Fundamentals and Processes for Photovoltaic Devices

MICRO 655

Prof. Christophe Ballif and Dr. Franz-Josef Haug

Preface

This script is intended to accompany the course "Fundamentals and processes for photovoltaic devices" which is taught in the MSc program of EPFL since 2009. It was created by Johannes Seif in 2012 and extended by him over several years. In 2016 and 2017 maintenance was taken over by Jean Cattin and in 2018 by Franz-Josef Haug. In 2018 the introduction was revised and the material on optics was extended and collected into a separate chapter. On this occasion also the appendix chapters on Schottky contacts, the effects of high doping, and on isotype junctions were added. In 2019 the chapters on the *p*-*n*-junction and on advanced device architectures were revised. The material on finite cell thickness was regrouped and partially moved to the appendix. Many of the figures were redrawn during that process. In 2020, the chapters on generation and recombination were fully revised, material on surface recombination was moved to the appendix, and a bibliography with links to key publications was added. During 2021 and 2022, the appendix on the efficiency limit was added and the text was reviewed for typos and minor issues. In 2023, a short section on sustainability, energy payback-time and carbon footprint was added. Moreover, the discussion of effective masses in silicon was moved to the appendix, thus allowing a more detailed description of the underlying band structures and energy surfaces. In 2024, the section on defect recombination was shortened by moving the material relevant for c-Si to the appendix where it is now discussed together with a brief introduction to data analysis and a presentation of typical examples.

Franz-Josef Haug, Christophe Ballif
Neuchâtel, Spring 2025

Contents

List of abbreviations

General

AMx D/G	Air Mass (x: atmospheres to go through) direct/global
ARC	Anti Reflection Coating
CB	Conduction Band
CBM	Conduction Band Minimum
DOS	Density of States
EQE	External Quantum Efficiency
ETM	Electron Transport Material
HTM	Hole Transport Material
IQE	Internal Quantum Efficiency
LCA	Life Cycle Analysis
LCOE	Levelised Cost of Electricity
MPP	Maximum Power Point
PR	Performance Ratio
PV	Photovoltaic (obviously)
QFL	Quasi Fermi-level
SCR	Space Charge Region (depletion region)
SR	Spectral Response
STC	Standard Testing Conditions
SRV	Surface recombination velocity
VB	Valence Band
VBM	Valence Band Maximum

Symbols in equations / parameters of the I(V) characteristic

A	Cell area
A^*	Richardson constant for Schottky barriers
C	Capacitance
D_{it}	Density of interface states
Δn	Excess carrier density
E_C	Energy of the conduction band edge
E_F	Fermi energy, may split into $E_{F,n}$ and $E_{F,p}$ for electrons and holes
E_V	Energy of the valence band edge
η	Solar cell efficiency
FF	Fill factor
G	Generation rate

G_F	Geometry factor
H (H')	Bulk (emitter) thickness
χ	Electron affinity (energy) of a semiconductor
ϵ	Permittivity of a material, e.g. ϵ_{Si}
I_{MPP} (j_{MPP})	MPP current (density)
I_L (j_L)	Light generated current (density)
I_{sc} (j_{sc})	Short circuit current (density)
I_0 (j_0)	Reverse saturation current (density)
iV_{oc}, iFF	implied iV_{oc} and implied FF as determined from a lifetime measurement
k	Wave vector, in vacuum $k_0 = 2\pi/\lambda_0$
kT	Thermal energy where k is the Boltzmann constant
λ	Wavelength, in vacuum λ_0
L_n (L_p)	Diffusion length of electrons (holes)
κ	Extinction coefficient, imaginary part of the refractive index n
m_e^* (m_h^*)	Effective mass of electrons (holes)
μ_e (μ_h)	Electron (hole) mobility
n	Band index
n	Diode quality factor, usually n_1 and n_2
n	Electron density; the most common meaning of n in this text, n_0 in equilibrium
n	Refractive index, usually n_{Si} and complemented by the extinction coefficient κ
N_A	Acceptor density
N_C	Effective density of states of the CB
n_i	Intrinsic carrier density
N_D	Donor density
N_V	Effective density of states of the VB
p	Hole density, p_0 in equilibrium
pFF	pseudo pFF, determined without series resistance from a SunsVoc measurement
Φ	Electrostatic potential
Ψ_n (Ψ_n)	Electron (hole) potential from midgap
Q_f	Charge density of fixed charges
Q_{it}	Charge density of interface states
ρ	Volume charge density, e.g. $\rho = qN_D^+$ for ionised donors
R	Reflectivity
R	Recombination rate, e.g. R_{rad} or R_{Auger}
R_p	Parallel resistance (shunt resistance)
R_s	Serial resistance
σ	Surface charge density, e.g. $\sigma_{it} = qQ_{it}$ for the interface charge
S, S_{eff}	Surface recombination velocity
T	Absolute temperature
τ_{eff}	Effective lifetime of minority carriers
τ_r	Relaxation time of majority carriers (semiconductors) or valence electrons (metals)
V_{bi}	Built-in voltage
V_{MPP}	Voltage at MPP
V_{oc}	Open circuit voltage
w (w_n, w_p)	Depletion layer width (on n- and p-side, respectively)
W_m	Work function (energy) of a metal

Useful constants and numerical values

Fundamental constants

Boltzmann constant	k	=	$1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$
Electron mass	m_e	=	$9.1 \times 10^{-31}\,\mathrm{kg}$
Elementary charge	q	=	$1.6 \times 10^{-19}\,\mathrm{A\,s}$
Planck constant	h	=	$6.63 \times 10^{-34}\,\mathrm{kg\,m^2\,s^{-1}}$
Speed of light	c	=	$2.99 \times 10^{8}\,\mathrm{m\,s^{-1}}$
Vacuum permittivity	ϵ_0	=	$8.85 \times 10^{-12}\,\mathrm{F\,m^{-1}}$

Energies

British thermal unit	1 BTU	=	$1055\,\mathrm{J}$
Calorie	1 cal	=	$4.184\,\mathrm{J}$
Electronvolt	$1\,\mathrm{eV}$	=	$1.6 \times 10^{-19}\,\mathrm{J}$
kWh	$1\,\mathrm{kW\,h}$	=	$3.6 \times 10^{6}\,\mathrm{J}$
Solar irradiation (outside atmosphere)	ϕ_{AM0}	=	$1366\,\mathrm{W\,m^{-2}}$
Ton of coal equivalent	1 tce	=	$29.3\,\mathrm{GJ}$
Ton of oil equivalent	1 toe	=	$41.8\,\mathrm{GJ}$
Thermal energy ($300\,\mathrm{K}$)	kT	=	$0.026\,\mathrm{eV}$

Numerical values for Si (at RT)

Lattice constant	a	=	$0.543\,\mathrm{nm}$
Density	ρ	=	$2.33\,\mathrm{g\,cm^{-3}}$
Relative permittivity (DC)	ϵ_{Si}	=	11.56
Atomic density (8 atoms per unit cell)	N	=	$5 \times 10^{22}\,\mathrm{cm^{-3}}$
Electron affinity	χ	=	$4.05\,\mathrm{eV}$
Bandgap	E_g	=	$1.12\,\mathrm{eV}$
Effective density of states (CB)	N_C	=	$2.86 \times 10^{19}\,\mathrm{cm^{-3}\,eV^{-1}}$
Effective density of states (VB)	N_V	=	$3.1 \times 10^{19}\,\mathrm{cm^{-3}\,eV^{-1}}$
Effective mass			
- electron conductivity	m_e^*/m_e	=	0.28
- hole conductivity	m_h^*/m_e	=	0.41
- electron DOS	m_e^*/m_e	=	1.09
- hole DOS	m_h^*/m_e	=	1.15
Intrinsic carrier density (undoped)	n_i	=	$9.65 \times 10^{9}\,\mathrm{cm^{-3}}$
Thermal velocity of electrons	$v_{th,e}$	=	$\sqrt{3kT/m_e} = 2.0 \times 10^{7}\,\mathrm{cm\,s^{-1}}$
Thermal velocity of holes	$v_{th,h}$	=	$\sqrt{3kT/m_h} = 1.7 \times 10^{7}\,\mathrm{cm\,s^{-1}}$

Chapter 1

Fundamentals and general knowledge

The sun provides us with an ample amount of energy. In just a little over one hour, the equivalent to the world's annual energy consumption is radiated to the earth's surface. Even though the irradiation varies with location, latitude, elevation, and climate, solar energy is available all over the globe. To tap into this source, Photvoltaics (PV) is unique as it converts sunlight directly into electricity without moving parts, and because it works equally well for small applications like calculators and large-scale installations like PV power farms.

1.1 Solar irradiance and air mass

Let us start by looking at the energy source of PV, the sun. The sun generates a power of $3.9 \times 10^{26}\,\mathrm{W}$ by fusion processes in its core. Various convective and radiative processes transport the power to the surface where the photosphere emits visible and IR radiation into space. Figure 1.1 shows that the emitted spectrum resembles the one of a black body at 5800 K (the sun's surface temperature) except for slight modifications due to the chemical composition of the photoshpere. To calculate the power density arriving on earth, we divide this power by the area of a spherical shell with the average radius of the earth's orbit. This yields a value of $1366\,\mathrm{W\,m^{-2}}$ and is equal to the integral over the solar spectrum in figure 1.1 with respect to the wavelength.

