

Part II

PV Fundamentals

4

Electrodynamic basics

In this chapter we introduce the basics of electrodynamics that are required for solar cell physics. First, we introduce the electromagnetic wave equations. The existence of these equations explains the existence of electromagnetic waves, such as light. From there we develop the equations describing the interaction of an electromagnetic wave with interfaces between two materials; in this way we naturally derive the basics of optics.

Later in the chapter we introduce the Poisson equation and the continuity equations that are very important for semiconductor physics, which we discuss in chapter 6.

4.1 The electromagnetic theory

While electricity and magnetism have been known since ancient times, it took until the nineteenth century to realise that these two phenomena are two sides on one medal, namely *electromagnetism*. We can easily see this by recalling that electric fields are generated by charges while magnetic fields are generated by currents, *i.e.* moving charges. Let us now assume that we are within an array of charges. Since charges create an electric field, we will experience such a field. Now we start moving with a constant velocity. This is equivalent to saying that the array of charges moves with respect to us. Since moving charges are a current, we now experience an magnetic field. Thus, when changing from one

frame of reference into another one that moves with respect to the first one with a constant velocity, electric fields are transformed into magnetic fields and *vice versa*.

Between 1861 and 1862, the Scottish physicist James Clerk Maxwell published works in that he managed to formulate the complete electromagnetic theory by a set of equations, the *Maxwell equations*. A modern formulation of these equations is given in Appendix A.1. The transformation of the electric and magnetic fields between different frames of reference is correctly described by Albert Einstein's theory of special relativity, published in 1905.

One of the most important prediction of the Maxwell equations is the presence of electromagnetic waves. A derivation is given in Appendix A.2. Maxwell soon realised that the speed of these waves is (within experimental accuracy) the same as the speed of light that then already was known. He brilliantly concluded that light is an electromagnetic wave.

In the 1880s the German physicist Heinrich Hertz could experimentally confirm that electromagnetic waves can be generated and have the same speed as light. His work laid the foundation for modern radiocommunication that has shaped the modern world.

The electromagnetic theory can perfectly describe how light propagates. However, it fails in explaining how

light is emitted and absorbed by matter. For this purpose, quantum mechanics is required.

4.2 Electromagnetic waves

As shown in Appendix A.2, electromagnetic waves are described by

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \mathbf{E} - \frac{n^2}{c_0^2} \left(\frac{\partial^2 \mathbf{E}}{\partial t^2} \right) = 0 \quad (4.1a)$$

for the electric field $\mathbf{E}(\mathbf{r}, t)$, where c_0 denotes the speed of light *in vacuo* and n is the refractive index of the material. In a similar manner we can derive the wave equation for the *magnetic field*,

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \mathbf{H} - \frac{n^2}{c_0^2} \left(\frac{\partial^2 \mathbf{H}}{\partial t^2} \right) = 0. \quad (4.1b)$$

The simplest solution to the wave equations (4.1) is the plane harmonic wave, where light of constant wavelength λ propagates in one direction. Without loss of generality, we assume that the wave travels along the z direction. The electric and magnetic fields in this case are

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{ik_z z - i\omega t}, \quad (4.2a)$$

$$\mathbf{H}(\mathbf{r}, t) = \mathbf{H}_0 e^{ik_z z - i\omega t}, \quad (4.2b)$$

where \mathbf{E}_0 and \mathbf{H}_0 are constant vectors (the amplitudes), k_z is the *wave number* and ω is the *angular frequency*. By substituting Eq. (4.2a) into Eq. (4.1a) we find that k_z and ω are connected to each other via

$$k_z^2 - \frac{n^2}{c_0^2} \omega^2 = 0. \quad (4.3)$$

Thus,

$$k_z = \frac{n\omega}{c}. \quad (4.4)$$

The angular frequency, measured in *radians per second* is related to the *frequency* of the wave v , measured in *Hertz*, via

$$\omega = 2\pi v = \frac{2\pi}{T}, \quad (4.5)$$

where $T = 1/v$ is the period, measured in *seconds*. The wave number k_z has the unit of an inverse length. It is related to the *wavelength* λ via

$$\lambda = \frac{2\pi}{k_z} = \frac{2\pi c}{n\omega} = \frac{c}{nv}. \quad (4.6)$$

We note that v and ω are independent of n , while k and λ change when the wave travels to media with different n .

Electromagnetic waves have some extraordinary properties that are derived in detail in Appendix A.3:

- The electric and magnetic field vectors are perpendicular to each other and also perpendicular to the

propagation vector,

$$\mathbf{k} \cdot \mathbf{E}_0 = \mathbf{k} \cdot \mathbf{H}_0 = \mathbf{H}_0 \cdot \mathbf{E}_0 = 0. \quad (4.7)$$

- The electric and magnetic fields are proportional to the propagation direction, hence electromagnetic waves are *transverse waves*.
- The electric and magnetic vectors have a constant, material dependent ratio. If the electric field is along the x -direction and the magnetic field is along the y -direction, this ratio is given by

$$H_{y,0} = \frac{n}{c\mu_0} E_{x,0} = \frac{n}{Z_0} E_{x,0}, \quad (4.8)$$

where

$$Z_0 = c\mu_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 376.7 \Omega \quad (4.9)$$

is the *impedance of free space*.

4.3 Optics of flat interfaces

In this section we repeat the major relations that describe an electromagnetic wave traversing an interface between a medium 1 and a medium 2, as illustrated in Fig. 4.1. We assume the two media to be non-absorptive. Therefore, only the real parts of the refractive indices are present. We denote them by n_1 and n_2 .

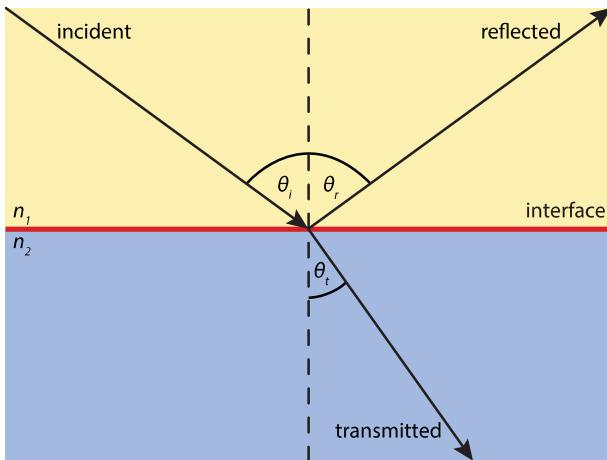


Figure 4.1: Scheme of light, reflected and refracted by a boundary.

A part of the incident light is *reflected*, where the angle of the scattered light θ_r is equal to the incident angle θ_i ,

$$\theta_r = \theta_i. \quad (4.10)$$

The other part enters medium 2, where the angle of the *refracted* light θ_t is related to θ_i via *Snell's law*,

$$n_1 \sin \theta_i = n_2 \sin \theta_t, \quad (4.11)$$

Relations between the magnitudes of the incident, reflected and refracted fields are given by *Fresnel equations*. We have to distinguish between *parallel* and *perpendicular polarised light*. Parallel or perpendicular polarised means that the *electric field* is parallel or perpendicular to the plane, respectively. For *perpendicular* polarised light the Fresnel equations are given by

$$t_s = \left(\frac{E_{0t}}{E_{0i}} \right)_s = \frac{2 n_1 \cos \theta_i}{n_1 \cos \theta_i + n_2 \cos \theta_t}, \quad (4.12a)$$

$$r_s = \left(\frac{E_{0r}}{E_{0i}} \right)_s = \frac{n_1 \cos \theta_i - n_2 \cos \theta_t}{n_1 \cos \theta_i + n_2 \cos \theta_t}, \quad (4.12b)$$

where the *s* stands for *senkrecht*, which is German for perpendicular. The relations for *parallel* polarised light are

$$t_p = \left(\frac{E_{0t}}{E_{0i}} \right)_p = \frac{2 n_1 \cos \theta_i}{n_1 \cos \theta_t + n_2 \cos \theta_i}, \quad (4.13a)$$

$$r_p = \left(\frac{E_{0r}}{E_{0i}} \right)_p = \frac{n_1 \cos \theta_i - n_2 \cos \theta_t}{n_1 \cos \theta_t + n_2 \cos \theta_i}. \quad (4.13b)$$

The intensities are proportional to the square of the electric field, $I \propto E^2$. For unpolarised light, we have to take the mean values of the two polarisations. For the reflectivity R we thus obtain

$$R = \frac{1}{2} (r_s^2 + r_p^2). \quad (4.14)$$

For normal incidence this leads to

$$R(\theta_i = 0) = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2. \quad (4.15)$$

Because of *conservation of energy* the sum of R and the transmittance T must be 1,

$$R + T = 1. \quad (4.16)$$

By combining Eqs. (4.14) with (4.16) and doing some calculations we find

$$T = 1 - R = \frac{n_2 \cos \theta_t}{n_1 \cos \theta_t} \frac{1}{2} (t_s^2 + t_p^2), \quad (4.17)$$

which leads to

$$T(\theta_i = 0) = \frac{4n_1 n_2}{(n_1 + n_2)^2} \quad (4.18)$$

for normal incidence.

A very important consequence of the Snell's law is *total reflection*. If $n_2 > n_1$, there is a *critical angle* at which no

light can leave the layer with n_2 any more,

$$\sin \theta_{\text{crit}} = \frac{n_1}{n_2}. \quad (4.19)$$

Hence if $\theta_2 \geq \theta_{\text{crit}}$, no light will be transmitted, but everything will be reflected back into the layer. For a silicon-air interface ($n_{\text{Si}} \approx 4.3$), we find $\theta_{\text{crit}} = 13.4^\circ$. For the supporting layers used in solar cells, the critical angle is much larger. For a silicon-glass interface ($n_{\text{glass}} \approx 1.5$), we $\theta_{\text{crit}} = 20.4^\circ$. And for an interface between silicon and zinc oxide, which is a transparent conducting oxide often used in solar cell technology, ($n_{\text{ZnO}} \approx 2$) the critical angle would be $\theta_{\text{crit}} = 30.3^\circ$

4.4 Optics in absorptive media

Let us recap what we have seen that far in this chapter: Starting from the Maxwell equations we derived the wave equations and looked at their properties for the special case of plane waves. After that we looked at the behaviour of electromagnetic waves at the interfaces between new media. For the whole discussion so far we implicitly assumed that the media is non-absorbing.

The working principle of solar cells is based on the fact that light is *absorbed* in an absorber material and that the absorbed light is used for exciting charge carriers that can be used to drive an electric circuit. Therefore

we will use this section to discuss how absorption of light in a medium can be described mathematically.

In general, the optical properties of an absorbing medium are described by an *complex electric permittivity* $\tilde{\epsilon}$,

$$\tilde{\epsilon} = \epsilon' + i\epsilon''. \quad (4.20)$$

Since the refractive index is given as the square root of $\tilde{\epsilon}$, it also is complex,

$$\tilde{n} = \sqrt{\tilde{\epsilon}} = n + i\kappa, \quad (4.21)$$

where κ denotes the imaginary part of the refractive index. From Eq. (4.4) it becomes clear that in our case also the wavenumber becomes complex,

$$\tilde{k}_z = \frac{\tilde{n}\omega}{c} = \frac{n\omega}{c} + i\frac{\kappa\omega}{c} = k'_z + ik''_z. \quad (4.22)$$

Let us now substitute Eq. (4.22) into Eq. (4.2a),

$$E_x(z, t) = E_{x,0} \cdot e^{i\tilde{k}_z z - i\omega t} = E_{x,0} \cdot e^{-k''_z z} e^{ik'_z z - i\omega t}. \quad (4.23)$$

We thus see that the electric field is attenuated exponentially, $\exp(-k''_z z)$ when travelling through the absorbing medium. The intensity of the electromagnetic field is proportional to the square of the electric field,

$$I(\mathbf{r}, t) \propto |\mathbf{E}(\mathbf{r}, t)|^2. \quad (4.24)$$

Therefore we find for the attenuation of the intensity of the electromagnetic field

$$I(z) = I_0 \exp(-2k''_z z) = I_0 \exp(-\alpha z), \quad (4.25)$$

where α is the *absorption coefficient*. It is related to the other properties via

$$\alpha = 2k''_z = 2\frac{\kappa\omega}{c} = \frac{4\pi\kappa}{\lambda_0}, \quad (4.26)$$

where $\lambda_0 = 2\pi c/\omega$ is the wavelength *in vacuo*.

Equation (4.25) is known as the *Lambert-Beer law*. A magnitude that is often used to judge the absorptivity of a material at a certain wavelength, is the *penetration depth* δ_p ,

$$\delta_p = \frac{1}{\alpha}. \quad (4.27)$$

At this depth, the intensity has decayed to a fraction of $1/e$ of the initial value.

In general, the complex refractive index and hence the absorption coefficient are no material constants but vary with the frequency. Especially α may change across several orders of magnitude across the spectrum, making the material very absorptive at one wavelength but almost transparent at other wavelength. Absorption spectra will be discussed thoroughly later on when looking at various photovoltaic materials in Part III.

4.5 Continuity and Poisson equations

At the end of this chapter we want to mention two equations that are very important for our treatise of semiconductor physics in Chapter 6.

4.5.1 Poisson equation

The first equation is the *Poisson equation* that relates the density of electric charges $\rho(\mathbf{r})$ to the electrical potential $U(\mathbf{r})$. For its derivation, we start with the first Maxwell equation (A.1a). Using Eq. (A.2a) we obtain

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \frac{\rho}{\epsilon\epsilon_0}, \quad (4.28)$$

where E_x , E_y and E_z are the components of the electric field vector, $\mathbf{E} = (E_x, E_y, E_z)$. Further, we here assume that we are in an electrostatic situation, *i.e.* there are no moving charges. From the second Maxwell equation (A.1b) we know that in that case the electric field is *rotation free*. Vector calculus teaches us that then the electric field is connected to the electric potential via

$$\mathbf{E} = - \left(\frac{\partial U}{\partial x}, \frac{\partial U}{\partial y}, \frac{\partial U}{\partial z} \right). \quad (4.29)$$

By combining Eqs. (4.28) with (4.29) we find the *Poisson equation*.

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) U = -\frac{\rho}{\epsilon\epsilon_0}. \quad (4.30)$$

In Chapter 6 we only will use the one-dimensional form given by

$$\frac{\partial^2 U}{\partial z^2} = -\frac{\rho}{\epsilon\epsilon_0}. \quad (4.31)$$

4.5.2 Continuity equation

Charge is a conserved quantity. The total amount of charge inside a volume V only can be changed via charges flowing through the boundary surface S of this volume. This can be expressed mathematically by the equation

$$\frac{dQ_V}{dt} + \iint_S \mathbf{J} dS = 0, \quad (4.32)$$

where \mathbf{J} is the current density vector and Q_V is the total charge contained within the volume V . It is given by

$$Q_V = \iiint_V \rho dV. \quad (4.33)$$

Equation (4.32) is the *integral* formulation of the continuity equation. It is equivalent to the *differential* for-

mulation that is given by

$$\frac{\partial \rho}{\partial t} + \left(\frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z} \right) = 0, \quad (4.34)$$

where J_x , J_y and J_z are the components of the current density vector, $\mathbf{J} = (J_x, J_y, J_z)$.

5

Solar Radiation

5.1 The Sun

We begin this chapter on Solar Radiation with a short introduction about *the Sun*. The Sun is the central star of our solar system. It consists mainly of hydrogen and helium. Some basic facts are summarised in table 5.1 and its structure is sketched in Fig. 5.1. The mass of the Sun is so large that it contributes to 99.68% of the total mass of the solar system. In the center of the Sun the pressure-temperature conditions are such that *nuclear fusion* can take place. In the major nuclear reaction, the *proton-proton reaction*, via a number of steps four protons react into

- a helium core (two protons and two neutrons),

- 2 positrons (the anti-particles of electrons),
- 2 neutrinos,
- electromagnetic radiation.

The positrons annihilate with electrons leading to additional radiation. The mass of the helium core is 0.635% less than that of four protons, the energy difference is converted into energy according to Einstein's equation

$$E = mc^2. \quad (5.1)$$

Every second thus, approx, 4 million tons of mass are converted into energy. However, the power density at the center of the Sun is estimated by theoretical assumptions only to be about 275 W/m^3 .

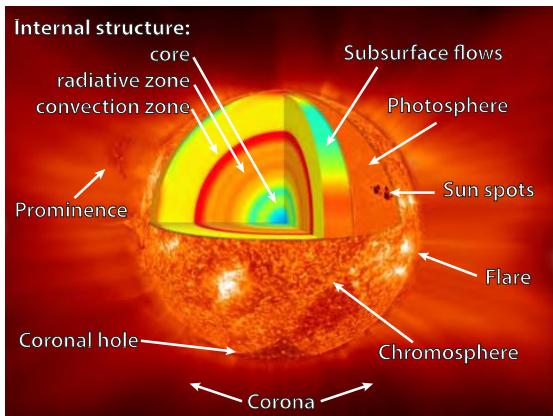


Figure 5.1: The Sun with its layer structure depicted [24].

The neutrinos hardly interact with matter and thus can leave the solar core without any hinder. Every second, about $6.5 \cdot 10^{10}$ per cm^2 pass through the Earth and hence also through our bodies. Neutrinos carry about 2% of the total energy radiated by the Sun.

The remainder of the radiation is released as electromagnetic radiation. The core of the Sun is so dense that radiation cannot travel freely but is continuously absorbed and re-emitted, such that it takes the radiation between 10 000 and 170 000 years to travel to the solar surface. The surface of the Sun is called the photo-

sphere. It has a temperature of about 6000 K. It behaves quite perfectly as a blackbody (see section 5.3) and is the source of the solar radiation that hits the Earth. The total power density of the solar radiation at the mean Earth-sun distance on a plane perpendicular to the direction of the Sun, outside the Earth's atmosphere, is referred to as the *solar constant*. Its value is approximately 1361 W/m^2 .

5.2 Radiometric properties

Radiometry is the branch of optics concerned with the measurement of light. Since photovoltaics deals with sunlight that is converted into electricity it is very important to discuss how the "amount of energy" of the light can be expressed physically and mathematically.

In solar science, not the total amount of the energy is important, but the amount of energy per unit time. We thus will use the *power* that is given by

$$P = \frac{dE}{dt}. \quad (5.2)$$

For our discussion we assume a surface A that is irradiated by light, as illustrated in Fig. 5.2 (a). For obtaining the total power that is incident on the surface, we have to integrate over the whole surface. Further we

Table 5.1: Some facts on the Sun

Mean distance from the Earth:	149 600 000 km (the astronomic unit, AU)
Diameter:	1 392 000 km (109 × that of the Earth)
Volume:	1 300 000 × that of the Earth
Mass:	1.993×10^{27} kg (332 000 times that of the Earth)
Density (at its center):	$>10^5$ kg m ⁻³ (over 100 times that of water)
Pressure (at its center):	over 1 billion atmospheres
Temperature (at its center):	about 15 000 000 K
Temperature (at the surface):	6 000 K
Energy radiation:	3.8×10^{26} W
The Earth receives:	1.7×10^{18} W

have to take into account that light is incident from all the different directions, which we parameterise with the spherical coordinates (θ, ϕ) . The *polar angle* θ is defined with respect to the normal of the surface element dA and ϕ is the *azimuth*, as sketched in Fig. 5.2 (b). Thus, we have to integrate over the hemisphere, from that light can be incident on the surface element dA , as well. We therefore obtain

$$P = \int_A \int_{2\pi} L_e \cos \theta d\Omega dA. \quad (5.3)$$

The quantity L_e is called the *radiance* and it is one of the most fundamental radiative properties. Its physical dimension is

$$[L_e] = \text{W}\cdot\text{m}^{-2}\cdot\text{sr}^{-1}.$$

The factor $\cos \theta$ expresses the fact that not the surface

element dA itself is the relevant property but the projection of dA to the normal of the direction (θ, ϕ) . This is also known as the *Lambert cosine law*.

We can express Eq. (5.3) as integrals of the surface coordinates (ξ, η) and the direction coordinates (θ, ϕ) , which reads as

$$P = \int_A \int_{2\pi} L_e(\xi, \eta; \theta, \phi) \cos \theta \sin \theta d\theta d\phi d\xi d\eta. \quad (5.4)$$

Since sunlight consists of a spectrum of different frequencies (or wavelengths), it is useful to use *spectral*

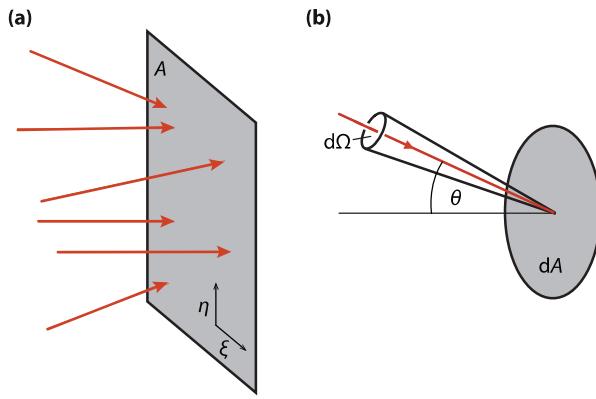


Figure 5.2: (a) Illustrating a surface A irradiated by light from various directions and (b) a surface element dA that receives radiation from a solid angle element $d\Omega$ under an angle θ with respect to the surface normal.

properties. These are given by

$$P_\nu = \frac{dP}{d\nu}, \quad P_\lambda = \frac{dP}{d\lambda}, \quad (5.5)$$

$$L_{e\nu} = \frac{dL_e}{d\nu}, \quad L_{e\lambda} = \frac{dL_e}{d\lambda}, \quad (5.6)$$

etc. Their physical dimensions are

$$\begin{aligned} [P_\nu] &= W \cdot Hz^{-1} = W \cdot s, & [P_\lambda] &= W \cdot m^{-1}, \\ [L_{e\nu}] &= W \cdot m^{-2} \cdot sr^{-1} \cdot s, & [L_{e\lambda}] &= W \cdot m^{-2} \cdot sr^{-1} \cdot m^{-1}, \end{aligned}$$

Since wavelength and frequency are connected to each other via $\nu\lambda = c$, P_ν and P_λ are related via

$$P_\nu = \frac{dP}{d\nu} = \frac{dP}{d\lambda} \frac{d\lambda}{d\nu} = P_\lambda \cdot \left(-\frac{c}{\nu^2} \right), \quad (5.7)$$

and similarly for $L_{e\nu}$ and $L_{e\lambda}$. The $-$ sign is because of the changing direction of integration when switching between ν and λ and usually is omitted.

The spectral power in wavelength thus can be obtained via

$$P_\lambda = \int_A \int_{2\pi} L_{e\lambda} \cos \theta d\Omega dA, \quad (5.8)$$

and analogously for P_ν . The radiance is given by

$$L_e = \frac{1}{\cos \theta} \frac{\partial^4 P}{\partial A \partial \Omega}, \quad (5.9)$$

and similarly for $L_{e\nu}$ and $L_{e\lambda}$.

Another very important radiometric property is the *irradiance* I_e that tells us the power density at a certain point (ξ, η) of the surface. It often also is called the *(spectral) intensity* of the light. It is given as the integral of the radiance over the solid angle,

$$I_e = \int_{2\pi} L_e \cos \theta d\Omega = \int_{2\pi} L_e(\xi, \eta; \theta, \phi) \cos \theta \sin \theta d\theta d\phi. \quad (5.10)$$

The *spectral irradiance* I_{ev} or $I_{e\lambda}$ is calculated similarly. The physical dimensions are

$$[I_e] = \text{W}\cdot\text{m}^{-2}, \quad [I_{ev}] = \text{W}\cdot\text{m}^{-2}\cdot\text{s}, \quad [I_{e\lambda}] = \text{W}\cdot\text{m}^{-2}\cdot\text{m}^{-1}.$$

The irradiance also is given as

$$I_e = \frac{\partial^2 P}{\partial A}, \quad (5.11)$$

and similarly for I_{ev} and $I_{e\lambda}$. Irradiance refers to radiation, that is received by the surface. For radiation emitted by the surface, we instead speak of *radiant emittance*, M_e , M_{ev} , and $M_{e\lambda}$.

As we discussed earlier, the energy of a photon is proportional to its frequency, $E_{ph} = h\nu = hc/\lambda$. Thus, the spectral power P_λ is proportional to the *spectral photon flow* $N_{ph,\lambda}$,

$$P_\lambda = N_{ph,\lambda} \frac{hc}{\lambda}, \quad (5.12)$$

and similarly for P_ν and $N_{ph,\nu}$. The total photon flow N

is related to the spectral photon flow via

$$N_{ph} = \int_0^\infty N_{ph,\nu} d\nu = \int_0^\infty N_{ph,\lambda} d\lambda. \quad (5.13)$$

The physical dimensions of the (spectral) photon flow are

$$[N_{ph}] = \text{s}^{-1}, \quad [N_{ph,\nu}] = 1, \quad [N_{ph,\lambda}] = \text{s}^{-1}\cdot\text{m}^{-1}.$$

The *(spectral) photon flux* Φ_{ph} is defined as the photon flow per area,

$$\Phi_{ph} = \frac{\partial^2 N_{ph}}{\partial A}, \quad (5.14)$$

and similarly for $\Phi_{ph,\nu}$ and $\Phi_{ph,\lambda}$. The physical dimensions are

$$[\Phi_{ph}] = \text{s}^{-1}\text{m}^{-2}, \quad [\Phi_{ph,\nu}] = \text{m}^{-2}, \quad [\Phi_{ph,\lambda}] = \text{s}^{-1}\text{m}^{-2}\cdot\text{m}^{-1}.$$

By comparing Eqs. (5.11) and (5.14) and looking at Eq. (5.12), we find

$$I_{e\lambda} = \Phi_{ph,\lambda} \frac{hc}{\lambda}, \quad (5.15)$$

and analogously for I_{ev} and $\Phi_{ph,\nu}$.

5.3 Blackbody radiation

If we take a piece of *e.g.* metal and start heating it up, it will start to glow, first in reddish colour getting more

and more yellowish if we increase temperature even further. It thus emits electromagnetic radiation that we call *thermal radiation*. Understanding this phenomenon theoretically and correctly predicting the emitted spectrum was one of the most important topics of physics in the late nineteenth century.

For discussing thermal radiation, the concept of the *blackbody* is very useful. A blackbody, which does not exist in nature, absorbs all the radiation that is incident on it, regardless of wavelength and angle of incidence. Its reflectivity therefore is 0. Of course, since it also will emit light according to its equilibrium temperature, it does not need to appear black to the eye.

Two approximations for the blackbody spectrum were presented around the turn of the century: First, in 1896, Wilhelm Wien empirically derived the following expression for the spectral blackbody radiance:

$$L_{e\lambda}^W(\lambda; T) = \frac{C_1}{\lambda^5} \exp\left(-\frac{C_2}{\lambda T}\right), \quad (5.16)$$

where λ and T are the wavelength and the temperature, respectively. While this approximation gives good results for short wavelengths, it fails to predict the emitted spectrum at long wavelengths, thus in the infrared.

Secondly, in 1900 and in a more complete version in 1905, Lord Rayleigh and James Jeans, respectively, derived

$$L_{e\lambda}^{RJ}(\lambda; T) = \frac{2ck_B T}{\lambda^4}, \quad (5.17)$$

where $k_B \approx 1.381 \cdot 10^{-23}$ J/K is the Boltzmann constant. The derivation of this equation was based on electrodynamic arguments. While $L_{e\lambda}^{RJ}$ is in good agreement to measured values at long wavelengths, it diverges to infinity for short wavelength. Further, the radiant emittance, which is obtained via integration over all wavelength, diverges towards infinity. This so called *ultraviolet catastrophe* demonstrates that Rayleigh and Jeans did not succeed in developing a model that can adequately describe thermal radiation.

