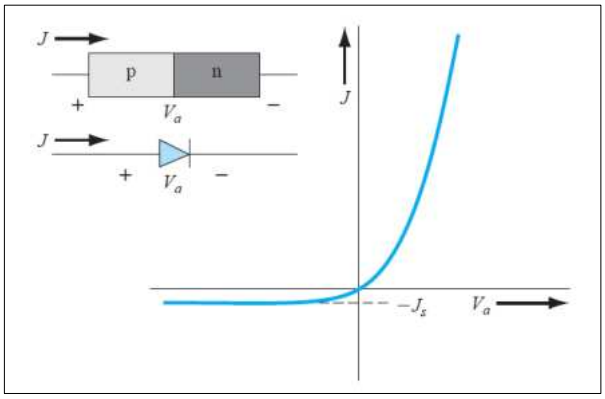

Figure 3.21: I-V characteristic of a pn junction diode.

Problem a): Estimate the contact resistance for the TLM structure in figure X, if the current through the structure is 1 mA and the voltages measured are V1=, V2=, V3=, V4=

Solution a):

Problem b): Estimate the voltage that is required to deplete the active layer of the reversely biased SOA-structure, for use as a detector. What is the effect if it is not completely depleted?

Solution b):

Problem 3.6:

states, the contribution to the recombination (forward) current can then be derived as (Neamen, Ch.8.2[102])

$$j^{rec} = j_{r0} e^{\frac{eV_a}{2k_B T}}, \tag{3.49}$$

where j_{r0} is the corresponding reverse-biased saturation current density, which is, in contrast to the ideal current case, not independent of the applied voltage, since it depends on the depletion width W, which in turn depends on the reverse bias voltage (see Eq. 3.35). In reality, the reverse-bias saturation current can be orders of magnitude larger than the ideal current J_S (see Neamen [102], Ch. 8.2.1). Usually, the total forward-bias current density is cast in a general form, similar to Eq. 3.48,

$$J_{tot} = J_S^{tot}(e^{\frac{eV_a}{nk_B T}} - 1), \tag{3.50}$$

where the parameter n is called the ideality factor. For large forward-bias voltage, $n = 1$ and for low forward-bias voltage, $n = 2$.

3.6 Quantum-well double heterostructures

We will now study a situation where a direct-bandgap semiconductor W (the “well” material), is sandwiched between a semiconductor B (the “barrier” material) at both

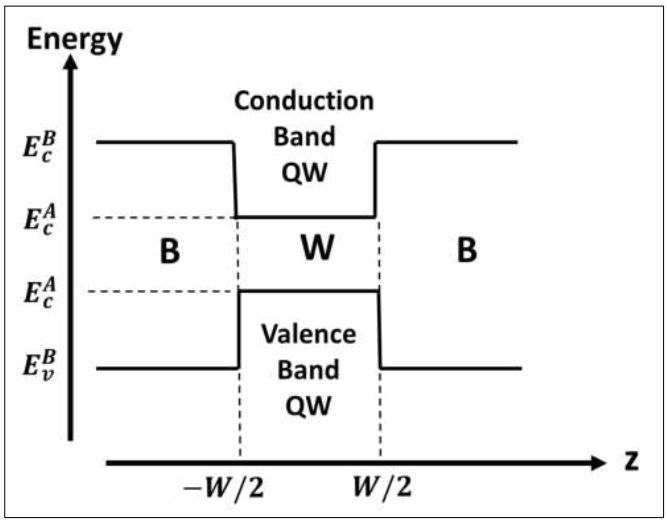

Figure 3.22: Schematic of the conduction and valence band-edge function $E_n^{edge}(z)$ for a Type I double heterostructure quantum well (QW). **B** stands for barrier material and **W** for well material.

sides, where the bandgap energies satisfy $E_g^W < E_g^B$, and the respective conduction and valence band edges E_c^W, E_c^B and E_v^W, E_v^B connect as sketched in Fig. 3.22. Such a band-edge situation is referred to as Type I line up. Well/Barrier combinations such as GaAs/AlGaAs, InGaAs/InP and GaN/AlGaN all have type I line up. Once the band line up is known, one has to decide how the electronic states in the heterostructure should be described. An approach that has been quite successfully applied to heterostructures is, in its simplest form, the replacement of the original Schrödinger equation (see Eq. 3.3: $\left(\left[-\frac{\hbar^2}{2m}\nabla^2 + U_c(\boldsymbol{r})\right]\psi_{n,k}(\boldsymbol{r}) = E_n(\boldsymbol{k})\psi_{n,k}(\boldsymbol{r})\right)$ by

$$\left[-\frac{\hbar^2}{2m_n}\nabla^2 + E_n^{edge}(z)\right]\Phi_n(\boldsymbol{r}) = E\Phi_n(\boldsymbol{r}) \tag{3.51}$$

where the crystal potential $U_c(r)$ is replaced by the band-edge energy $E_n^{edge}(z)$ with n indicating the conduction ($n = c$) or valence band ($n = v$), while the effect of the background potential is contained in the corresponding effective mass m_n. In the simple approximation of Eq. 3.51), the Schrödinger equation for the electron states in the conduction-band quantum well can be written

$$[-\frac{\hbar^2}{2m_c}\nabla^2 + V(z)]\Phi(r) = E\Phi(r), \qquad V(z) = \begin{cases} 0 & , \quad |z| \le \frac{W}{2} \\ E_c^B - E_c^W & , \quad |z| > \frac{W}{2} \end{cases} . \tag{3.52}$$

Since the potential V depends only on z, the wave function $\Phi(r)$ can be separated as follows:

$$\Phi_(r) = e^{ik_x.x}e^{ik_y.y}f(z), \tag{3.53}$$

where $f(z)$ satisfies the 1D Schrödinger equation

$$[-\frac{\hbar^2}{2m_c}\frac{d^2}{dz^2} + V(z)]f(z) = E_n f(z) \tag{3.54}$$

and the total energy E is given by

$$E = \frac{\hbar^2}{2m_c}(k_x^2 + k_y^2) + E_n. \tag{3.55}$$

Hence, each level E_n defines a parabolic subband in the x, y plane. Assuming the barrier infinitely high, the wavefunction solutions are confined to the well. In that case, the corresponding electron is confined to the well, but can move freely in the x,y plane with effective mass m_c. The wave functions $f(z)$ are given by

$$f(z) = \begin{cases} cos(\frac{\pi n z}{W}) & , \quad n = 1,3,5,\ldots \\ sin(\frac{\pi n z}{W}) & , \quad n = 2,4,6,\ldots \end{cases} , \tag{3.56}$$

with the corresponding well energies

$$E_n = \frac{\hbar^2}{2m_c}\frac{\pi^2 n^2}{W^2}. \tag{3.57}$$

If the barrier potential is not infinite, only the energy levels that satisfy $E_n < E_c^B - E_c^W$ correspond to electrons bound to the well, but with exponentially decaying wavefunction tails in the barrier regions. This is schematically indicated in Fig. 3.23.