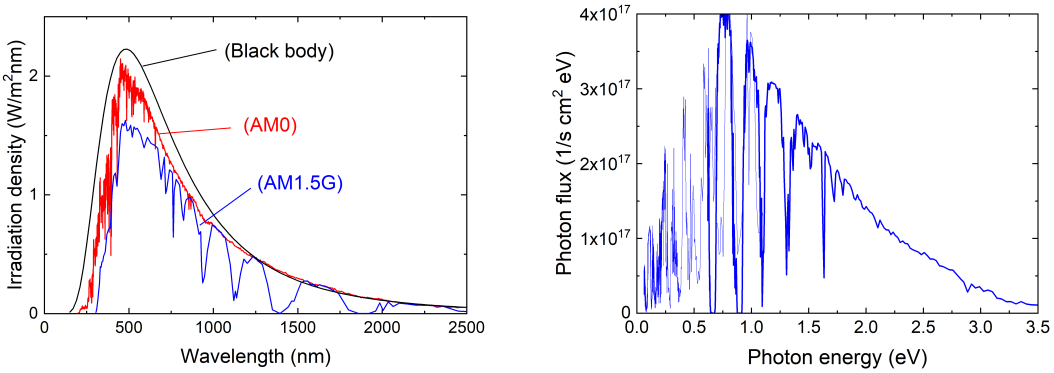


Figure 1.1: The left panel compares the characteristic of black body radiation (black) to the solar spectra measured in earth orbit (red) and on the surface of the earth (blue). The right panel illustrates the photon flux density of the latter with respect to the incident photon energy.

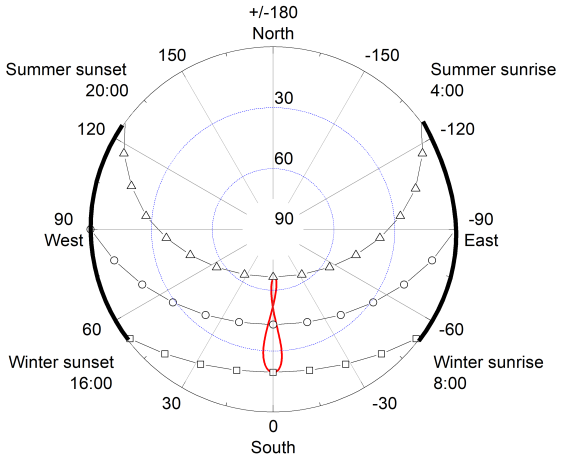


Figure 1.2: Solar chart for a latitude of 47° North, an average value for Switzerland (left). The symbols refer to full hours on the days of winter solstice (squares), equinox (circles), and summer solstice (triangles). The red characteristic illustrates the analemma for the true noon position throughout the year. The image to the right shows the sun-disk of Nebra, a bronze-age artifact from 2000 BC. Its golden arcs at the left and at the right reproduce the azimuth angles of sunrise and sunset (image: WikiCommons).

On the surface of the earth, the spectrum is changed because of absorption on its way through the atmosphere, e.g. by water vapour, ozone, CO_2, etc. Additional changes are due to scattering at water droplets and dust particles. Scattering processes are most pronounced for short wavelengths, i.e. blue light, thus explaining the blue colour of our sky. On average, the scattering phenomena are proportional to the length of the light path through the atmosphere. Since this length can be related to the angle α between the sun and the latitude of the observer, the *air mass* (AM) was introduced to classify illumination spectra.

$$\mathrm{AM}x \quad \text{with} \quad x = 1/\cos(\alpha) \tag{1.1}$$

Accordingly, AM0 is the spectrum in earth orbit, AM1.0 the one at the equator. For solar energy research, standardisation agencies in Europe and the US defined AM1.5 for a latitude of 48° North. As light gets scattered in the atmosphere, we must distinguish between direct and diffuse irradiation. The former comes directly from the sun, the latter comes uniformly from all directions above the horizon. Thus, we distinguish for a latitude of 48°:

- AM1.5D (direct): the light coming directly from the sun
- AM1.5G (global): all received light, i.e. 90% of direct light and 10% of diffuse radiation coming from all directions after scattering from the atmosphere and aerosols or reflection from the ground

Figure 1.1 shows the standardised spectral illumination intensities for the two cases of AM0 and AM1.5G. In the former, the integrated power density is $1366\,\mathrm{W\,m^{-2}}$ as described above, in the latter it is $1000\,\mathrm{W\,m^{-2}}$. The right panel of figure 1.1 shows the AM1.5G spectrum after scaling to the photon flux with respect to the energy of the photons. Note that the maximum of the incident power density is assumed in the visible at 500 nm whereas the maximum photon flux occurs at 0.75 eV which corresponds to an IR wavelength of 1650 nm.

In addition to the latitude of the observer, the angle of incidence varies during the day and the year. As this angle goes into the air mass, it changes the irradiation that a solar cell receives. Thus, we need a *solar chart* as shown in figure 1.2 for 47° North, an average value for Switzerland. The outermost circle corresponds to an elevation of 0° above the horizon, i.e. conditions of sunrise and sunset. The centre of the graph is at 90°, i.e. vertically above the observer. For a given day, the maximum elevation angles are observed around noon along the axis between the centre and the southern horizon. On this line, the elevation angles are located within a band of 23.5° above and below the latitude angle of the observer.

When the position of the sun is checked at a given hour throughout the year, say at noon and without daylight saving, it traces the 8-shaped characteristic shown in red which is called *analemma*. Thus, the highest position above the horizon is not always observed exactly at 12:00, but up to 15 min ahead or behind according the *equation of time*. The deviation is due to two effects, the elliptic orbit of the earth around the sun, and the inclination of the earth-axis with respect to the plane of its orbit around the sun.

1.2 Standard Test Conditions

In order to compare the results obtained at different labs, solar cells and modules are measured under Standard Test Conditions (STC). In addition to an incident spectrum of AM1.5G with an intensity of $1000\,\mathrm{W\,m^{-2}}$ ($= 0.1\,\mathrm{W\,cm^{-2}}$),[1] these assume additionally a device temperature of 25 °C. Much effort goes into the optimisation of solar cells for exactly these conditions, but in real situations out in the field, a solar cell is hardly ever exposed to any of them. First of all, the cell will likely be part of a module where the spectrum is changed by the encapsulation material, a glass cover, and likely a layer of dust. Next, the spectrum and the illumination intensity depend on the air mass which changes continuously throughout the day and for every day of the year according to the elevation angle shown in figure 1.2. Moreover, typical operation temperatures differ widely from 25 °C. Under full illumination in summer and without wind chill, modules can heat up to temperatures of 60 - 80 °C. The effect of temperature will be discussed in section 1.8.

1.3 Units and Rating of cells

A variety of physical quantities show up in PV. It is important to distinguish their units.

- The STC illumination spectrum is generally tabulated as areal power density and with respect to wavelength interval. Thus, $1000\,\mathrm{W\,m^{-2}}$ means that one square meter receives a power of 1000 W or an energy of 1 kW h every hour.

- In figure 1.1 we saw and alternative representation in terms of photon flux per energy-interval ($\mathrm{m^{-2}\,eV^{-1}}$). The choice of the interval is important as it adds a scaling factor, and as a result the maxima occur either in the visible or in the IR as discussed above.

- The solar cell efficiency is usually denoted by the symbol η. It is the ratio between the power density that a solar cell with given size delivers at its maximum power point, and the incoming power of the illumination under STC.

[1] Note that it is not obvious to obtain a good spectrum with a homogeneous power density across a certain area. This is one of the challenges for suppliers of characterisation equipment like sun simulators.

$$\eta_{STC} = \frac{P_{max}}{1000\,\mathrm{W\,m^{-2}}} \tag{1.2}$$

- Installations are usually not defined by their surface area, but by the output power that they can deliver under ideal conditions of STC. This power is normally specified as *Watt peak* and denoted with the unit W_p.
- In the context of PV, energies are seldom measured in Joule where $1\,\mathrm{J} = 1\,\mathrm{W\,s}$, but more often in kW h, a unit of energy that we also know from our electricity bill. Thus, $1\,\mathrm{kW\,h} = 3.6 \times 10^6\,\mathrm{J}$.
- The energy yield of a system is obtained by measuring how much power it delivers in a certain period of time, normally a year. As a *rule of thumb*, at the latitude of Central Europe a solar installation of $1\,\mathrm{kW_p}$ can be expected to yield 1000 - 3000 kW h. In other words, there are approximately 1000 - 2000 h of ideal sunshine throughout the year.

1.4 Basic properties of solar cells

In 2020, the most commonly used design for solar cells on the market was the one shown in figure 1.3. It consists of a crystalline p-type silicon wafer, therefore holes are the *majority carriers*. Nevertheless, a few electrons will still be present by thermal excitation. As their density is low, we call them *minority carriers*. Normally, an n-doped region is formed at the front side by diffusion of phosphorus atoms, and the majority carriers in that region are electrons. In more modern cells designs the polarities are interchanged, but the following remains valid also for those cells.

We can imagine that the majority carriers move like an ideal gas between the ionized dopants (and the atoms of the lattice), such that globally the two materials are neutral. However, in the region close to the junction the two types of majority carriers feel a concentration gradient that drives them towards the oppositely doped region. As they move towards the interface, they leave behind their ionised dopants, and as soon as they cross the interface, they will recombine with carriers of the opposite type. Thus, positive phosphorus ions remain at the interface on the n-side and negative boron ions on the p-side. The region of the ionized dopants is called *space charge region (SCR)* and the electric field generated between them exactly balances the diffusion process of the carriers. In figure 1.3 the SCR is illustrated in green to denote the fact that it is depleted of majority carriers. We have to keep in mind that in typical silicon solar cells it extends only over a few hundreds of nanometers whereas most of the crystal is neutral and therefore free of any electric field.