In 1900, Max Planck found an equation, that interpolates between the Wien approximation and the Rayleigh-Jeans law,

$$L_{e\lambda}^{BB}(\lambda; T) = \frac{2hc^2}{\lambda^5} \frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right) - 1}, \quad (5.18a)$$

where $c \approx 2.998 \cdot 10^8$ m/s is the speed of light *in vacuo* and $h \approx 6.626 \cdot 10^{-34}$ m²kg/s is the nowadays called *Planck constant*. Via Eq. (5.7) we find the *Planck law* expressed as a function of the frequency ν ,

$$L_{ev}^{BB}(\nu; T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp\left(\frac{h\nu}{k_B T}\right) - 1}, \quad (5.18b)$$

It is remarkable to see that the Planck law contains three fundamental constants, c , k_B , and h , which are amongst the most important constants in physics.

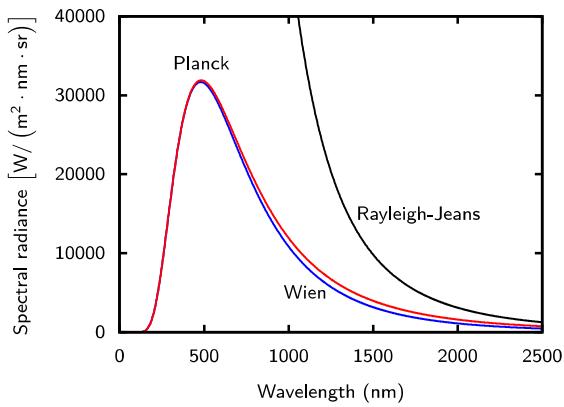


Figure 5.3: The blackbody spectrum at 6000 K as calculated with the Wien approximation, the Rayleigh-Jeans law and the Planck law.

Figure 5.3 shows the spectrum of a blackbody of 6000 K temperature and the Wien approximation and the Rayleigh-Jeans law. We indeed see that the Wien approximation fits well at short wavelengths, while the Rayleigh-Jeans law matches well at long wavelengths but completely fails at short wavelengths.

Both the Wien approximation [Eq. (5.16)] and the Rayleigh-Jeans law [Eq. (5.17)] can be directly derived from the Planck law:

For short wavelength,

$$\exp\left(\frac{hc}{\lambda k_B T}\right) \gg 1,$$

such that the -1 can be ignored and we arrive at the Wien approximation with $C_1 = 2hc^2$ and $C_2 = hc/k_B$.

For long wavelength we can use the approximation

$$\exp\left(\frac{hc}{\lambda k_B T}\right) - 1 \approx \frac{hc}{\lambda k_B T},$$

which directly results in the Rayleigh-Jeans law.

The total radiant emittance of a black body is given by

$$M_e^{BB}(T) = \int_0^\infty \int_{2\pi} L_{e\lambda}^{BB}(\lambda; T) \cos \theta \sin \theta d\theta d\phi d\lambda = \sigma T^4, \quad (5.19)$$

where

$$\sigma = \frac{2\pi^5 k_B^4}{15c^2 h^3} \approx 5.670 \cdot 10^{-8} \frac{\text{J}}{\text{sm}^2 \text{K}^4} \quad (5.20)$$

is the *Stefan-Boltzmann* constant. Equation (20.5) is known as the *Stefan-Boltzmann law*. As a matter of fact, it already was discovered in 1879 and 1884 by Jozef Stefan and Ludwig Boltzmann, respectively, *i.e.* about twenty years prior to the derivation of Planck's law. This law is very important because it tells us that if the temperature of a body (in K) is doubled, it emits 16 times as much power. Little temperature variations thus have a large influence on the total emitted power.

Another important property of blackbody radiation is Wien's displacement law, which states that the wavelength of maximal radiance is indirectly proportional to the temperature,

$$\lambda_{\max} T = b \approx 2.898 \cdot 10^{-3} \text{ m} \cdot \text{K}. \quad (5.21)$$

Figure 5.4 shows the spectra for three different temperatures. Note the strong increase in radiance with temperature and also the shift of the maximum to shorter wavelengths.

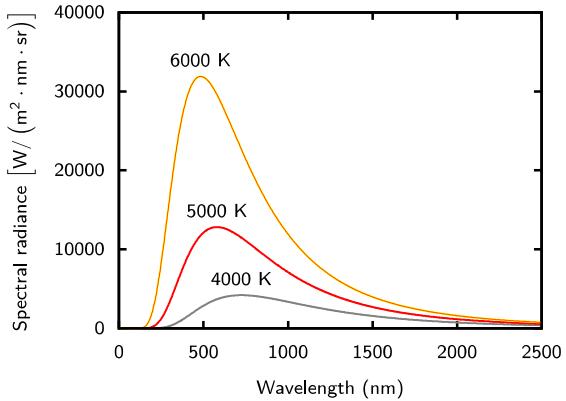


Figure 5.4: The blackbody spectrum at three different temperatures

5.4 Wave-particle duality

In Planck's law, as stated in Eqs. (5.18), the constant h appeared for the first time. Its product with the frequency, $h\nu = hc/\lambda$ has the unit of an energy. Planck

himself did not see the implications of \hbar . It was Einstein, who understood in 1905 that Planck's law actually has to be interpreted such that light comes in quanta of energy with the size

$$E_{\text{ph}} = h\nu. \quad (5.22)$$

Nowadays, these quanta are called *photons*. In terms of classical mechanics we could say that *light shows the behaviour of particles*.

On the other hand, we have seen in Chapter 4 that light also shows *wave character* which becomes obvious when looking at the propagation of light through space or at reflection and refraction at a flat interface. It also was discovered that other particles, such as electrons, show wave-like properties.

This behaviour is called *wave-particle duality* and is a very intriguing property of *quantum mechanics* that was discovered and developed in the first quarter of the twentieth century. Many discussion was held on how this duality has to be interpreted - but this is out of the focus of this book. So we just will accept that depending on the situation light might behave as wave or as particle.

5.5 Solar spectra

As we already mentioned in chapter 3, only photons of appropriate energy can be absorbed and hence generate electron-hole pairs in a semiconductor material. Therefore, it is important to know the spectral distribution of the solar radiation, *i.e.* the number of photons of a particular energy as a function of the wavelength λ . Two quantities are used to describe the solar radiation spectrum, namely the *spectral irradiance* $I_{e\lambda}$ and the *spectral photon flux* $\Phi_{\text{ph}}(\lambda)$. We defined these quantities already in section 5.2.

The surface temperature of the Sun is about 6000 K. If it would be a perfect black body, it would emit a spectrum as described by Eqs. (5.18), which give the spectral radiance. For calculating the spectral *irradiance* a blackbody with the size and position of the Sun would have on Earth, we have to multiply the spectral radiance with the solid angle of the Sun as seen from Earth,

$$I_{e\lambda}^{BB}(T; \lambda) = L_{e\lambda}^{BB}(T; \lambda)\Omega_{\text{Sun}}. \quad (5.23)$$

We can calculate Ω_{Sun} with

$$\Omega_{\text{Sun}} = \pi \left(\frac{R_{\text{Sun}}}{\text{AU} - R_{\text{Earth}}} \right)^2. \quad (5.24)$$

Using $R_{\text{Sun}} = 696\,000$ km, an astronomical unit $\text{AU} = 149\,600\,000$ km, and $R_{\text{Earth}} = 6370$ km, we find

$$\Omega_{\text{Sun}} \approx 68 \mu\text{sr}. \quad (5.25)$$

The blackbody spectrum is illustrated in Fig. 5.5. The spectrum outside the atmosphere of the Earth is already very different. It is called the AM0 spectrum, because no (or “zero”) atmosphere is traversed. AM0 also is shown in Fig. 5.5. The irradiance at AM0 is $I_e(\text{AM0}) = 1361 \text{ W}\cdot\text{m}^{-2}$.

When solar radiation passes through the atmosphere of the Earth, it is attenuated. The most important parameter that determines the solar irradiance under clear sky conditions is the distance that the sunlight has to travel through the atmosphere. This distance is the shortest when the Sun is at the zenith, *i.e.* directly overhead. The ratio of an actual path length of the sunlight to this minimal distance is known as the *optical air mass*. When the Sun is at its zenith the optical air mass is unity and the spectrum is called the air mass 1 (AM1) spectrum. When the Sun is at an angle θ with the zenith, the air mass is given by

$$\text{AM} := \frac{1}{\cos \theta}. \quad (5.26)$$

For example, when the Sun is 60° from the zenith, *i.e.* 30° above the horizon, we receive an AM2 spectrum. Depending on the position on the Earth and the position of the Sun in the sky, terrestrial solar radiation varies both in intensity and the spectral distribution. The attenuation of solar radiation is due to scattering and absorption by air molecules, dust particles and/or aer-

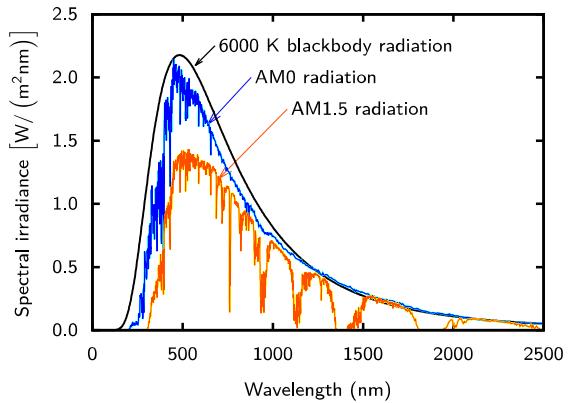


Figure 5.5: Different solar spectra: the blackbody spectrum of a blackbody at 6000 K, the extraterrestrial AM0 spectrum and the AM1.5 spectrum.

osols in the atmosphere. Especially, steam (H_2O), oxygen (O_2) and carbon dioxide (CO_2) cause absorption. Since this absorption is wavelength-selective, it results in gaps in the spectral distribution of solar radiation as apparent in Fig. 5.5. Ozone absorbs radiation with wavelengths below 300 nm. Depletion of ozone from the atmosphere allows more ultra-violet radiation to reach the Earth, with consequent harmful effects upon biological systems. CO_2 molecules contribute to the absorption of solar radiation at wavelengths above 1 μm . By changing the CO_2 content in the atmosphere the absorption in the infrared, which has consequences for our climate.

Solar cells and photovoltaic modules are produced by many different companies and laboratories. Further, many different solar cell technologies are investigated and sold. It is therefore of utmost importance to define a *reference solar spectrum* that allows a comparison of all the different solar cells and PV modules. The industrial standard is the AM1.5 spectrum, which corresponds to an angle of 48.2°. While the “real” AM1.5 spectrum corresponds to a total irradiance of $827 \text{ W} \cdot \text{m}^{-2}$, the industrial standard corresponds to $I_e(\text{AM1.5}) = 1000 \text{ W} \cdot \text{m}^{-2}$ and is close to the maximum received at the surface of the Earth. The power generated by a PV module under these conditions is thus expressed in the unit Watt peak, W_p .

The actual amount of solar radiation that reaches a par-

ticular place on the Earth is extremely variable. In addition to the regular daily and annual variation due to the apparent motion of the Sun, irregular variations have to be taken into account that are caused by local atmospheric conditions, such as clouds. These conditions particularly influence the direct and diffuse components of solar radiation. The direct component of solar radiation is that part of the sunlight that directly reaches the surface. Scattering of the sunlight in the atmosphere generates the diffuse component. A part of the solar radiation that is reflected by the Earth’s surface, which is called albedo, may be also present in the total solar radiation. We use a term global radiation to refer to the total solar radiation, which is made up of these three components.

The design of an optimal photovoltaic system for a particular location depends on the availability of the solar insolation data at the location. Solar irradiance integrated over a period of time is called solar irradiation. For example, the average annual solar irradiation in the Netherlands is $1\,000 \text{ kWh/m}^2$, while in Sahara the average value is $2\,200 \text{ kWh/m}^2$, thus more than twice as high. We will discuss these issues in more detail in Chapter 15.

6

Basic Semiconductor Physics

6.1 Introduction

With this chapter we start with the discussion of some important concepts from *semiconductor physics*, which are required to understand the operation of solar cells. After giving a brief introduction into semiconductor physics in this chapter, we will discuss the most important *generation and recombination* mechanisms in Chapter 7. Finally, we will focus on the physics of *semiconductor junctions* in Chapter 8.

The first successful solar cell was made from crystalline silicon (c-Si), which still is by far the most widely used PV material. Therefore we shall use c-Si as an example to explain the concepts of semiconductor physics that

are relevant to solar cell operation. This discussion will give us a basic understanding of how solar cells based on other semiconductor materials work.

The *central semiconductor parameters* that determine the design and performance of a solar cell are:

1. Concentrations of doping atoms, which can be of two different types: *donor atoms*, which donate free electrons or *acceptor atoms*, which accept electrons. The concentrations of donor and acceptor atoms are denoted by N_D and N_A , respectively, and determine the width of the space-charge region of a junction, as we will see in Chapter 8.
2. The mobility μ and the diffusion coefficient D of charge carriers is used to characterise the transport

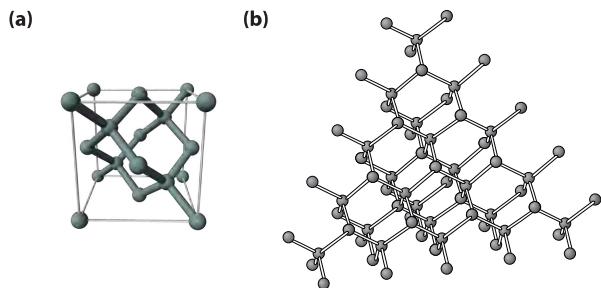


Figure 6.1: (a) A diamond lattice unit cell representing a unit cell of single crystal Si [25], (b) the atomic structure of a part of single crystal Si.

of carriers transport due to drift and diffusion, respectively, which we will discuss in Section 6.5.

3. The lifetime τ and the diffusion length L of the *excess carriers* characterise the recombination-generation processes, discussed in Chapter 7.
4. The band gap energy E_g , and the complex refractive index $n - ik$, where k is linked to the absorption coefficient α , characterise the ability of a semiconductor to absorb electromagnetic radiation.

6.2 Atomic structure

The *atomic number* of silicon is 14, which means that there are 14 electrons orbiting the nucleus. In ground state configuration, two electrons are in the first shell, both living in the 1s orbital. Further, eight electrons are in the second shell, two in the 2s and six in the 2p orbitals. Hence, four electrons live in the third shell, which is the outermost shell for a Si atom. Only these four electrons interact with other atoms, for example via forming chemical bonds. They are called the *valence electrons*.

Two Si atoms are bonded together when they share each other's valence electron. This is the so called *covalent bond* that is formed by two electrons. Since Si atoms have four valence electrons, they can be covalently bonded to four other Si atoms. In the crystalline form each Si atom is covalently bonded to four neighbouring Si atoms, as illustrated in Fig. 6.1.

In the ground state, two valence electrons live in the 3s orbital and the other two are present in the three 3p orbitals (p_x , p_y and p_z). In that state only the two electrons in the 3p orbitals can form bond as the 3s orbital is full. In a silicon crystal, where every atom is *symmetrically* connected to four others, the Si atoms are present as so-called *sp₃ hybrids*. The 3p and 3s orbitals are mixed forming 4 sp₃ orbitals. Every of these four orbitals is occupied by one electron that can form a covalent bond with a valence electron from a neighbour-

ing atom.

All bonds have the same length and the angles between the bonds are equal with 109.5° . The number of bonds that an atom has with its immediate neighbours in the atomic structure is called the *coordination number* or *coordination*. Thus, in single crystal silicon, the coordination number for all Si atoms is four, we can also say that Si atoms are fourfold coordinated. A *unit cell* can be defined, from which the crystal lattice can be reproduced by duplicating the unit cell and stacking the duplicates next to each other. Such a regular atomic arrangement is described as a structure with *long range order*.

A diamond lattice unit cell represents the real lattice structure of monocrystalline silicon. Figure 6.1 (a) shows the arrangement of the unit cell and Fig. 6.1 (b) the atomic structure of single crystal silicon. One can determine from Fig. 6.1 (a) that there are eight Si atoms in the volume of the unit cell. When a lattice constant of c-Si is 5.4 \AA one can easily calculate that the density of atoms is approximately $5 \times 10^{22} \text{ cm}^{-3}$. Figure 6.1 (a) shows the crystalline Si atomic structure with no foreign atoms. In practice, a semiconductor sample always contains some impurity atoms. When the concentration of impurity atoms in a semiconductor is insignificant we refer to such semiconductor as an *intrinsic semiconductor*.

At practical operational conditions, e.g. at room tem-

perature,¹ there are always some of the covalent bonds broken. The breaking of the bonds results in liberating the valence electrons from the bonds and making them mobile through the crystal lattice. We refer to these electrons as free electrons (henceforth simply referred to as electrons). The position of a missing electron in a bond, which can be regarded as positively charged, is referred to as a hole. This situation can be easily visualised by using the *bonding model* illustrated in Fig. 6.2.

In the bonding model the atomic cores (atoms without valence electrons) are represented by circles and the valence or bonding electrons are represented by lines interconnecting the circles. In case of c-Si one Si atom has four valence electrons and four nearest neighbours. Each of the valence electron is equally shared with the nearest neighbour. There are therefore eight lines terminating on each circle. In an ideal Si crystal at 0 K all valence electrons take part in forming covalent bonds between Si atoms and therefore no free electrons are present in the lattice. This situation is schematically shown in Fig. 6.2 (a).

At temperatures higher than 0 K the bonds start to break due to absorbing thermal energy. This process results in creation of mobile electrons and holes. Figure 6.2 (b) shows a situation when a covalent bond is broken and one electron departs from the bond leav-

¹In semiconductor physics most of the time a temperature of 300 K is assumed.

ing a so-called *hole* behind. A single line between the atoms in Fig. 6.2 (b) represents the remaining electron of the broken bond. When a bond is broken and a hole created, a valence electron from a neighbouring bond can “jump” into this empty position and restore the bond. The consequence of this transfer is that at the same time the jumping electron creates an empty position in its original bond. The subsequent “jumps” of a valence electron can be viewed as a motion of the hole, a positive charge representing the empty position, in the direction opposite to the motion of the valence electron through the bonds.

Since breaking of a covalent bond leads to the formation of an electron-hole pair, in intrinsic semiconductors the concentration of electrons is equal to the concentration of holes. In intrinsic silicon at 300 K approximately $1.5 \times 10^{10} \text{ cm}^{-3}$ broken bonds are present. This number then gives also the concentration of holes, p , and electrons, n . Hence, at 300 K, $n = p = 1.5 \times 10^{10} \text{ cm}^{-3}$. This concentration is called the *intrinsic carrier concentration* and is denoted as n_i .

6.3 Doping

The concentrations of electrons and holes in c-Si can be manipulated by *doping*. Doping of silicon means that atoms of other elements substitute Si atoms in the

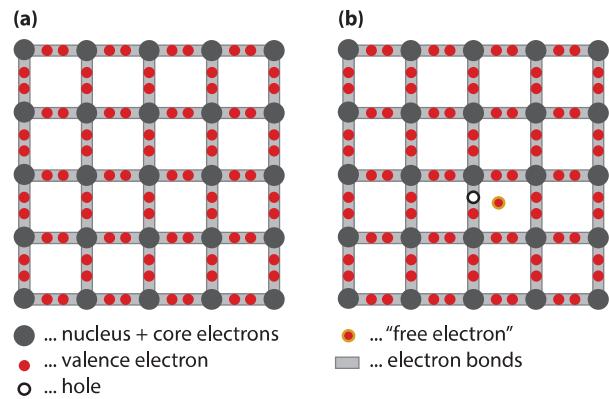


Figure 6.2: The bonding model for c-Si. (a) No bonds are broken. (b) A bond between two Si atoms is broken resulting in a free electron and hole.

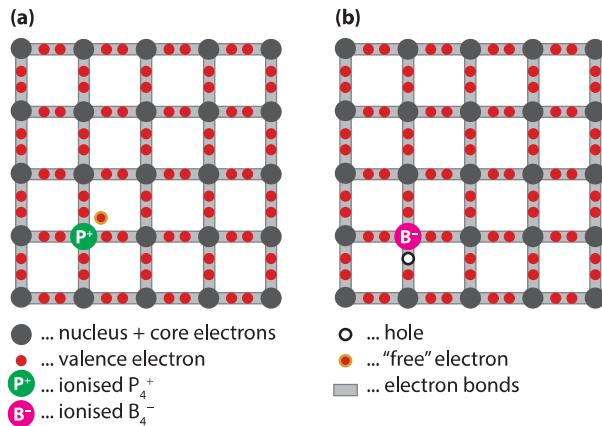


Figure 6.3: The doping process illustrated using the bonding model. (a) A phosphorus (P) atom substitutes a Si atom in the lattice resulting in the positively-ionised P atom and a free electron, (b) A boron (B) atom substitutes a Si atom resulting in the negatively ionised B atom and a hole.

crystal lattice. The substitution has to be carried out by atoms with three or five valence electrons, respectively. The most used elements to dope c-Si are boron (B) and phosphorus (P), with atomic numbers of 5 and 15, respectively.

The process of doping action can best be understood with the aid of the bonding model and is illustrated in

Fig. 6.3. When introducing phosphorus atom into the c-Si lattice, four of the five phosphorus atom valence electrons will readily form bonds with the four neighbouring Si atoms. The fifth valence electron cannot take part in forming a bond and becomes rather weakly bound to the phosphorus atom. It is easily liberated from the phosphorus atom by absorbing the thermal energy, which is available in the c-Si lattice at room temperature. Once free, the electron can move throughout the lattice. In this way the phosphorus atom that substitutes a Si atom in the lattice “*donates*” a free (mobile) electron into the c-Si lattice. The impurity atoms that enhance the concentration of electrons are called *donors*. We denote the concentration of donors by N_D .

An atom with three valence electrons such as boron cannot form all bonds with four neighbouring Si atoms when it substitutes a Si atom in the lattice. However, it can readily “*accept*” an electron from a nearby Si-Si bond. A thermal energy that the c-Si lattice contains at room temperature is sufficient to enable an electron from a nearby Si-Si bond to attach itself to the boron atom and complete the bonding to the four Si neighbours. In this process a hole is created that can move around the lattice. The impurity atoms that enhance the concentration of holes are called *acceptors*. We denote the concentration of acceptors by N_A .

Note that by substituting Si atoms with only one type of impurity atoms, the concentration of only one type

of mobile charge carriers is increased. Charge neutrality of the material is nevertheless maintained because the sites of the bonded and thus fixed impurity atoms become charged. The donor atoms become positively ionised and the acceptor atoms become negatively ionised.

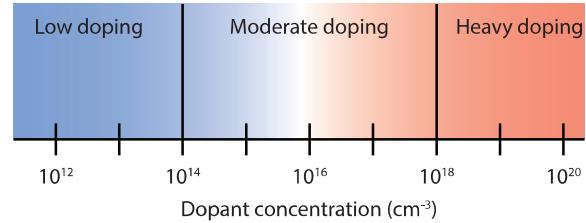


Figure 6.4: The range of doping levels used in c-Si.

The possibility to control the electrical conductivity of a semiconductor by doping is one of most important semiconductor features. The electrical conductivity in semiconductors depends on the concentration of electrons and holes as well as their mobility. The concentration of electrons and holes is influenced by the amount of the impurity atoms that are introduced into the atomic structure of semiconductor. Figure 6.4 shows the range of doping that is used in case of c-Si. We denote a semiconductor as *p*-type or *n*-type when holes or electrons, respectively, dominate its electrical conductivity. In case that one type of charge carriers has a higher concentration than the other type these carriers are called majority carriers (holes in the *p*-type and electrons in the *n*-type), while the other type with lower concentration are then called minority carriers (electrons in the *p*-type and holes in the *n*-type).

6.4 Carrier concentrations

6.4.1 Intrinsic semiconductors

Any operation of a semiconductor device depends on the concentration of carriers that transport charge inside the semiconductor and hence cause electrical currents. In order to determine and to understand device operation it is important to know the precise concentration of these charge carriers. In this section the concentrations of charge carriers inside a semiconductor are derived assuming the semiconductor is under *equilibrium*. The term equilibrium is used to describe the unperturbed state of a system, to which no external voltage, magnetic field, illumination, mechanical stress, or other perturbing forces are applied. In the equilibrium state the observable parameters of a semiconductor do not change with time.

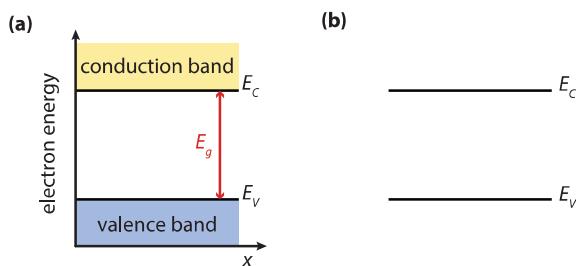


Figure 6.5: (a) The basic energy band diagram and (b) widely used simplified version of the energy diagram.

In order to determine the carrier concentration one has to know the function of density of allowed energy states of electrons and the occupation function of the allowed energy states. The density of energy states function, $g(E)$, describes the number of allowed states per unit volume and energy. The occupation function is the well known Fermi-Dirac distribution function, $f(E)$, which describes the ratio of states filled with an electron to total allowed states at given energy E . In an isolated Si atom, electrons are allowed to have only discrete energy values. The periodic atomic structure of single crystal silicon results in the ranges of allowed energy states for electrons that are called *energy bands* and the excluded energy ranges, forbidden gaps or *band gaps*. Electrons that are liberated from the bonds determine the charge transport in a semiconductor. There-

fore, we further discuss only those bands of energy levels, which concern the valence electrons. Valence electrons, which are involved in the covalent bonds, have their allowed energies in the *valence band* (VB) and the allowed energies of electrons liberated from the covalent bonds form the *conduction band* (CB). The valence band is separated from the conduction band by a band of forbidden energy levels. The maximum attainable valence-band energy is denoted E_V , and the minimum attainable conduction-band energy is denoted E_C . The energy difference between the edges of these two bands is called the band gap energy or band gap, E_g , and it is an important material parameter.

$$E_G = E_C - E_V. \quad (6.1)$$

At room temperature (300 K), the band gap of crystalline silicon is 1.12 eV.² A plot of the allowed electron energy states as a function of position is called the energy band diagram and for intrinsic c-Si is shown in Fig. 6.5.

The density of energy states at an energy E in the conduction band close to E_C and in the valence band close

²An electron-volt is equal to the energy, which an electron acquires when it passes through a potential of 1 volt in vacuum ($1\text{eV} = 1.602 \times 10^{-19} \text{ J}$).

to E_V are given by

$$g_c(E) = \left(\frac{4\sqrt{2}\pi m_n^*}{h^3} \right)^{\frac{3}{2}} \sqrt{E - E_C}, \quad (6.2a)$$

$$g_v(E) = \left(\frac{4\sqrt{2}\pi m_p^* t}{h^3} \right)^{\frac{3}{2}} \sqrt{E - E_V}, \quad (6.2b)$$

where m_n^* and m_p^* is the *effective mass* of electrons and holes, respectively. As the electrons and holes move in the periodic potential of the c-Si crystal, the mass has to be replaced by the effective mass, which takes the effect of a periodic force into account. The effective mass is also averaged over different directions to take anisotropy into account.