In Sec. 3.3, we derived the number of electron states with energy between E and $E+dE$ (see Eq. 3.12) as $N_c(E)dE = \frac{2V_c}{(2\pi)^2}(\frac{2m_c}{\hbar^2})^{3/2}(E-E_c)^{1/2}dE$. In the quantum-well situation, however, this number is different and can be derived following a similar reasoning as in the derivation of Eq. 3.12, but now for each quantum level n the 2-dimensional parabola given by Eq. 3.55 must be considered. Thus the number of electron states in the conduction band with energy between E and $E+dE$ is given by the number of k-points in the circle band with radius $k = \sqrt{2m_c(E - E_c^W - E_n)}/\hbar$ and thickness dk, or $N_n(E)dE = \frac{2V_c}{(2\pi)^2 W} 2\pi k \frac{dk}{dE} dE$, with W the QW width. Hence, substituting for k and performing the differentiation, we obtain

$$N_n(E)dE = \begin{cases} \frac{V_c m_c}{\pi\hbar^2 W} dE & , \quad E > E_c^W + E_n \\ 0 & , \quad E < E_c^W + E_n \end{cases} , \; n = 1,2,\ldots. \tag{3.58}$$

Thus for each quantum number n, the density of states per unit volume is constant when $E > E_c^W + E_n$. The overall density of states is the sum of the densities for all values of n, so that it exhibits the staircase distribution shown in Fig. 3.24. Each step of the staircase corresponds to a different quantum number n and may be regarded as a subband within the conduction band. The bottoms of these subbands move progressively higher for higher quantum numbers. It is easily shown by substituting $E = E_c^W + E_n$ in Eq. 3.12, and by using Eq. **??**, that at $E = E_c^W + E_n$, the quantum-well density of states is the same as that for the bulk. The density of states in the valence band has a similar staircase distribution. In contrast with bulk semiconductor, the quantum-well structure exhibits a substantial density of states at its lowest allowed conduction-band energy level and at its highest allowed valence-band energy level.

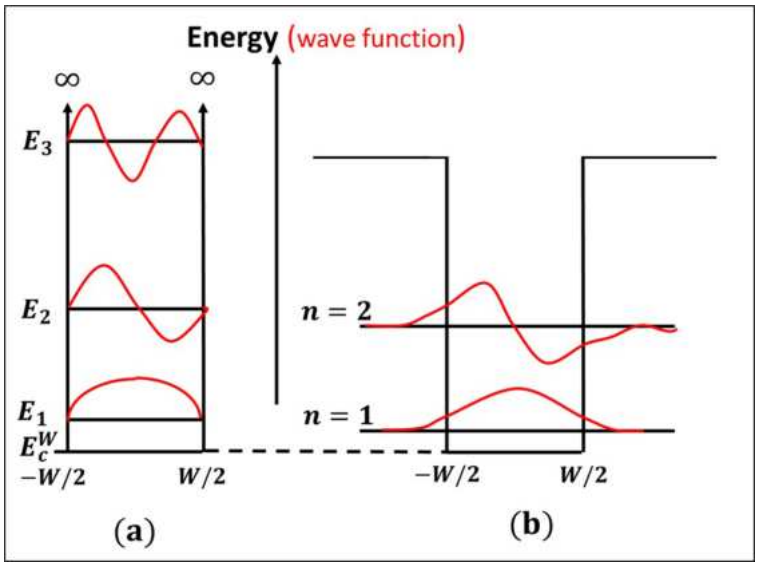

Figure 3.23: Schematic picture of the conduction-band QW energy levels comparing infinite (a) and finite (b) barriers. In case (b), only a finite number of confined states exists.

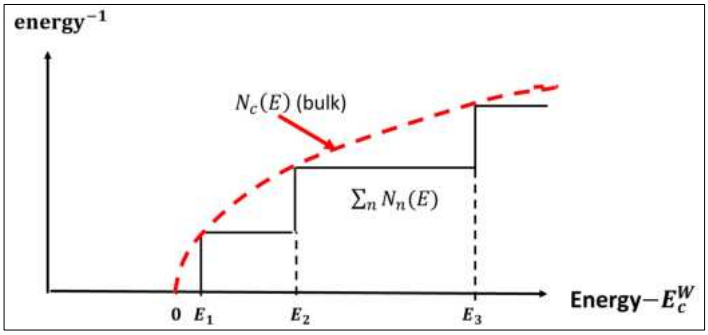

Figure 3.24: Schematic of density of states in a one-dimensional quantum well.

Problem a): Estimate for a reverse bias voltage of 5 V over the standard PHM waveguide structure the magnitude of

1. the Pockels effect
2. the band-filling effect
3. the Kerr effect
4. The Franz-Keldysh effect (?)
5. ??

Solution a):

Problem 3.7:

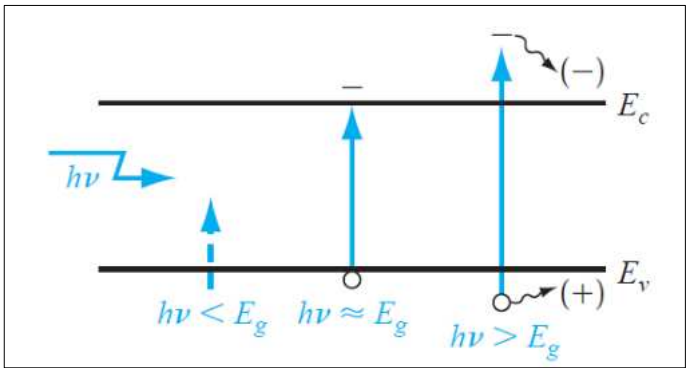

Figure 3.25: Optically generated e-h pair formation in a semiconductor when $h\nu > E_g$. When $h\nu < E_g$, no absorption occurs.

3.7 Electro-optic effects

3.8 Light emission and absorption in semiconductor devices

3.8.1 Optical absorption

When a semiconductor is illuminated with light, the photons may be absorbed or they may propagate through the semiconductor, depending on the photon energy and on the bandgap energy E_g. If the photon energy is less than E_g, the photons are not absorbed, but the light is transmitted through the material; the semiconductor appears to be transparent. Photons with $E = h\nu > E_g$ can interact with valence electrons and elevate the electrons into the conduction band. The valence band contains many electrons and the conduction band contains many empty states, so this interaction occurs with high probability, each photon creating one electron in the conduction band and one hole in the valence band, i.e., an electron–hole (e-h) pair. The basic absorption processes for different values of $h\nu$ are shown in Figure 3.25. When $h\nu > E_g$, an electron–hole pair is created, and the excess energy may give the electron or hole additional kinetic energy, which will be dissipated as heat in the semiconductor.

electron–hole (e-h) pair

As a result of the absorption process, the intensity of the photon flux, denoted by $I_\nu(x)$ and expressed in terms of $energy\, cm^{-2} s^{-1}$, decreases exponentially with distance x through the semiconductor material as

$$I_\nu(x) = I_\nu(0) e^{-\alpha x}. \tag{3.59}$$

This is shown in Fig. 3.26 as a function of x for two different values of the absorption coefficient α, the relative number of photons absorbed per unit distance, given in units of cm^{-1}. If the absorption coefficient is large, all photons are absorbed over a relatively short distance. The absorption coefficient in the semiconductor is heavily dependent on the photon energy and bandgap energy, and on the nature of the bandgap. Indirect-bandgap semiconductors, like silion, have much lower absorption than direct semiconductors, like most III-V semiconductors. This is illustrated in Fig. 3.27 where the absorption coefficient is plotted as a function of wavelength $\lambda = c/\nu$ for several semiconductor materials. The absorption coefficient α increases very rapidly

absorption coefficient

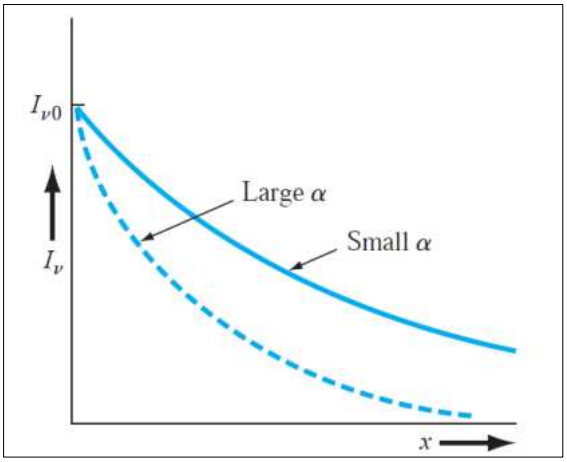

Figure 3.26: Photon intensity versus distance for two absorption coefficients.

for $h\nu > E_g$, or for $\lambda\,(\mu m) < \frac{1.24}{(E_g(eV)}$. The absorption becomes very small for $h\nu < E_g$, so the semiconductor appears transparent to photons in this energy range.