When light impinges on the semiconductor, energetic photons can excite electrons from filled states in the valence band (VB) into empty states of the conduction band (CB), leaving behind holes in the VB. Thus, every absorbed photon creates an electron and a hole that are added to the carriers already present by thermal excitation and doping. Normally, the density of these photogenerated excess carriers is less than the doping density in the solar cell and therefore the additional carriers hardly change the majority carrier densities from their equilibrium values. The situation is very different for the minority carriers because photogeneration surpasses their equilibrium density by several orders of magnitude. The additional minority carriers have a tendency to recombine, such that once the illumination stops, their densities will return to their equilibrium densities within a period called *lifetime*. Under continuous illumination there is a *steady state* that exactly balances generation and recombination (and ideally extraction), resulting in minority carrier densities higher

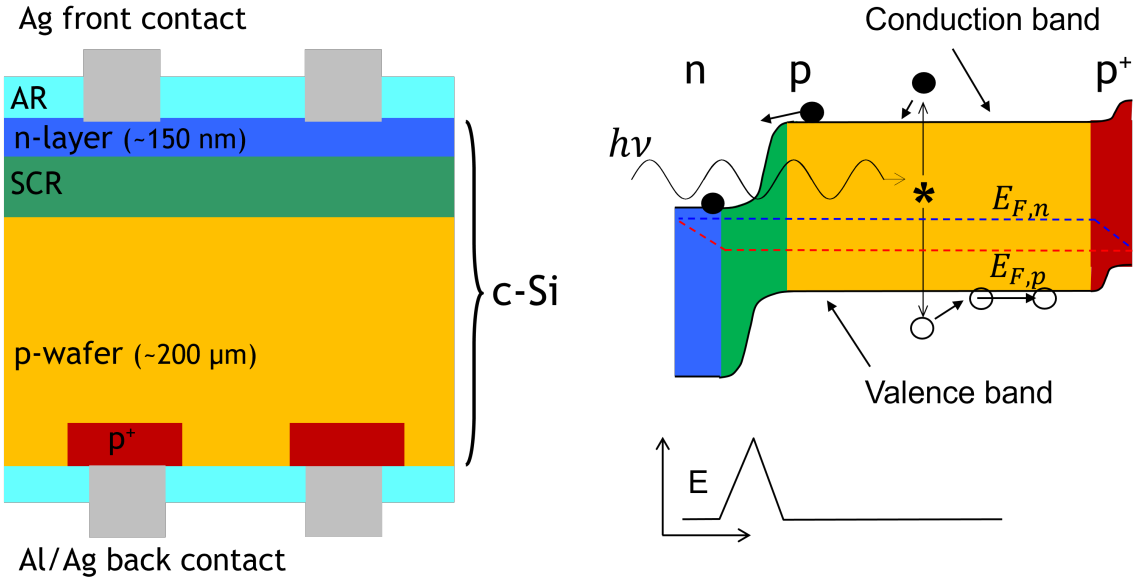


Figure 1.3: Sketch of typical solar cell structure (in cross-section, left) and of a band diagram across a p-n junction in operation conditions, i.e. under forward bias and illumination (right).

than their equilibrium densities in dark. The deviation from equilibrium can be described by introducing two different *Quasi Fermi levels (QFLs)*, one for the majority carriers which will not differ much from the equilibrium value, and another one for the minority carriers.

During their lifetime, the minority carriers are also free to diffuse within the solar cell. If one of them, say an electron in the p-doped region, crosses the junction within its lifetime, it joins the pool of majority carriers on the n-doped side and as such it is much less prone to recombination. The art of solar cell design is thus to ensure that the minority carriers feel an incentive to diffuse across the junction, and that they can do so within their lifetime. Note that the field of the junction has no impact on the movement of the carriers as long as they diffuse in neutral bulk material. The only way to impose a direction upon the random movement of diffusion, is by providing a concentration gradient.[2] This means that we have to continuously extract electrons from the n-side and holes from the p-side. At the same time, we should avoid that electrons get collected by the contact of the p-side (or holes by the contact of the n-side). For the electrons, this is typically achieved by a highly p^+-doped region as illustrated in figure 1.3, but it could also be some other material as long as it has a high work function and preferentially transports holes instead of electrons.

1.5 Current voltage characteristic

The $I(V)$-characteristic of a p-n-junction is usually called *diode equation* and was presented in 1949 by Shockley [1]. The steps of the derivation will be discussed in section 6.2, here we just refer to the final result. In addition to the exponential term of the original diode-equation, the description of

[2]Some types of solar cell use a built-in field to collect carriers by drift. In that case, the absorbing material is undoped and sandwiched between doped films to form a p-i-n junction. The design can only be applied to thin films because the potential difference between typical p- and n-materials is in the order of 1 V which cannot create any appreciable field across a full wafer.

solar cells requires the addition of a term that accounts for the photocurrent I_{ph} which is generated by the solar cell:

$$I(V) = I_0 \left[\exp\left(\frac{qV}{nkT}\right) - 1 \right] - I_{ph} \tag{1.3}$$

For a typical diode in electronics, $I_{ph} = 0$ and the equation contains only the applied voltage V, the thermal energy kT, and the saturation current I_0. The diode quality factor n is an empirical addition that will be discussed in more detail in section 6.5. For the time being, we will set it to unity. Let us first look at the solar cell in dark where it behaves simply as a diode. The left panel of figure 1.4 shows that there are two possibilities to connect the diode to a voltage source such as a battery. The blocking direction of the diode is defined as **reverse bias**. We associate this type of polarity with negative voltages. As the blocking is not perfect, there is a small negative current and when the reverse bias is increased, it approaches a constant value called *reverse saturation current* I_0. In this type of operation the diode consumes power, albeit only a small amount, and we associate it with the third quadrant of the $I(V)$-diagram. When the polarity of the applied voltage source is inverted, we operate the diode under **forward bias**. Thus, the diode is conductive and the current increases exponentially with the applied voltage. This type of operation also consumes power, but we associate it with the first quadrant of the $I(V)$-diagram where the bias voltage as well as the current are positive. The $P(V)$ diagram in the lower part of figure 1.4 illustrates the sign-convention which identifies positive power with consumption in the device.

When the diode is *illuminated*, we have to include the photocurrent I_{ph} in eq. (1.3). In this case the solar cell is an active element that works as source of power and the circuit diagram becomes a little more complicated. Since the solar cell must be described by a combination of a diode and a current source, figure 1.4 illustrates it by the gray box that delivers power to an external load. Compared to the dark $I(V)$-characteristic in the left panel, the presence of the current source shifts the characteristic downwards by an amount equal to the photocurrent I_{ph}. Under operation the solar cell provides a bias which we count positive, but the flow of current into the load resistance is counted negative. Thus, the producing part of the $I(V)$ characteristic is in the fourth quadrant and the $P(V)$ diagram in the lower part of figure 1.4 shows a negative branch associated with power generation. By varying the load resistance between zero (short circuit) and infinity (open circuit), we can measure the $I(V)$ characteristic over the fourth quadrant as illustrated by the part of the characteristic that is highlighted in red. If we want to measure the pink parts of the characteristic in the first and in the third quadrant, we have to replace the load by a variable source that can override the voltage generated by the solar cell.

1.5.1 Special points on the characteristic

The current voltage characteristic of eq. (1.3) allows us to define the key parameters of a solar cell. They are illustrated in right panel of figure 1.4.

- The **open circuit voltage** V_{oc} is the voltage that builds up under open circuit conditions, i.e. when no current is extracted from the cell.

- The **short circuit current** I_{sc} is the current that flows when the cell is shorted, i.e. no voltage can build up and $U = 0\,\mathrm{V}$. In first approximation, I_{sc} corresponds to the photocurrent I_{ph}.

- The **maximum power point** P_{MPP} is represented by a voltage V_{MPP} and a current I_{MPP}. Their product is the maximal power that can be delivered by the cell.

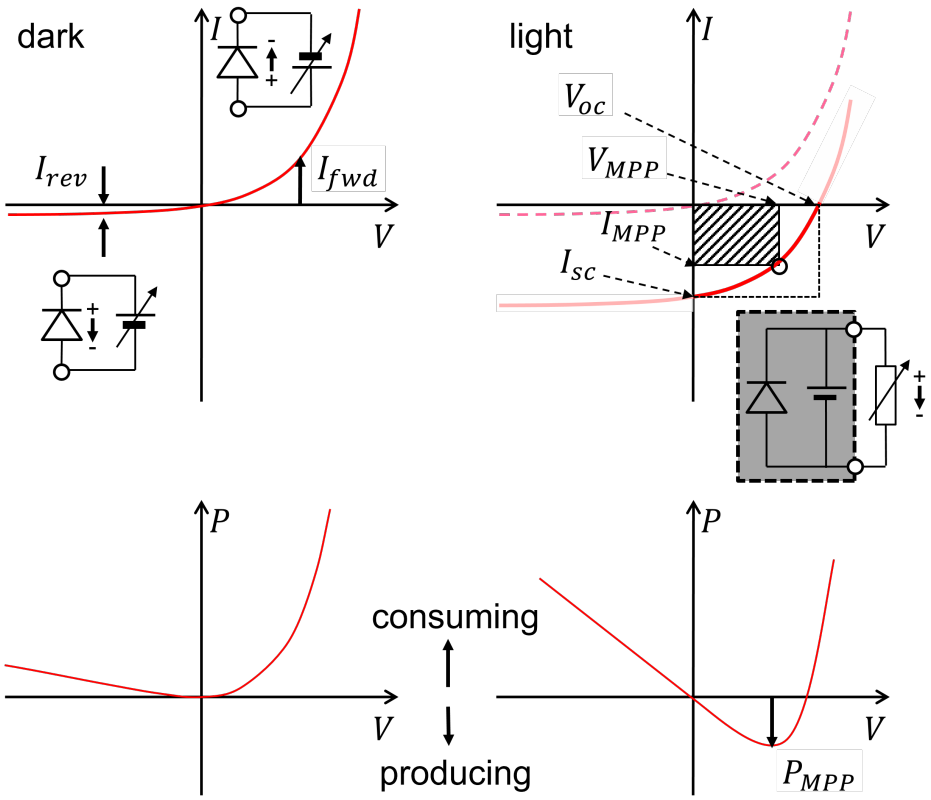


Figure 1.4: The top row shows the I-V curves of a solar cell in dark (left) and under illumination (right) with the respective equivalent circuits. The bottom row illustrates where the respective power is positive (consumed) and negative (produced).