The Fermi-Dirac distribution function is given by

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)}, \quad (6.3)$$

where k is Boltzmann's constant ($k = 1.38 \times 10^{-23}$ J/K) and E_F is the so-called Fermi energy. kT is the *thermal energy*, at 300 K it is 0.0258 eV. The *Fermi energy* is the electrochemical potential of the electrons in a material and in this way it represents the averaged energy of electrons in the material. The Fermi-Dirac distribution function for different temperatures is shown in Fig. 6.6.

The carriers that contribute to charge transport are electrons in the conduction band and holes in the valence

band. The total concentration of electrons in the conduction band and the total concentration of holes in the valence band is obtained by multiplying density of states function with the distribution function and integrating across the whole energy band,

$$n = \int_{E_C}^{E_{top}} g_c(E)f(E)dE, \quad (6.4a)$$

$$p = \int_{E_{bottom}}^{E_V} g_v(E)[1 - f(E)]dE, \quad (6.4b)$$

Substituting the density of states and the Fermi-Dirac distribution function into Eq. (6.4) the resulting expressions for n and p are obtained after solving the equations. The full derivation can be found for example in Reference [20].

$$n = N_C \exp\left(\frac{E_F - E_C}{kT}\right) \quad \text{for } E_C - E_F \geq 3kT, \quad (6.5a)$$

$$p = N_V \exp\left(\frac{E_V - E_F}{kT}\right) \quad \text{for } E_F - E_V \geq 3kT, \quad (6.5b)$$

where N_C and N_V are the effective densities of the conduction band states and the valence band states, respectively. For crystalline silicon, we have at 300 K

$$N_C = 3.22 \times 10^{19} \text{ cm}^{-3}, \quad (6.6a)$$

$$N_V = 1.83 \times 10^{19} \text{ cm}^{-3}. \quad (6.6b)$$

When the requirement that the Fermi level lies in the band gap more than $3kT$ from either band edge is sat-

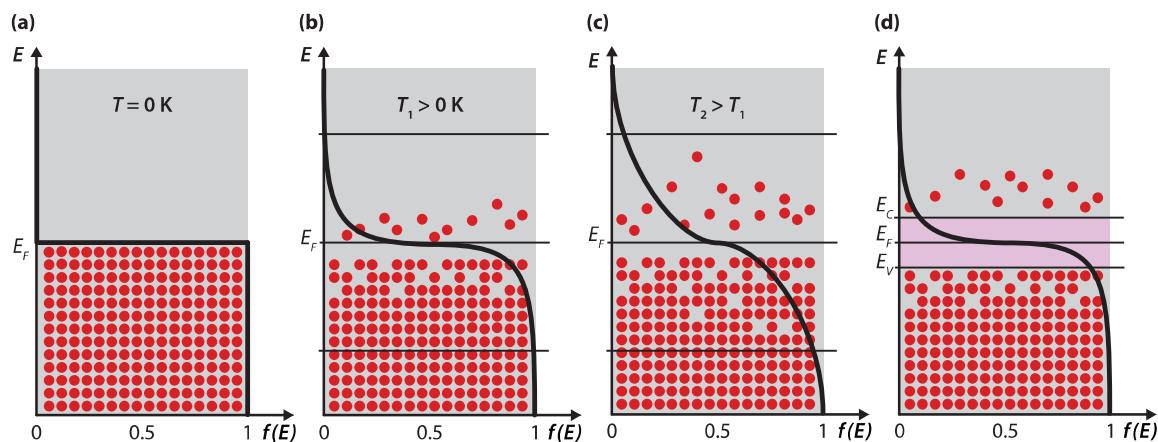


Figure 6.6: The Fermi-Dirac distribution function. (a) For $T = 0 \text{ K}$, all allowed states below the Fermi level are occupied by two electrons. (b, c) At $T > 0 \text{ K}$ not all states below the Fermi level are occupied and there are some states above the Fermi level that are occupied. (d) In an energy gap between bands no electrons are present.

isified the semiconductor is referred to as an *nondegenerate* semiconductor.

If an intrinsic semiconductor is in equilibrium, we have $n = p = n_i$. By multiplying the corresponding sides of Eqs. (6.5) we obtain

$$\begin{aligned} np &= n_i^2 = N_C N_V \exp\left(\frac{E_V - E_C}{kT}\right) \\ &= N_C N_V \exp\left(-\frac{E_g}{kT}\right), \end{aligned} \quad (6.7)$$

which is independent of the position of the Fermi level and thus valid for doped semiconductors as well. When we denote the position of the Fermi level in the intrinsic material E_{Fi} we may write

$$\begin{aligned} n_i &= N_C \exp\left(\frac{E_{Fi} - E_C}{kT}\right) \\ &= N_V \exp\left(\frac{E_V - E_{Fi}}{kT}\right). \end{aligned} \quad (6.8)$$

From Eq. (6.8) we can easily find the position of E_{Fi} to be

$$\begin{aligned} E_{Fi} &= \frac{E_C + E_V}{2} + \frac{kT}{2} \ln\left(\frac{N_V}{N_C}\right) \\ &= E_C - \frac{E_g}{2} + \frac{kT}{2} \ln\left(\frac{N_V}{N_C}\right). \end{aligned} \quad (6.9)$$

The Fermi level E_{Fi} lies close to the midgap $[(E_C + E_V)/2]$; a slight shift is caused by the difference in the effective densities of the valence and conduction band.

6.4.2 Doped semiconductors

It has been already mentioned in Section 6.3 that the concentrations of electrons and holes in c-Si can be manipulated by doping. The concentration of electrons and holes is influenced by the amount of the impurity atoms that substitute silicon atoms in the lattice. Under assumption that the semiconductor is uniformly doped and in equilibrium a simple relationship between the carrier and dopant concentrations can be established. We assume that at room temperature the dopant atoms are ionised. Inside a semiconductor the local charge density is given by

$$\rho = e (p + N_D^+ - n - N_A^-), \quad (6.10)$$

where e is the elementary charge ($e = 1.602 \times 10^{-19}$ C). N_D^+ and N_A^- denote the density of the *ionised donor* and *acceptor* atoms, respectively. As every ionised atom corresponds to a free atom (hole), N_D^+ and N_A^- tell us the concentration of electrons and holes due to doping, respectively.

Under equilibrium conditions, the local charge of the uniformly doped semiconductor is zero, which means that the semiconductor is charge-neutral everywhere. We thus can write:

$$p + N_D^+ - n - N_A^- = 0. \quad (6.11)$$

As previously discussed, the thermal energy available at room temperature is sufficient to ionise almost all the

dopant atoms. We therefore may assume

$$N_D^+ \approx N_D \quad \text{and} \quad N_A^- \approx N_A, \quad (6.12)$$

and hence

$$p + N_D - n - N_A = 0, \quad (6.13)$$

which is the common form of the *charge neutrality equation*.

Let us now consider an *n*-type material. At room temperature almost all donor atoms N_D are ionised and donate an electron into the conduction band. Under the assumption that $N_A = 0$, Eq. (6.13) becomes

$$p + N_D - n = 0, \quad (6.14)$$

Under the assumption that

$$N_D \approx N_D^+ \approx n \quad (6.15)$$

we can expect that the concentration of holes is lower than that of electrons, and becomes very low when N_D becomes very large. We can calculate more accurate the concentration of holes in the *n*-type material from Eq. (6.7).

$$p = \frac{n_i^2}{n} \approx \frac{n_i^2}{N_D} \ll n. \quad (6.16)$$

In case of a *p*-type material almost all acceptor atoms N_A are ionised at room temperature and accept an electron and leaving a hole in the valence band. Under the

assumption that $N_A = 0$, Eq. (6.13) becomes

$$p - n - N_A = 0. \quad (6.17)$$

Under the assumption that

$$N_A \approx N_A^- \approx p \quad (6.18)$$

we can expect that the concentration of electrons is lower than that of holes. We can calculate more accurate the concentration of electrons in the *p*-type material from Eq. (6.7).

$$n = \frac{n_i^2}{p} \approx \frac{n_i^2}{N_A} \ll p. \quad (6.19)$$

Inserting donor and acceptor atoms into the lattice of crystalline silicon introduces allowed energy levels into the forbidden bandgap. For example, the fifth valence electron of the P atom does not take part in forming a bond, is rather weakly bound to the atom and easily liberates from the P atom. The energy of the liberated electron lies in the CB. The energy levels, which we denote E_D , of the weakly-bound valence electrons of the donor atoms have to be positioned close the CB. Notice that a dashed line represents the E_D . This means that an electron that occupies the E_D level, is localised to the vicinity of the donor atom.

Similarly, the acceptor atoms introduce allowed energy levels E_A close to the VB. Doping also influences the

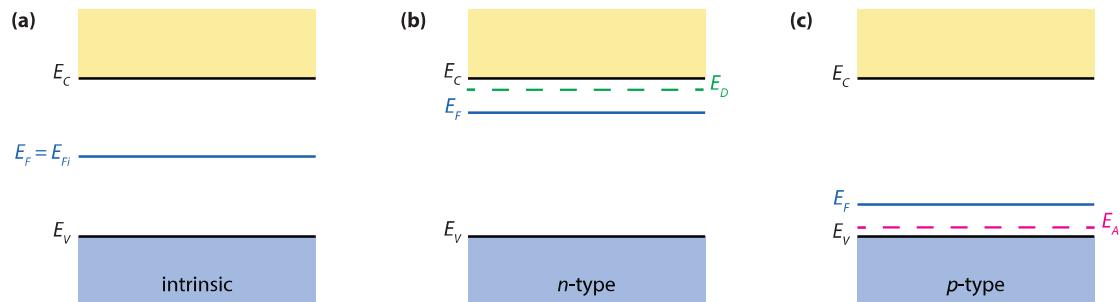


Figure 6.7: A shift of the position of the Fermi energy in the band diagram and the introduction of the allowed energy level into the bandgap due to the doping.

position of the Fermi energy. When increasing the concentration of electrons by increasing the concentration of donors the Fermi energy will increase, which is represented by bringing the Fermi energy close to the CB in the band diagram. In the *p*-type material the Fermi energy is moving closer the VB. A change in the Fermi-energy position and the introduction of the allowed energy level into the bandgap due to the doping is illustrated Fig. 6.7.

The position of the Fermi level in an *n*-type semiconductor can be calculated using Eqs. (6.5a) and (6.15) and in a *p*-type semiconductor using Eqs. (6.5b) and

(6.18), respectively,

$$E_C - E_F = kT \ln \left(\frac{N_C}{N_D} \right) \quad \text{for } n\text{-type}, \quad (6.20a)$$

$$E_F - E_V = kT \ln \left(\frac{N_V}{N_A} \right) \quad \text{for } p\text{-type}. \quad (6.20b)$$

6.5 Transport properties

In contrast to the equilibrium conditions, under operational conditions a net electrical current flows through a semiconductor device. The electrical currents are generated in a semiconductor due to the transport of charge

Example

This example demonstrates how much the concentration of electrons and holes can be manipulated by doping. A c-Si wafer is uniformly doped with $1 \times 10^{17} \text{ cm}^{-3}$ P atoms. P atoms act as donors and therefore at room temperature the concentration of electrons is almost equal to the concentration of donor atoms:

$$n = N_D^+ \approx N_D = 10^{17} \text{ cm}^{-3}.$$

The concentration of holes in the n-type material is calculated from Eq. (6.15),

$$p = \frac{n_i^2}{n} = \frac{(1.5 \times 10^{10})}{10^{17}} = 2.22 \times 10^3 \text{ cm}^{-3}.$$

We notice that there is a difference of 14 orders between n (10^{17} cm^{-3}) and p ($2.25 \times 10^3 \text{ cm}^{-3}$). It is now obvious why electrons in n-type materials are called the majority carriers and holes the minority carriers. We can calculate the change in the Fermi energy due to the doping. Let us assume that the reference energy level is the bottom of the conduction band, $E_C = 0 \text{ eV}$. Using Eq. (6.9) we calculate the Fermi energy in the intrinsic c-Si.

$$E_{Fi} = E_C - \frac{E_g}{2} + \frac{kT}{2} \ln \left(\frac{N_V}{N_C} \right) = -\frac{1.12}{2} + \frac{0.0258}{2} \ln \left(\frac{1.83 \times 10^{19}}{3.22 \times 10^{19}} \right) = -0.57 \text{ eV}.$$

The Fermi energy in the n-type doped c-Si wafer is calculated from Eq. (6.5a)

$$E_F = E_C + kT \ln \left(\frac{n}{N_C} \right) = 0.0258 \times \ln \left(\frac{10^{17}}{3.22 \times 10^{19}} \right) = -0.15 \text{ eV}.$$

We notice that the doping with P atoms has resulted in the shift of the Fermi energy towards the CB. Note that when $n > N_C$, $E_F > E_C$ and the Fermi energy lies in the CB.

by electrons and holes. The two basic transport mechanisms in a semiconductor are *drift* and *diffusion*.

6.5.1 Drift

Drift is charged-particle motion in response to an electric field. In an electric field the force acts on the charged particles in a semiconductor, which accelerates the positively charged holes in the direction of the electric field and the negatively charged electrons in the direction opposite to the electric field. Because of collisions with the thermally vibrating lattice atoms and ionised impurity atoms, the carrier acceleration is frequently disturbed. The resulting motion of electrons and holes can be described by average drift velocities \mathbf{v}_{dn} and \mathbf{v}_{dp} for electrons and holes, respectively. In case of low electric fields, the average drift velocities are directly proportional to the electric field as expressed by

$$\mathbf{v}_{dn} = -\mu_n \xi, \quad (6.21a)$$

$$\mathbf{v}_{dp} = \mu_p \xi. \quad (6.21b)$$

The proportionality factor is called mobility μ . It is a central parameter that characterises electron and hole transport due to drift. Although the electrons move in the opposite direction to the electric field, because the charge of an electron is negative the resulting electron

drift current is in the same direction as the electric field. This is illustrated in Fig. 6.8.

The electron and hole drift-current densities are then given as

$$\mathbf{J}_{n, \text{drift}} = -en\mathbf{v}_{dn} = en\mu_n \xi, \quad (6.22a)$$

$$\mathbf{J}_{p, \text{drift}} = ep\mathbf{v}_{dp} = ep\mu_p \xi. \quad (6.22b)$$

Combining Eqs. (6.22a) and (6.22b) leads to the total drift current,

$$\mathbf{J}_{\text{drift}} = e(p\mu_p + n\mu_n)\xi. \quad (6.23)$$

Mobility is a measure of how easily the charge particles can move through a semiconductor material. For example, for c-Si with a doping concentration N_D or N_A , respectively, at 300 K, the mobilities are

$$\mu_n \approx 1360 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1},$$

$$\mu_p \approx 450 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}.$$

As mentioned earlier, the motion of charged carriers is frequently disturbed by collisions. When the number of collisions increases, the mobility decreases. Increasing the temperature increases the collision rate of charged carriers with the vibrating lattice atoms, which results in a lower mobility. Increasing the doping concentration of donors or acceptors leads to more frequent collisions with the ionised dopant atoms, which results in a lower mobility as well. The dependence of mobility on

doping and temperature is in more detailed discussed in standard textbooks for semiconductor physics and devices, such as Reference [20].

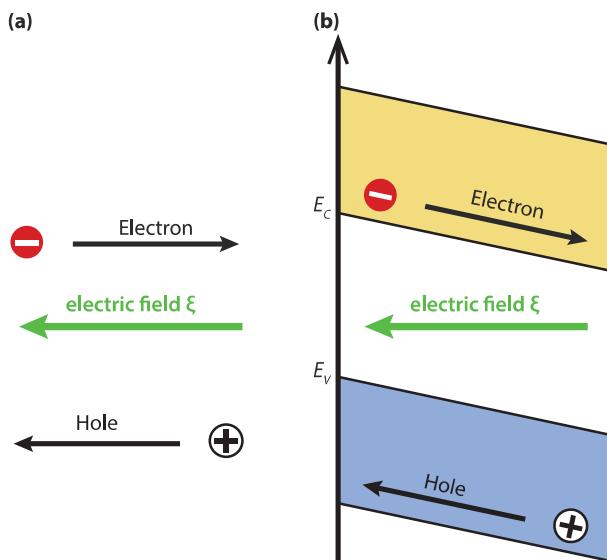


Figure 6.8: Visualisation of (a) the direction of carrier fluxes due to an electric field and (b) the corresponding band diagram.

6.5.2 Diffusion

Diffusion is a process whereby particles tend to spread out from regions of high particle concentration into regions of low particle concentration as a result of random thermal motion. The driving force of diffusion is a *gradient* in the particle concentration. In contrast to the drift transport mechanism, the particles need not be charged to be involved in the diffusion process. Currents resulting from diffusion are proportional to the gradient in particle concentration. For electrons and holes, they are given by

$$\mathbf{J}_{n,\text{diff}} = eD_n \nabla n, \quad (6.24a)$$

$$\mathbf{J}_{p,\text{diff}} = -eD_p \nabla p, \quad (6.24b)$$

Combining Eqs. (6.24a) and (6.24b) leads to the total diffusion current,

$$\mathbf{J}_{\text{diff}} = e(D_n \nabla n - D_p \nabla p). \quad (6.25)$$

The proportionality constants, D_n and D_p are called the electron and hole *diffusion coefficients*, respectively. The diffusion coefficients of electrons and holes are linked with the mobilities of the corresponding charge carriers

Example

To obtain some idea about values of diffusion coefficients, let us assume an *c-Si* wafer at room temperature, doped with donors, $N_D = 10^{14} \text{ cm}^{-3}$. According to Eq. (6.26),

$$D_N = \frac{kT}{e} \mu_n = 0.0258 \text{ V} \times 1360 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1} = 35 \text{ cm}^2 \text{s}^{-1}.$$

by the *Einstein relationship* that is given by

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{e}. \quad (6.26)$$

Figure 6.9 visualises the diffusion process as well as the resulting directions of particle fluxes and current.

Combining Eqs. (6.23) and (6.25) leads to the total current,

$$\begin{aligned} \mathbf{J} &= \mathbf{J}_{\text{drift}} + \mathbf{J}_{\text{diff}} \\ &= e(p\mu_p + n\mu_n)\xi + e(D_n \nabla n - D_p \nabla p). \end{aligned} \quad (6.27)$$

6.5.3 Continuity equations

As we have discussed in Section 4.5, charge is a conserved quantity. Now we will have a look at how the

continuity equations for the electrons and holes can be formulated. We have to take drift, diffusion, and recombination as well as generation processes into account; the latter are discussed in Chapter 7.

For electrons, the rate with that the concentration changes is given by

$$\frac{\partial n}{\partial t} = \left. \frac{\partial n}{\partial t} \right|_{\text{drift}} + \left. \frac{\partial n}{\partial t} \right|_{\text{diff}} + \left. \frac{\partial n}{\partial t} \right|_R + \left. \frac{\partial n}{\partial t} \right|_G, \quad (6.28a)$$

where R and G denotes recombination and generation, respectively. For holes, we obtain

$$\frac{\partial p}{\partial t} = \left. \frac{\partial p}{\partial t} \right|_{\text{drift}} + \left. \frac{\partial p}{\partial t} \right|_{\text{diff}} + \left. \frac{\partial p}{\partial t} \right|_R + \left. \frac{\partial p}{\partial t} \right|_G. \quad (6.28b)$$

In these equations $\partial n / \partial t$ ($\partial p / \partial t$) is the time rate change in the electron (hole) concentration

Without any generation or recombination, the number of electrons and holes is conserved, *i.e.* Eq. 4.34 can be applied. We thus obtain

$$\left. \frac{\partial n}{\partial t} \right|_{\text{drift}} + \left. \frac{\partial n}{\partial t} \right|_{\text{diff}} = \frac{1}{e} \nabla \times \mathbf{J}_n, \quad (6.29a)$$

$$\left. \frac{\partial p}{\partial t} \right|_{\text{drift}} + \left. \frac{\partial p}{\partial t} \right|_{\text{diff}} = -\frac{1}{e} \nabla \times \mathbf{J}_p, \quad (6.29b)$$

where \mathbf{J}_n (\mathbf{J}_p) is the electron (hole) current density.

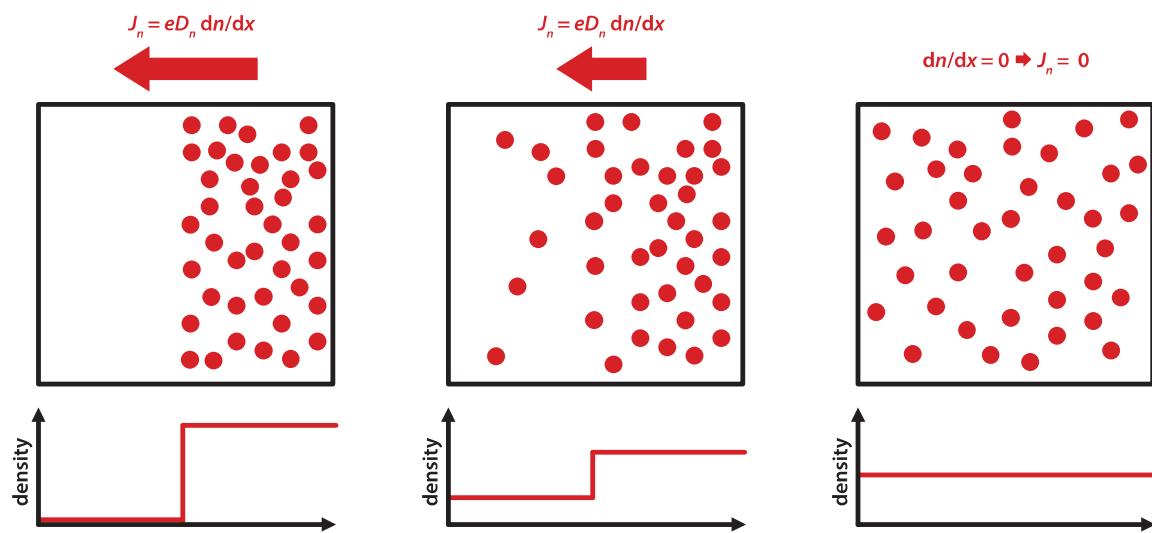


Figure 6.9: Visualisation of electron diffusion.

The equations can be written in a more compact form when introducing substitutions,

$$\left. \frac{\partial n}{\partial t} \right|_R = -R_n, \quad (6.30a)$$

$$\left. \frac{\partial p}{\partial t} \right|_R = -R_p, \quad (6.30b)$$

$$\left. \frac{\partial n}{\partial t} \right|_G = G_n, \quad (6.31a)$$

$$\left. \frac{\partial p}{\partial t} \right|_G = G_p. \quad (6.31b)$$

R_n (R_p) denotes the net thermal recombination-generation rate of electrons (holes), G_n (G_p) is the generation rate of electrons (holes) due to other processes, such as photogeneration. We will discuss these processes in detail in Chapter 7. Substituting Eqs. (6.29), (6.30) and (6.31) into Eq. (6.28) finally leads to

$$\frac{\partial n}{\partial t} = \frac{1}{e} \nabla \times \mathbf{J}_n - R_n + G_n, \quad (6.32a)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{e} \nabla \times \mathbf{J}_p - R_p + G_p. \quad (6.32b)$$

7

Generation and Recombination of Electron-Hole Pairs

7.1 Introduction

Assume that a piece of a semiconductor is illuminated by a light pulse, which leads to excitations of electrons from the valence band to the conduction band and consequently to the creation of holes in the valence band. This illumination will disturb the semiconductor from the state of thermal equilibrium. In the valence band an excess concentration of holes $p > p_0$ is present, where p_0 denotes the equilibrium concentration. Similarly, in the conduction band the electron concentration is larger than the equilibrium concentration, $n > n_0$, which means that *excess carriers are present*. It is clear that in this *non-equilibrium* state Eq. (6.7) is not fulfilled any-

more. Instead, now an inequality is valid,

$$np > n_i^2. \quad (7.1)$$

After the pulse stops, the excess electrons will *recombine* with holes, until the equilibrium state is reached again. Depending on the properties of the semiconductor different types of recombination can and will occur. In this chapter we will discuss the most important of these mechanisms.

The recombination rate strongly determines the performance of the solar cells. On the one hand, it will reduce the current that can be collected and hence utilised from the solar cell. However, the photo-generation rate often is several magnitudes higher than the recombination rate, such that the effect of recombination on

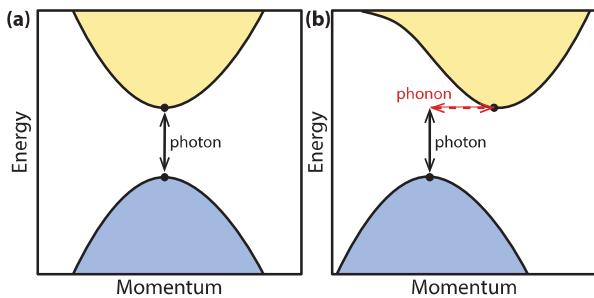


Figure 7.1: Illustrating the dispersion diagram of (a) an direct bandgap semiconductor and (b) a indirect bandgap semiconductor.

the solar cell current is negligible. On the other hand, the recombination rate strongly determines the saturation current density; the more recombination, the higher the saturation current density. As we will see in Chapter 8, a high saturation current density has a detrimental effect on the solar cell voltage, and hence on the energy conversion efficiency.

Before we can actually start with the treatment on the different generation and recombination mechanisms, we mention that we have to distinguish between *direct* and *indirect* semiconductors. Figure 7.1 shows the energy-momentum space of the electrons, which also is called the *electronic dispersion diagram*. On the vertical axis the

energy state in the electronic bands is plotted. On the horizontal axis the momentum of the charge carrier is shown. This momentum is also called the *crystal momentum*, it is also related to the *wave vector \mathbf{k}* of the electron. It is important to realise that the position of the valence and conduction band may differ in different directions of the lattice coordination. We can understand this by realising that the crystal can look very different if we look at it from different directions. Hence, also the energy levels in that electrons and holes can propagate across the crystal depend on the direction. The dispersion diagram of silicon is discussed in more detail in Chapter 12.

For a direct band gap material the highest point of the valence band is vertically aligned with the lowest point of the conduction band, as shown in Fig. 7.1 (a). This means that exciting an electron from the valence to the conduction band requires only the energy provided by a photon without any additional momentum transfer. In contrast, for an indirect band gap the highest point of the valence band is not aligned with the lowest point of the conduction band, as shown in Fig. 7.1 (b). Therefore, exciting an electron from the valence to conduction band requires energy provided by a photon and momentum, provided by a phonon.

Therefore, the excitation of an electron induced by photon absorption is more likely to happen for direct band gap materials than for indirect band gap materials

and hence the absorption coefficient for direct band gap materials is significantly higher than for indirect band gap materials. The same principle makes the reverse process of radiative recombination more likely to happen in a direct band gap material. In an indirect band gap material additional momentum is required to make the electron and hole recombine.

Crystalline silicon is an indirect band gap material. In such a material, the radiative recombination is inefficient and recombination will be dominated by the Auger recombination mechanism discussed in Section 7.4. For direct band gap materials such as gallium arsenide under moderate illumination conditions, radiative recombination will be the dominant loss mechanism of charge carriers. For very high illumination conditions, Auger recombination starts to play a role as well. Gallium arsenide based solar cells are discussed in Chapter 13.

7.2 Bandgap-to-bandgap processes

Generation and recombination processes that happen from bandgap to bandgap are also called *direct* generation and recombination. They are much more likely to happen in direct bandgap materials, as there no change in momentum is required for an electron that is excited

into the conduction band. These processes most usually are *radiative*, which means that a photon is absorbed when an electron-hole pair is created, and a photon is emitted if electron-hole pairs recombine directly.

We will use this section also to introduce general concepts important for generation and recombination, such as the *minority carrier lifetime*.