In addition to the interband processes in Fig. 3.25, absorption of light can also occur within a band (intrabandprocess). It occurs in doped semiconductors and involves carriers that can move freely, such as electrons in the conduction band and/or holes in the valence band. This process is called free-carrier absorption (FCA, also called plasma effect) and plays a significant role in lowering the efficiency of light emission in semiconductor lasers (see Chapter 29). The optical power absorption coefficient α_{FC} due to free carrier absorption is obtained as (see Numai[103], App. F).

interband absorption
intraband absorption
plasma effect
free-carrier absorption

$$\alpha_{FC} = \frac{ne^2}{m^*\omega^2\epsilon_0 n_\tau c\tau}, \tag{3.60}$$

where n is the carrier density, m^* the effective mass of the carrier, ω the angular frequency of the light, ϵ_0 the permittivity in vacuum, n_τ the refractive index, c the speed of light in vacuum and τ the mean free time between carrier collisions.

3.8.2 Photodetectors

There are several semiconductor devices that can be used to detect the presence of photons. These devices are known as photodetectors; they convert optical signals into electrical signals. When excess electrons and holes are generated by the light in a semiconductor, there is an increase in the conductivity of the material. This change in conductivity is the basis of the photoconductor, perhaps the simplest type of photodetector. If electrons and holes are generated within the space charge region of a pn-junction, then they will be separated by the electric field and a current will be produced in the external circuit. The pn-junction is the basis of several photodetector devices including the photodiode and the phototransistor.

pn-Photodiode A pn-photodiode is a pn-junction diode operated with an applied reverse-biased voltage (Fig. 3.28(a)). We will consider a broad (i.e., in the x-direction) diode under reverse bias voltage, in which excess carriers are generated uniformly throughout the semiconductor device. Fig. 3.28(a) shows the reverse biased diode in

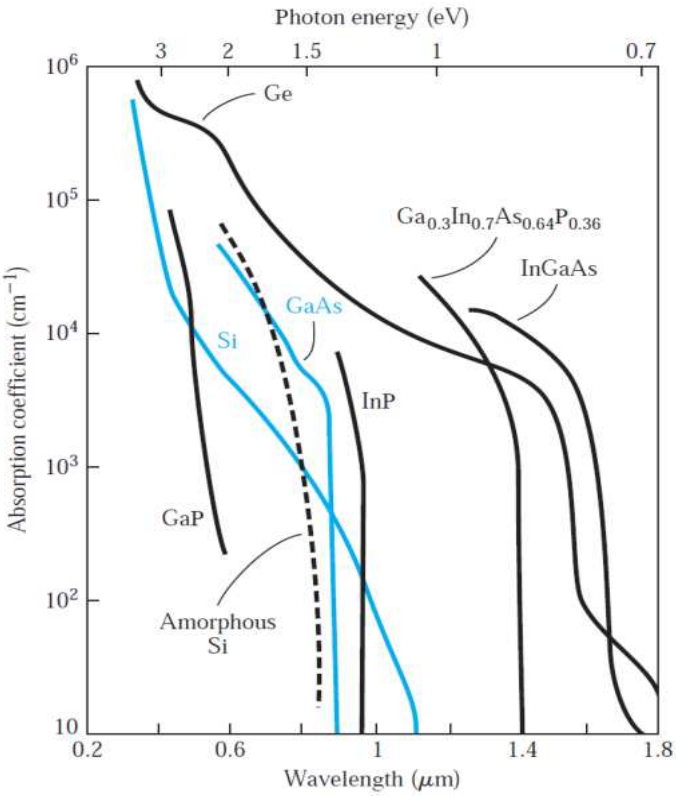

Figure 3.27: Absorption coefficient α versus wavelength for some semiconductors.

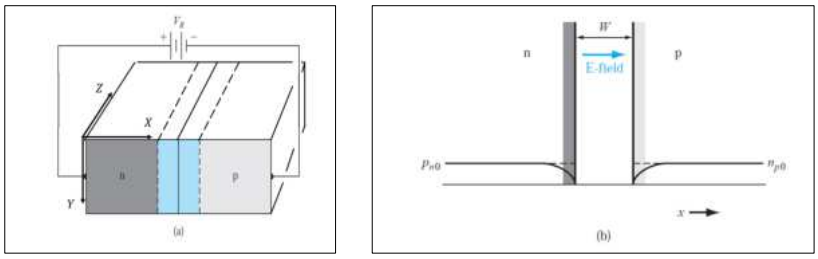

Figure 3.28: (a) Sketch of reverse-biased pn-junction. (b) Corresponding minority carrier concentrations

an optical wave-guide configuration. Fig. 3.28(b) shows the minority carrier distribution in the reverse biased junction prior to photon illumination.

Under the influence of a flux $\Phi(x, y, z)$ of photons propagating in the z-direction, excess carriers will be generated within the space charge region (blue area, width W_i, Fig. 3.28(a)). Then assuming each absorbed photon produces one electron and one hole, the generation rate of excess carriers is

$$G_(x, y, z) = 2\alpha\Phi(x, y, z), \tag{3.61}$$

with α the absorption coefficient (per unit of length). These excess carriers are swept out of the depletion region very quickly by the electric field, the electrons are swept

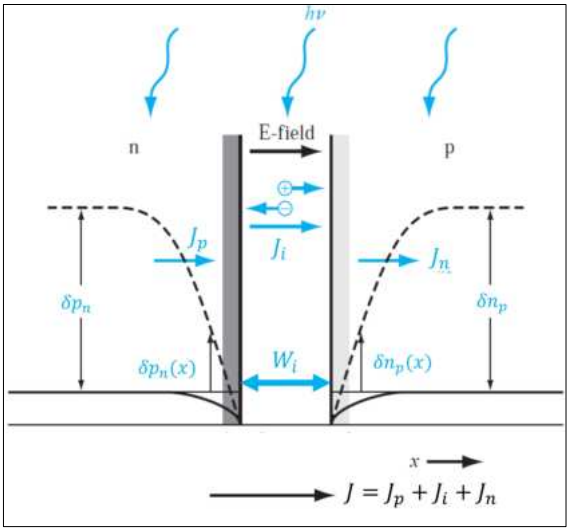

Figure 3.29: Steady-state, photo-induced minority carrier contributions $\delta p_n, \delta n_p$ and photocurrents J_p, J_i, J_n in a broad reverse-biased pn-junction.

into the n-region and the holes into the p-region.