Apart from these parameters that can directly be measured, there are two additional parameters that can be deduced from them. The **fill factor** FF is the ratio between the maximum power and the product of the V_{oc} and the I_{sc}. To get a high FF, one strives to get the $I(V)$-curve as rectangular as possible.

$$FF = \frac{P_{MPP}}{I_{sc}V_{oc}} = \frac{I_{MPP}V_{MPP}}{I_{sc}V_{oc}} \tag{1.4}$$

For the exponential relationship of eq. 1.3, an ideal value FF_0 can determined analytically by the Lambert W-function, or in a good approximation by $FF_0 \approx qV_{oc}/(qV_{oc} + 4.7nkT)$ [2].[3] The **cell efficiency** η is then calculated from the cell area A, the incident power density P (which is $1\,\mathrm{kW\,m^{-2}} = 0.1\,\mathrm{W\,cm^{-2}}$ for STC) and the maximal power P_{MPP}. The power incident on the cell is given by $P_{in} = A \cdot P$.

$$\eta = \frac{P_{MPP}}{P_{in}} = \frac{I_{MPP}V_{MPP}}{P_{in}} = \frac{I_{sc}V_{oc} \cdot FF}{P_{in}} \tag{1.5}$$

[3]The subscript denotes that FF_0 is an idealized value. Ref. [2] also gives approximations to determine FF values that contain the influence of series- and parallel-resistance. Section 6.5 discusses in more detail how the $j(V)$ characteristic is modified by these resistances.

Regarding the photocurrent, we could naively assume that every photon in the solar spectrum generates an electron-hole pair. However, photons can only be absorbed when they have sufficient energy to excite electrons across the bandgap, i.e when $h\nu > E_g$. Thus, we would have an interest to use a semiconductor with low bandgap in order to generate a high photocurrent.

Typical drawings such as figure 1.3 could lead us to a similarly naive assumption regarding the open circuit voltage as the photogenerated electrons and holes are normally at the band edges where they have a potential difference equivalent to the energy of the bandgap. Unfortunately, in terms of device operation, we have to consider the statistical average of their electro-chemical potential which is given by their QFLs. Thus, the V_{oc} is related to the splitting of the QFLs upon illumination with sunlight. This splitting depends on the density of minority carriers in steady-state which in turn depends on the recombination processes in the device. In essence, the V_{oc} is not directly related to the bandgap, but we can still say that high we should use a semiconductor with high bandgap if we want to generate a high voltage.

Unfortunately, there is an additional loss mechanism in semiconductors. To understand it, imagine a photon with an energy that exceeds the bandgap; it can excite an electron from a state deeper than the VB edge, and carry it into a state high above the CB edge. After that, the excess energy is quickly lost by multiple interactions with phonons. This is called *thermalisation* and means that the electron and the hole will lose their excess energy and end up in states close to the respective band edges as illustrated in figure 1.3. In this case the useful energy is equivalent to the bandgap, the excess is lost to heating.

In the end we have to find an optimum for this trade-off; on the one hand, semiconductors with a high bandgap can have a high voltage, but they collect a low photocurrent because of transmission losses. On the other hand, semiconductors with low bandgap can have high photocurrent but at a low voltage, and they make poor use of energetic photons because of thermalisation losses.

There are a few more interesting features that we can observe after a bit of math from the simple diode equation of eq. (1.3):

1. First of all, the current I is often divided by the cell area A to obtain the current density j.

2. The **photogenerated current density** j_{ph} is the same as the *short circuit current density*. Strictly speaking this is only true if we neglect the influence of series- and shunt resistance ($R_s = 0$ and $R_p = \infty$) and if there are no overly serious recombination losses. If these assumptions apply, an increase of the illumination intensity by a factor x will increase the current by the same factor:

$$j_{ph} \approx j_{sc} \propto \text{illumination} \tag{1.6}$$

3. By setting the current density to zero, we can determine a relation for the V_{oc}:

$$V_{oc} = \frac{kT}{q} \ln\left(\frac{j_{ph}}{j_0} + 1\right) \approx \frac{kT}{q} \ln\left(\frac{j_{ph}}{j_0}\right) \tag{1.7}$$

   In most circumstances, the photocurrent is much larger than the saturation current and the approximation holds. We can conclude that an increase of the photocurrent will increase the V_{oc} logarithmically, for example by concentrating the light intensity. On the other hand, indoor operation or cloud cover will substantially reduce the V_{oc}.

4. The eq. (1.7) shows that a low j_0 is needed for high V_{oc}, one of the ingredients for high cell efficiency.

1.6 Spectral response (SR) and External Quantum Efficiecy (EQE)

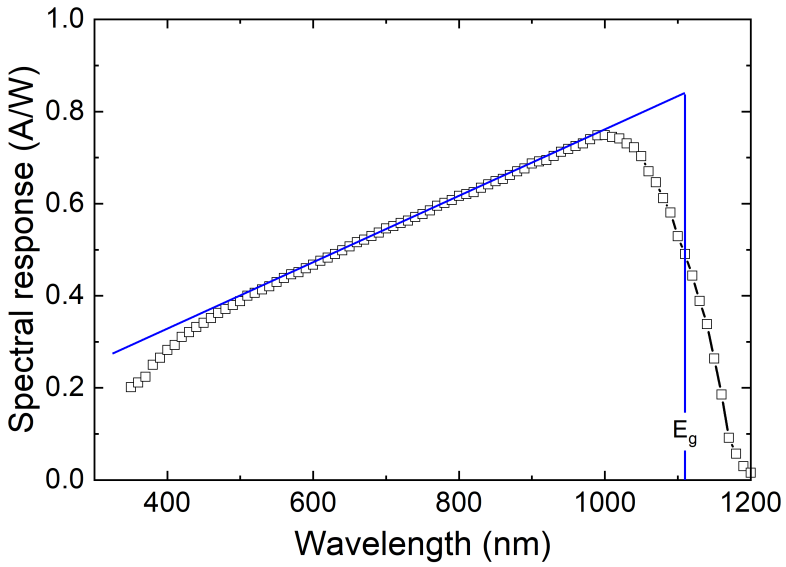


Figure 1.5: Spectral response of a c-Si solar cell (open squares) and idealized characteristic (blue line). The drop at 1110 nm corresponds to the bandgap of c-Si.

The *spectral response* SR(λ) is measured in short circuit condition and gives the spectrally resolved current per incident power (A/W) as function of the wavelength λ of the incident light. If we know additionally the spectrally resolved illumination intensity $\Phi(\lambda)$ (in units of $\mathrm{W\,m^{-2}\,nm^{-1}}$), we can calculate the current that is generated by the device. For a cell area A and the spectral range relevant for silicon, we obtain:

$$I_{\text{cell, spec}} = A \int_{300}^{1200} \Phi(\lambda) \cdot \mathrm{SR}(\lambda) d\lambda \tag{1.8}$$

The SR curve of an ideal cell is triangular as shown in Figure 1.5. The linear increase for short wavelengths is explained as follows; let us assume that we illuminate an ideal cell with a power of 1 W concentrated into a single wavelength. As the ideal cell converts all incident light into current, we just need to know the number of photons in the illuminating beam. For a given wavelength, say 620 nm which corresponds to an energy of 2.0 eV, this number is given by $1\,\mathrm{W}/2.0\,\mathrm{eV} = 3.1 \times 10^{18}\,\mathrm{s^{-1}}$. Multiplying with the elementary charge, a current of 0.5 A is obtained. Other energies can be treated accordingly. We can thus expect that the characteristic rises linearly. When we reach the wavelength corresponding to the bandgap energy, it should drop to zero.

For a real cell the SR deviates from the triangular form as illustrated in figure 1.5 for the example of a c-Si solar cell. At short wavelengths, some photons are reflected, or they are absorbed parasitically in the the layers of the anti-reflection coating and the front contact. Throughout the visible, the curve follows the ideal characteristic fairly well, but towards the bandgap it deviates more and more because silicon absorbs these photons only weakly. At longer wavelengths we note that the experimental characteristic extends beyond the expected drop at the bandgap energy. This is particular for indirect semiconductors such as silicon where absorption can be assisted by

a phonons. When the lattice provides the phonon, it can absorb a photon with energy less than bandgap, but the lattice gets cooled in the process.