7.2.1 Radiative Generation

When light penetrates into a material it will be (partially) absorbed as it propagates through the material. If the photon energy is higher than the bandgap energy of the semiconductor, it is sufficient to break bonds and to excite a valence electron into the conduction band and leaving a hole behind in the valance band. This process is illustrated in Fig. 7.2 (a).

The absorption profile in the material depends on the absorption coefficient of the material, which is wavelength dependent. The most frequent approach to calculate the absorption profile of photons in semiconductor devices is by using *Lambert-Beer's law* that we already introduced in Eq. (4.25), where it was formulated for the decay of intensity. Here we formulate it with the *photon flux density* $\Phi(x, \lambda)$, which decreases exponentially with the distance x that it travelled through

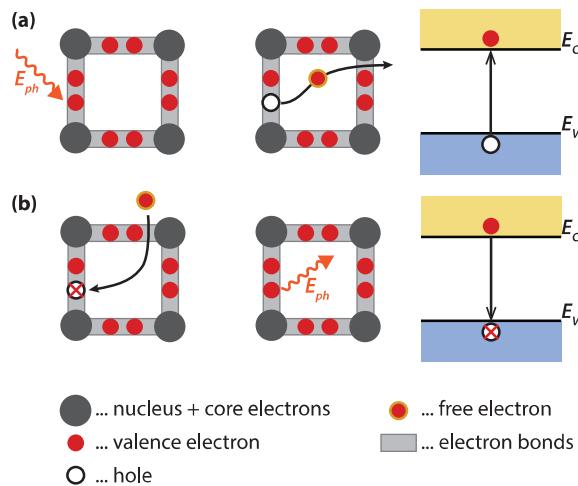


Figure 7.2: Visualisation of bandgap-to-bandgap (a) generation and (b) recombination processes using the bonding model and the energy band diagram.

the absorber,

$$\Phi(x, \lambda) = \Phi_0(\lambda) \exp[-\alpha(\lambda)x], \quad (7.2)$$

where Φ_0 is the incident photon flux density and $\alpha(\lambda)$ is the absorption coefficient. The photon flux density is defined as the number of photons per unit area, unit time and unit wavelength. It is related to the spectral power density $P(\lambda)$ associated with the solar radiation via

$$\Phi_0(\lambda) = P(\lambda) \frac{\lambda}{hc}. \quad (7.3)$$

The spectral generation rate $g_{sp}(x, \lambda)$, which is the number of electron-hole pairs generated at a depth x in the film per second unit volume and unit wavelength, by photons of wavelength λ , is calculated according to

$$g_{sp}(x, \lambda) = \eta_g \Phi_0(\lambda) \alpha(\lambda) \exp[-\alpha(\lambda)x], \quad (7.4)$$

where we assumed zero reflection. η_g is the generation quantum efficiency, usually assumed equal unity. This assumption means that every photon generates one and only one electron-hole pair. The optical generation rate $G_L(x)$ is calculated from the spectral generation rate by integrating over a desired wavelength spectrum,

$$G_L(x) = \int_{\lambda_1}^{\lambda_2} g_{sp}(x, \lambda) d\lambda. \quad (7.5)$$

It has the unit $[G_L] = \text{cm}^{-3}\text{s}^{-1}$. The optical generation rate is related to the absorption profile $A(x)$ in the film

via

$$G_L(x) = \eta_g A(x). \quad (7.6)$$

Hence,

$$A(x) = \int_{\lambda_1}^{\lambda_2} \Phi_0(\lambda) \alpha(\lambda) \exp[-\alpha(\lambda)x] d\lambda. \quad (7.7)$$

Because of the photogeneration excess electrons and holes will be created. The rates of the generated concentration of generated excess electrons and holes per second are the same, therefore we can write

$$\frac{\partial n}{\partial t} \Big|_{\text{light}} = \frac{\partial p}{\partial t} \Big|_{\text{light}} = G_L. \quad (7.8)$$

7.2.2 Direct recombination

We now will discuss direct recombination, that mainly occurs in direct bandgap semiconductors, such as gallium arsenide. It is illustrated in Fig. 7.2 (b). In this section we roughly follow the derivation by Sze [26].

Let us first look at the situation at *thermal equilibrium*. If the temperature is higher than 0 K, the crystal lattice is vibrating. This vibrational energy will be sufficient to break bonds from time to time, which leads to the generation of electron-hole pairs at a generation rate G_{th} , where the *th* stands for *thermal*. As we are in thermal

equilibrium, the expression

$$np = n_i^2 \quad (7.9)$$

must be valid. Hence, also recombination must take place at the same rate as the generation,

$$R_{th} = G_{th}. \quad (7.10)$$

We may assume that the direct recombination rate is proportional to the concentration of electrons in the conduction band and to the concentration of the available holes in the valence band,

$$R^* = \beta np, \quad (7.11)$$

where β is a proportionality factor. For the thermal recombination we have

$$R_{th} = \beta n_0 p_0. \quad (7.12)$$

We now look at a situation where the semiconductor is illuminated such that a constant generation rate G_L is present throughout the volume of the semiconductor. In this situation excess electrons and holes are created. As the electron and hole concentrations are increasing, also the recombination rate will increase according to Eq. (7.11). At some point, the generation and recombination rates will be the same, such that n and p do not change any more. This situation is called the *steady state*

Example

Let us calculate the total absorption in a $d = 300 \mu\text{m}$ thick c-Si wafer for light with a wavelength of 500 nm. The optical constants of c-Si at this wavelength are: refractive index is $n = 4.293$, extinction coefficient $k = 0.045$ and absorption coefficient $\alpha = 1.11 \times 10^4 \text{ cm}^{-1}$. The incident irradiance is 1000 W m^{-2} .

First we calculate the photon flux density at 500 nm corresponding to the irradiance of 1000 W m^{-2} . Using Eq. (7.3) we obtain

$$\Phi_0(\lambda) = P(\lambda) \frac{\lambda}{hc} = \frac{1000 \text{ W m}^{-2} \times 500 \times 10^{-9} \text{ m}}{6.625 \times 10^{-34} \text{ Js} \times 2.998 \times 10^8 \text{ ms}^{-1}} = 2.5 \times 10^{21} \text{ m}^{-2} \text{s}^{-1}.$$

Using Eq. (4.15) we calculate how many incident photons are reflected from the surface,

$$R = \left| \frac{\tilde{n}_0 - \tilde{n}_1}{\tilde{n}_0 + \tilde{n}_1} \right|^2 = \left| \frac{1.0 - 0.0i - (4.293 - 0.045i)}{1.0 - 0.0i + 4.293 - 0.045i} \right|^2 = 0.38.$$

Using Lambert-Beer's law [Eq. (7.2)] we calculate the photon flux density at the backside of the wafer, i.e. at $300 \mu\text{m}$ distance from the surface. We take the reflected light into account by adapting Eq. (7.2),

$$\Phi(d) = \Phi_0(1 - R) \exp[-\alpha(\lambda)d] = 2.5 \times 10^{21} \text{ m}^{-2} \text{s}^{-1} \times (1 - 0.38) \exp(-1.11 \times 10^6 \text{ m}^{-1} \times 300 \times 10^{-6} \text{ m}) \approx 0.$$

The total absorption in the wafer is the difference between the photon flux density at the surface after reflection and the photon flux density at the back of the wafer,

$$A = \Delta\Phi = \Phi_0(1 - R) - \Phi(d) = 2.5 \times 10^{21} \text{ m}^{-2} \text{s}^{-1} \times (1 - 0.38) - 0 = 1.55 \times 10^{21} \text{ m}^{-2} \text{s}^{-1}.$$

When we assume that all the absorbed photons generate one electron-hole pair ($\eta_g = 1$), we can calculate the photocurrent density corresponding to the absorbed photon flux,

$$J_{ph} = eA = 1.602 \times 10^{-19} \text{ C} \times 1.55 \times 10^{21} \text{ m}^{-2} \text{s}^{-1} = 248.31 \text{ C m}^{-2} \text{s}^{-1} = 248.31 \text{ A m}^{-2}.$$

situation. The total recombination and generation rates are given by

$$R^* = \beta np = \beta(n_0 + \Delta n)(p_0 + \Delta p), \quad (7.13)$$

$$G = G_{th} + G_L, \quad (7.14)$$

where n_0 and p_0 are the equilibrium concentrations. Δn and Δp are the excess carrier concentrations that are given by

$$\Delta n = n - n_0, \quad (7.15a)$$

$$\Delta p = p - p_0. \quad (7.15b)$$

In steady state R^* and G are equal, hence

$$G_L = R^* - G = R_d, \quad (7.16)$$

where R_d denotes the *net* radiative recombination rate. By substituting Eqs. (7.11) and (7.12) into Eq. (7.16), we obtain

$$G_L = R_d = \beta(np - n_0 p_0). \quad (7.17)$$

We now assume the semiconductor to be *n*-type and *low level injection*, which means that $\Delta n \ll n$ and $p \ll n$. Under these assumptions the recombination rate becomes

$$R_d \approx \beta n_0(p - p_0) = \frac{p - p_0}{\tau_{pd}}, \quad (7.18)$$

where

$$\tau_{pd} = \frac{1}{\beta n_0} \quad (7.19)$$

is the *lifetime of the minority holes* in the *n*-type semiconductor. Clearly, if no excess carriers are present, $R_d = 0$. The excess carrier concentration is given as the sum of the generation rate and the lifetime,

$$p - p_0 = G_L \tau_{pd}. \quad (7.20)$$

To understand the meaning of the *lifetime*, we consider a situation where the light and hence generation at the rate G_L is suddenly shut off. Without loss of generality we may assume that the light is shut off at the instant $t = 0$. As no generation is present any more, the excess carrier concentration will change according to the differential equation

$$\frac{dp}{dt} = -\frac{p(t) - p_0}{\tau_{pd}}. \quad (7.21)$$

If we solve this equation with the boundary condition $p(t = 0) = p_0 + G_L \tau_{pd}$, we find

$$p(t) = p_0 + G_L \tau_p \exp\left(-\frac{t}{\tau_{pd}}\right). \quad (7.22)$$

We therefore see that the minority carrier lifetime is the time constant at which an excess carrier concentration decays exponentially, if no external generation is taking place any more.

For a *p*-type semiconductor at low-level injection ($\Delta p \ll$

p and $n \ll p$ we find similar expressions,

$$R_d \approx \beta p_0(n - n_0) = \frac{n - n_0}{\tau_{nd}}, \quad (7.23)$$

where the lifetime of the electrons is given by

$$\tau_{nd} = \frac{1}{\beta p_0}. \quad (7.24)$$

Let us assume that in a semiconductor several recombination mechanisms are present, with recombination rates R_1, R_2, \dots . The total recombination rate then is given by

$$R_{\text{tot}} = R_1 + R_2 + \dots \quad (7.25)$$

If we have an n -type semiconductor under low injection we may assume

$$R_{\text{tot}} = \frac{p - p_0}{\tau_{p1}} + \frac{p - p_0}{\tau_{p2}} + \dots = \frac{p - p_0}{\tau_{p,\text{tot}}}. \quad (7.26)$$

We therefore have an overall lifetime that is related to the lifetimes of the different processes via

$$\frac{1}{\tau_{p,\text{tot}}} = \frac{1}{\tau_{p1}} + \frac{1}{\tau_{p2}} + \dots \quad (7.27)$$

The more recombination mechanisms are present, the shorter the overall lifetime of the excess minority carriers.

The last aspect that we want to discuss in this section is a situation where the excess of generated carriers is not uniform throughout the semiconductor. In this case diffusion of the excess carriers takes place. The excess carriers will diffuse in the semiconductor till they recombine. We define the average distance that the minority carriers can diffuse among majority carries before being annihilated as *minority-carrier diffusion length*. They are defined as:

$$L_n = \sqrt{D_n \tau_n} \quad \begin{array}{l} \text{for electrons in a} \\ p\text{-type material,} \end{array} \quad (7.28a)$$

$$L_p = \sqrt{D_p \tau_p} \quad \begin{array}{l} \text{for holes in an} \\ n\text{-type material,} \end{array} \quad (7.28b)$$

where D_n and D_p are the diffusion coefficients as introduced in Section 6.5.

7.3 Shockley-Read-Hall Recombination

In the *Shockley-Read-Hall* (SRH) recombination process, which is illustrated in Fig. 7.3, the recombination of electrons and holes does not occur directly from bandgap to bandgap. It is facilitated by a *impurity atom* or *lattice defects*. Their concentration is usually small compared to the acceptor or donor concentrations. These recombination centres introduce allowed

Example

To get an idea about the diffusion lengths, let us assume room temperature, mobility of electrons in a p-type c-Si wafer to be $\mu_n \approx 1250 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$, which corresponds to doping of $N_A = 10^{14} \text{ cm}^{-3}$, and $\tau_n = 10^{-6} \text{ s}$. For the given conditions, the electron diffusion length in the p-type c-Si can be calculated from Eq. (7.28a).

$$L_n = \sqrt{D_N \tau_n} = \sqrt{\frac{kT}{e}} \mu_n \tau_n \\ = \sqrt{0.0258 \text{ V} \times 1250 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1} \times 10^{-6} \text{ s}} = 57 \text{ nm.}$$

energy levels (E_T) within the forbidden gap, so-called *trap states*. An electron can be *trapped* at such a defect and consequently recombines with a hole that is attracted by the trapped electron. Though this process seems to be less likely than the direct thermal recombination, it is the dominant recombination-generation process in semiconductors at most operational conditions. The process is typically non-radiative and the excess energy is dissipated into the lattice in form of heat. The name is a reverence to William Shockley, William T. Read and Robert N. Hall, who published the theory of this recombination mechanism in 1952 [27, 28].

We distinguish between two kinds of traps: first, *donor-type* traps that are neutral when they contain an elec-

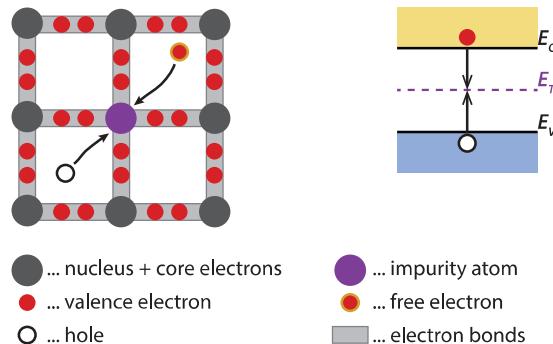


Figure 7.3: Visualisation of Shockley-Read-Hall recombination using the bonding model and the energy band diagram.

tron and positively charged when they do not contain an electron. Secondly, *acceptor-type* traps that are negatively charged when they contain an electron and neutral when they do not contain an electron.

The SRH statistics is based on four processes that are involved in recombination in a single-electron trap:

- r_1 : capture of an electron from the conduction band,
- r_2 : emission of an electron to the conduction band,
- r_3 : capture of a hole from the valence band, and

- r_4 : emission of a hole to the valence band.

These processes are illustrated in Fig. 7.4 for both donor- and acceptor-type traps. The electron and hole capture rates are proportional to the free carrier concentration, n or p , respectively, the thermal velocity v_{th} , the trap density N_T , the trap occupancy by electron, f , or holes, $1 - f$, and the electron and hole capture cross-section of the traps, σ_n and σ_p . The emission rates are proportional to the trap density and the electron or hole occupancy of the traps, as well as the emission coefficient for electron or holes, e_n or e_p , respectively. All processes and their rates are listed in Table 7.1.

The *thermal velocity* is the average velocity of the electrons and holes due to thermal movement. It can be obtained by setting the thermal and the kinetic energy equal. Since electrons and holes have three degrees of freedom, we obtain

$$\frac{1}{2}m_n^*v_{th,n}^2 = \frac{3}{2}kT \quad \frac{1}{2}m_p^*v_{th,p}^2 = \frac{3}{2}kT, \quad (7.29)$$

where m_n^* and m_p^* are the effective masses of the electrons and holes, respectively. For electrons in silicon and gallium arsenide, the thermal velocity is about 10^7 cm/s. For the following derivation we assume that v_{th} is the same for electrons and holes.

The electron capture cross section σ_n describes the effectiveness of the trap state to capture an electron. It is a measure of how close an electron has to come to the

Table 7.1: Processes associated with single-electron trapping and their rates.

Donor-like traps		
Process	Rate	
r_1 electron capture	$nv_{th}\sigma_n^+ N_T(1 - f)$	
r_2 electron emission	$e_n^- N_T f$	
r_3 hole capture	$pv_{th}\sigma_p^0 N_T f$	
r_4 hole emission	$e_p^+ N_T(1 - f)$	
Acceptor-like traps		
Process	Rate	
r_1 electron capture	$nv_{th}\sigma_n^0 N_T(1 - f)$	
r_2 electron emission	$e_n^- N_T f$	
r_3 hole capture	$pv_{th}\sigma_p^- N_T f$	
r_4 hole emission	$e_p^0 N_T(1 - f)$	

trap to be captured. It has the unit of area, cm². Similarly, σ_p describes the effectiveness of a trap state to capture a hole.

The derivation of the recombination efficacy, below, is valid for both donor- and acceptor-like traps. Therefore, the capture cross-section and emission coefficients in Table 7.1 are generalised by omitting their charge state. Depending on the type of trap considered the appropriate cross-sections and emission coefficients need to be substituted.

According to the Fermi-Dirac statistics the carrier dis-

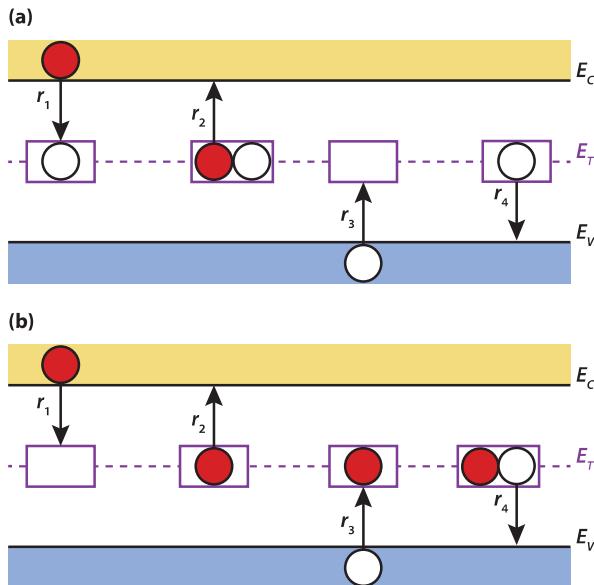


Figure 7.4: Schematic illustration of the processes involved with SRH recombination in a single-electron trap state for (a) a donor-type and (b) an acceptor-type trap.

tribution in a semiconductor in thermal equilibrium depends on the chemical potential of the carriers, which is referred to as the Fermi level E_F . When the device is illuminated or a bias voltage is applied, the carriers on either side of the mobility gap are no longer in equilibrium. Yet, they do relax to a state of quasi-equilibrium with their respective bands. This leads to the definition of the quasi-Fermi levels for electrons and holes, E_{Fn} and E_{Fp} , which determine the carrier concentrations under non-equilibrium conditions. Note that in thermal equilibrium $E_{Fn} = E_{Fp} = E_F$. General expressions for the free electron and hole concentrations n and p , respectively, both under equilibrium and non-equilibrium conditions, read

$$n = N_C \exp\left(\frac{E_{Fn} - E_C}{kT}\right), \quad (7.30a)$$

$$p = N_V \exp\left(\frac{E_V - E_{Fp}}{kT}\right), \quad (7.30b)$$

where E_C (E_V) is the conduction (valence) band edge and N_C (N_V) the effective density of states in the conduction (valence) band, respectively. According to the Fermi-Dirac statistics the occupation function in thermal equilibrium is given by

$$f(E_T) = \frac{1}{1 + \exp\left(\frac{E_T - E_F}{kT}\right)}, \quad (7.31)$$

where E_T is the trap energy.

In thermal equilibrium no net recombination occurs, such that $r_1 = r_2$ and $r_3 = r_4$. Substituting the rate equations from Table 7.1 and Eqs. (7.30–7.31) yields the following expressions for the emission coefficients:

$$e_n = v_{th}\sigma_n N_C \exp\left(\frac{E_T - E_C}{kT}\right), \quad (7.32a)$$

$$e_p = v_{th}\sigma_p N_V \exp\left(\frac{E_V - E_T}{kT}\right). \quad (7.32b)$$

By substituting N_C and N_V by the intrinsic carrier concentration n_i times an exponential according to Eq. (6.8), we obtain

$$e_n = v_{th}\sigma_n n_i \exp\left(\frac{E_T - E_{Fi}}{kT}\right), \quad (7.33a)$$

$$e_p = v_{th}\sigma_p n_i \exp\left(\frac{E_{Fi} - E_T}{kT}\right). \quad (7.33b)$$

Consider now a non-equilibrium steady-state situation. Under equilibrium, it is assumed that the emission coefficients are approximately equal to the emission coefficients. As recombination involves exactly one electron and one hole, at steady state the rate at which the electrons leave the conduction band equals the rate at which the holes leave the valence band. The recombination rate is therefore equal to

$$R_{SRH} = \frac{dn}{dt} = \frac{dp}{dt} = r_1 - r_2 = r_3 - r_4. \quad (7.34)$$

By substituting the rates from Table 7.1, the expression for the steady-state occupation function can be determined to be

$$f(E_T) = \frac{v_{th}\sigma_n n + e_p}{V_{th}\sigma_n n + v_{th}\sigma_p p + e_n + e_p}. \quad (7.35)$$

Finally, the recombination rate is obtained by substituting Eq. (7.35) in the rate equations as in Eq. (7.34), yielding

$$R_{SRH} = v_{th}^2 \sigma_p \sigma_n N_T \frac{np - n_i^2}{v_{th}\sigma_n n + v_{th}\sigma_p p + e_n + e_p}, \quad (7.36)$$

where n_i is the intrinsic carrier concentration as in Eq. (6.7).

We can simplify the general expression of Eq. (7.36) when we assume the same capture cross sections for electrons and holes, $\sigma_n = \sigma_p \equiv \sigma_0$, which yields

$$e_n + e_p = 2v_{th}\sigma_0 n_i \cosh\left(\frac{E_T - E_{Fi}}{kT}\right), \quad (7.37)$$

and hence

$$R_{SRH} = v_{th}\sigma N_T \frac{np - n_i^2}{n + p + 2n_i \cosh\left(\frac{E_T - E_{Fi}}{kT}\right)}. \quad (7.38)$$

We now look at an *n*-type semiconductor at low injection rate, meaning that the amount concentration of excess electrons is small compared to the total electron

concentration, $n \approx n_0$, where n_0 is the electron concentration under thermal equilibrium. Further, we may assume $n \gg p$. By applying these assumptions to Eq. 7.38 we obtain

$$\begin{aligned} R_{\text{SRH}} &= v_{th}\sigma N_T \frac{p - p_0}{1 + 2\frac{n_i}{n_0} \cosh\left(\frac{E_T - E_{Fi}}{kT}\right)} \\ &= c_p N_T (p - p_0) = \frac{p - p_0}{\tau_{p,\text{SRH}}}, \end{aligned} \quad (7.39)$$

where c_p is called the *hole capture coefficient*. $\tau_{p,\text{SRH}}$ is the lifetime of holes in an n -type semiconductor.

In a similar manner, we can derive for a p -type semiconductor at a low injection rate

$$\begin{aligned} R_{\text{SRH}} &= v_{th}\sigma N_T \frac{n - n_0}{1 + 2\frac{n_i}{p_0} \cosh\left(\frac{E_T - E_{Fi}}{kT}\right)} \\ &= c_n N_T (n - n_0) = \frac{n - n_0}{\tau_{n,\text{SRH}}}, \end{aligned} \quad (7.40)$$

with the electron capture coefficient c_n and the electron lifetime $\tau_{n,\text{SRH}}$.

We see that the lifetime is related to the capture coefficients via

$$\tau_{p,\text{SRH}} = \frac{1}{c_p N_T} \quad \text{and} \quad \tau_{n,\text{SRH}} = \frac{1}{c_n N_T}. \quad (7.41)$$

The lifetime of the minority carriers due to Shockley-Read-Hall recombination therefore is indirectly propor-

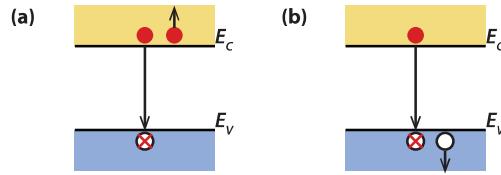


Figure 7.5: Schematic illustration of Auger recombination with (a) two electrons and (b) two holes involved.

tional to the trap density N_T . Hence, for a good semiconductor device it is crucial to keep N_T low.

The values of the minority-carrier lifetimes can vary a lot. When the trap concentration in c-Si is very low, τ_n (τ_p) can achieve values around 1 ms. On the other hand, the intentional introduction of gold atoms into Si, which introduce efficient traps into Si, can decrease τ_n (τ_p) to values around 1 ns. Typical minority-carrier lifetimes in most c-Si devices are usually around 1 μ s. For an efficient collection of photo-generated carriers in c-Si solar cells the minority carrier lifetimes should be in range of tens of milliseconds.

7.4 Auger Recombination

We already mentioned that direct recombination is not or very limited possible for indirect semiconductors, as

both transfer in energy *and* momentum must occur for an electron in the conduction band to recombine with a hole in the valence band. In indirect semiconductors, *Auger recombination* becomes important. In difference to direct and SRH recombination, which involve two particles, *i.e.* an electron and a hole, Auger recombination is a *three particle process*, as illustrated in Fig. 7.5.

In Auger recombination, momentum and energy of the recombining hole and electron is conserved by transferring energy and momentum to an another electron (or hole). If the third particle is an electron, it is excited into higher levels in the electronic band. This excited electron relaxes again, such that the energy is transferred vibrational energy of the lattice, or *phonon modes*, and finally heat. Similarly, if the third particle is a hole, it is excited into deeper levels of the valence band, from where it rises back to the valence band edge via transferring its energy to phonon modes.

As Auger recombination is a three-particle process, the Auger recombination rate R_{Aug} strongly depends on the charge carrier densities for the electrons n and holes p . The recombination rates for electron-electron-hole (eeh) and electron-hole-hole (ehh) processes are given by

$$R_{\text{eeh}} = C_n n^2 p, \quad (7.42\text{a})$$

$$R_{\text{ehh}} = C_p n p^2, \quad (7.42\text{b})$$

respectively, where C_n and C_p are the proportionality constants that are strongly dependent on the temper-

ature [26]. R_{eeh} is dominant when the electrons are the majority charge carriers, while R_{ehh} is dominant when the holes are the majority charge carriers. Adding them leads to the total Auger recombination rate,

$$R_{\text{Aug}} = R_{\text{eeh}} + R_{\text{ehh}} = C_n n^2 p + C_p n p^2, \quad (7.43)$$

In strongly doped *n*-type silicon with a donor concentration N_D under low injection we can assume that $n \approx N_D$ and hence that the eeh process is dominant. We then can write

$$R_{\text{eeh}} = C_n N_D^2 p. \quad (7.44)$$

Hence, the lifetime can be approximated with

$$\tau_{\text{eeh}} = \frac{1}{C_n N_D^2}. \quad (7.45)$$

Similarly, for strongly doped *p*-type silicon with acceptor concentration N_A we may assume $p \approx N_A$ and hence the ehh process being dominant,

$$R_{\text{ehh}} = C_p N_A^2 n. \quad (7.46)$$

The lifetime then is

$$\tau_{\text{ehh}} = \frac{1}{C_p N_A^2}. \quad (7.47)$$

As the Auger recombination under these conditions is proportional to the square of the doping levels, it becomes the more important the higher the doping is.