The photon-generated current density from the space charge region is given by

$$J_i(y,z) = e\int_{W_i} dx G_(x,y,z) = 2e\alpha\int_{W_i} dx\Phi(x,y,z), \tag{3.62}$$

where the integral is over the space charge region width W_i. We may note that J_i is in the reverse-biased x-direction through the pn-junction. This photocurrent density responds very quickly to the photon illumination and is known as the prompt photocurrent. In fact, the speed of the photodiode is limited by the carrier transport through the space charge region, which can be estimated as follows. If we assume that the saturation drift velocity is 10^7 cm/s and the depletion width is 2 μm, the transit time is $\tau_t \sim 20$ ps. Hence, the cut-off frequency is $f_{max} \sim 50$ GHz. This frequency response is substantially higher than that of photoconductors (≤ 1 GHz). A photoconductor utilizes "across-the-gap" electron-hole pair creation and/or ionization of doped impurities in the material so as to exhibit light-induced conductivity.

There are two other, much slower, but larger, contributions to the photon-generated current density. These are due to excess carriers generated in the neutral n and p-regions of the diode. With these contributions the total steady-state diode photocurrent density for the broad diode is now

$$J(y,z) = 2e\alpha\sum_{j=i,n,p}\int_{W_j} dx\Phi(x,y,z), \tag{3.63}$$

where $W_{n,p}$ are the diffusion lengths of the photo-induced minority carriers in the n, p-regions. Representative values are $W_i \sim 1$ μm, $W_n \sim 30$ μm, $W_p \sim 10$ μm. In most pn-junction structures, the assumption of a broad diode will not be valid, so modifications of the photocurrent expression 3.63 are possible.

The time response of the diffusion components of the photocurrent is relatively slow since these currents are the results of the diffusion of minority carriers toward the de-

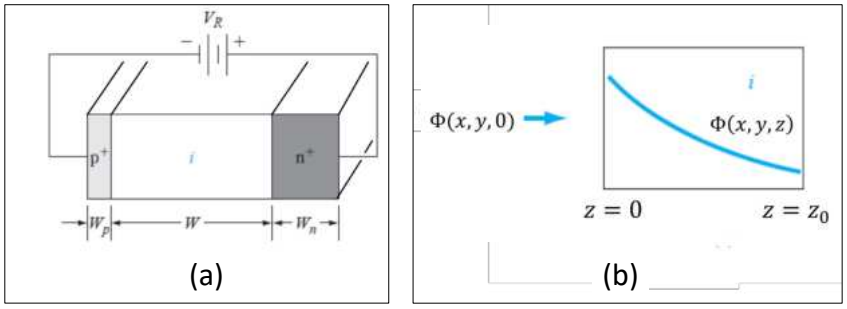

Figure 3.30: (a) A reverse-bias PIN photodiode. (b) Geometry showing nonuniform photon absorption.

Problem a): Estimate the voltage that is required to deplete the active layer of the reversely biased SOA-structure, for use as a detector. What is the effect if it is not completely depleted?

Solution b):

Problem 3.8:

pletion region. The diffusion components of photocurrent are referred to as the delayed photocurrent. The photo-induced currents situation is schematically summarized in Fig. 3.29.

PIN-photodiode In many photodetector applications, the speed of response is important; therefore, the prompt photocurrent, generated in the space charge region, is the only photocurrent of interest. To increase the photodetector sensitivity, the depletion region width should be made as large as possible. This can be achieved in a PIN photodiode, which consists of a p-region and an n-region separated by an intrinsic region i. A sketch of a PIN diode is shown in Fig. 3.30a. The intrinsic region width W is much larger than the space charge width of a normal pn-junction. If a reverse bias is applied to the PIN diode, the space charge region extends completely through the intrinsic region.

Assume that a photon flux density $\Phi_(x,y,0)$ in the z-direction is incident on the i-region. If we disregard the carrier generation in the p and n-regions, then the photon flux, as a function of distance, in the intrinsic region is $\Phi(x,y,z) = \Phi(x,y,0)e^{-\alpha z}$, where α is the photon absorption coefficient. This nonlinear photon absorption is shown in Fig. 3.30(b). The reverse biased photocurrent density generated in the intrinsic region is found as

$$J(y,z_0) = 2e\alpha \int_W dx \int_0^{z_0} dz\Phi(x,y,z) = 2e \int_W dx\Phi(x,y,0)(1-e^{-\alpha z_0}). \tag{3.64}$$

This equation assumes that there is no electron–hole recombination within the p and n regions and that each photon absorbed creates one electron–hole pair.

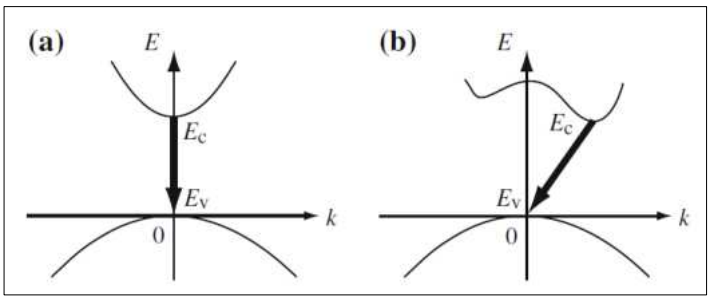

Figure 3.31: Direct (a) and indirect (b) transitions.

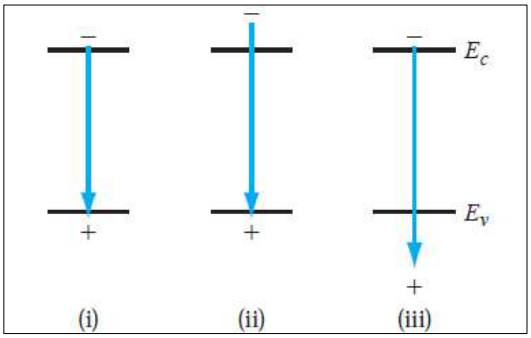

Figure 3.32: (a) The basis interband transitions: (i) intrinsic emission with energy close to the bandgap energy, (ii) emission from energetic electron and (ii) emission to energetic hole.

3.8.3 Light emission

In semiconductors, the transitions of electrons from high energy states to low energy states are designated recombinations of the electrons and the holes. We distinguish between radiative and nonradiative recombinations. A radiative recombination is accompanied by the emission of a photon whose energy corresponds to the energy difference between the initial and final states. In contrast, in a nonradiative recombination no photon is emitted, but the energy difference in the transition is transferred in a different way, e.g., by acoustic vibrations of the crystal lattice, called phonons or, more general, transformed into forms other than photons or phonons.

radiative recombination
nonradiative recombination
phonon

Let us now consider the transitions of electrons from the bottom of the conduction band down to the top of the valence band. If the bottom of the conduction band and the top of the valence band occur at the same wave vector $\boldsymbol{k}_0$ (often $\boldsymbol{k}_0 = 0$), this situation is referred to as a direct transition. A transition, in which the bottom of the conduction band and the top of the valence band have different $\boldsymbol{k}$-values, is referred to as indirect transition. These direct and indirect transitions are schematically shown in Fig. 3.31. Light emission due to radiative recombinations, also called luminescence, is only possible for direct transitions. The reason is conservation of momentum and since a photon carries a negligeable amount of momentum, this implies equal momentum for the initial and final states, or $\Delta\boldsymbol{k} \approx 0$. Indirect transitions, on the other hand, need emission of a momentum-carrying phonon.

direct transition
indirect transition
luminescence

Some recombination processes may result in photon emission from direct bandgap materials. If a recombination occurs with the emission of a photon, one refers to