The **external quantum efficiency** (EQE) is more often used as alternative representation to the spectral response. It is defined as ratio between the number of carriers extracted under short circuit condition and the number of *photons impinging on the cell*. In other words, it is the probability that an electron-hole pair is created and collected from an incident photon of a specific wavelength. It can be calculated from the spectral response $SR(\lambda)$:[4]

$$\mathrm{EQE}(\lambda) = \mathrm{SR}(\lambda) \cdot \frac{E(\lambda)}{q} = \frac{\mathrm{SR}(\lambda)}{q} \cdot \frac{1240}{\lambda[\mathrm{nm}]} \tag{1.9}$$

The EQE is very useful to identify lossy regions within the cell. This relates to the fact that high-energy photons (blue) are absorbed in the first 10 - 100 nm of the cell whereas the low-energy photons (red) can travel all the way through the cell to the rear-side. Thus, low EQE at short wavelengths indicates parasitic absorption in the front region. Losses in the red part of the spectrum indicate collection losses because of short diffusion length or recombination at the rear contact. A discussion of these effects requires a bit of math and can be found in appendix D.

As low EQE could also be due to reflection losses, the **internal quantum efficiency** (IQE) gives the ratio between the number of carriers extracted under short circuit condition and the number of photons *absorbed within the cell*. The IQE can be calculated from the EQE by the following formula, where $R(\lambda)$ is the reflection of the cell at a given wavelength.

$$\mathrm{IQE}(\lambda) = \frac{\mathrm{EQE}(\lambda)}{1 - R(\lambda)} \tag{1.10}$$

1.7 Thermodynamic efficiency limit

The thermodynamic efficiency limit of single-junction solar cells was derived by Shockley and Queisser in 1961 [3]. The derivation is based on the following assumptions:

- **Transmission and thermalisation are the only spectral losses.** A perfect anti-reflection ensures that every photon with energy above the bandgap is absorbed and contributes an electron-hole pair.

- **Radiative recombination is the only recombination loss.** Since the solar cell is assumed to absorb ideally all photons with energy exceeding the bandgap, it fulfils the definition of a black body with perfect absorptivity in this spectral region. By thermodynamic reciprocity, this corresponds to ideal emissivity in that spectral region and therefore it will emit black-body radiation in that same spectral region according to its operating temperature of 298 K.[5]

[4] Here, $E(\lambda) = hc/\lambda$ is the energy of a photon at a given wavelength.

[5] Without concentration, this process creates entropy because sunlight enters the cell from a small cone whereas black-body radiation is emitted into the full solid angle. A the majority of this radiation remain within the semi-conductor by light trapping and will eventually be re-absorbed. The light that is emitted into the angular range of escape cone subsequently gets refracted into the full hemisphere and therefore a much larger angular domain than the incident sunlight. In an ideal concentrator, the cone of the incoming sunlight can be widened up to 23'000 times until it fills the hemisphere above the cell (46'000 times for widening to the full solid angle around an ideal bifacial device). In that case, the cones of incoming and exiting radiation would be the same. Effectively, concentration increases the cell efficiency because it reduces the generation of optical entropy.

- **An exponential relationship between current density and voltage.** The prefactor j_0 is defined by radiative recombination effects and under illumination the superposition principle applies with j_{sc}.

Figure 1.6 illustrates a limiting efficiency 33.7% for a bandgap of 1.34 eV and shows also the impact of various loss mechanisms [4]. Note that the efficiency limit can be much higher for "exotic" spectra such as indoor light. For the case of purely monochromatic illumination at an energy equal to the bandgap, the efficiency limit is 68% without concentration and 87% with concentration.

For the bandgap of c-Si, the limit would be 32% under AM1.5 illumination. However, in c-Si the absorption of the indirect bandgap is so weak that even an ideal optical design requires a device thickness of 100 µm or more, resulting in a comparatively low generation rate over most of the bulk. Even for large lifetimes this translates into a low excess carrier density and consequently low voltage. Another consequence of the low absorption coefficient is a low rate of radiative recombination, and therefore c-Si is limited by Auger recombination. The determination of the limiting efficiency for c-Si is outlined in appendix C, resulting in values between 29.4 and 29.6% [5, 6].

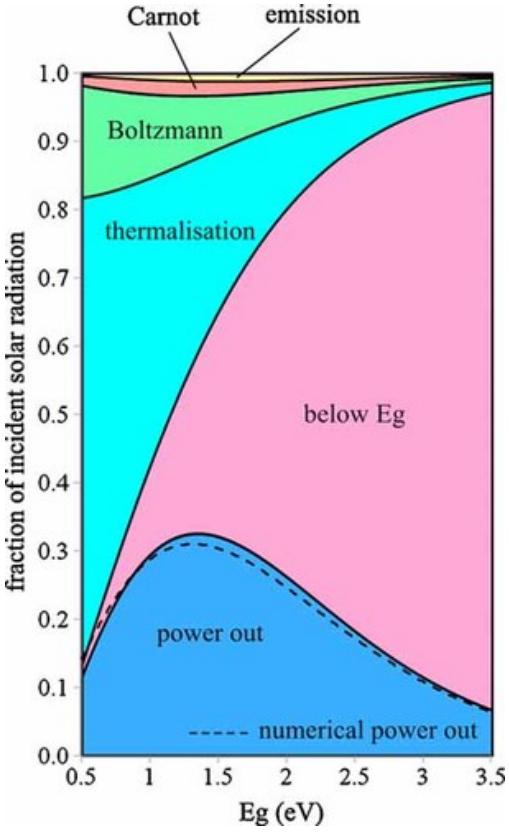


Figure 1.6: SQ-limit as a function of E_g (blue), taking into account the AM1.5G spectrum (reproduced from [4], with permission 5581960494158 from John Wiley & Sons).

An experimentally proven way beyond the Shockley-Queisser limit is the use of more than one junction. The limit for *tandem cells* with two junctions is 47% without concentration. For devices with an infinite number of stacked junctions it is 68% without concentration and 86.8% under concentration [7], the same values that were quoted above for single junctions illuminated by monochromatic photons with energy equal to the bandgap energy. In multi-junction devices, the cell with the highest bandgap should be on top in order to absorb the high-energy photons. Photons with lower energy are transmitted to lower lying cells with increasingly lower bandgap energies. Thus, *multi-junction cells* absorb exactly the spectral region as the cell with the lowest bandgap would. The advantage is related to the absorption of high-energy photons in the cells with higher bandgap where reduced thermalisation losses allow a better conversion of their energy.

The stacking of the cells can be done monolithically (cells sequentially grown on top of each other and connected in series), or mechanically (cells stacked on top of each other with cells, each with independent circuitry). Monolithic integration is widely used for III-V solar cells, but only a few combinations can be manufactured without *lattice mismatch*. For example, high-efficiency cells for space applications use lattice matched triple junction devices, starting with c-Ge that serves as bottom-cell and as growth template for a GaAs middle cell and a InP top cell. In terms of its bandgap, c-Si could make very efficient bottom cells for two-junction tandem cells, but there is no lattice-matched partner with suitable bandgap to grow monolithic tandem cells. However, c-Si can be combined with polycrystalline materials such as perovskite solar cells.

For monolithic interconnection, *current-matching* of all the sub-cells in the stack is mandatory

since device performance will be limited by the sub-cell with the lowest current (in a series connection, the current must be the same in all cells). A design may be matched very well under the AM1.5 spectrum, but outdoors it will be prone to mismatch under the red-rich spectrum of the morning/evening hours or in winter; likewise, mismatch will also be observed under the blue-rich spectrum at noon in summer or at locations closer to the equator. Besides spectral mismatch, another issue of outdoor operation arises from the different temperature coefficients of the component cells as discussed in the next section.

1.8 Temperature coefficient of solar cells

In typical outdoor operation, the cell temperature is likely to differ from the one defined by STC. Therefore, it has become common to include in data sheets the performance parameters under **nominal operating cell temperature (NOCT)**. The NOCT is obtained for a mounting configuration with open rear, an air temperature of 20 °C, a wind speed of $1\,\mathrm{m\,s^{-1}}$ and irradiation with a reduced illumination intensity of $800\,\mathrm{W\,cm^{-2}}$. For typical module designs, the NOCT values are 30 - 45 °C.

A more detailed description of the solar cell performance can be obtained through the temperature coefficients which are defined not only for the efficiency, but also for the V_{oc}, the j_{sc}, and the FF. Many of the changes relate to the reduction of the bandgap with increasing temperature.[6] The main effect is a small gain in j_{sc} and a reduced V_{oc} due to an increase in the intrinsic carrier density n_i. On top of that there is a broad variety of additional physical phenomena [8].

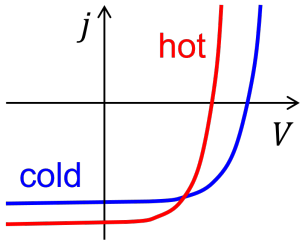


Figure 1.7: I(V) curve vs. temperature

The gain in j_{sc} is due to reducing bandgap with increasing temperature, thus extending the absorption further into the region of long wavelengths. Whereas this is a generic feature of almost all semiconductors, in silicon the absorption is additionally enhanced as higher temperatures also provide more of the phonons needed in the indirect absorption process. To illustrate the impact of the temperature on the V_{oc}, let us revert to eq. (1.7). The term of j_0 can be often be expressed as $j_0 = j_{00} \cdot e^{-E_g/kT}$, and thus we obtain:

$$V_{oc} \approx \frac{E_g}{q} - \frac{kT}{q} \ln\left(\frac{j_{00}}{j_{ph}}\right) \tag{1.11}$$

The first term contributes a slight reduction due to the temperature dependence of the bandgap. A more obvious temperature dependence arises from the prefactor kT/q in the second term and an additional dependence arises through j_{ph} within the logarithm.

The temperature coefficient of the FF is yet more complicated. For the $j(V)$ characteristic of the ideal solar cell, there is an analytic relationship to the V_{oc} through Lambert's W-function. To a good approximation the relation is linear, thus contributing the same temperature dependence as the V_{oc}. In real solar cells, the FF picks up additional dependencies through the series resistance, and to a lesser extent through the parallel resistance [2]. The resistive losses are associated to Schottky barriers between the semiconductor and the the contacts as well as to Ohmic losses in the contact materials themselves; with increasing temperature the former decrease whereas the latter increase and eventually become dominant.