For an *n*-type semiconductor under high-injection conditions, as it occurs for example for concentrator solar cells with very high irradiance values the lifetime can be approximated with

$$\tau_{\text{Auger, hi}} = \frac{1}{(C_n + C_p)\Delta n^2} = \frac{1}{C_a \Delta n^2}, \quad (7.48)$$

where $\Delta n = n - n_0 = p - p_0$ is the excess carrier density [29]. Under these circumstances Auger recombination also might become important for direct bandgap materials such as gallium arsenide.

7.5 Carrier concentration in non-equilibrium

When a semiconductor is illuminated additional electrons and holes are generated in the material by the absorption of photons. The *photo-generated carriers* interact with the semiconductor lattice. The extra energy that the electron-hole pairs receive from the photons with energies larger than the band gap of the semiconductor is released into the lattice in form of heat. After this so called *thermalisation* process, which is very fast and takes approximately 10^{-12} s, the carrier concentrations achieve a steady-state. In this non-equilibrium state the electron and hole concentrations are different than those in the equilibrium state. In non-equilibrium states

two Fermi distributions are used to describe the electron and hole concentrations. One Fermi distribution with the *quasi-Fermi energy for electrons*, E_{FC} , describes the occupation of states in the conduction band with electrons. Another Fermi distribution with the *quasi-Fermi energy for holes*, E_{FV} , describes the occupation of states in the valence band with electrons, and therefore determines also the concentration of holes. Using the band diagram with the quasi-Fermi levels the process of creation of electron-hole pairs and their subsequent thermalisation that describe the carrier concentration under illumination is illustrated in Fig. 7.6. The difference between the quasi-Fermi levels is the electrochemical energy, μ_{eh} , of the generated electron-hole pairs which represents the measure for the conversion efficiency of solar radiation.

The density of electrons and holes under non-equilibrium conditions is described by

$$n = N_C \exp\left(\frac{E_{FC} - E_C}{kT}\right), \quad (7.49a)$$

$$p = N_V \exp\left(\frac{E_V - E_{FV}}{kT}\right), \quad (7.49b)$$

It then follows that under non-equilibrium conditions

$$\begin{aligned} np &= N_C N_V \exp\left(\frac{E_V - E_C}{kT}\right) \exp\left(\frac{E_{FC} - E_{FV}}{kT}\right) \\ &= n_i^2 \exp\left(\frac{E_{FC} - E_{FV}}{kT}\right). \end{aligned} \quad (7.50)$$

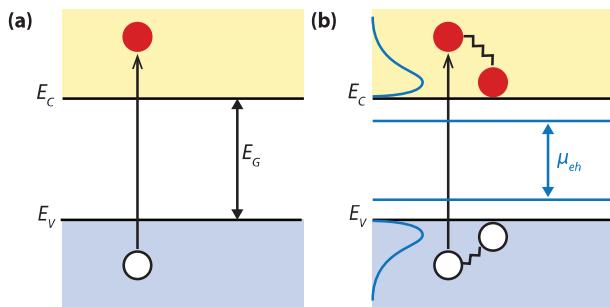


Figure 7.6: Thermalisation of photo-generated electron-hole pairs resulting in non-equilibrium charge-carrier concentrations described by the quasi-Fermi levels.

By using the quasi-Fermi level formalism for describing the concentration of charge carriers in non-equilibrium conditions the electron and hole current densities inside a semiconductor, \mathbf{J}_N and \mathbf{J}_P , can be expressed with

$$\mathbf{J}_N = n\mu_n \nabla E_{FC}, \quad (7.51a)$$

$$\mathbf{J}_P = p\mu_p \nabla E_{FV}. \quad (7.51b)$$

One can notice from Eqs. (7.51) that when a quasi-Fermi level varies with position,

$$\frac{dE_{FC}}{dx} \neq 0 \quad \text{or} \quad \frac{dE_{FV}}{dx} \neq 0,$$

the current is flowing inside the semiconductor. By checking the position dependence of the quasi-Fermi

levels in an energy band diagram, one can easily determine whether current flows inside the semiconductor.

8

Semiconductor Junctions

Almost all solar cells contain junctions between (different) materials of different doping. Since these junctions are crucial to the operation of the solar cell, we will discuss their physics in this chapter.

A *p-n* junction fabricated in the same semiconductor material such as c-Si is an example of an *p-n homojunction*. There are also other types of junctions: A *p-n* junction that is formed by two chemically different semiconductors is called a *p-n heterojunction*. In a *p-i-n* junctions, the region of the internal electric field is extended by inserting an intrinsic, *i*, layer between the *p*-type and the *n*-type layers. The *i*-layer behaves like a capacitor and it stretches the electric field formed by the *p-n* junction across itself. Another type of the junction is a junction between a *metal* and a *semiconductor*, a so-called

MS junction. The Schottky barrier formed at the metal-semiconductor interface is a typical example of the MS junction.

8.1 *p-n* homojunctions

8.1.1 Formation of a space-charge region in the *p-n* junction

Figure 8.1 shows schematically isolated pieces of a *p*-type and an *n*-type semiconductor and their corresponding band diagrams. In both isolated pieces the charge neutrality is maintained. In the *n*-type semicon-

ductor the large concentration of negatively-charged free electrons is compensated by positively-charged ionised donor atoms. In the *p*-type semiconductor holes are the majority carriers and the positive charge of holes is compensated by negatively-charged ionised acceptor atoms. For the isolated *n*-type semiconductor we can write

$$n = n_{n0} \approx N_D, \quad (8.1a)$$

$$p = p_{n0} \approx n_i^2 / N_D. \quad (8.1b)$$

For the isolated *p*-type semiconductor we have

$$p = p_{p0} \approx N_A, \quad (8.2a)$$

$$n = n_{p0} \approx n_i^2 / N_A. \quad (8.2b)$$

When a *p*-type and an *n*-type semiconductor are brought together, a very large difference in electron concentration between *n*- and *p*-type regions causes a diffusion current of electrons from the *n*-type material across the *metallurgical junction* into the *p*-type material. The term "metallurgical junction" denotes the interface between the *n*- and *p*-type regions. Similarly, the difference in hole concentration causes a diffusion current of holes from the *p*- to the *n*-type material. Due to this diffusion process the region close to the metallurgical junction becomes almost completely depleted of mobile charge carriers. The gradual depletion of the charge carriers gives rise to a space charge created by the charge

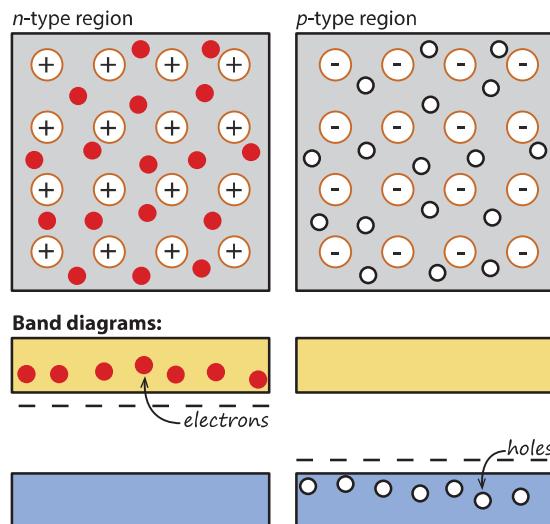


Figure 8.1: Schematic representation of an isolated *n*-type and *p*-type semiconductor and corresponding band diagrams.

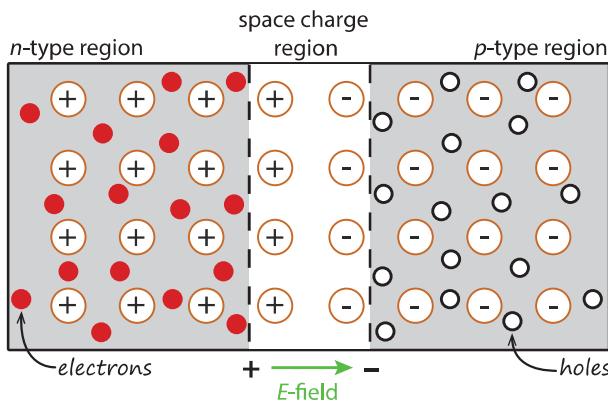


Figure 8.2: Formation of a space-charge region, when *n*-type and *p*-type semiconductors are brought together to form a junction. The coloured part represents the space-charge region.

of the ionised donor and acceptor atoms that is not compensated by the mobile charges any more. This region of the space charge is called the *space-charge region* or *depleted region* and is schematically illustrated in Fig. 8.2. Regions outside the depletion region, in which the charge neutrality is conserved, are denoted as the quasi-neutral regions.

The space charge around the metallurgical junction results in the formation of an internal electric field which forces the charge carriers to move in the opposite dir-

ection than the concentration gradient. The diffusion currents continue to flow until the forces acting on the charge carriers, namely the concentration gradient and the internal electrical field, compensate each other. The driving force for the charge transport does not exist any more and no net current flows through the *p-n* junction.

8.1.2 The *p-n* junction under equilibrium

The *p-n* junction represents a system of charged particles in diffusive equilibrium in which the electro-chemical potential is constant and independent of position. The electro-chemical potential describes an average energy of electrons and is represented by the Fermi energy. It means that under equilibrium conditions the Fermi level has constant position in the band diagram of the *p-n* junction. Figure 8.3 shows the energy-band diagram of a *p-n* junction under equilibrium. The distance between the Fermi level and the valence and/or conduction bands does not change in the quasi-neutral regions and is the same as in the isolated *n*- and *p*-type semiconductors. Inside the space-charge region, the conduction and valence bands are not represented by straight horizontal lines any more but they are curved. This indicates the presence of an electric field in this region. Due to the electric field a difference in the electrostatic potential is created between the boundaries of the space-charge region. Across the depletion re-

gion the changes in the carriers concentration are compensated by changes in the electrostatic potential. The electrostatic-potential profile is also drawn in Fig. 8.3

The concentration profile of charge carriers in a *p-n* junction is schematically presented in Fig. 8.4. In the quasi-neutral regions the concentration of electrons and holes is the same as in the isolated doped semiconductors. In the space-charge region the concentrations of majority charge carriers decrease very rapidly. This fact allows us to use the assumption that the space-charge region is depleted of mobile charge carriers. This assumption means that the charge of the mobile carriers represents a negligible contribution to the total space charge in the depletion region. The space charge in this region is fully determined by the ionised dopant atoms fixed in the lattice.

The presence of the internal electric field inside the *p-n* junction means that there is an electrostatic potential difference, ψ_0 , across the space-charge region. We shall determine a profile of the internal electric field and electrostatic potential in the *p-n* junction. First we introduce an approximation, which simplifies the calculation of the electric field and electrostatic-potential. This approximation (*the depletion approximation*) assumes that the space-charge density, ρ , is zero in the quasi-neutral regions and it is fully determined by the concentration of ionised dopants in the depletion region. In the depletion region of the *n*-type semiconductor it

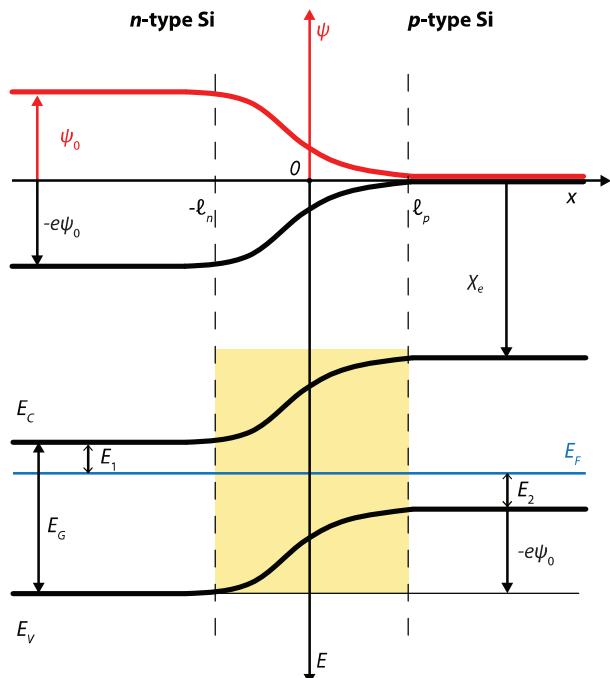


Figure 8.3: Energy-band diagram of the *p-n* junction. The electrostatic potential profile (red curve) is also presented in the figure.

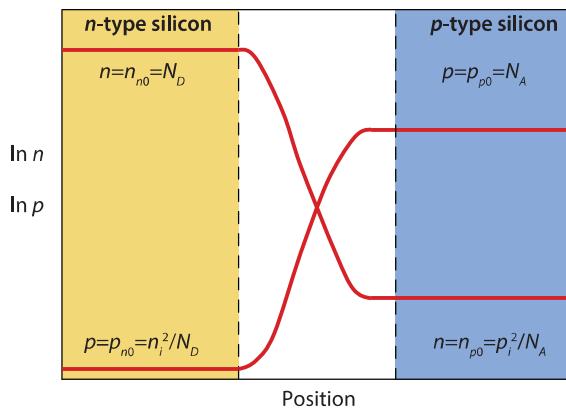


Figure 8.4: Concentrations profile of mobile charge carriers in a p-n junction under equilibrium.

is the concentration of positively charged donor atoms, N_D , which determines the space charge in this region. In the p-type semiconductor, the concentration of negatively charged acceptor atoms, N_A , determines the space charge in the depletion region. This is illustrated in Fig. 8.5. Further, we assume that the p-n junction is a step junction; it means that there is an abrupt change in doping at the metallurgical junction and the doping concentration is uniform both in the p-type and the n-type semiconductors.

According to Fig. 8.5 the position of the metallurgical junction is placed at zero, the width of the space-charge region in the n-type material is denoted as ℓ_n and the width of the space-charge region in the p-type material is denoted as ℓ_p . The space-charge density is described by

$$\rho(x) = eN_D \quad \text{for} \quad -\ell_n \leq x \leq 0, \quad (8.3a)$$

$$\rho(x) = -eN_A \quad \text{for} \quad 0 \leq x \leq \ell_p, \quad (8.3b)$$

where N_D and N_A is the concentration of donor and acceptor atoms, respectively. Outside the space-charge region the space-charge density is zero. The electric field is can be calculated from the Poisson's equation, in one dimension can be written as

$$\frac{d^2\psi}{dx^2} = -\frac{d\xi}{dx} = -\frac{\rho}{\epsilon_r \epsilon_0}, \quad (8.4)$$

where ψ is the electrostatic potential, ξ is the electric field, ρ is the space-charge density, ϵ_r is the semicon-

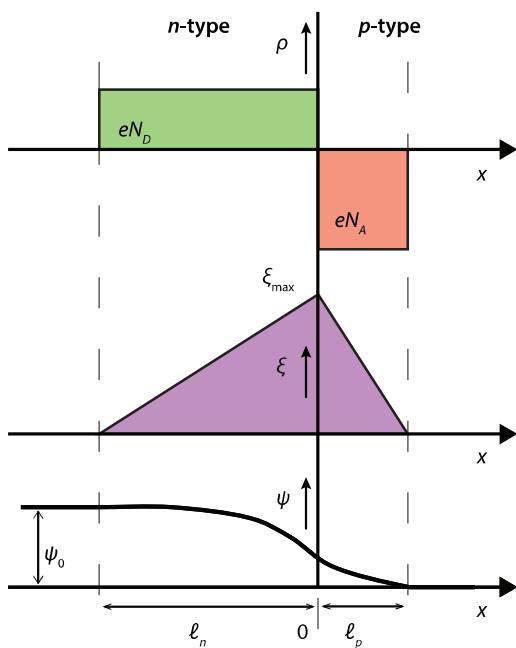


Figure 8.5: The space-charge density $\rho(x)$, the electric field $\xi(x)$, and the electrostatic potential $\psi(x)$ across the depletion region of a p - n junction under equilibrium.

ductor dielectric constant and ϵ_0 is the vacuum permittivity. The vacuum permittivity is $\epsilon_0 = 8.854 \cdot 10^{-14}$ F/cm and for crystalline silicon $\epsilon_r = 11.7$. The electric field profile can be found by integrating the space-charge density across the space-charge region,

$$\xi = \frac{1}{\epsilon_r \epsilon_0} \int \rho \, dx. \quad (8.5)$$

Substituting the space-charge density with Eqs. (8.3) and using the boundary conditions

$$\xi(-\ell_n) = \xi(\ell_p) = 0, \quad (8.6)$$

we obtain as solution for the electric field

$$\xi(x) = \frac{e}{\epsilon_r \epsilon_0} N_D (\ell_n + x) \quad \text{for } -\ell_n \leq x \leq 0, \quad (8.7a)$$

$$\xi(x) = \frac{e}{\epsilon_r \epsilon_0} N_A (\ell_p - x) \quad \text{for } 0 \leq x \leq \ell_p. \quad (8.7b)$$

At the metallurgical junction, $x = 0$, the electric field is continuous, which requires that the following condition has to be fulfilled

$$N_A \ell_p = N_D \ell_n. \quad (8.8)$$

Outside the space-charge region the material is electrically neutral and therefore the electric field is zero there.

The profile of the electrostatic potential is calculated by integrating the electric field throughout the space-charge region and applying the boundary conditions,

$$\psi = - \int \xi dx. \quad (8.9)$$

We define the zero electrostatic potential level at the outside edge of the *p*-type semiconductor. Since we assume no potential drop across the quasi-neutral region the electrostatic potential at the boundary of the space-charge region in the *p*-type material is also zero,

$$\psi(\ell_p) = 0. \quad (8.10)$$

Using Eqs. (8.7) for describing the electric field in the *n*-type and *p*-type parts of the space-charge region, respectively, and taking into account that at the metallurgical junction the electrostatic potential is continuous, the solution for the electrostatic potential can be written as

$$\left. \begin{aligned} \psi(x) = & -\frac{e}{2\varepsilon_r\varepsilon_0} N_D (x + \ell_n)^2 \\ & + \frac{e}{2\varepsilon_r\varepsilon_0} (N_D \ell_n^2 + N_A \ell_p^2) \end{aligned} \right\} \quad \text{for } -\ell_n \leq x \leq 0, \quad (8.11a)$$

$$\psi(x) = \frac{e}{2\varepsilon_r\varepsilon_0} N_A (x - \ell_p)^2 \quad \text{for } 0 \leq x \leq \ell_p. \quad (8.11b)$$

Under equilibrium a difference in electrostatic potential, ψ_0 , develops across the space-charge region. The

electrostatic potential difference across the *p-n* junction is an important characteristic of the junction and is denoted as the *built-in voltage* or diffusion potential of the *p-n* junction. We can calculate ψ_0 as the difference between the electrostatic potential at the edges of the space-charge region,

$$\psi_0 = \psi(-\ell_n) - \psi(\ell_p) = \psi(-\ell_n). \quad (8.12)$$

Using Eq. (8.11a) we obtain for the built-in voltage

$$\psi_0 = \frac{e}{2\varepsilon_r\varepsilon_0} (N_D \ell_n^2 + N_A \ell_p^2) \quad (8.13)$$

Another way to determine the built-in potential ψ_0 is to use the energy-band diagram presented in Fig. 8.3.

$$e\psi_0 = E_G - E_1 - E_2. \quad (8.14)$$

Using Eqs. (6.1) and (6.20), which determine the band gap, and the positions of the Fermi energy in the *n*- and *p*-type semiconductor, respectively,

$$\begin{aligned} E_G &= E_C - E_V, \\ E_1 &= E_C - E_F = kT \ln(N_C/N_D), \\ E_2 &= E_F - E_V = kT \ln(N_V/N_A). \end{aligned}$$

We can write

$$\begin{aligned} e\psi_0 &= E_G - kT \ln\left(\frac{N_V}{N_A}\right) - kT \ln\left(\frac{N_C}{N_D}\right) \\ &= E_G - kT \ln\left(\frac{N_V N_C}{N_A N_D}\right). \end{aligned} \quad (8.15)$$

Using the relationship between the intrinsic concentration, n_i and the band gap, E_G [see Eq. (6.7)],

$$n_i^2 = N_C N_V \exp \left[-\frac{E_G}{kT} \right],$$

we can rewrite Eq. (8.15) and obtain

$$\psi_0 = \frac{kT}{e} \ln \left(\frac{N_A N_D}{n_i^2} \right) \quad (8.16)$$

This equation allows us to determine the built-in potential of a p - n junction from the standard semiconductor parameters, such as doping concentrations and the intrinsic carrier concentration. With the built-in potential we can calculate the width of the space charge region of the p - n junction in the thermal equilibrium. Substituting ψ_0 using Eq. (8.16) into Eq. (8.12) and taking the boundary condition [Eq. (8.6) into account, the resulting expressions for ℓ_n and ℓ_p are obtained. The full derivation can be found for example Ref. [20].

$$\ell_n = \sqrt{\frac{2\epsilon_r \epsilon_0}{q} \psi_0 \frac{N_A}{N_D} \left(\frac{1}{N_A + N_D} \right)}, \quad (8.17a)$$

$$\ell_p = \sqrt{\frac{2\epsilon_r \epsilon_0}{q} \psi_0 \frac{N_D}{N_A} \left(\frac{1}{N_A + N_D} \right)}. \quad (8.17b)$$

The total space-charge width, W , is the sum of the partial space-charge widths in the n - and p -type semicon-

ductors and can be calculated with

$$W = \ell_n + \ell_p = \sqrt{\frac{2\epsilon_r \epsilon_0}{e} \psi_0 \left(\frac{1}{N_A} + \frac{1}{N_D} \right)}. \quad (8.18)$$

The space-charge region is not uniformly distributed in the n - and p -regions. The widths of the space-charge region in the n - and p -type semiconductor are determined by the doping concentrations as illustrated by Eqs. (8.17). Knowing the expressions for ℓ_n and ℓ_p we can determine the maximum value of the internal electric field, which is at the metallurgical junction. By substituting ℓ_p from expressed by Eq. (8.17b) into Eq. (8.7b) we obtain the expression for the maximum value of the internal electric field,

$$\xi_{\max} = \sqrt{\frac{2e}{\epsilon_r \epsilon_0} \psi_0 \left(\frac{N_A N_D}{N_A + N_D} \right)}. \quad (8.19)$$

Example

A crystalline silicon wafer is doped with 10^{16} acceptor atoms per cubic centimetre. A 1 micrometer thick emitter layer is formed at the surface of the wafer with a uniform concentration of 10^{18} donors per cubic centimetre. Assume a step p-n junction and that all doping atoms are ionised. The intrinsic carrier concentration in silicon at 300 K is $1.5 \cdot 10^{10} \text{ cm}^{-3}$.

Let us calculate the electron and hole concentrations in the p- and n-type quasi-neutral regions at thermal equilibrium. We shall use Eqs. (8.1) and (8.2) to calculate the charge carrier concentrations.

$$\text{P-type region: } p = p_{p0} \approx N_A = 10^{16} \text{ cm}^{-3}.$$

$$n = n_{p0} = n_i^2 / p_{p0} = (1.5 \cdot 10^{10})^2 / 10^{16} = 2.25 \cdot 10^4 \text{ cm}^{-3}$$

$$\text{N-type region: } n = n_{n0} \approx N_A = 10^{18} \text{ cm}^{-3}.$$

$$p = p_{n0} = n_i^2 / n_{n0} = (1.5 \cdot 10^{10})^2 / 10^{18} = 2.25 \cdot 10^2 \text{ cm}^{-3}$$

We can calculate the position of the Fermi energy in the quasi-neutral n-type and p-type regions, respectively, using Eq. (6.20a). We assume that the reference energy level is the bottom of the conduction band, $E_C = 0 \text{ eV}$.

$$\text{N-type region: } E_F - E_C = -kT \ln(N_C/n) = -0.0258 \ln(3.32 \cdot 10^{19} / 10^{18}) = -0.09 \text{ eV}.$$

$$\text{P-type region: } E_F - E_C = -kT \ln(N_C/n) = -0.0258 \ln(3.32 \cdot 10^{19} / 2.24 \cdot 10^4) = -0.90 \text{ eV}.$$

The minus sign tells us that the Fermi energy is positioned below the conduction band.

The built-in voltage across the p-n junction is calculated using Eq. (8.16),

$$\psi_0 = \frac{kT}{e} \ln \left(\frac{N_A N_D}{n_i^2} \right) = 0.0258 \text{ V} \left[\frac{10^{16} 10^{18}}{(1.5 \cdot 10^{10})^2} \right] = 0.81 \text{ V}.$$

The width of the depletion region is calculated from Eq. (8.18),

$$W = \sqrt{\frac{2\epsilon_r \epsilon_0}{e} \psi_0 \left(\frac{1}{N_A} + \frac{1}{N_D} \right)} = \sqrt{\frac{2 \cdot 11.7 \cdot 8.854 \cdot 10^{-14}}{1.602 \cdot 10^{-19}} \cdot 0.81 \left(\frac{1}{10^{16}} + \frac{1}{10^{18}} \right)} = 3.25 \cdot 10^{-5} \text{ cm} = 0.325 \mu\text{m}.$$

A typical thickness of c-Si wafers is 300 μm . The depletion region is 0.3 μm which represents 0.1% of the wafer thickness. It is important to realise that almost the whole bulk of the wafer is a quasi-neutral region without an internal electrical field.

The maximum electric field is at the metallurgical junction and is calculated from Eq. (8.19).

$$\xi_{\max} = \sqrt{\frac{2e}{\epsilon_r \epsilon_0} \psi_0 \left(\frac{N_A N_D}{N_A + N_D} \right)} = \frac{2 \cdot 1.602 \cdot 10^{-19}}{11.7 \cdot 8.854 \cdot 10^{-14}} \cdot 0.81 \left(\frac{10^{16} 10^{18}}{10^{16} + 10^{18}} \right) = 5 \cdot 10^4 \text{ V cm}^{-1}.$$

8.1.3 The p-n junction under applied voltage

When an external voltage, V_a , is applied to a p-n junction the potential difference between the n- and p-type regions will change and the electrostatic potential across the space-charge region will become $(\psi_0 - V_a)$. Remember that under equilibrium the built-in potential is negative in the p-type region with respect to the n-type region. When the applied external voltage is negative with respect to the potential of the p-type region, the applied voltage will increase the potential difference across the p-n junction. We refer to this situation as p-n junction under *reverse-bias voltage*. The potential barrier across the junction is increased under reverse-bias voltage, which results in a wider space-charge region. The band diagram of the p-n junction under reverse-biased voltage is presented in Fig. 8.6 (a). Under external voltage the p-n junction is not under equilibrium any more and the concentrations of electrons and holes are described by the quasi-Fermi energy for electrons, E_{FC} , and the quasi-Fermi energy for holes, E_{FV} , respectively. When the applied external voltage is positive with respect to the potential of the p-type region, the applied voltage will decrease the potential difference across the p-n junction. We refer to this situation as p-n junction under *forward-bias voltage*. The band diagram of the p-n junction under forward-biased voltage is presented in Fig. 8.6 (b). The potential barrier across the

junction is decreased under forward-bias voltage and the space charge region becomes narrower. The balance between the forces responsible for diffusion (concentration gradient) and drift (electric field) is disturbed. Lowering the electrostatic potential barrier leads to a higher concentration of minority carriers at the edges of the space-charge region compared to the situation in equilibrium. This process is referred to as minority-carrier *injection*. This gradient in concentration causes the diffusion of the minority carriers from the edge into the bulk of the quasi-neutral region.