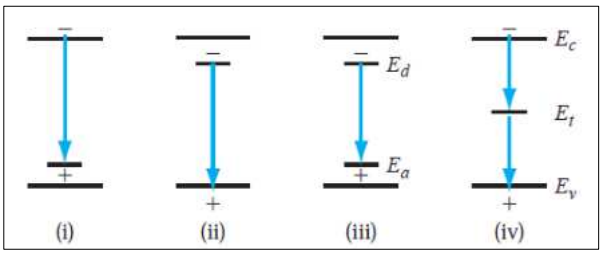

Figure 3.33: Recombination processes involving impurity or defect states: (i) conduction band to acceptor, (ii) the donor to valence-band transition, case (iii) the donor to acceptor transition, and case (iv) is the recombination due to a deep trap. Case (iv) is a non-radiative process.

spontaneous emission

interband transitions

emission spectrum

spontaneous emission in case the electron hole pair was formed by thermal excitation or luminescence when the pair is formed by other means, such as electrical (electroluminescence) or optical (photoluminescence). Figure 3.32 shows the basic interband transitions, where case (i) corresponds to an intrinsic emission very close to the bandgap energy of the material. Cases (ii) and (iii) correspond to energetic electrons or holes. If either of these recombinations result in the emission of a photon, the energy of the emitted photon will be slightly larger than the bandgap energy. There will then be an emission spectrum and a bandwidth associated with the emission (See Fig. 3.35).

The possible recombination processes involving impurity or defect states are shown in Figure 3.33. Case (i) is the conduction band to acceptor transition, case (ii) the donor to valence-band transition, case (iii) the donor to acceptor transition, and case (iv) is the recombination due to a deep trap. Case (iv) is a non-radiative process; the other recombination processes may or may not result in the emission of a photon. Figure 3.34 shows the Auger recombination process, which can become important in direct bandgap materials with high doping concentrations. The Auger recombination process is a nonradiative process. In case (i), it is a recombination between an electron and a hole, accompanied by the transfer of energy to another free hole. Similarly, in case (ii), the recombination between an electron and hole results in the transfer of energy to a free electron. The third particle involved in this process will eventually lose its energy to the lattice in the form of heat. The process involving two holes and an electron would occur predominantly in heavily doped p-type materials, and the process involving two electrons and a hole would occur primarily in a heavily doped n-type material.

Auger recombination

As argued around Fig. 3.32, the emission of a photon is not necessarily at a single, discrete energy, but can occur over a range of energies. The spontaneous emission rate generally has the form (71)

$$I(\nu) \propto \nu^2 (h\nu - E_g)^{\frac{1}{2}} e^{-\frac{h\nu - E_g}{k_B T}}, \tag{3.65}$$

where E_g is the bandgap energy. Figure 3.35 shows the emission spectra from gallium arsenide.

Luminescent efficiency We have shown that not all recombination processes are radiative. An efficient luminescent material is one in which radiative transitions predominate. The quantum efficiency η_q is defined as the ratio of the radiative recombination

quantum efficiency

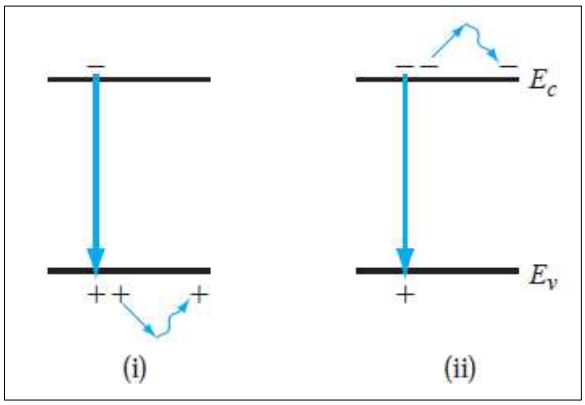

Figure 3.34: The Auger recombination process (nonradiative): (i), recombination between an electron and hole, accompanied by the transfer of energy to another free hole, (ii) recombination between an electron and hole with transfer of energy to a free electron.

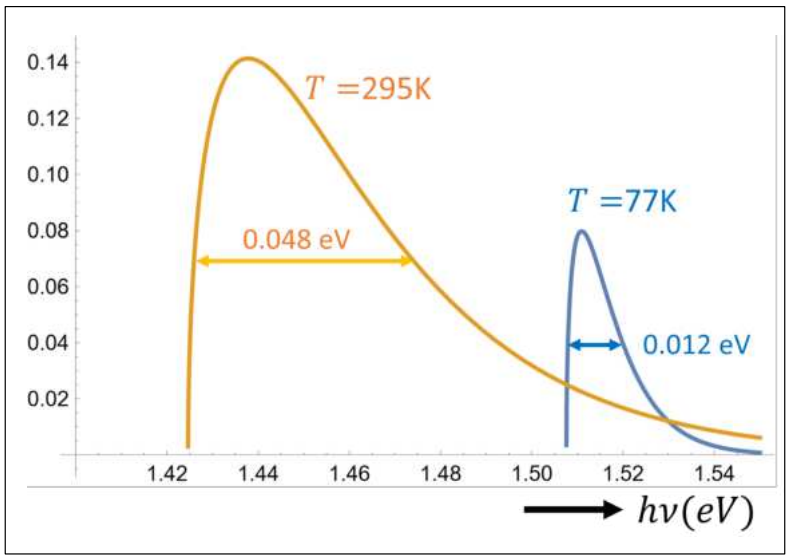

Figure 3.35: GaAs diode emission spectra at T = 300 K and T = 77 K.

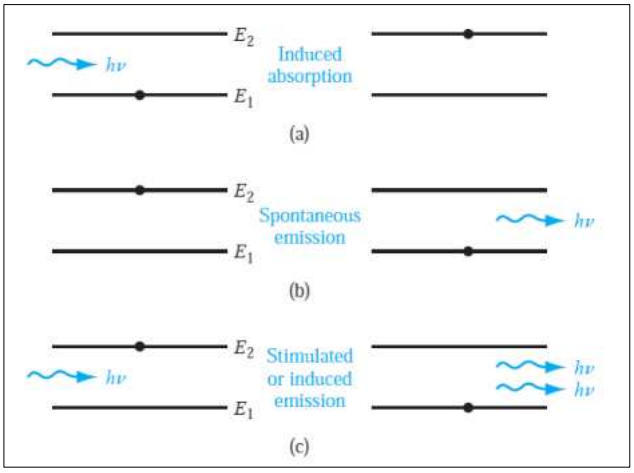

Figure 3.36: Schematic diagram showing (a) induced absorption, (b) spontaneous emission, and (c) stimulated emission processes.

rate to the total recombination rate for all processes, or

$$\eta_q = R_r / R, \tag{3.66}$$

radiative recombination rate lifetime

with R_r the radiative recombination rate and R the total recombination rate of the excess carriers. Since the recombination rate is inversely proportional to lifetime, we can write the quantum efficiency in terms of lifetimes as

$$\eta_q = \frac{\tau_n r}{\tau_r + \tau_n r}, \tag{3.67}$$

where $\tau_n r$ is the nonradiative lifetime and τ_r is the radiative lifetime. For a high luminescent efficiency, the nonradiative lifetime must be large; i.e., the probability of a nonradiative recombination is small compared to the radiative recombination. The interband recombination rate of electrons and holes is directly proportional to the number of electrons available and directly proportional to the number of available empty states (holes). We can write

$$R_r = Bnp, \tag{3.68}$$

where R_r is the band-to-band radiative recombination rate and B is the constant of proportionality. The values of B for direct-bandgap materials are on the order of 100 times larger than for indirect bandgap materials. The probability of a direct band-to-band radiative recombination transition in an indirect bandgap material is very unlikely.