[6] For silicon, the decrease is fairly linear around RT and can be approximated with $-0.273\,\mathrm{meV\,°C^{-1}}$, more details are discussed in appendix A.1.

Overall, the contribution of the V_{oc} accounts for $\simeq 90\,\%$ of the efficiency temperature coefficient. Generally cells with higher bandgap have a lower efficiency temperature coefficient, but even within a given material the cell designs that yield high V_{oc} have a low temperature coefficient. For example, the efficiency of c-Si cells in BSF design may reduce by $\sim -0.5\%/°\mathrm{C}$ whereas c-Si heterojunction cells perform much better at $\sim -0.3\%/°\mathrm{C}$. In a technology such as a-Si with its higher bandgap of 1.7 eV the temperature coefficient and be as little as $\sim -0.2\%/°\mathrm{C}$.

To account for the changes at different operating temperatures, we can define the *performance ratio* (*PR*). It is the ratio of the average efficiency η (%) in operation and the quoted STC efficiency (%):

$$PR = \frac{\eta_{\mathrm{avg}}}{\eta_{\mathrm{STC}}} \tag{1.12}$$

It can be more than 100%, when the cell is operated in conditions that are more favourable than STC. This can happen in winter when the module temperature is less than 25 °C, or when reflection from snow on surrounding surfaces provides additional irradiation.

1.9 Sustainability and carbon footprint

To assess the sustainability of PV we need to carry out a *life cycle assessment (LCA)* to determine the amount of energy that is needed for the procurement of raw materials and how much is consumed during the individual manufacturing steps. A more complete LCA would include also transport to the installation site, details of deployment such as the mounting structure, the cabling and the inverters, and maintenance such as the regular cleaning needed for modules in dusty desert environments. Finally, it should also include the recycling.

Let us assess the values for the steps up to module manufacturing. We assume a wafer thickness of 170 µm and a kerf loss of 65 µm. This leads to a use of silicon around $600\,\mathrm{g\,m^{-2}}$ which is the prime input at the bottom of the second column in table 1.1. Starting from there, we have to move upwards in the second column through the different loss mechanisms along the chain of silicon production, arriving at a requirement for silicon of so-called metallurgical grade (MG-Si) of around $760\,\mathrm{g\,m^{-2}}$.

The energy and the fuel required for each of these steps is normally given per kg of Si and they include the enthalpy of the chemical reactions and other items such as generation of process heat. For example, the first item represents the reduction of silicon oxide to MG-Si. It proceeds along the general reaction $SiO_2 + C \rightarrow Si + CO_2$ as explained in more detail in section 7.1. Since 1 kg of silicon contains 34 mol of silicon, the process consumes at least $13.4\,\mathrm{MJ\,kg^{-1}}$ by burning 34 mol of carbon. Table 1.1 shows that in reality it is almost double that amount because less energetic carbon sources such as wood chips and because fuel may be carried from the furnace by the exhaust. On top of that, a significant amount of electricity is required to run the arc discharge that powers the furnace. Surprisingly, the next lines in table 1.1 shows that almost five times as much fuel and electricity is needed for the purification to solar grade silicon, and yet another three times of the electricity for ingot pulling.

In columns 4 and 6 of table 1.1, these values are scaled to the module area. The last column gives an estimate of the equivalent amount of primary energy. To this end, we'd have to know the exact electricity mix used for the different fabrication steps. The entries in the table represent a worst-case scenario where the electricity is assumed to come from a conventional thermal power plant with an efficiency of 30%. Finally, the last line contains some additional inputs such as the module assembly. For example, encapsulation needs polymers and either one or two panels of glass,

weighing 10 or 20 $\mathrm{kg\,m^{-2}}$, respectively. The energy needed for the production of glass adds thus 260 or 520 $\mathrm{MJ\,m^{-2}}$, respectively. A typical value for the weight of the aluminium frame is 2.6 kg, adding another 100 $\mathrm{MJ\,m^{-2}}$. Obviously, frame-less designs or modules with glass-backsheet configuration can significantly reduce the demand of primary energy.

Table 1.1: Energy used during the individual steps of c-Si module fabrication, [9, 10, 11].

	Si	Fuel		Electricity		Primary energy
Step	$\mathrm{kg/m^2}$	MJ/kg	$\mathrm{MJ/m^2}$	kWh/kg	$\mathrm{kWh/m^2}$	$\mathrm{MJ/m^2}$
MG Si	0.76	23	13.6	11	6.6	120
SoG Si	0.67	122	72	49	29	515
Cz pull	0.64	-	-	32	19	245
Wafering	**0.6**		2.4		4.8	60
Cell			5.9		17.7	220
Module			5.4		14	800
Total						1950

We can now estimate the **energy pay-back time** (EPBT), i.e. the time the module has to produce energy in order to yield the same amount as was used for its manufacturing. Adding all items of the primary energy for the worst-case scenario, we obtain 1950 $\mathrm{MJ\,m^{-2}}$ or 530 $\mathrm{kW\,h\,m^{-2}}$. To assess the energy harvest, we need to know the module efficiency and the place of the installation. For the former, studies of 2021 assume efficiencies between 19.8 and 20.5% for c-Si modules [10, 11], for the latter these studies considered representative locations with low, medium and high irradiation such as Northern Europe, Southern Europe and locations in the Earth's sun belt with respective irradiation intensities of 1000, 1700 and 2300 $\mathrm{kW\,h\,m^{-2}\,y^{-1}}$. Sticking to Northern and Southern Europe, we can expect to harvest up to 205 and 350 $\mathrm{kW\,h\,m^{-2}\,y^{-1}}$, respectively. Finally, the definition of the EPBT takes into account the energy mix used for module manufacturing and the performance ratio of the installation.[7] Altogether, this yields EPBTs of 1.3 and 0.8 years for crystalline silicon modules installed at sites in Northern and Southern Europe, respectively.[8]

Based on the EPBT we can also estimate carbon footprint of photovoltaics, i.e. how much CO_2-equivalent a PV system releases for the production of a kW h. To this end we have to acquire a bit more information. First, we need the energy mix used for the module production to assess the amount of CO_2 that got released during manufacturing. Second, we need again the location of the installation to determine its annual production. Finally, we need the module lifetime to find the total amount of energy it will produce during its life cycle. There are some uncertainties going into this estimate, but taking again the examples of Northern and Southern Europe, for c-Si this amounts to 40 and 24 grams of CO_2 per kW h, respectively [10]. Note that this estimate for the equivalent CO_2 emission hinges on our assumption of the manufacturing process which means we could further reduce the carbon footprint by running the production with cleaner energy.

[7] For more details, see pp 12-13 of "Methodological Guidelines on Net Energy Analysis of Photovoltaic Electricity, 2nd Edition (2021)" at `https://iea-pvps.org/research-tasks/pv-sustainability/`

[8] Corresponding assessments of thin film technologies date a bit further back; in 2013 and for installation in Southern Europe, payback times of 0.7, 1.0 and 1.4 years were found for CdTe, CIGS and a-Si, respectively [12]. Interestingly, the fabrication of a-Si modules required the lowest amount of energy, but the lower efficiency compared to CIGS and CdTe modules resulted in the longest payback time. For comparison, the same report quotes EPBTs between 1.9 and 2.3 y for monocrystalline silicon and between 1.2 and 1.5 y for multicrystalline silicon. Compared to the more recent data of 2021 cited in the previous paragraph, this illustrates a substantial improvement for silicon technology within 8 years. The two main reasons are the use of much thinner wafers and the increased efficiency of recent cell designs.

For comparison, for a coal fired power plant the carbon footprint is about 800 grams per kW h and the CO_2 emission is inherent to the burning of the fuel. For for hydro power it is around 24 grams per kW h, largely due to the CO_2 released during the fabrication of cement that is used for the concrete in the dams. For electricity from wind turbines and from nuclear reactors the values are about 12 grams per kW h, in the former case largely due to the building materials of foundation, tower, generator and blades, and in the latter case due to mining and enrichment of uranium.[9]

[9]Numbers according to Wikipedia, "Life cycle greenhouse gas emissions of energy sources".

Chapter 2

PV state of the art

2.1 Photovoltaics market

In terms of PV technologies, the market is dominated by crystalline silicon which is a wafer-based technology. A small share is held by thin film technologies, most prominently CdTe. Finally, there is a niche-market for multijunction technologies based on costly III-V materials that serve space applications and a few terrestrial applications such as concentrator cells. Table 2.1 reproduces record efficiencies at cell- and module level.[1]

Table 2.1: Confirmed efficiency of the different PV technologies under STC [13]

	Lab scale Cell efficiency	**Champion module efficiency**
GaAs	29.1%	25.1
c-Si	27.4%	25.4%
CdTe	21.0%	19.9%
CIGS	23.35%	19.2%
Organic-PV	15.8%	13.1%
Perovskite	25.2%	19.2
III-V 6-junction	39.2 %	-
Perovskite/Si (monolithic)	31.3%	26.9
Micromorph (a-Si/μc-Si)	14.0%	12.3%

In most cases the decision for or against investment in a PV system is based on cost, idealistic or aesthetic arguments are normally secondary, and the choice of technology follows third. Therefore, one of the biggest challenges of PV is to produce the cells and modules in a more cost effective way. There are several possibilities to reduce the costs per Watt, namely:

- **Increasing the efficiency**, i.e. more Watts for the same production costs.
- **Optimization of the processes and materials**, i.e. lower production cost.[2]

[1] The latest record efficiencies measured at accredited certification laboratories around the world are collected in the *Solar Cell Efficiency Tables*. They are published twice a year in the Journal "Progress in Photovoltaics: Research and Applications". Alternatively, check the NREL efficiency chart at `http://www.nrel.gov/pv/assets/images/efficiency-chart.png`.