The diffusion of minority carriers into the quasi-neutral region causes a so-called recombination current, J_{rec} , since the diffusing minority carriers recombine with the majority carriers in the bulk. The recombination current is compensated by the so-called thermal generation current, J_{gen} , which is caused by the drift of minority carriers, which are present in the corresponding doped regions (electrons in the p-type region and holes in the n-type region), across the junction. Both, the recombination and generation currents have contributions from electrons and holes. When no voltage is applied to the p-n junction, the situation inside the junction can be viewed as the balance between the recombination and generation currents,

$$J = J_{rec} - J_{gen} = 0 \quad \text{for } V_a = 0 \text{ V.} \quad (8.20)$$

It is assumed that when a moderate forward-bias voltage is applied to the junction the recombination cur-

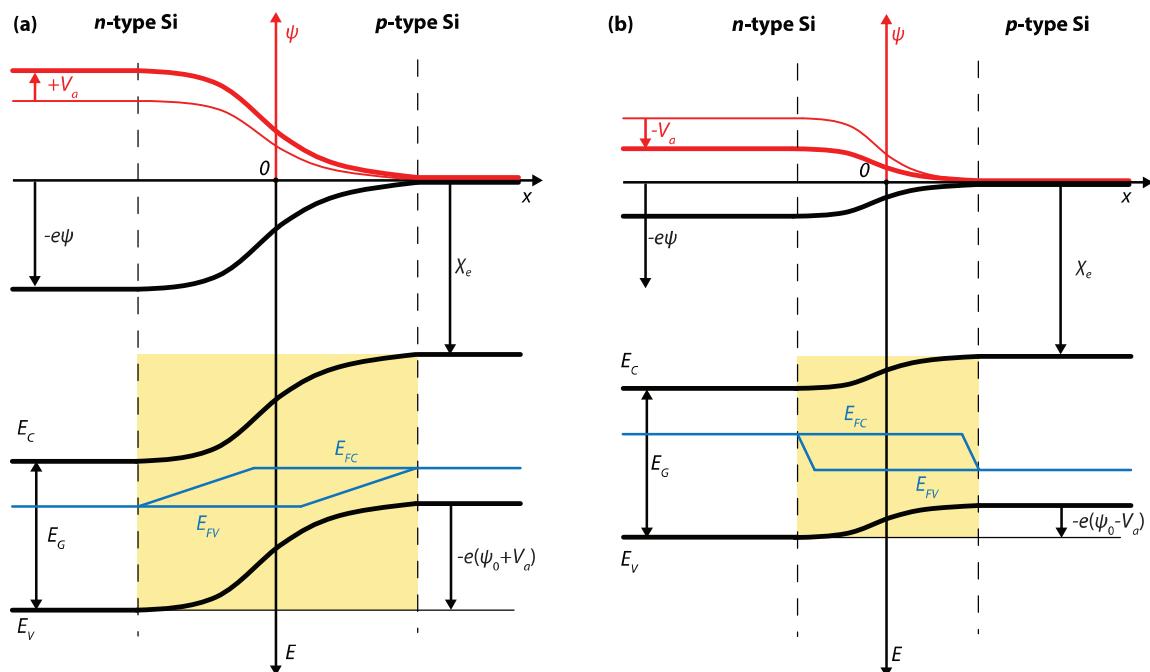


Figure 8.6: Energy band diagram and potential profile (in red colour) of a *p*-*n* junction under (a) reverse bias, and (b) forward bias.

rent increases with the Boltzmann factor $\exp(eV_a/kT)$,

$$J_{\text{rec}}(V_a) = J_{\text{rec}}(V_a = 0) \exp\left(\frac{eV_a}{kT}\right). \quad (8.21)$$

This assumption is called the *Boltzmann approximation*.

The generation current, on the other hand, is almost independent of the potential barrier across the junction and is determined by the availability of the thermally-generated minority carriers in the doped regions,

$$J_{\text{gen}}(V_a) \approx J_{\text{gen}}(V_a = 0). \quad (8.22)$$

The external net-current density can be expressed as

$$\begin{aligned} J(V_a) &= J_{\text{rec}}(V_a) - J_{\text{gen}}(V_a) \\ &= J_0 \left[\exp\left(\frac{eV_a}{kT}\right) - 1 \right], \end{aligned} \quad (8.23)$$

where J_0 is the saturation-current density of the p-n junction, given by

$$J_0 = J_{\text{gen}}(V_a = 0). \quad (8.24)$$

Equation (8.23) is known as the *Shockley equation* that describes the current-voltage behaviour of an ideal p-n diode. It is a fundamental equation for microelectronics device physics. The detailed derivation of the dark-current density of the p-n junction is carried out in Appendix B.1. The saturation-current density is given by

$$J_0 = e n_i^2 \left(\frac{D_N}{L_N N_A} + \frac{D_P}{L_P N_D} \right). \quad (8.25)$$

The saturation-current density depends in a complex way on the fundamental semiconductor parameters. Ideally the saturation-current density should be as low as possible and this requires an optimal and balanced design of the p-type and n-type semiconductor properties. For example, an increase in the doping concentration decreases the diffusion length of the minority carriers, which means that the optimal product of these two quantities requires a delicate balance between these two properties.

The recombination of the majority carriers due to the diffusion of the injected minority carriers into the bulk of the quasi-neutral regions results in a lowering of the concentration of the majority carriers compared to the one under equilibrium. The drop in the concentration of the majority carriers is balanced by the flow of the majority carriers from the electrodes into the bulk. In this way the net current flows through the p-n junction under forward-bias voltage. For high reverse-bias voltage, the Boltzmann factor in Eq. (8.23) becomes very small and can be neglected. The net current density is given by

$$J(V_a) = -J_0, \quad (8.26)$$

and represents the flux of thermally generated minority carriers across the junction. The current density-voltage (J-V) characteristic of an ideal p-n junction is schematically shown in Fig. 8.7.

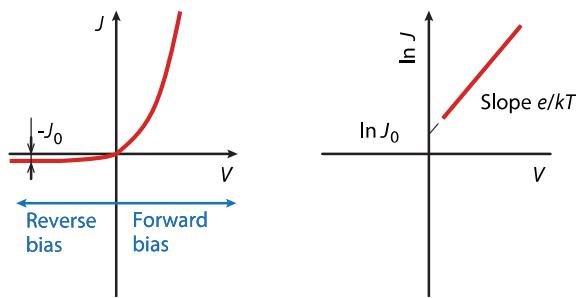


Figure 8.7: J - V characteristic of a p - n junction; (a) linear plot and (b) semi-logarithmic plot.

8.1.4 The p - n junction under illumination

When a p - n junction is illuminated the additional electron-hole pairs are generated in the semiconductor. The concentration of minority carriers (electrons in the p -type region and holes in the n -type region) strongly increases. This increase in the concentration of minority carriers leads to the flow of the minority carriers across the depletion region into the quasi-neutral regions. Electrons flow from the p -type into the n -type region and holes from the n -type into the p -type region. The flow of the photo-generated carriers causes the so-called *photo-generation current*, J_{ph} , which adds to the thermal-generation current, J_{gen} . When no external contact between the n -type and the p -type re-

gions is established, which means that the junction is in the open-circuit condition, no net current can flow inside the p - n junction. It means that the current resulting from the flux of photo-generated and thermally-generated carriers has to be balanced by the opposite recombination current. The recombination current will increase through lowering of the electrostatic potential barrier across the depletion region. This situation of the illuminated p - n junction under open-circuit condition using the band diagram is presented in Fig. 8.8 (a). The electrostatic-potential barrier across the junction is lowered by an amount of V_{oc} . We refer to V_{oc} as the *open-circuit voltage*. Under non-equilibrium conditions the concentrations of electrons and holes are described by the quasi-Fermi energy levels. It is illustrated in Fig. 8.8 (a) that the electrochemical potential of electrons, denoted by E_{FC} , is higher in the n -type region than in the p -type region by an amount of eV_{oc} . This means that a voltmeter will measure a voltage difference of V_{oc} between the contacts of the p - n junction. Under illumination, when the n -type and p -type regions are short circuited, the photo-generated current will also flow through the external circuit. This situation is illustrated in Fig. 8.8 (b). Under the short-circuit condition the electrostatic-potential barrier is not changed, but from a strong variation of the quasi-Fermi levels inside the depletion region one can determine that the current is flowing inside the semiconductor.

When a load is connected between the electrodes of the

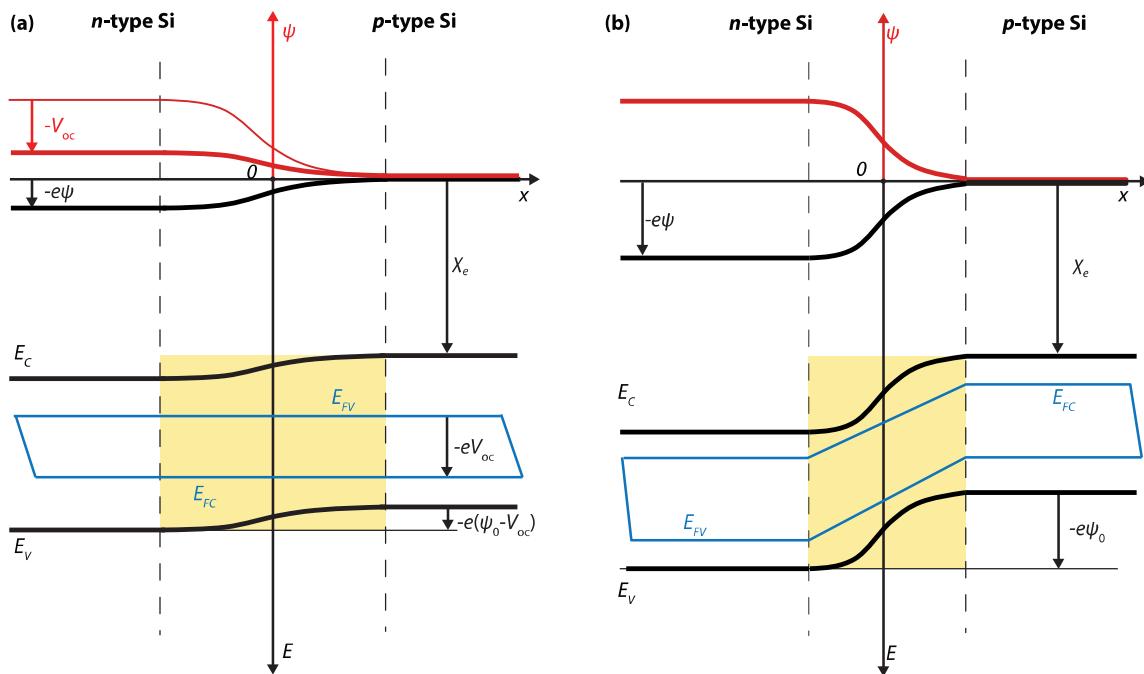


Figure 8.8: Energy band diagram and electrostatic-potential (in red colour) of an illuminated *p-n* junction under the (a) open-circuit and (b) short-circuit conditions.

illuminated *p-n* junction, only a fraction of the photo-generated current will flow through the external circuit. The electro-chemical potential difference between the *n*-type and *p*-type regions will be lowered by a voltage drop over the load. This in turn lowers the electrostatic potential over the depletion region which results in an increase of the recombination current. In the *superposition approximation*, the net current flowing through the load is determined as the sum of the photo- and thermal generation currents and the recombination current. The voltage drop at the load can be simulated by applying a forward-bias voltage to the junction, therefore Eq. (8.23), which describes the behaviour of the junction under applied voltage, is included to describe the net current of the illuminated *p-n* junction,

$$\begin{aligned} J(V_a) &= J_{\text{rec}}(V_a) - J_{\text{gen}}(V_a) - J_{\text{ph}} \\ &= J_0 \left[\exp \left(\frac{eV_a}{kT} \right) - 1 \right] - J_{\text{ph}}. \end{aligned} \quad (8.27)$$

Both the dark and illuminated *J-V* characteristics of the *p-n* junction are represented in Fig. 8.9. Note, that in the figure the superposition principle is reflected. The illuminated *J-V* characteristic of the *p-n* junction is the same as the dark *J-V* characteristic, but it is shifted down by the photo-generated current density J_{ph} . The detailed derivation of the photo-generated current density of the *p-n* junction is carried out in Appendix

B.2. and its value under uniform generation rate, G , is

$$J_{\text{ph}} = eG (L_N + W + L_P), \quad (8.28)$$

where L_N and L_P is the minority-carrier-diffusion length for electrons and holes, respectively, and W is the width of the depletion region. It means only carriers generated in the depletion region and in the regions up to the minority-carrier-diffusion length from the depletion region contribute to the photo-generated current. When designing the thickness of a solar cell, Eq. (8.28) must be considered. The thickness of the absorber should not be thicker than the region from which the carriers contribute to the photo-generated current.

8.2 *p-n* heterojunctions

8.3 Metal-semiconductor junctions

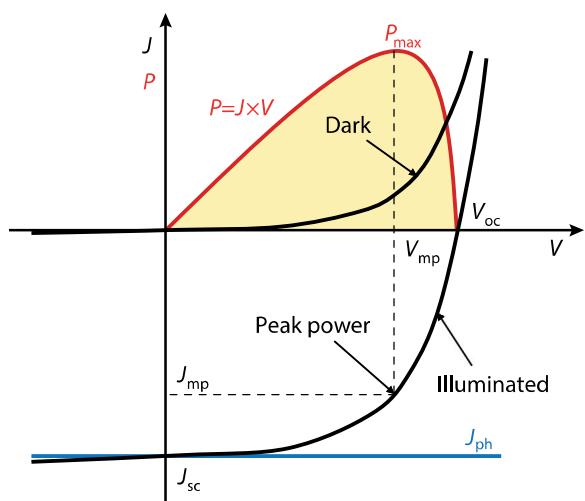


Figure 8.9: J - V characteristics of a p - n junction in the dark and under illumination.

9

Solar Cell Parameters and Equivalent Circuit

9.1 External solar cell parameters

The main parameters that are used to characterise the performance of solar cells are the *peak power* P_{\max} , the *short-circuit current density* J_{sc} , the *open-circuit voltage* V_{oc} , and the *fill factor* FF . These parameters are determined from the illuminated J - V characteristic as illustrated in Fig. 8.9. The *conversion efficiency* η can be determined from these parameters.

9.1.1 Short-circuit current density

The *short-circuit current* I_{sc} is the current that flows through the external circuit when the electrodes of the

solar cell are short circuited. The short-circuit current of a solar cell depends on the photon flux density incident on the solar cell, which is determined by the spectrum of the incident light. For a standard solar cell measurements, the spectrum is standardised to the AM1.5 spectrum. The I_{sc} depends on the area of the solar cell. In order to remove the dependence of the solar cell area on I_{sc} , often the *short-circuit current density* is used to describe the maximum current delivered by a solar cell. The maximum current that the solar cell can deliver strongly depends on the optical properties of the solar cell, such as absorption in the absorber layer and reflection.

In the ideal case, J_{sc} is equal to J_{ph} as can be easily derived from Eq. (8.27). J_{ph} can be approximated by Eq.

(8.28), which shows that in case of an ideal diode (for example no surface recombination) and uniform generation, the critical material parameters that determine J_{ph} are the diffusion lengths of minority carriers. Crystalline silicon solar cells can deliver under an AM1.5 spectrum a maximum possible current density of 46 mA/cm². In laboratory c-Si solar cells the measured J_{sc} is above 42 mA/cm², while commercial solar cell have an J_{sc} exceeding 35 mA/cm².

9.1.2 Open-circuit voltage

The *open-circuit voltage* is the voltage at which no current flows through the external circuit. It is the maximum voltage that a solar cell can deliver. V_{oc} corresponds to the forward bias voltage, at which the dark current compensates the photocurrent. V_{oc} depends on the photo-generated current density and can be calculated from Eq. (8.27) assuming that the net current is zero,

$$V_{oc} = \frac{kT}{e} \ln \left(\frac{J_{ph}}{J_0} + 1 \right). \quad (9.1)$$

This equation shows that V_{oc} depends on the saturation current of the solar cell and the photo-generated current. While J_{ph} typically has a small variation, the key effect is the saturation current, since this may vary by orders of magnitude. The saturation current density, J_0 ,

depends on the recombination in the solar cell. Therefore, V_{oc} is a measure of the amount of recombination in the device. Laboratory crystalline silicon solar cells have a V_{oc} of up to 720 mV under the standard AM1.5 conditions, while commercial solar cells typically have V_{oc} exceeding 600 mV.

9.1.3 Fill factor

The fill factor is the ratio between the maximum power ($P_{max} = J_{mp}V_{mp}$) generated by a solar cell and the product of V_{oc} with J_{sc} (see Fig. 8.9),

$$FF = \frac{J_{mp}V_{mp}}{J_{sc}V_{oc}}. \quad (9.2)$$

Assuming that the solar cell behaves as an ideal diode the fill factor can be expressed as a function of open-circuit voltage V_{oc} ¹,

$$FF = \frac{v_{oc} - \ln(v_{oc} + 0.72)}{v_{oc} + 1}, \quad (9.3)$$

where $v_{oc} = V_{oc} \cdot e/(kT)$ is a normalised voltage. Eq. (9.3) is a good approximation of the ideal value of FF for $v_{oc} > 10$. The FF as a function of V_{oc} is illustrated in Fig. 9.1. This figure shows that FF does not change

¹M.A. Green, Solar Cells; Operating Principles, Technology and System Applications, Prentice-Hall, 1982.

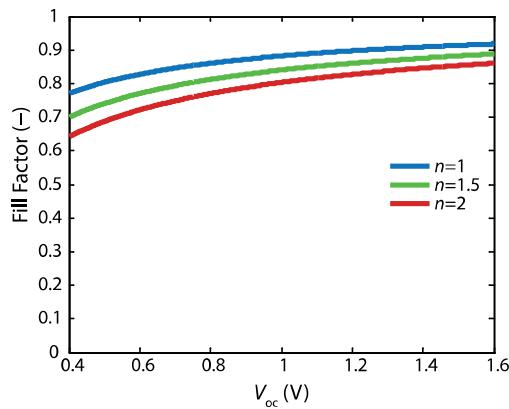


Figure 9.1: The FF as a function of V_{oc} for a solar cell with ideal diode behaviour.

drastically with a change in V_{oc} . For a solar cell with a particular absorber, large variations in V_{oc} are not common. For example, at standard illumination conditions, the difference between the maximum open-circuit voltage measured for a silicon laboratory device and a typical commercial solar cell is about 120 mV, giving a maximal FF of 0.85 and 0.83, respectively. However, the variation in maximum FF can be significant for solar cells made from different materials. For example, a GaAs solar cell may have a FF approaching 0.89.

However, in practical solar cells the dark diode current Eq. (8.23) does not obey the Boltzmann approximation. The non-ideal diode is approximated by introducing an *ideality factor* n , into the Boltzmann factor,

$$\exp \frac{eV_a}{nkT}.$$

Fig. 9.1 also demonstrates the importance of the diode ideality factor when introduced into the normalised voltage in Eq. (9.3). The ideality factor is a measure of the junction quality and the type of recombination in a solar cell. For the ideal junction where the recombination is represented by the recombination of the minority carriers in the quasi-neutral regions the n is equal to 1. However, when other recombination mechanisms occur, the n can have a value of 2. A high n value not only lowers the FF, but since it signals a high recombination, it leads to a low V_{oc} . Eq. (9.3) describes a maximum achievable FF. In practice the FF is often lower due to the presence of parasitic resistive losses.

9.1.4 Conversion efficiency

The *conversion efficiency* is calculated as the ratio between the maximal generated power and the incident power. The irradiance value P_{in} of 1000 W/m² for the AM1.5 spectrum has become a standard for measuring

the conversion efficiency of solar cells,

$$\eta = \frac{P_{\max}}{P_{in}} = \frac{J_{mp} V_{mp}}{P_{in}} = \frac{J_{sc} V_{oc} FF}{P_{in}}. \quad (9.4)$$

Typical external parameters of a crystalline silicon solar cell as shown are; $J_{sc} \approx 35 \text{ mA/cm}^2$, V_{oc} up to 0.65 V and FF in the range 0.75 to 0.80. The conversion efficiency lies in the range of 17 to 18%.

Example

A crystalline silicon solar cell generates a photo-current density of $J_{\text{ph}} = 35 \text{ mA/cm}^2$. The wafer is doped with 10^{17} acceptor atoms per cubic centimetre and the emitter layer is formed with a uniform concentration of 10^{19} donors per cubic centimetre. The minority-carrier diffusion length in the p-type region and n-type region is $500 \cdot 10^{-6} \text{ m}$ and $10 \cdot 10^{-6} \text{ m}$, respectively. Further, the intrinsic carrier concentration in silicon at 300 K is $1.5 \cdot 10^{10} \text{ cm}^{-3}$, the mobility of electrons in the p-type region is $\mu_n = 1000 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$ and holes in the n-type region is $\mu_p = 100 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$. Assume that the solar cell behaves as an ideal diode. Calculate the built-in voltage, the open-circuit voltage and the conversion efficiency of the cell.

$$J_{\text{ph}} = 350 \text{ Am}^{-2}.$$

$$N_A = 10^{17} \text{ cm}^{-3} = 10^{23} \text{ m}^{-3}.$$

$$N_D = 10^{19} \text{ cm}^{-3} = 10^{25} \text{ m}^{-3}.$$

$$L_N = 500 \cdot 10^{-6} \text{ m}.$$

$$L_P = 10 \cdot 10^{-6} \text{ m}.$$

$$D_N = (kT/e)\mu_n = 0.0258 \text{ V} \cdot 1000 \cdot 10^{-4} \text{ cm}^2 \text{V}^{-1} \text{s}^{-1} = 2.58 \cdot 10^{-3} \text{ m}^2 \text{s}^{-1}.$$

$$D_P = (kT/e)\mu_p = 0.0258 \text{ V} \cdot 100 \cdot 10^{-4} \text{ cm}^2 \text{V}^{-1} \text{s}^{-1} = 2.58 \cdot 10^{-4} \text{ m}^2 \text{s}^{-1}.$$

Using Eq. (8.16) we calculate the built-in voltage of the cell,

$$\psi_0 = \frac{kT}{e} \ln \left(\frac{N_A N_D}{n_i^2} \right) = 0.0258 \text{ V} \cdot \ln \left[\frac{10^{23} 10^{25}}{(1.5 \cdot 10^{16})^2} \right] = 0.93 \text{ V}.$$

According to the assumption that the solar cell behaves as an ideal diode, the Shockley equation describing the J-V characteristic is applicable. Using Eq. (8.25) we determine the saturation-current density,

$$\begin{aligned} J_0 &= e n_i^2 \left(\frac{D_N}{L_N N_A} + \frac{D_P}{L_P N_D} \right) = 1.602 \cdot 10^{-19} \text{ C} (1.5 \cdot 10^{16})^2 \text{ m}^{-6} \left(\frac{2.58 \cdot 10^{-3} \text{ m}^2 \text{s}^{-1}}{500 \cdot 10^{-6} \text{ m} 10^{23} \text{ m}^{-3}} + \frac{2.58 \cdot 10^{-4} \text{ m}^2 \text{s}^{-1}}{100 \cdot 10^{-6} \text{ m} 10^{25} \text{ m}^{-3}} \right) \\ &= 1.95 \cdot 10^{-9} \frac{\text{C}}{\text{m}^2 \text{s}} = 1.95 \cdot 10^{-9} \frac{\text{A}}{\text{m}^2}. \end{aligned}$$

Using Eq. (9.1) we determine the open-circuit voltage,

$$V_{\text{oc}} = \frac{kT}{e} \ln \left(\frac{J_{\text{ph}}}{J_0} + 1 \right) = 0.0258 \text{ V} \ln \left(\frac{350 \text{ Am}^{-2}}{1.95 \cdot 10^{-9} \text{ Am}^{-2}} + 1 \right) = 0.67 \text{ V}.$$

The fill factor of the cell can be calculated from Eq. (9.3). First, we normalise V_{oc} ,

$$v_{\text{oc}} = V_{\text{oc}} / \frac{kT}{e} = \frac{0.67 \text{ V}}{0.0258 \text{ V}} = 26.8.$$

Hence,

$$FF = \frac{v_{\text{oc}} - \ln(v_{\text{oc}} + 0.72)}{v_{\text{oc}} + 1} = 0.84.$$

Finally, the conversion efficiency is determined using Eq. (9.4),

$$\eta = \frac{J_{\text{sc}} V_{\text{oc}} FF}{P_{\text{in}}} = \frac{350 \text{ Am}^{-2} 0.67 \text{ V} 0.84}{1000 \text{ W m}^{-2}} = 0.197 = 19.7\%.$$

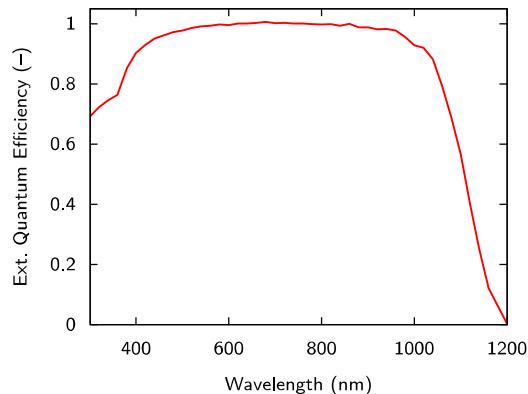


Figure 9.2: The external quantum efficiency of a high-quality crystalline silicon based solar cell.

9.2 The external quantum efficiency

The external quantum efficiency $\text{EQE}(\lambda_0)$ is the fraction of photons incident on the solar cell that create electron-hole pairs in the absorber, which are successfully collected. It is wavelength dependent and is usually measured by illuminating the solar cell with monochromatic light of wavelength λ_0 and measuring the photocurrent I_{ph} through the solar cell. The external quantum effi-

ciency is then determined as

$$\text{EQE}(\lambda_0) = \frac{I_{\text{ph}}(\lambda_0)}{e \Phi_{\text{ph}}(\lambda_0)}, \quad (9.5)$$

where e is the elementary charge and Φ_{ph} is the photon flux incident on the solar cell. Since I_{ph} is dependent on the bias voltage, the bias voltage must be fixed. The photon flux is usually determined by measuring the EQE of a calibrated photodiode under the same light source.

Figure 9.2 illustrates a typical EQE for a high quality crystalline silicon based solar cell. We can identify the major optical loss mechanisms for such a solar cell: For short wavelengths only a small fraction of the light is converted into electron-hole pairs. Most photons are already absorbed in the layers that the light traverses prior to the absorber layer. For long wavelengths, the penetration depth² of the light exceeds the optical thickness of the absorber. Then the absorber itself becomes transparent so that most of the light leaves the solar cell before it can be absorbed. We can see that for this type of solar cells the EQE is close to 1 for a broad wavelength band. Hence, in this band almost all absorbed photons are converted into electron-hole pairs that can leave the solar cell.

²According to Lambert-Beer's law, the intensity of light in an absorbing layer decays exponentially, $I(z) \propto \exp(-\alpha z)$, where α is the absorption coefficient. The penetration depth is then defined as $d_{\text{pen}}(\lambda_0) = 1/\alpha(\lambda_0)$. The absorption coefficient α is related to the imaginary part k of the complex refractive index via $\alpha(\lambda_0) = 4\pi k/\lambda_0$.

When a bias voltage of 0 V is applied, the measured photocurrent density equals the short circuit current density. When applying a sufficiently large reversed bias voltage, it can be assured that nearly all photo-generated charge carriers in the intrinsic layer are collected. Thus, this measurement can be used to study the optical effectiveness of the design, *i.e.* light trapping and light absorption in inactive layers, such as the TCO layer, doped layers and the back reflector.

Measuring the EQE

EQE spectra are measured using an EQE-setup that also called *spectral response setup*. For this measurement, usually a wavelength selective light source, a calibrated light detector and a current meter are required. Usually, the used light source is a *xenon gas discharge lamp* that as a very broad spectrum covering all the wavelengths important for the solar cell performance. With the help of filters and monochromators a very narrow wavelength band of photon energies can be selected that then can be incident on the solar cell.