3.9 Stimulated emission and amplification

3.9.1 Stimulated emission

Suppose an electron occupies a state in the conduction band, while a hole in the valence band is present with the same momentum $\boldsymbol{k}$. If the electron recombines with the hole without any external trigger, a photon will be emitted. This is *spontaneous emission*, discussed in the previous section. However, if an incident photon with nearly the

spontaneous emission

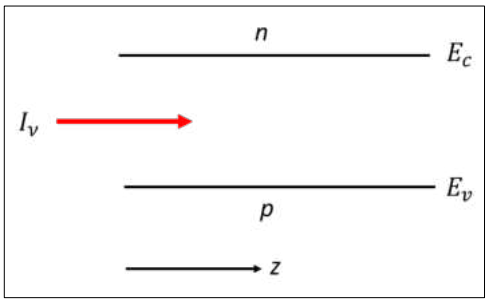

Figure 3.37: Light propagating in the z-direction through a material with conduction and valence bands.

same energy as the transition triggers the transition, we end up with two identical photons. This is stimulated emission, a process that can be fully understood by quantum theory. Stimulated emission leads to amplification of the light, preserving the optical phase. Fig. 3.36 illustrates the basic processes of induced absorption, spontaneous emission, and stimulated emission.

stimulated emission amplification

Now, consider propagation of light in a material with density n_c of electrons in the conduction band and density n_v of electrons in the valence band, as illustrated in Fig. 3.37.

In case that $n_c > n_v$, there will be more stimulated emissions than absorptions, that is, the intensity will increase on propagation, or, equivalently, the beam is amplified. The change in intensity as a function of z can be written as

$$\frac{dI}{dz} = \sigma_i (n_c - n_v) I, \tag{3.69}$$

where σ_i is the cross section for an incident photon to induce a transition. It is noted that σ_i depends on the frequency of the light and has the dimension of an area (length × length). In this equation we assume no loss mechanisms and disregard the spontaneous emission. The solution of Eq. 3.69 can be written as

$$I_(z) = I(0) e^{gz}, \tag{3.70}$$

where $g \equiv \sigma_i (n_c - n_v)$ is the net amplification or gain factor. Amplification occurs when $g > 0$ and absorption when $g < 0$. Hence, amplification is obtained when $n_c > n_v$, a situation called population inversion, which can be realized in a forward-biased pn-junction diode, if both sides of the junction are degenerately doped (see Sec. 3.3). Figure 3.38(a) shows the energy-band diagram of such a junction in thermal equilibrium. The Fermi level is in the conduction band in the n-region and in the valence band in the p-region.

Figure 3.38(b) shows the energy bands of the pn-junction when the forward bias is applied. We see that in the depletion region a situation is realized where electrons and holes are both present. There are large numbers of electrons in the conduction band directly above many holes. These electrons and holes can recombine under emission of a photon with energy $E_g < h\nu \leq E_{Fn} - E_{Fp} = eV_a$, where V_a is the applied (forward) voltage. A more elaborate analysis taking account of the density of states in the conduction and valence bands yields that the gain factor is given by

$$g(\nu) = \sigma_i (n_c - n_v) \propto \sqrt{h\nu - E_g} \left\{ 1 - e^{\frac{h\nu + E_{Fp} - E_{Fn}}{k_B T}} \right\}. \tag{3.71}$$

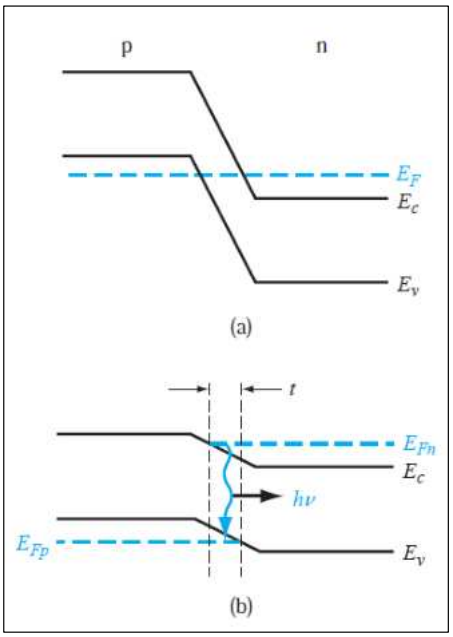

Figure 3.38: (a) Degenerately doped pn-junction at zero bias. (b) Degenerately doped pn-junction under forward bias with photon emission.

Problem a):

1. Estimate the carrier density in the basic cell under an injection current density of: a. 0.1 kA/cm2 b. 1 kA/cm2
2. Estimate the transparency current density for the standard SOA structure.
3. Estimate the gain of the standard SOA structure for an injection current density of 2 kA/cm2

Solution a):

Problem 3.9:

For $g(\nu) > 0$, we must have $h\nu > E_{Fn} - E_{Fp}$, which implies that the junction must be degenerately doped since we also have the requirement that $h\nu \geq E_g$. Near the junction, there is a region in which the population inversion occurs. Under these above-mentioned conditions, the pn-junction serves as an amplifier for light, a Semiconductor Optical Amplifier (SOA).

3.10 Quantum Dots

In section 3.6 the quantum well was introduced: 2-dimensional confinement of electrons and holes between two barriers. Similarly, confinement of carriers can be realized by encompassing the well material W in all dimensions by the barrier material B, like the pit in a cherry, as sketched in Fig. 3.39. Typically, self-assembled quantum dots

Figure 3.39: Illustrating a Quantum Dot (QD) by the pit in a cherry, where the pit resembles the well and the flesh the barrier material.

are between 5 and 50 nm in size. Quantum dots defined by lithographical methods, for instance by etching on two-dimensional electron gases (such as obtained in the quantum-well double heterostructures) can have lateral dimensions between 20 and 100 nm.

From Wikipedia: wikipedia.org/wiki/Quantum_dot (16-09-2024):

Quantum dots

Quantum dots (QDs) or semiconductor nanocrystals are semiconductor particles, a few nanometres in size with optical and electronic properties that differ from those of larger particles via quantum mechanical effects. They are a central topic in nanotechnology and materials science. When a quantum dot is illuminated by UV light, an electron in the quantum dot can be excited to a state of higher energy. In the case of a semiconducting quantum dot, this process corresponds to the transition of an electron from the valence band to the conductance band. The excited electron can drop back into the valence band releasing its energy as light. This light emission (photoluminescence) is illustrated in the figure on the right. The color of that light depends on the energy difference between the conductance band and the valence band, or the transition between discrete energy states when the band structure is no longer well-defined in QDs.

Nanoscale semiconductor materials tightly confine either electrons or electron holes. The confinement is similar to a three-dimensional particle in a box model. The quantum dot absorption and emission features correspond to transitions between discrete quantum mechanically allowed energy levels in the box that are reminiscent of atomic spectra. For these reasons, quantum dots are sometimes referred to as artificial atoms, emphasizing their bound and discrete electronic states, like naturally occurring atoms or molecules. It was shown that the electronic wave functions in quantum dots resemble the ones in real atoms. By coupling two or more such quantum dots, an artificial molecule can be made, exhibiting hybridization even at room temperature. Precise assembly of quantum dots can form superlattices that act as artificial solid-state materials that exhibit unique optical and electronic properties.