[2] For example, c-Si technology has achieved an enormous cost reduction by the development of sawing technology. An ingot can yield more wafers by sawing thinner wafers, sawing with thinner wires that reduce the kerf loss, and

- **Industrialisation**, i.e. higher cell/module throughput to reduce the cost.

In a PV system the module should always operate at the MPP. Consequently, solar cells or modules are usually operated with a *maximum power point tracker* (MPPT). This is an electronic device that contains a variable DC load whose "resistance" is regularly adapted to operate the cell at the MPP. Often it is coupled to a DC-DC converter that converts the generated power into a stable DC output. If we want to run AC appliances, we have to use an *inverter*, and if we want to feed solar power into the electricity grid, the output must be carefully synchronised with the phase of the grid voltage. Typical inverters can handle 3 - 500 kW and feed two or three phases, but there are also models that reach 1 MW.

Terrestrial PV installations showed a spectacular growth over the last three decades; starting from isolated systems installed by a few enthusiasts, they became a standard for buildings and commercial electricity generation. Initially, the growth was stimulated by governmental regulations and incentives such as the feed-in-tariffs that Germany introduced in the late 1990s and the early 2000s. In 2009, Germany elected a conservative government that massively curbed the incentives, but by this time the cost of PV modules had substantially decreased, such that the number of PV installations kept on growing in other countries, notably in China. By 2016, the installation volume for large scale PV installations in sun-belt countries exceeded 30 GW_p per year, at bidding prices of €0.1, and more recently below €0.05 per kW h.

Besides driving down the costs of an installation, further challenges lie in the transport of electricity over long distances and in the storage for times without sunlight. Transport losses may be surmountable by DC power lines that exhibit only 3% on 1000 km. Storage is possible with various technologies, e.g. H_2 storage, compressed air, hydro storage, flywheels, etc., but each of these options has its advantages and disadvantages and it is not yet clear which one will prevail.

2.2 Crystalline Si

Borrowing knowledge from microelectronics manufacturing and benefiting from its abundance and reliability, c-Si cells developed into the most widely used PV technology. In a nutshell, c-Si cells use single-crystal wafers with etched surface texture. After a diffusion step to create the junction at the front and deposition of a Si_3N_4 anti-reflection coating, metallic electrodes are screen-printed at the front and at the rear, and a firing step is used to contact the wafer and to sinter the metals. Advanced designs aim at minimizing, or avoiding altogether, the direct contact between the metallisation and the active parts of the semiconductor. To avoid shadowing by the front electrode, both contact polarities can be applied to the rear-side of the cell, obviously at the cost of a more complicated manufacturing process.

2.3 Thin film PV

Compared to cell production on the basis of crystalline wafers, thin films bear the advantage of lower material usage and the possibility to apply them to flexible substrates. Their drawbacks are less standardized equipment and the fact that the deposited materials are generally polycrystalline or amorphous which (normally) results poorer electric performance due to recombination at grain boundaries.

---

sawing with less surface damage to reduce the depth of the saw damage etch (SDE). Alternatively, thin film cells use less material in the first place, but experience tells that customers do not accept lower efficiencies.

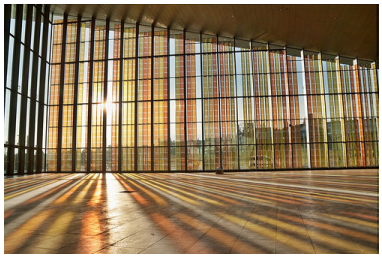
2.3.1 CdTe

The compound CdTe is a II-IV semiconductor. It has a bandgap of 1.44 eV which is a perfect match to our solar spectrum. The compound evaporates and condensates congruently which means that evaporated material forms absorber layers with the correct composition. It is naturally p-type and forms an excellent heterojunction with n-type CdS. This is exploited in fabrication where typically glass is coated with a transparent front contact such as fluorine doped SnO_2 (FTO), followed by the evaporation of CdS and CdTe. The real challenge comes with the rear contact as there are no metals that form an Ohmic contact to CdTe. Thus, an additional p-type layer such ZnTe or Sb_2Te_3 in combination with graphite is often used. Modules based on CdTe often have issues with public acceptance. On this topic we note that Cd is indeed toxic when available in digestible form. However, CdTe is a stable compound and leakage experiments with broken modules remained below environmental limits.

2.3.2 CIGS

The acronym CIGS refers to the family of $Cu(In,Ga)Se_2$ semiconductors which are of the type I-III-IV_2. For a long time it was the thin film material that achieved the highest efficiencies, but its fabrication process is comparatively complicated as the best material is obtained by co-evaporation of four elements. Usually it deposited on glass coated with molybdenum which is fairly stable in the corrosive atmosphere of the Se-vapour. The formation of an interfacial layer of $MoSe_2$ is actually beneficial as it form an Ohmic rear contact to the p-type CIGS absorber. The cell process normally proceeds with the deposition of n-type CdS in a chemical bath and the deposition of the transparent ZnO front contact by sputtering.

2.3.3 Dye sensitized solar cells



In 1991, Brian O'Regan and Michael Grätzel published an innovative concept of a photovoltaic cell that is inspired by photosynthesis [14]. This cell works with a porous TiO_2 matrix into which a dye is infiltrated by solution processing. Absorption of a photon excites an electron into an excited state of the dye from where it is transferred extremely fast to the CB of TiO_2. Subsequently, the charge is replenished from a redox-couple, usually an iodine-based electrolyte in contact with a Pt counter-electrode. This technology reached efficiencies of approx 11.9% at the lab-level. The big issue is long term stability, notably the sealing of the liquid electrolyte. The advantages are potentially low cost, ease of fabrication, and the possibility of using dyes with different colours. This can be interesting for building integration with semi-transparent modules as demonstrated by the facade-installation on EPFL's Swiss Tech Convention Centre.

2.3.4 Perovskites

Perovskites solar cells developed from dye sensitized solar cells when researchers at Yokohama University tried to replace dyes by nanoparticles of an organic-inorganic metal halide. Starting from a device efficiency of 3.8% shown in 2009, this technology has seen unprecedented efficiency growth to 25.2% in 2019. The materials have exceptional optoelectronic properties, close to III-V materials, but they can be grown with low-cost fabrication techniques, such as solution processes, printing, evaporation, etc. Currently, the main issue lies in up-scaling and stability. Moreover, they

contain small amounts of Pb which raises concerns about toxicity, similarly to Cd. The topic is discussed in more detail in section 11.6.

2.4 Ultra-high efficiency and concentrators

2.4.1 III-V Multi-junction cells

The highest conversion efficiencies of all PV technologies are achieved with **multi-junction cells** on the basis of III-V materials. The most commonly used design uses a c-Ge wafer that provides a bottom cell for IR radiation and also serves as epitaxy-template for a lattice-matched stack of a GaAs middle-cell and a (Ga,In)P top cell. More recent designs don't use Ge any more. In 2019, a mechanically stacked 6 junction-cell reached a confirmed efficiency of 47.1% under a concentration of 143x.

One big issue of those bandgap engineered cells is that the best bandgap combination is not necessarily made up of materials that can be grown on top of each other very easily. There might be *lattice mismatch* that leads to defects and dislocations which in turn can lead to peeling of the layers as they grow thicker. This can be circumvented by introducing buffer layers, which mediate between two materials. The principle can also be applied to match refractive indices of two adjacent layers.

Another possibility is to mechanically stack a wide band gap III-V cell on top of a low band gap cell such as silicon. This concept now reached efficiencies above 32.8% for tandems and 35.9% for a triple cell consisting of a III-V tandem on a c-Si bottom cell.

Multi-junction solar cells based on III-V materials are rather expensive due to scarce elements such as In and Ga, and because of their complex processing techniques. They are preferably applied in space and for some terrestrial application using concentrators.

2.4.2 Tandem cells with c-Si

Until a few years ago, silicon was never considered for tandem solar cells as there is no lattice-matched material with suitable bandgap. This changed with the advent of perovskite materials as they can be processed without the constraint of lattice matching. See section 11.6.

2.4.3 Concentrator cells

Increasing the light intensity is a logical way to increase the power output of a solar cell. The idea here is to use lenses or curved mirrors to focus the light onto a small but very efficient cell (usually III-V multi-junction). The main problem is the **small angle tolerance**, i.e. the cells have to face the sun exactly in order to work, and they cannot use any diffused light. This makes **tracking essential**, ideally on two axes. Tracking faces a variety of mechanical challenges because conditions with direct sunlight and without cloud-cover are often found in desert climates. Consequently, the moving parts must tolerate wind and sand.

2.5 Novel or emerging technologies

There are several concepts that have been proposed to circumvent the Shockley-Queisser limit. The most successful one is the use of more than one semiconductor in tandem and multi-junction configurations. Others have not yet made it into working devices, but it is interesting to understand their operation principles.

2.5.1 Hot carrier solar cells

Here the idea is to extract charge carriers before they loose excess energy by thermalization. To this end, collection should happen faster than the time of phonon interactions, i.e. within 1 - 10 ps. Alternatively, there are efforts to engineer the density of phonon states and thereby slow down thermalisation.