As already seen in Eq. (9.5), $\text{EQE}(\lambda)$ is proportional to the current divided by the photon flux. While the current can be easily determined using an Ampere meter, the photon flux must be determined indirectly. This is done by performing a measurement with a calibrated photodetector (or solar cell), whose EQE is

known. Via this measurement we find

$$\Phi_{\text{ph}}(\lambda_0) = \frac{I_{\text{ph}}^{\text{ref}}(\lambda_0)}{e \text{EQE}^{\text{ref}}(\lambda_0)}, \quad (9.6)$$

By combining Eqs. (9.5) and (9.6) we therefore find

$$\text{EQE}(\lambda_0) = \text{EQE}^{\text{ref}}(\lambda_0) \frac{I_{\text{ph}}(\lambda_0)}{I_{\text{ph}}^{\text{ref}}(\lambda_0)}. \quad (9.7)$$

The EQE therefore can be determined by performing two current measurements. Of course it is very important that the light source is sufficiently stable during the whole measurement as we assume that the photon flux in the reference measurement and the actual measurement is unchanged.

If we perform the EQE measurement under short circuit conditions, the measurement can be used to determine the *short circuit current density* J_{sc} . Determining J_{sc} via the EQE has the advantage that it is independent of the spectral shape of the used light source, in contrast to determining the J_{sc} via an *JV* measurement. Secondly, on lab scale the real contact area of solar cells is not accurately determined during *JV* measurements. When using shading masks, the EQE measurement is independent on the contact area. Hence, for accurately measuring the sort circuit current density, it is not sufficient to rely on *JV* measurements only, but EQE setups have to be used.

For determining J_{sc} we combine that the photon flux at a certain wavelength with the EQE at this wavelength, leading to the flow of electrons leaving the solar cell at this wavelength. J_{sc} then is obtained by integrating across all the relevant wavelength,

$$J_{sc} = -e \int_{\lambda_1}^{\lambda_2} \text{EQE}(\lambda) \Phi_{\text{AM}1.5}(\lambda) d\lambda. \quad (9.8)$$

For crystalline silicon, the important range would be from 300 to 1200 nm.

9.3 The equivalent circuit

The $J-V$ characteristic of an illuminated solar cell that behaves as the ideal diode is given by Eq. (8.27),

$$\begin{aligned} J(V_a) &= J_{\text{rec}}(V_a) - J_{\text{gen}}(V_a) - J_{\text{ph}} \\ &= J_0 \left[\exp \left(\frac{eV_a}{kT} \right) - 1 \right] - J_{\text{ph}}. \end{aligned}$$

This behaviour can be described by a simple equivalent circuit, illustrated in Fig. 9.3 (a), in which a diode and a current source are connected in parallel. The diode is formed by a $p-n$ junction. The first term in Eq. (8.27) describes the dark diode current density while the second term describes the photo-generated current density. In practice the FF is influenced by a series resistance R_s , and a shunt resistance R_p . The influence of

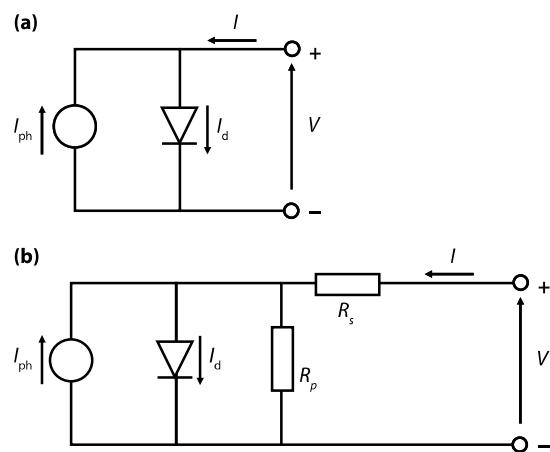


Figure 9.3: The equivalent circuit of an (a) ideal solar cell and a (b) solar cell with a series resistance R_s and a shunt resistance R_p .

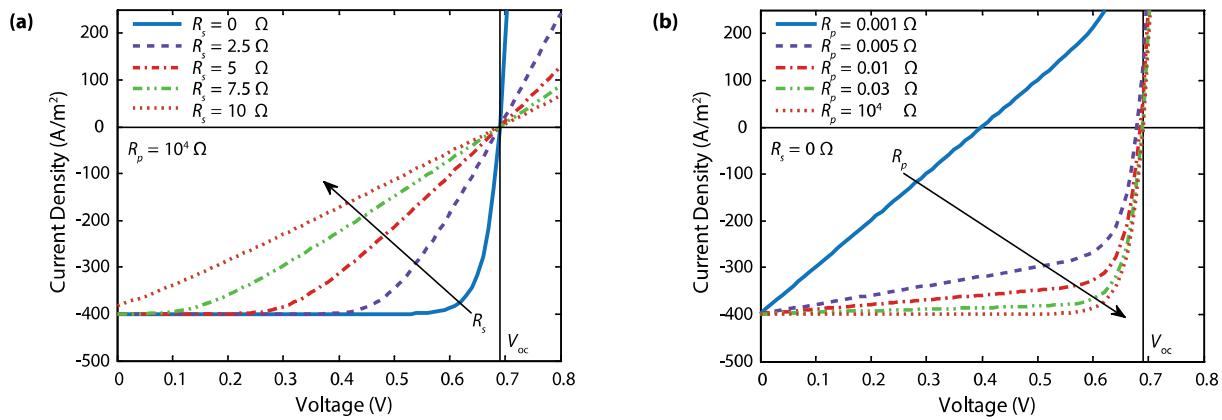


Figure 9.4: Effect of the (a) series resistance and (b) parallel resistance on the J - V characteristic of a solar cell.

these parameters on the J - V characteristic of the solar cell can be studied using the equivalent circuit presented in Fig. 9.3 (b). The J - V characteristic of the one-diode equivalent circuit with the series resistance and the shunt resistance is given by

$$J = J_0 \left\{ \exp \left[\frac{e(V - AJR_s)}{kT} \right] - 1 \right\} + \frac{V - AJR_s}{R_p} - J_{ph}, \quad (9.9)$$

where A is the area of the solar cell. The effect of R_s and R_p on the J - V characteristic is illustrated in Fig. 9.4.

In real solar cells the FF is influenced by additional recombination occurring in the p - n junction. This non-ideal diode is often represented in the equivalent circuit by two diodes, an ideal one with an ideality factor equal to unity and a non-ideal diode with an ideality factor larger than one. The equivalent circuit of a real solar cell is presented in Fig. 9.5. The J - V characteristic of the two-diode equivalent circuit is given by

$$\begin{aligned} J &= J_{01} \left[\exp \left(\frac{e(V - AJR_s)}{n_1 kT} \right) - 1 \right] \\ &+ J_{02} \left[\exp \left(\frac{e(V - AJR_s)}{n_2 kT} \right) - 1 \right] \\ &+ \frac{V - AJR_s}{R_p} - J_{ph}, \end{aligned} \quad (9.10)$$

where J_{01} and J_{02} are the saturation currents of the two diodes, respectively. n_1 and n_2 are the ideality factors of the two diodes.

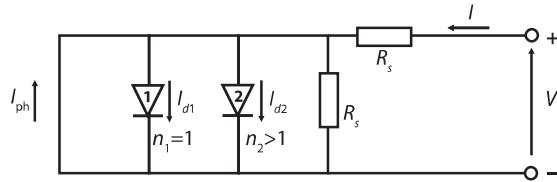


Figure 9.5: The equivalent circuit of a solar cell based on the two-diode model.

10

Losses and Efficiency Limits

In the previous chapters we have learned the basic physical principles of solar cells. In this chapter we will bring the different building blocks together and analyse, how efficient a solar cell theoretically can be.

After discussing different efficiency limits and the major loss mechanisms, we will finalise this chapter with the formulation of *three design rules* that should always be kept in mind when designing solar cells.

It is very important to understand, why a solar cell cannot convert 100% of the incident light into electricity. Different efficiency limits can be formulated, each taking different effects into account.

10.1 The thermodynamic limit

The most general efficiency limit is the *thermodynamic efficiency limit*. In this limit, the photovoltaic device is seen as a thermodynamic *heat engine*, as illustrated in Fig. 10.1. Such a heat engine operates between two heat reservoirs; a hot one with temperature T_H and a cold one temperature T_C . For the heat engine, three energy flows are relevant. First, the heat flow \dot{Q}_H from the hot reservoir to the engine. Secondly, the *work* \dot{W} that is performed by the engine and thirdly, heat flowing from the engine to the cold reservoir that serves as a *heat sink*, \dot{Q}_C . Clearly, the third energy flow is a loss and consequently, the efficiency of the heat engine is given

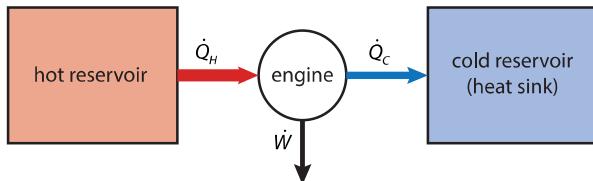


Figure 10.1: Illustrating the major heat flows in a generic heat engine.

by

$$\eta = \frac{W}{Q_h}. \quad (10.1)$$

The *second law of thermodynamics* teaches us that the entropy of an independent system never decreases. It only increases or stays the same. While the heat flows \dot{Q}_H and \dot{Q}_C carry entropy, the performed work \dot{W} is an entropy-free form of energy. Thermodynamics teaches us that there is an efficiency limit for the transformation of heat into entropy-free energy. An (ideal) engine that has this maximal efficiency is called a *Carnot engine* and its efficiency is given by

$$\eta_{\text{Carnot}} = 1 - \frac{T_C}{T_H}. \quad (10.2)$$

For a Carnot engine, the entropy does not increase. Note that all the temperatures must be given in a temperature scale where the absolute zero takes the value

0. For example, the Kelvin scale is such a scale. From Eq. (10.2) we can already see two important trends that are basically true for every heat engine, *i.e.* also steam engines or combustion engines. The efficiency increases, if the higher temperature T_H is increased and/or the lower temperature T_C is decreased.

Let us now look at a solar cell that we imagine as a heat engine operating between an *absorber* of temperature T_A (this is our hot reservoir) and a cold reservoir, which is given by the surroundings and that we assume to be of temperature $T_C = 300$ K. What this heat engine actually does is that it converts the energy stored in the heat of the absorber into entropy-less chemical energy that is stored in the electron-hole pairs. We may here assume that the transformation of chemical energy into electrical energy happens lossless, *i.e.* with an efficiency of 1. Clearly, the efficiency of this thermodynamic heat engine is given by

$$\eta_{\text{TD}} = 1 - \frac{T_C}{T_A}. \quad (10.3)$$

The absorber will be heated by absorbing sunlight. As we look at the ideal situation, we assume the absorber to be a black body that absorbs all incident radiation. Further, we assume the sun to be a blackbody of temperature $T_S = 6000$ K. As we have seen in Chapter 5, the solar irradiance incident onto the absorber is given

by

$$I_e^S = \sigma T_S^4 \Omega_{\text{inc}}, \quad (10.4)$$

where Ω_{inc} is the solid angle covered by the incident sunlight. As the absorber is a black body of temperature T_A it also will emit radiation. The emittance of the absorber is given by

$$E_e^A = \sigma T_A^4 \Omega_{\text{emit}}. \quad (10.5)$$

Ω_{emit} is the solid angle into that the absorber can emit.

We can easily see that the efficiency of the absorption process is given by

$$\eta_A = \frac{I_e^S - E_e^A}{I_e^S} = 1 - \frac{E_e^A}{I_e^S} = 1 - \frac{\Omega_{\text{emit}} T_A^4}{\Omega_{\text{inc}}} \frac{T_A^4}{T_S^4}. \quad (10.6)$$

The absorber efficiency can be increased by increasing Ω_{inc} , which can be achieved by *concentrating* the incident sunlight. Under *maximal concentration* sunlight will be incident onto the absorber from all angles of the hemisphere, *i.e.* $\Omega_{\text{inc}}^{\text{max}} = 2\pi$. We assume the absorber to be open towards the surroundings and hence the sun on the top side. Its bottom side is connected to the heat engine such that radiative loss only can happen via the top side. Therefore, also $\Omega_{\text{emit}} = 2\pi$. The maximal absorber efficiency is therefore achieved under maximal concentration and it is given by

$$\eta_A^{\text{max}} = 1 - \frac{T_A^4}{T_S^4}. \quad (10.7)$$

Note that η_A is the larger the lower T_A while the efficiency of the heat engine η_{TD} is the larger the higher T_A .

For the total efficiency of the ideal solar cell we combine Eq. (10.3) with Eq. (10.7) and obtain

$$\eta_{\text{SC}} = \left(1 - \frac{T_A^4}{T_S^4}\right) \left(1 - \frac{T_C}{T_A}\right). \quad (10.8)$$

Figure 10.2 shows the absorber efficiency, the thermodynamic efficiency and the solar cell efficiency. We see that the solar cell efficiency reaches its maximum of about 85% for an absorber temperature of 2480 K. Please note that the solar cell model presented in this section does not resemble a real solar cell but that it is only intended to discuss the physical limit of converting solar radiation into electricity. Several much more detailed studies on the thermodynamic limit have been performed. We want to refer the interested reader to works by Würfel [21] and Markvart *et al.* [30–32].

10.2 The Shockley-Queisser Limit

We now will take a look at the theoretical limit for single-junction solar cells. This limit is usually referred to as the *Shockley-Queisser* (SQ) limit, as they where the

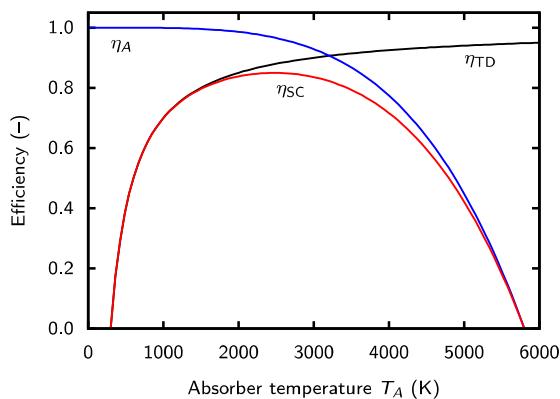


Figure 10.2: The absorber efficiency η_A , the thermodynamic efficiency η_{TD} and the combined solar cell efficiency η_{SC} under full concentration for a solar temperature of 5800 K and an ambient temperature of 300 K.

first ones to formulate this limit based purely on physical assumptions and without using empirically determined constants [23]. We will derive the SQ limit in a two-step approach. First, we will discuss the losses due to *spectral mismatch*. Secondly, we also will take into account that the solar cell will have a temperature different from 0 K which means that it emits electromagnetic radiation according to Planck's law. Just like Shockley and Queisser, we will do this with the *detailed balance* approach.

10.2.1 Spectral mismatch

There are two principal losses that strongly reduce the energy conversion efficiency of single-junction solar cells. As discussed in Chapter 8, an important part of a solar cell is the absorber layer, in which the photons of the incident radiation are efficiently absorbed resulting in a creation of electron-hole pairs. In most cases, the absorber layer is formed by a semiconductor material, which we characterise by its bandgap energy E_g . In principle, only photons with energy higher than the band gap energy of the absorber can generate electron-hole pairs. Since the electrons and holes tend to occupy energy levels at the bottom of the conduction band and the top of the valence band, respectively, the extra energy that the electron-hole pairs receive from the photons is released as heat into the semicon-

ductor lattice in the *thermalisation* process. Photons with energy lower than the band gap energy of the semiconductor absorber are in principle not absorbed and cannot generate electron-hole pairs. Therefore these photons are not involved in the energy conversion process. The *non-absorption* of photons carrying less energy than the semiconductor band gap and the *excess energy* of photons, larger than the band gap, are the two main losses in the energy conversion process using solar cells. Both of these losses are thus related to the spectral mismatch between the energy distribution of photons in the solar spectrum and the band gap of a semiconductor material.

Shockley and Queisser call the efficiency that is obtained when taking the spectral mismatch losses into account the *ultimate efficiency*, that is given according to the hypothesis that '*each photon with energy greater than $h\nu_g$ produces one electronic charge e at a voltage of $V_g = h\nu_g/e'$* ' [23].

Let us now determine the fraction of energy of the incident radiation spectrum that is absorbed by a single-junction solar cell. When we denote λ_g as the wavelength of photons that corresponds to the band gap energy of the absorber of the solar cell, only the photons with $\lambda \leq \lambda_g$ are absorbed. The fraction p_{abs} of the incident power that is absorbed by a solar cell and

used for energy conversion can be expressed as

$$p_{\text{abs}} = \frac{\int_0^{\lambda_g} \frac{hc}{\lambda} \Phi_{\text{ph}, \lambda} d\lambda}{\int_0^{\infty} \frac{hc}{\lambda} \Phi_{\text{ph}, \lambda} d\lambda}, \quad (10.9)$$

where $\Phi_{\text{ph}, \lambda}$ is the spectral photon flux of the incident light as defined in Chapter 5. The fraction of the absorbed photon energy exceeding the bandgap energy is lost because of thermalisation. The fraction of the absorbed energy that the solar can deliver as useful energy is then given by

$$p_{\text{use}} = \frac{E_g \int_0^{\lambda_g} \Phi_{\text{ph}, \lambda} d\lambda}{\int_0^{\infty} \frac{hc}{\lambda} \Phi_{\text{ph}, \lambda} d\lambda}. \quad (10.10)$$

By combining Eqs. (10.9) and (10.10), we can determine the *ultimate conversion efficiency*,

$$\eta_{\text{ult}} = p_{\text{abs}} p_{\text{use}} = \frac{E_g \int_0^{\lambda_g} \Phi_{\text{ph}, \lambda} d\lambda}{\int_0^{\infty} \frac{hc}{\lambda} \Phi_{\text{ph}, \lambda} d\lambda}. \quad (10.11)$$

Figure 10.3 illustrates the fraction of the AM1.5 spectrum that can be converted into a usable energy by a crystalline silicon solar cell. Figure 10.4 shows the ultimate conversion efficiency of a solar cells band gap of a semiconductor absorber for three different radiation spectra, black-body radiation at 6000 K, AM0 and AM1.5 solar radiation spectra. The figure demonstrates that in the case of a crystalline silicon solar cell

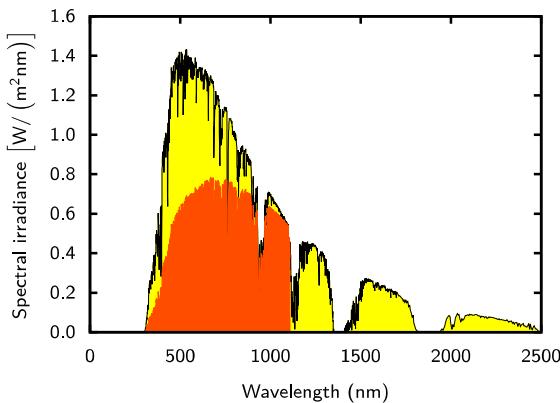


Figure 10.3: The fraction of the AM1.5 spectrum that can be converted into a usable energy by a crystalline silicon solar cell with $E_g = 1.12$ eV.

($E_g = 1.12$ eV) the losses due to the spectral mismatch account for almost 50%. It also shows that an absorber material for a single junction solar cell has an optimal band gap of 1.1 eV and 1.0 eV for the AM0 and AM1.5 spectra, respectively. Note that the maximum conversion efficiency for the AM1.5 spectrum is higher than that for AM0, while the AM0 spectrum has a higher overall power density. This is because of the fact that the AM1.5 spectrum has a lower power density in parts of the spectrum that are not contributing to the energy conversion process as can be seen in Fig. 10.3. The dips in the AM1.5 spectrum also result in the irregular shape of the conversion efficiency as function of the band gap.

10.2.2 Detail balance limit of the efficiency

Similar to *Shockley and Queisser* we now will formulate the *detail balance limit of the efficiency*. But before we start we briefly will discuss the reason that the ultimate efficiency formulated above is not physical for solar cells with temperatures higher than 0 K.

Let us estimated that the solar cell is embedded in an environment of ambient temperature of 300 K and that the solar cell temperature also is 300 K. As the solar cell will be in thermal equilibrium with its surroundings, it will absorb thermal radiation according to the ambient temperature and it will also emit the same amount

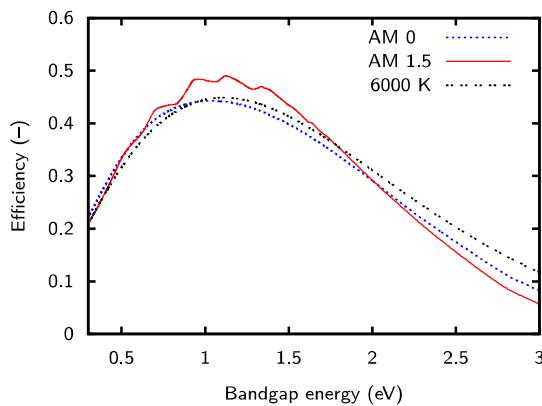


Figure 10.4: The ultimate conversion efficiency for the black body spectrum at 6000 K, the AM0 and AM1.5 solar radiation spectra, limited only by the spectral mismatch as a function of the band gap of a semiconductor absorber in single junction solar cells.

of radiation. Therefore recombination of electron-hole pairs will be present in the semiconductor leading to a recombination current different from zero. As we have seen in Eq. (9.1), the open-circuit voltage will be reduced with increasing recombination current, which is an efficiency loss.

We start the derivation of the detailed balance limit with recalling the definition of the efficiency from Eq. (9.4),

$$\eta = \frac{J_{\text{ph}} V_{\text{oc}} FF}{P_{\text{in}}} \quad (10.12)$$

For calculating η_{ult} we made the assumption that ‘each photon with energy greater than $h\nu_g$ produces one electronic charge e at a voltage of $V_g = h\nu_g/e$ ’. Under the same assumption, we obtain for the short circuit current density

$$J_{\text{ph}}(E_g) = -e \int_0^{\lambda_g} \Phi_{\text{ph}, \lambda} d\lambda \quad (10.13)$$

with $\lambda_g = hc/E_g$. Note that we here implicitly assumed that the photo-generated current J_{ph} is equivalent to the short circuit current. This assumption is valid as the recombination current originating from thermal emission is orders of magnitude lower than the photo-generated current. By combining Eqs. (10.11) and (10.13) we find

$$J_{\text{ph}} = -\frac{e}{E_g} P_{\text{in}} \eta_{\text{ult}} = -\frac{P_{\text{in}} \eta_{\text{ult}}}{V_g} \quad (10.14)$$

Let us now define the *bandgap utilisation efficiency* η_V

that is given by

$$\eta_V = \frac{V_{\text{oc}}}{V_g} \quad (10.15)$$

and tells us the fraction of the bandgap that can be used as open-circuit voltage (Shockley and Queisser use the letter v for this efficiency). We now combine Eqs. (10.12), (10.12) and (10.15) and find

$$\eta = \eta_{\text{ult}} \eta_V FF. \quad (10.16)$$

For determining the efficiency in the detailed balance limit, we therefore must determine the bandgap utilisation efficiency and the fill factor. Let us start with η_V .

According to Eq. (9.1), the open circuit voltage is given as

$$V_{\text{oc}} = \frac{kT}{e} \ln \left(\frac{J_{\text{ph}}}{J_0} + 1 \right). \quad (10.17)$$

The only unknown in this equation is the dark current J_0 . We assume the solar cell to be in *thermal equilibrium* with its surroundings at an ambient temperature of $T_a = 300$ K. Further, we assume that the solar cell absorbs and emits as a black body for wavelengths shorter than the bandgap wavelength of the solar cell absorber. For wavelengths longer than the bandgap we assume the solar cell to be completely transparent thus to neither absorb nor emit. This is the same assumption that we already used for the absorption of sunlight.

Using the equation for the *blackbody radiance* $L_{e\lambda}^{BB}$ as

given in Eq. (5.18a) we find for the radiative recombination current

$$\begin{aligned} J_0(E_g) &= -2e \int_0^{\lambda_g} \int_{2\pi} L_{e\lambda}^{BB}(\lambda; T_a) \cos \theta d\Omega d\lambda \\ &= -2e\pi \int_0^{\lambda_g} \frac{2hc^2}{\lambda^5} \left[\exp \left(\frac{hc}{\lambda k_B T_a} \right) - 1 \right]^{-1} d\lambda, \end{aligned} \quad (10.18)$$

where the factor 2 arises from the fact that we assume the solar cell to emit thermal radiation both at its front and back sides.

Combining Eqs. (10.15) with (10.17) we find

$$\eta_V(E_g) = kT/E_g \ln \left[\frac{J_{\text{ph}}(E_g)}{J_0(E_g)} + 1 \right]. \quad (10.19)$$

Figure 10.5 shows the bandgap utilisation efficiency for three different spectra of the incident sunlight. For a bandgap of 1.12 eV this efficiency is about $\eta_V \approx 77\%$.

For the fill factor we take the empirical but very accurate approximation

$$FF = \frac{v_{\text{oc}} - \ln(v_{\text{oc}} + 0.72)}{v_{\text{oc}} + 1} \quad (10.20)$$

with $v_{\text{oc}} = eV_{\text{oc}}/kT$. We already discussed this approximation in Eq. (9.3).

Figure 10.6 finally shows the Shockley-Queisser efficiency limit for three different spectra of the incident

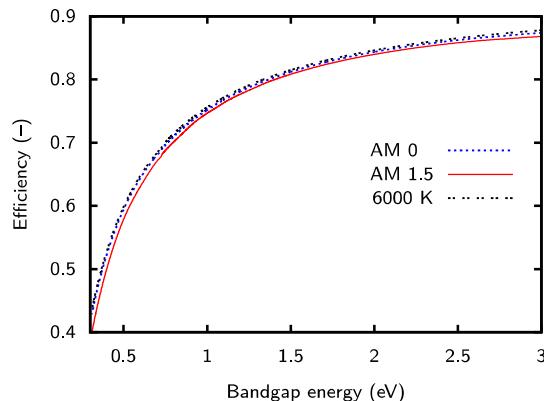


Figure 10.5: The bandgap utilisation efficiency η_V for the black body spectrum at 6000 K, and the AM0 and AM1.5 solar spectra.

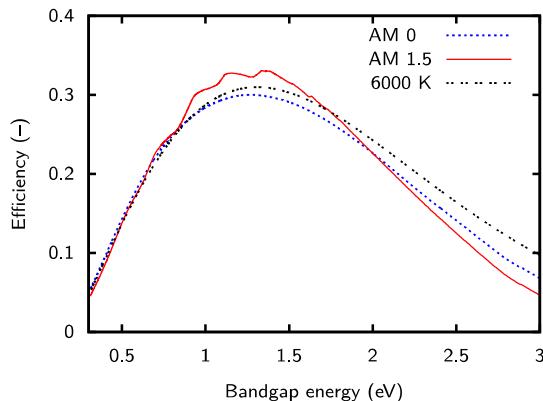


Figure 10.6: The Shockley-Queisser efficiency limit for the black body spectrum at 6000 K, and the AM0 and AM1.5 solar spectra.

light. For the AM1.5 spectrum the limit is about 33.1% at 1.34 eV. For AM0 it is 30.1% at 1.26 eV.

The major loss mechanisms that are taken into account in the Shockley-Queisser limit are illustrated in Fig. 10.7. The major losses are non-absorbed photons below the bandgap and thermalised energy of photons above the bandgap. The other losses are due to the voltage loss because of thermal radiation and the fill factor being different from 100%.