Quantum dots have properties intermediate between bulk semiconductors and discrete atoms or molecules. Their optoelectronic properties change as a function of both

size and shape. Larger QDs of 5–6 nm diameter emit longer wavelengths, with colors such as orange, or red. Smaller QDs (2–3 nm) emit shorter wavelengths, yielding colors like blue and green. However, the specific colors vary depending on the exact composition of the QD.

Potential applications of quantum dots include single-electron transistors, solar cells, LEDs, lasers, single-photon sources, second-harmonic generation, quantum computing, cell biology research, microscopy, and medical imaging. Their small size allows for some QDs to be suspended in solution, which may lead to their use in inkjet printing, and spin coating. They have been used in Langmuir–Blodgett thin films. These processing techniques result in less expensive and less time-consuming methods of semiconductor fabrication.

3.11 Excitons and Quantum confined Stark effect

exciton

The absorption of a photon can result in the formation of an electron and a hole at some distance from each other but which are nevertheless bound together by their mutual Coulomb interaction. This entity, which is much like a hydrogen atom but with a hole rather than a proton, is called an exciton. A photon may be emitted as a result of the electron and hole recombining, thereby annihilating the exciton.

In Sec. 3.3 the quantized wave vector $\boldsymbol{k}$ was introduced such that $\hbar\boldsymbol{k}$ is the quantized crystal momentum. This was a consequence of the periodic crystal potential experienced by each individual independent electron. However, the electron and hole pair, including their Coulomb interaction, experience the crystal periodicity. Therefore, the exciton does, and we must replace $\hbar\boldsymbol{k}$ by $\hbar\boldsymbol{K}_{ex}$, where $K_{ex} \equiv \boldsymbol{k}_c - \boldsymbol{k}_v$ is the quantized wave vector for the exciton (see Singh [104] and Dresselhaus [105]).

Then, in the approximation where the exciton is described as an electron in the central Coulomb field of a hole, the exciton energy levels are given by

$$E_n^{ex}(K_{ex}) = E_g + E_n + \frac{\hbar^2}{2(m_c + m_v)} K_{ex}^2, \tag{3.72}$$

where E_g is the gap energy of the semiconductor and the exciton levels are given by

$$E_n = -R_{ex}/n^2 \equiv -\frac{m_r e^4}{2(4\pi\epsilon)^2\hbar^2}\frac{1}{n^2}, \tag{3.73}$$

with $m_r = \frac{m_c m_v}{m_c+m_v}$ the reduced effective mass of the electron-hole system and $R_{ex} = \frac{m_r e^4}{2(4\pi\epsilon)^2\hbar^2}$ the binding energy of the exciton. The last term in the right-hand side of (87???) represents the kinetic energy of the exciton. The corresponding "Bohr radius", i.e., the effective extent of the exciton can be expressed as

$$a_{ex} = \frac{\epsilon m_0}{\epsilon_0 m_r} a_B \quad (a_B = Bohr radius = 0.529 \times 10^{-6} m), \tag{3.74}$$

where m_0 is the free-electron mass. The exciton radius is $\sim 1 \times 10^{-8}$ m (see Table 1). Thus the exciton is spread over a large number of unit cells. A sketch of the exciton band structure situation is presented in Fig. 3.40.

Since a photon carries only a negligeable amount of momentum, optical transitions from the valence band to an excitonic state can occur only at $K_{ex} \approx 0$. Excitonic effects have very dramatic consequences for the optical properties of semiconductors,

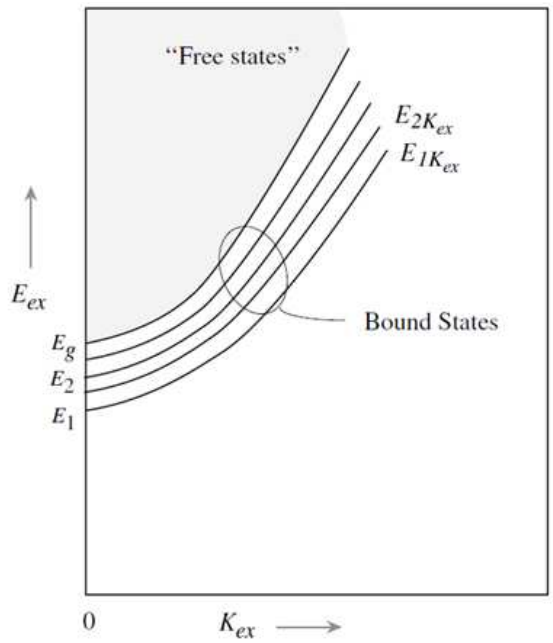

Figure 3.40: Sketch of the exciton band structure (87)

Table 3.1: Exciton binding energy (E_1) and exciton radius (a_{ex}) in direct bandgap bulk semiconductors with the zinc-blende structure (from Yu and Cardona [106]).

Semiconductor	$E_1(meV)$	$a_{ex}(\mathring{A})$
InP	4.9	112
GaAs	5.1	113
CdTe	11	12.2
ZnTe	13	11.5
ZnSe	19.9	10.7
ZnS	29	10.2

especially near the band edges. Below the band edge, there is a strong and sharp excitonic absorption/emission transition. Also just above the bandgap, there is a strong enhancement of the absorption process especially in 3D systems.

It can be shown (see Singh [104]) that near the band edge the absorption coefficient for optical radiation of frequency ω can be expressed as

$$\alpha_{ex}(\omega) = \alpha_F(\omega)\frac{2\pi R_{ex}}{(\hbar\omega - E_g)^{1/2})}, \tag{3.75}$$

where $\alpha_F(\omega)$ is the absorption coefficient without excitonic effects, i.e.,

$$\alpha_F(\omega) = C\frac{(\hbar\omega - E_g)^{1/2}}{\hbar\omega}. \tag{3.76}$$

In (91??) the term $(\hbar\omega - E_g)^{1/2}$ stems from the density of states corresponding to the transition from $E_v(k)$ to $E_c(k) = E_v(k) + \hbar\omega$ (compare with (13??)) and C is a constant.

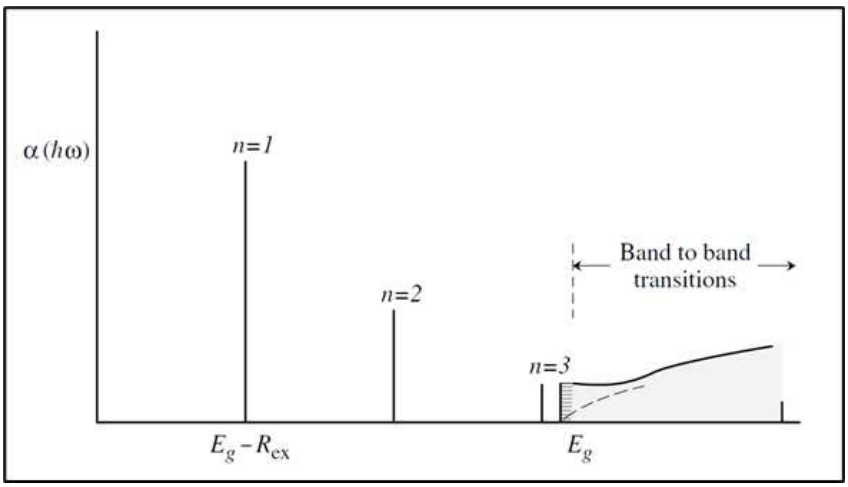

Figure 3.41: Schematic representation of the absorption spectrum with a few delta peaks (solid line, not on scale) in a semiconductor with excitonic effects versus photon energy. The dashed curve corresponds to the independent-electron picture (see (76)

By combining (91??) and (91??), we see that due to the excitonic effect the absorption coefficient assumes a non-zero constant value above and down to the band edge. Below the band edge, the density of discrete excitonic bound states increases proportional to $1/n^3$ so that the absorption coefficient may show a narrow peak with width $\sim 1meV$ just below the band edge. Therefore, apart from a few delta peaks, corresponding to the first lowest discrete exciton bound-state levels, the absorption coefficient versus $\hbar\omega$ assumes a non-zero value around the bandgap energy, i.e., dramatically larger than in the independent-electron picture (91??). This is schematically indicated in Fig. 3.41.

It should be noted that in InP and GaAs (see Table 3.1), the exciton binding energy R_{ex} is on the order of $5\,meV$, which corresponds to a thermal energy at $\sim 57K$. Therefore, the effects of discrete excitonic levels are visible only at temperatures far below room temperature. This is demonstrated in Fig. 3.42. Usually, at room temperature excitons don't play significant roles in bulk semiconductors. This is very different in quantum wells, where exciton effects are significantly more pronounced, as can be understood from the following considerations: When the width of the quantum well is less than the diameter of the excitonic Bohr orbit, the electron–hole separation will be limited by the quantum well width, rather than by the larger "Bohr" radius, thereby significantly increasing the Coulomb binding energy of the excitons. Thus small quantum well widths enhance exciton effects. Normally sharp exciton peaks in bulk GaAs are observed only at low temperature, but in quantum well structures excitonic effects can be observed at room temperature.

The quantum-confined Stark effect (QCSE) describes the effect of an external electric field upon the light absorption spectrum or emission spectrum of a quantum well (QW). In the absence of an external electric field, electrons and holes within the quantum well may only occupy states within a discrete set of energy sub bands. Only a discrete set of frequencies of light may be absorbed or emitted by the system. When an external electric field is applied, the electron states shift to lower energies, while the hole states shift to higher energies. This reduces the permitted light absorption or

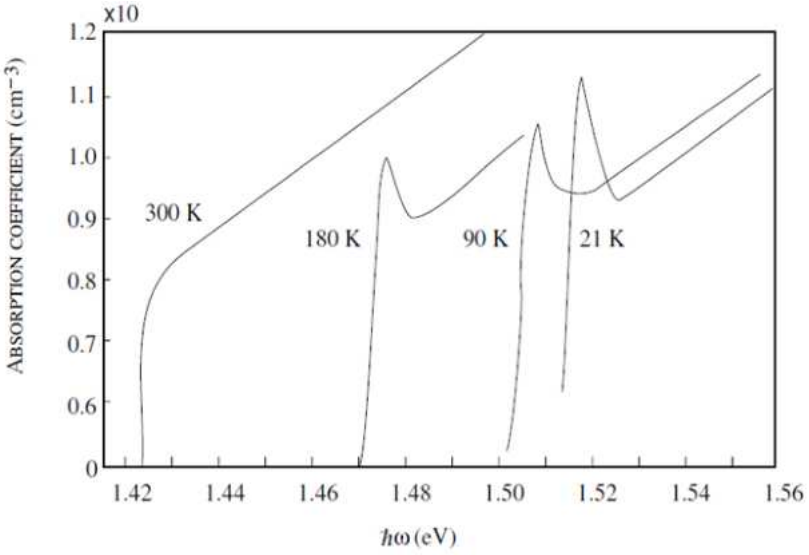

Figure 3.42: Typical measured absorption spectra for GaAs at various temperatures, as indicated. For each temperature, the band-to-band absorption edge corresponds to the band-gap energy at that particular temperature. Excitonic effects are no longer observable at room temperature. (From Dresselhaus???).

emission frequencies. Additionally, the external electric field shifts electrons and holes to opposite sides of the well, decreasing the overlap integral, which in turn reduces the recombination efficiency (i.e. fluorescence quantum yield) of the system. The spatial separation between the electrons and holes is limited by the presence of the potential barriers around the quantum well, meaning that excitons are able to exist in the system even under the influence of an electric field. The quantum-confined Stark effect is used in QCSE optical modulators, which allow optical communications signals to be switched on and off rapidly.

3.12 Appendices

Appendix 3A Derivation of electron and hole density

We start with the last expression in Eq. 3.13, which can be expressed in the Boltzmann approximation as

$$n_e \cong \frac{1}{2\pi^2}(\frac{2m_c}{\hbar^2})^{3/2}\int_{E_c}^{\infty} dE\,(E-E_c)^{1/2} exp[-\frac{E-E_F}{k_B T}] \tag{3.77}$$

Note that at room temperature, $T = 300$ K, we have $k_B T = 0.026$ eV, whereas the energy gap in semiconductors is $\lessapprox 1$ eV. Hence, if E_F is in the energy gap and not too close to E_c, Eq. 3.77 is an excellent approximation. The integral in (. 3.77) can be solved by introducing

$$\eta = \frac{E-E_c}{k_B T}, \tag{3.78}$$

leading to

$$n_e \cong \frac{1}{2\pi^2}(\frac{2m_c k_B T}{\hbar^2})^{3/2} exp[-\frac{E_c-E_F}{k_B T}]\int_0^{\infty} d\eta \eta^{1/2} e^{-\eta}, \tag{3.79}$$

where the integral is tabulated and has the value $\frac{1}{2}\sqrt{\pi}$. Therefore, Eq. 3.79 becomes

$$n_e \cong \frac{1}{4}(\frac{2m_c k_B T}{\pi\hbar^2})^{3/2} exp[-\frac{E_c - E_F}{k_B T}] \tag{3.80}$$

The thermal electron concentration can thus be written as

$$n_e = N_c exp[-\frac{E_c - E_F}{k_B T}] \tag{3.81}$$

where N_c is called the effective density of states in the conduction band,

$$N_c \cong \frac{1}{4}(\frac{2m_c k_B T}{\pi\hbar^2})^{3/2} \tag{3.82}$$

Now we can follow a similar procedure for the holes in the valence band. Since a hole is a non-occupied electron state, the hole equilibrium distribution is not $f_F(E)$ but rather

$$1 - f_F(E) = \frac{1}{1 + exp(\frac{E_F - E}{k_B T})} \approx e^{-\frac{E_F - E}{k_B T}} \tag{3.83}$$

where the last step is again the above-introduced Boltzmann approximation, valid when $E_F - E_v \gg k_B T$ and $E \leq E_v$). Like Eq. 3.77 we can express the thermal-equilibrium concentration of holes in the valence band as

$$n_h \cong \frac{1}{2\pi^2}(\frac{2m_v}{\hbar^2})^{3/2} \int_{-\infty}^{E_v} dE\,(E_v - E)^{1/2} exp[-\frac{E_F - E}{k_B T}] \tag{3.84}$$

where the lower integration limit is taken as minus infinity since the exponential decays fast enough to validate this approximation. Following the same procedure as for n_e above, we now find for the thermal hole concentration

$$n_h = N_v exp[-\frac{E_F - E_v}{k_B T}] \tag{3.85}$$

where N_v is called the effective density of states in the valence band,

$$N_v \cong \frac{1}{4}(\frac{2m_v k_B T}{\pi\hbar^2})^{3/2} \tag{3.86}$$