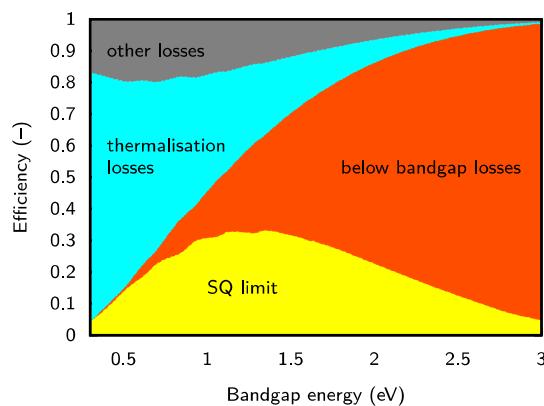


Figure 10.7: The major loss mechanisms in the Shockley Queisser limit. For this calculation the AM1.5 spectrum was used as incident light.

10.2.3 Efficiency limit for silicon solar cells

It is very important to note that the Shockley-Queisser (SQ) limit is not directly applicable to solar cells made from crystalline silicon. The reason for this is that silicon is a so-called *indirect bandgap* semiconductor as we will discuss in detail in Chapter 12. This means that Auger recombination, which is a non-radiative recombination mechanism, is dominant. For the derivation of the SQ limit we assumed that only radiative recombination is present. Clearly, this assumption cannot be valid for crystalline silicon solar cells. Several attempts to calculate the efficiency limit while taking radiative recombination mechanisms into account were performed in the past. A study from 2013 by Richter *et al.* derives an efficiency limits of 29.43% for silicon solar cells .

As the Shockley-Queisser limit only considers radiative recombination, it is most valid for direct band gap materials such as GaAs. Because of its direct band gap, radiative recombination is the limiting recombination mechanism for GaAs.

10.3 Other Losses

The Shockley-Queisser limit is a very idealised model. For example all optical losses are neglected. Now we

will discuss several other loss mechanisms that were not taken account above. At the end of this section we will derive at an equation for the efficiency were all these important losses are taken into account.

10.3.1 Optical losses

For the Shockley-Queisser limit, we only took the bandgap energy E_g into account for deriving the efficiency limit. However, the real performance is also strongly influenced by the optical properties given as the complex refractive index $\tilde{n} = n - ik$, which is a function of the wavelength.

As we already discussed in Section 4.3, a part of the light is reflected from and the other part is transmitted when light arrives on an interface between two media. The interface is therefore characterised by the wavelength dependent reflectance $R(\lambda)$ and transmittance $T(\lambda)$. All the reflections and transmissions at the different interfaces in the solar cell result in a total reflectance between the solar cell and the surrounding air. Hence, a part of the incident energy that can be converted into a usable energy by the solar cell is lost by reflection. We shall denote the total *effective* reflectance in the wavelength range of interest as R^* .

As we will discuss in more detail in Chapter 12, in most c-Si solar cells thin metal strips are placed on the

front side of the solar cell that serve as front electrode. The metal-covered area does not allow the light to enter the solar cell because it reflects or slightly absorbs the incident light. The area that is covered by the electrode effectively decreases the active area of the solar cell. When we denote the total area of the cell as A_{tot} and the cell area that is not covered by the electrode as A_f , the fraction of the active area of the cell is determined by the ratio

$$C_f = \frac{A_f}{A_{\text{tot}}}, \quad (10.21)$$

which is called the coverage factor C_f . These loss is called the *shading loss*. The design of the front electrode is of great importance since it should one the one hand minimise losses due to the series resistance of the front electrode, *i.e.* should be designed with sufficient cross-section. The optimal design of the front electrode is therefore a trade-off between a high coverage factor and a sufficiently low series resistance of the front electrode.

When light penetrates into a material, it will be partially absorbed as it propagates through the material. The absorption of light in the material depends on its absorption coefficient and the layer thickness, as we have seen in Section 4.4. In general, light is absorbed in all layers of the solar cell. All the absorption in layers different from the absorber layer is loss. It is called the *parasitic absorption*. Further, due to the limited thickness of the absorber layer, not all the light entering the absorber layer is absorbed. Incomplete absorption in

the absorber due to its limited thickness is an additional loss that lowers the energy conversion efficiency. The incomplete absorption loss can be described by the internal optical quantum efficiency IQE_{op} , which is defined as the probability of a photon being absorbed in the absorber material. Since there is a chance that a highly energetic photon can generate more than one electron-hole pair, we also define the *quantum efficiency for carrier generation* η_g which represents the number of electron-hole pairs generated by one absorbed photon. Usually η_g is assumed to be unity.

10.3.2 Solar cell collection losses

Not all charge carriers that are generated in a solar cell are collected at the electrodes. The photo-generated carriers are the excess carriers with respect to the thermal equilibrium and are subjected to the recombination. The carriers recombine in the bulk, at the interfaces, and/or at the surfaces of the junction. The recombination is determined by the electronic properties of materials that form the junction, such as density of states introduced into the band gap by the *R-G* centers. The concentration of *R-G* centers strongly influences the minority-carrier lifetimes as discussed in Chapter 7.

The contributions of both the electronic and optical properties of the solar cell materials to the photovoltaic performance are taken into account in the absolute

external quantum efficiency that we already defined in Chapter 9. The EQE can be approximated by

$$\text{EQE}(\lambda) = (1 - R^*) \text{IQE}_{\text{op}}(\lambda) \eta_g(\lambda) \text{IQE}_{\text{el}}(\lambda), \quad (10.22)$$

where the IQE_{el} denotes the electrical quantum efficiency and is defined as the probability that a photo-generated carrier is collected.

When we take the shading losses and the EQE into account, we find the short-circuit current density to be

$$J_{\text{sc}} = J_{\text{ph}} (1 - R^*) \text{IQE}_{\text{op}}^* \eta_g^* \text{IQE}_{\text{el}}^* C_f, \quad (10.23)$$

where the $*$ denote averages across the relevant wavelength range and J_{ph} is as defined in Eq. (10.13).

10.3.3 Additional limiting factors

We have seen in Section 10.2 that the V_{oc} of a solar cell depends on the saturation current J_0 and the photo-generated current J_{sc} of the solar cell. The saturation current density depends on the recombination in the solar cell. Recombination cannot be avoided and depends on the doping of the different regions (*n*-type and *p*-type regions) of a junction and the electronic quality of materials forming the junction. The doping levels and the recombination determine the bandgap utilisation efficiency η_V that we already defined in Section 10.2.

For determining the Shockley-Queisser limit we assumed for the FF that the solar cell behaves as an ideal diode. In a real solar cell, however, the FF is lower than the ideal value because of the following reasons:

- The voltage drop due to the *series resistance* R_s of a solar cell, which is introduced by the resistance of the main current path through which the photo-generated carriers arrive to the external circuit. The contributions to the series resistance come from the bulk resistance of the junction, the contact resistance between the junction and electrodes, and the resistance of the electrodes themselves.
- The voltage drop due to *leakage currents*, which is characterised by the shunt resistance R_p of a solar cell. The leakage current is caused by the current through local defects in the junction or due to the shunts at the edges of solar cells.

- The *recombination* in a non-ideal solar cell results in a decrease of the FF.

10.3.4 Conversion efficiency

Again, we start with the expression for the efficiency

$$\eta = \frac{J_{sc} V_{oc} FF}{P_{in}}. \quad (10.24)$$

Using Eqs. (10.23) and (10.14) and we find

$$\eta = \frac{\eta_{ult} V_{oc} FF}{V_g} (1 - R^*) IQE_{op}^* \eta_g^* IQE_{el}^* C_f \\ = p_{abs} p_{use} (1 - R^*) IQE_{op}^* \eta_g^* IQE_{el}^* C_f \eta_V FF, \quad (10.25)$$

where we used Eqs. (10.9) and (10.10). By filling in the definitions for p_{abs} , p_{use} , η_V and C_f , we obtain [33].

$$\eta = \underbrace{\frac{\int_0^{\lambda_g} \frac{hc}{\lambda} \Phi_{ph,\lambda} d\lambda}{\int_0^{\infty} \frac{hc}{\lambda} \Phi_{ph,\lambda} d\lambda}}_1 \underbrace{\frac{E_g \int_0^{\lambda_g} \Phi_{ph,\lambda} d\lambda}{\int_0^{\lambda_g} \frac{hc}{\lambda} \Phi_{ph,\lambda} d\lambda}}_2 \underbrace{(1 - R^*)}_3 \underbrace{IQE_{op}^*}_4 \underbrace{\eta_g^*}_5 \underbrace{IQE_{el}^*}_6 \underbrace{\frac{A_f}{A_{tot}}}_7 \underbrace{\frac{eV_{oc}}{E_g}}_7 \underbrace{FF}_8. \quad (10.26)$$

This describes the conversion efficiency of a solar cell in terms of components that represent particular losses in energy conversion.

1. Loss due to non-absorption of long wavelengths,
2. Loss due to thermalisation of the excess energy of

- photons,
3. Loss due to the total reflection,
 4. Loss by incomplete absorption due to the finite thickness,
 5. Loss due to recombination,
 6. Loss by metal electrode coverage, shading losses,
 7. Loss due to voltage factor,
 8. Loss due to fill factor.

10.4 Design Rules for Solar Cells

Now, as we extensively have discussed all the factors that limit the efficiency of a solar cell we are able to distill three *design rules*. We will use this design rules in Part III when we discuss different PV technologies. The three design rules are

1. Utilisation of the band gap energy,
2. Spectral utilisation,
3. Light trapping.

We now will take a closer look at each of these design rules.

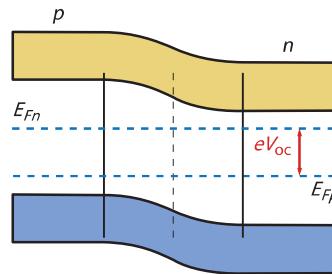


Figure 10.8: The open circuit voltage of a solar cell is determined by the splitting of the quasi-Fermi levels.

10.4.1 Bandgap utilisation

As we have seen earlier in this chapter, the open circuit voltage V_{oc} is always below the voltage V_g corresponding to the bandgap, which we characterised with the bandgap utilisation efficiency η_V . The open circuit voltage is determined by the extent to which the quasi-Fermi levels are able to split, which is limited by the charge carrier recombination mechanisms. We will discuss that various PV materials have different recombination mechanism that limit the utilisation of the band gap energy.

Let us now discuss the first design rule, *bandgap utilisation*. Figure 10.8 shows a *p-n*-junction, with the *p*-doped region on the left and the *n*-doped region right. Fur-

ther the quasi-Fermi levels are depicted. The extent of splitting between the quasi-Fermi levels determines the open circuit voltage.

The open circuit voltage, for example expressed in Eq. (10.17) also can be expressed in terms of the generation rate G_L , the life time τ_0 of the minority charge carriers and the intrinsic density of the charge carriers n_i in the semiconductor material,

$$V_{oc} = \frac{2kT}{e} \ln \left(\frac{G_L \tau_0}{n_i} \right). \quad (10.27)$$

The derivation of this equation is out of the scope of this book.

Let us now take a closer look on Eq. (10.27). If we increase the irradiance, or in other words, the generation rate of charge carriers, the open circuit voltage is increased. This is a welcome effect which is utilised in concentrator photovoltaics, which we will briefly discuss in Section 13.2. Secondly, we see that the lifetime τ_0 plays an important role. The larger the lifetime of the minority charge carrier, the larger the open circuit voltage can be. Or in other words, the longer the lifetime, the larger the possible splitting between the quasi-Fermi levels and the larger the fraction of the bandgap energy that can be utilised.

The lifetime of the minority charge carrier is determined by the recombination rate. As discussed in

Chapter 7, we have to consider three different recombination mechanisms: radiative, Shockley-Read-Hall, and Auger recombination. While radiative and Auger recombination depend on the semiconductor itself, SRH recombination is proportional to the density of traps or impurities in the semiconductor. In the three recombination mechanisms energy and momentum are transferred from charge carriers to phonons or photons.

The efficiency of the different recombination processes depends on the nature of the band gap of the used semiconductor material used. We distinguish between *direct* and *indirect* bandgap semiconductors. Crystalline silicon is an indirect band gap material. The radiative recombination in an indirect band gap material is inefficient and recombination will be dominated by the Auger mechanism. For direct band gap materials such as GaAs under moderate illumination conditions, radiative recombination will be the dominant loss mechanism of charge carriers. For very high illumination conditions, Auger recombination starts to play a role as well.

To summarise, we find that in the defect rich solar cells, the open circuit voltage is limited by the SRH recombination. In low-defect solar cells based on indirect band gap materials, the open circuit voltage is limited by Auger recombination. In low-defect solar cells based on direct band gap materials, the open circuit voltage is limited by radiative recombination.

Besides band gap utilisation, it also is important to discuss the relationship between the maximum thickness for the absorber layer of a solar cell and the dominant recombination mechanism. As we have seen in Chapter 7, the recombination mechanism also affects the diffusion length of the minority charge carrier. The diffusion length L_n of minority electrons is given by

$$L_n = \sqrt{D_n \tau_n}, \quad (10.28)$$

where D_n is the diffusion coefficient and τ is the lifetime of the minority charge carrier. Similarly, we can formulate the diffusion length for minority holes, L_p .

It is important to realise that the thickness of the absorber layer should not exceed the diffusion length. To understand this requirement, we consider photons that would penetrate far into the absorber layer being absorbed. We want these charge carriers to be separated at the p - n junction or at the back contact. If the distance of these charge carriers from the p - n junction or the back contact is exceeding the diffusion length, the excited charge carriers will recombine within the typical diffusion length before arriving at the p - n junction or back contact. In other words, their lifetime is too short. This means that all charge carriers generated at a distance larger than the diffusion length from the p - n junction or the back contact cannot be collected and hence are lost. If the charge carriers are generated with diffusion length, they can be collected. This means that the

diffusion length of the minority charge carrier, limits the maximum thickness of the solar cell.

To summarise this section, the open circuit voltage is limited by the dominant recombination mechanism. The dominance of radiative, Auger or Shockley-Reed-Hall recombination depends on the type of semiconductor materials used in the solar cell and the illumination conditions. We will discuss several different cases on Part III on PV technology.

10.4.2 Spectral utilisation

The spectral utilisation is mainly determined by the choice of materials from which the solar cell is made of. As we have seen in Section 10.2, and mainly Eq. (10.13), the photocurrent is determined by the bandgap of the material. For a bandgap of 0.62 eV corresponding to a wavelength of 2000 nm, we could theoretically generate a short circuit current density of 62 mA/cm². If we consider c-Si, having a band gap of 1.12 eV (1107 nm), we arrive at a theoretical current density of 44 mA/cm².

The optimal bandgap for single-junction solar cells is determined by the Shockley-Queisser limit, as illustrated in Fig. 10.6. For single junction solar cells, semiconductor material as such silicon, gallium arsenide and cadmium telluride have a band gap close to the optimum.

In Part III we will discuss various concepts that allow to surpass the Shockley-Queisser limit. Here, we will briefly discuss the concept of *multiplication solar cells*. In these devices, solar cells with different bandgaps are stacked on top of each other. As illustrated in 10.9, the excess energy can be reduced significantly, and the spectral utilisation will improve.

10.4.3 Light trapping

The third and last design rule that we discuss is *light trapping*. In an ideal solar cell, all light that is incident on the solar cell should be absorbed in the absorber layer. As we have discussed in Section 4.4, the intensity of light decreases exponentially as it travels through an absorptive medium. This is described by the Lambert-Beer law that we formulated in Eq. (4.25),

$$I(d) = I_0 \exp(-\alpha d). \quad (10.29)$$

From the Lambert-Beer law it follows that at the side, at which the light is entering the film, more light is absorbed in reference to the back side. The total fraction of the incident light absorbed in the material is equal to the light intensity entering the absorber layer minus the intensity transmitted through the absorber layer,

$$I^{\text{abs}}(d) = I_0[1 - \exp(-\alpha d)]. \quad (10.30)$$

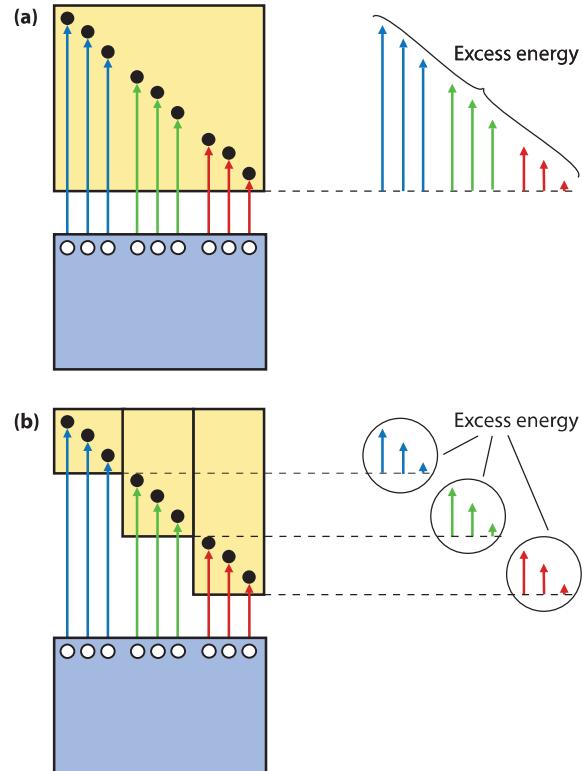


Figure 10.9: Illustrating the lost excess energy in (a) a single-junction and (b) a multi-junction solar cells.

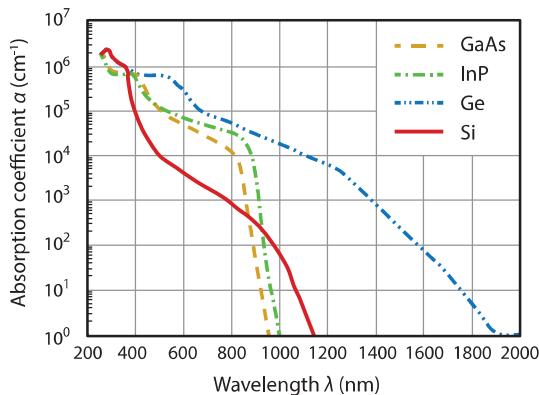


Figure 10.10: Absorption coefficients of different semiconductors.

Ideally, we would like to absorb a solar cell 100% of the incident light. Such an absorber is called *optically thick* and has a transmissivity very close to 0. As we can see from the Lambert-Beer law, this can be achieved by either absorbers with a large thickness d or with very large absorption coefficients α .

Figure 10.10 shows the absorption coefficients for four different semiconductor materials: germanium (Ge), silicon (Si), gallium arsenide (GaAs) and indium phosphide (InP). We notice that germanium has the lowest band gap. It starts to absorb at long wavelengths,

which corresponds to a low photon energy. GaAs has the highest band gap, as it starts to absorb light at the smallest wavelength, or highest photon energy. Secondly, if we focus on the visible spectral part from 300 nm up 700 nm, we see that the absorption coefficients of InP and GaAs are significantly higher than for silicon. This is related to the fact that InP and GaAs are direct band gap materials as discussed earlier. Materials with an indirect band gap have smaller absorption coefficients. Only in the very blue and ultraviolet part below 400 nm, Si has a direct band gap transition. Silicon is a relatively poor absorber. Therefore for the same fraction of light thicker absorber layers are required in comparison to GaAs.

In general, for all semiconductor materials the absorption coefficient in the blue is orders of magnitude higher than in the red. Therefore the penetration depth of blue light into the absorber layer is rather small. In crystalline silicon, the blue light is already fully absorbed within a few nanometers. The red light requires an absorption path length of 60 μm to be fully absorbed. The infrared light is hardly absorbed, and after an optical path length of 100 μm only about 10% of the light intensity is absorbed.

As the absorption of photons generates excited charge carriers, the wavelength dependence of the absorption coefficient determines the local generation profile of the charge carriers. At the front side where the light enters

the absorbing film, the generation of charge carriers is significantly higher than at the back side. It follows that the EQE values measured in the blue correspond to charge carriers generated close to the front of the solar cell, whereas the EQE in the red part represents charge carriers generated throughout the entire absorber layer.

Further, it is important to reduce the optical loss mechanisms such as *shading losses*, *reflection*, and *parasitic absorption* that we already discussed in Section 10.3.1.

For reducing the reflection, anti-reflective coatings (ARC) can be used. Light that is impinging onto a surface between two media with different refractive indices will always be partly reflected and partly transmitted. In order to reduce losses, it is important to minimise these reflective losses.

The first method is based on a clever utilisation of the Fresnel equations that we discussed in Section 4.3. For understanding how this can work, we first will take a look at interfaces with silicon, the most common used material for solar cells. Let us consider light of 500 nm wavelengths falling onto an air-silicon interface perpendicularly. At 500 nm, the refractive index of air is $n_0 = 1$ and that of silicon is $n_s = 4.3$. With the Fresnel equations we hence find that the optical losses due to reflection are significant with 38.8%.

The reflection can be significantly reduced by introducing an interlayer with a refractive index n_1 with a

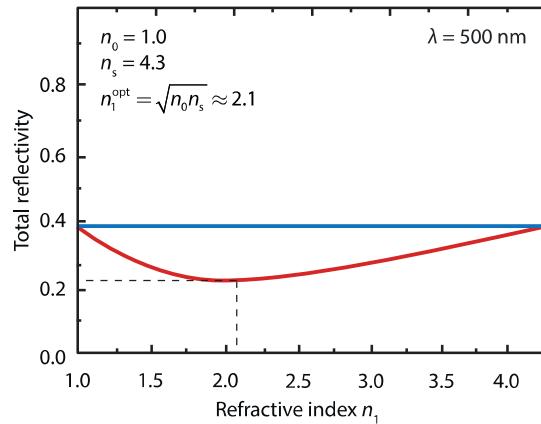


Figure 10.11: Illustrating the effect of an interlayer with refractive index n_1 in between n_0 and n_s on the reflectivity.

value in between that of n_0 and n_s . If no multiple reflection or interference is taken into account, it can easily be shown that the reflectivity becomes minimal if

$$n_1 = \sqrt{n_0 n_s}. \quad (10.31)$$

This is also seen in Fig. (10.11), where n_1 takes all the values in between n_0 and n_s . In this example, including a single interlayer can reduce the reflection at the interface from 38.8% down to 22.9%. If more than one interlayers are used, the reflection can be reduced even further. This technique is called *refractive index grading*.

In another approach constructive and destructive interference of light is utilised. In Chapter 4 we already discussed that light can be considered as an electromagnetic wave. Waves have the interesting properties that they can interfere with each-other, they can be *superimposed*. For understanding this we look at two waves A and B that have the same wavelength and cover the same portion of space,

$$A(x, t) = A_0 e^{ikx - i\omega t}, \quad (10.32a)$$

$$B(x, t) = B_0 e^{ikx - i\omega t + i\phi}. \quad (10.32b)$$

The letter ϕ denotes the *phase shift* between A and B and is very important. We can superimpose the two

waves by simply adding them

$$\begin{aligned} C(x, t) &= A(x, t) + B(x, t) \\ &= A_0 e^{ikx - i\omega t} + B_0 e^{ikx - i\omega t + i\phi} \\ &= A_0 e^{ikx - i\omega t} + B_0 e^{ikx - i\omega t} e^{i\phi} \\ &= (A_0 + B_0 e^{i\phi}) e^{ikx - i\omega t}. \end{aligned} \quad (10.33)$$

The amplitude of the superimposed wave is thus

$$C_0 = A_0 + B_0 e^{i\phi}. \quad (10.34)$$

Depending of the phase shift, the superimposed wave will be stronger or weaker than A and B . If ϕ is *in phase*, i.e. a multiple of 2π , i.e. $0, 2\pi, 4\pi, \dots$, we will have maximal amplification of waves,

$$C_0 = A_0 + B_0 \cdot 1 = A_0 + B_0. \quad (10.35)$$

This situation is called *constructive interference*.

But if ϕ is *in antiphase*, i.e. from the set $\pi, 3\pi, 5\pi, \dots$ we have maximal attenuation of the waves, or *destructive interference*,

$$C_0 = A_0 + B_0 \cdot (-1) = A_0 - B_0. \quad (10.36)$$

Based on this principle, we can design an anti-reflection coating, as illustrated in Fig. 10.12. The green wave shows the reflection back from the first interface and the red wave shows the wave which is reflected back

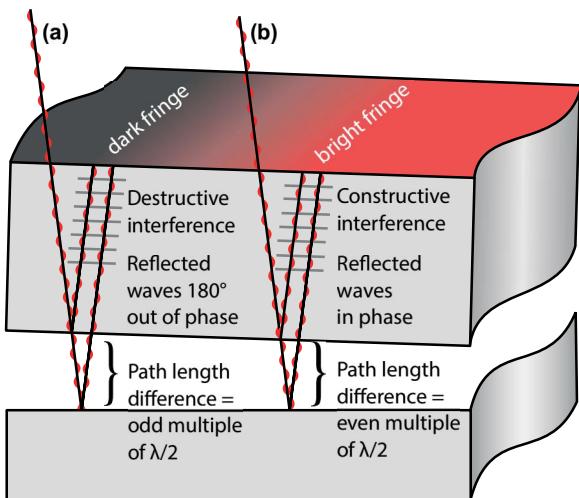


Figure 10.12: Illustrating the working principle of an anti-reflective coating based on interference [34].

from the second interface. If we look at the two waves coupled out of this system, they appear to be in anti-phase. As a result the total amplitude of the electric field of the outgoing wave is smaller and hence the total irradiance coupled out of the system is smaller as well.

It can be shown easily that they have antiphase, when the product of the refractive index and thickness of the interlayer is equal to the wavelength divided by four,

$$n_1 d = \frac{\lambda_0}{4}. \quad (10.37)$$

Using an antireflection coating based on interference demands that the typical length scale of the interlayer thickness must be in the order of the wavelength.

The last approach that we discuss for realising anti-reflective coating is using *textured interfaces*. Here, we consider the case where the typical length scales of the surface features are larger than the typical wavelength of light. In this case, which is also called the *geometrical limit*, the reflection and transmission of the light rays are fully determined by the Fresnel equations and Snell's law.

The texturing helps to enhance the coupling of light into the layer. For example, for light that is perpendicular incident, light that is reflected at one part of the textured surface can be reflected into angles in which the trajectory of the light ray is incident a second time

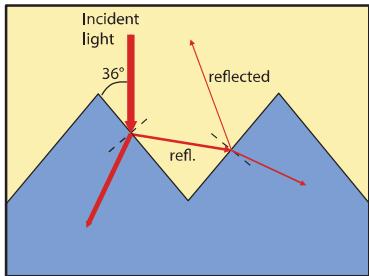


Figure 10.13: Illustrating the effect of texturing.

somewhere else on the interfaces, as illustrated in Fig. 10.13. Here, another fraction of the light will be transmitted into the layer and effectively less light will be reflected, when compared to a flat interface in-between the same materials.

In summary, we have discussed three types of anti-reflective coatings: Rayleigh films with intermediate refractive indexes, anti-reflection coatings based on destructive interference and enhanced in-coupling of light due to scattering at textured interfaces.

At the end of this chapter, we want to discuss something that is very important for thin-film solar cells. Because of reducing production cost but also because of reducing bulk recombination it is desirable to have the absorber layer as thin as possible. On the other hand,

it should be *optically thick* in order to absorb as much light as possible. In principle, if the light can be reflected back and forth inside the absorber until everything is absorbed. However, at every internal reflection part of the light is transmitted out of the film.

But if the light would travel through the layer at an angle larger than the critical angles of the front and back interfaces of the absorber, it could stay there until everything is absorbed without any loss. Unfortunately, it is very difficult if not impossible to design such a solar cell.

Note that textured interfaces do not help in this case. For the same reason more light can be coupled into the absorber using such a texturisation, also more light is coupled out of the absorber!