2.5.2 Multiple exciton generation

A high-energy photon can create more than one electron-hole pair as demonstrated with some quantum dot materials. However, in most cases the multiple exciton generation becomes only effective when the photon energy is higher than three times the bandgap energy. For our solar spectrum, that would mean to use a material with bandgap in the IR, resulting in a severe limitation of the output voltage.

2.5.3 Intermediate band cells

The idea of an intermediate band (IB) within a large bandgap is that a low energy photon could be used to excite an electron from the VB to the IB and a second low energy photon would take it to the CB. High energy photons could excite directly from the VB to the CB.

A naturally occurring IB was postulated for some highly mismatched alloys like $GaN_xP_yAs_{1-x-y}$, an effective IB was also reported for (Zn,Cd)Te-(Zn,Cd)Se multi-quantum wells. If successful, the red part of the solar spectrum could be used more efficiently.

2.5.4 Up- and down-conversion

The principle of this approach aims again at the reduction of the thermalisation losses and on the promotion of low energetic photons.

- **Down-conversion** at the front side: One high-energy photon is absorbed and two low-energy photons are emitted. Their energy ideally corresponds to the band gap of the absorber material.
- **Up-conversion** in the middle or at the rear of the cell: Two low energy-photons are absorbed and one highenergy photon is re-emitted.

Most up- and down-conversion processes are second order interactions and therefore very inefficient, i.e. *on average only one photon in 1000 leads to such a process* . But on the other hand this might be interesting for concentrator cells, as the probability of the process scales with intensity.

2.5.5 Down-shifting

Different from the conversion effects discussed in the previous two sections, **down shifting** is a fluorescence processes that maintains the number of photons. Therefore, it does not circumvent the Shockley-Queisser limit, but it could help to avoid parasitic absorption in supporting layers of the front contact. For example, embedding fluorescent particles in the front encapsulation could absorb UV light and emit it into visible wavelengths that are efficiently converted by the solar cell. The fluorescence effect can have efficiencies up to 90%, but as the emission is omnidirectional, the radiative losses towards the front have to be taken into account.

Chapter 3

Basic semiconductor properties

3.1 A short introduction to band theory

Solids can be considered as entities where atoms are *closely packed* next to each other, unlike in gases. When the interatomic distance is short enough, the wavefunctions of the valence-electrons can overlap. Thus, the discrete electronic states of the atoms will broaden due to the overlap and eventually form *bands*. We can very roughly associate insulators with a state where wavefunctions are still localized to atoms, and metals to a state where wavefunctions extend over many atoms, in extreme cases even over the full dimensions of the solid. This behaviour is shown in figure 3.1.

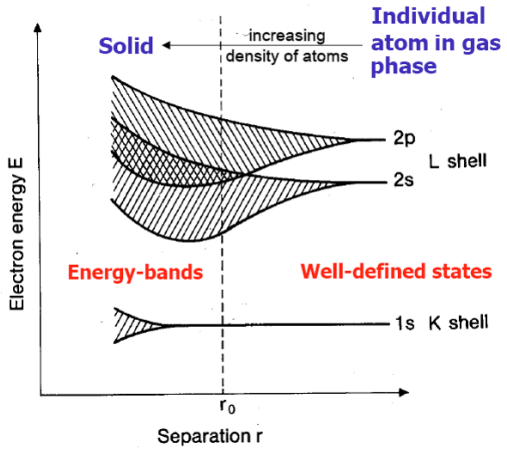


Figure 3.1: Formation of the energy bands as the atoms get closer to each other.

To describe a a solid, say a cube of edge-length L, we recognize that the electrons will be confined by some sort of potential that is binding throughout the solid, i.e. negative. For the region outside of the solid we can assume a large positive value or even infinity like an ideal potential well. Alternatively, and mathematically even simpler, we can consider a periodic continuation (called Born-von Karmann boundary conditions). Thus, the electronic wavefunctions can be defined as follows:

$$\Psi_{\vec{k}}(\vec{r}) = \frac{1}{\sqrt{V}} \cdot e^{i\vec{k}\vec{r}} \tag{3.1}$$

Here, the prefactor ensures that the wavefunction is normalized when integrating its modu-

lus $\Psi\Psi^*$ over the volume $V = L^3$. To accommodate the condition of periodicity, we define the wavevector $\vec{k}$ by $\vec{k} = 2\pi/L \cdot (n_x, n_y, n_z)^T$ where n_x, n_y, and n_z are integers that count how may full wavelengths we can fit into L. The endpoints of all these $\vec{k}$-vectors form a lattice which we call *reciprocal lattice*. For a large crystal the spacing becomes very dense and we may treat $\vec{k}$ as if it were a continuous variable. However, we should always keep in mind the discrete nature of the wavevectors. When we put the wavefunction into the free-particle Schrödinger equation, we find the energy E as function of the wavevector. This is called *dispersion relation*:

$$E = \frac{\hbar^2}{2m_e}\vec{k}^2 \tag{3.2}$$

In essence, this is just an expression for the kinetic energy in terms of the momentum $\vec{p} = \hbar\vec{k}$, but by means of the the Pauli exclusion principle and additionally taking into account two possible spin directions, every k-state can be associated with two electrons. The actual occupation of such states can then be described with statistics; for metals it is often a good approximation to use Fermi-Dirac statistics at 0 K. Thus, all the low-energy states are filled consecutively until all valence electrons are taken care of and the ensemble of filled states resembles a sphere in k-space. The electron with the highest energy defines the Fermi energy and the radius of the sphere is related to the Fermi-momentum. Dividing the volume of the Fermi-sphere by the volume of the elementary cell of the reciprocal space gives the number of occupied states N. Finally, if we divide by the volume of the crystal in real space, we obtain the electron density n, a quantity that is also accessible to optical measurements that yield the plasma frequency or other macroscopic measurements such as the Hall effect.

To describe semiconductors, we have to add a description of the crystal structure and we have to describe the binding potential on atomic level, the global description in terms of a quantum well is no longer sufficient. Let us consider a crystal where the positions of the atoms form a regular lattice denoted by $\vec{r}_n$. We can then assume that the valence electrons move in a *periodic potential* that is formed by the nuclei and the strongly bound electrons on the inner shells:

$$V(\vec{r}) = V(\vec{r} + \vec{r}_n) \tag{3.3}$$

The eq. (3.3) states that we could select any point within the crystal, move by an arbitrary lattice vector $\vec{r}_n$, and find exactly the same same potential as the one at the starting point. Even if we do not know the exact nature of this potential, the periodicity of the potential implies that the Schrödinger equation is no longer solved by simple plane waves, but by *Bloch waves*. As above, they contain a phase factor that contains the wavevector $\vec{k}$, but now they is an additional prefactor that has the same periodicity as the potential:

$$\Psi_{n,\vec{k}}(\vec{r} + \vec{R}) = u_{n,\vec{k}}(\vec{r})e^{i\vec{k}\vec{r}} \tag{3.4}$$

To assess the impact of the periodic crystal structure, let us start by looking at a simple cubic (sc) crystal with lattice spacing a. This means that the $\vec{k}$-vectors of eq. (3.4) are not only constrained by the dimensions of the solid, but additionally we have to fit multiples of their periods into the elementary cell of the crystal. In reciprocal space this introduces a superstructure for the $\vec{k}$-vectors because now they must additionally comply with a grid whose spacing is defined by $G = 2\pi/a$. For the simple cubic lattice, the reciprocal lattice is again a simple cubic lattice.

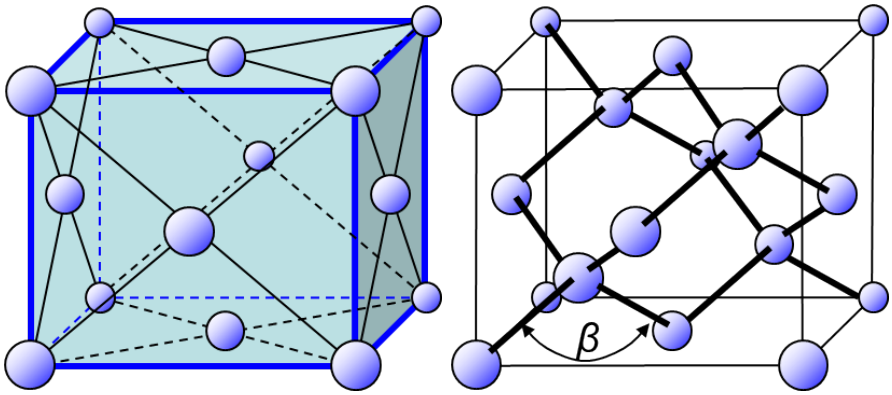


Figure 3.2: Illustration of the face centred cubic lattice (left) and the elementary cell of the diamond lattice (right).

At this point we are still free to choose an elementary cell. In semiconductor science we generally use the *Wigner-Seitz cell* and we call it the *first Brillouin zone*. To find it, we select one point at the origin of the reciprocal lattice. From there, we draw lines to all neighbouring points and design a perpendicular plane at half way. For the sc lattice, this yields a cube whose six faces are perpendicular to the x-, y-, and z-axes and located at distances of $\pm G/2$ away from the origin. Silicon does not crystallise in the sc lattice, but in the face centred cubic (fcc) lattice shown in figure 3.2 and silicon actually has additional atoms in the unit cell. The reciprocal lattice of fcc is the body centred cubic (bcc) lattice shown in figure 3.3, but due to the additional points on the faces, now the cube has an edge length of $G = 4\pi/a$ and the first Brillouin zone becomes a bit more complex. Like the sc lattice, it has six faces perpendicular to the main axes, but there are eight additional faces that are perpendicular to the diagonals of the cube, resulting in the octahedrally cropped cube shown in figure 3.3.

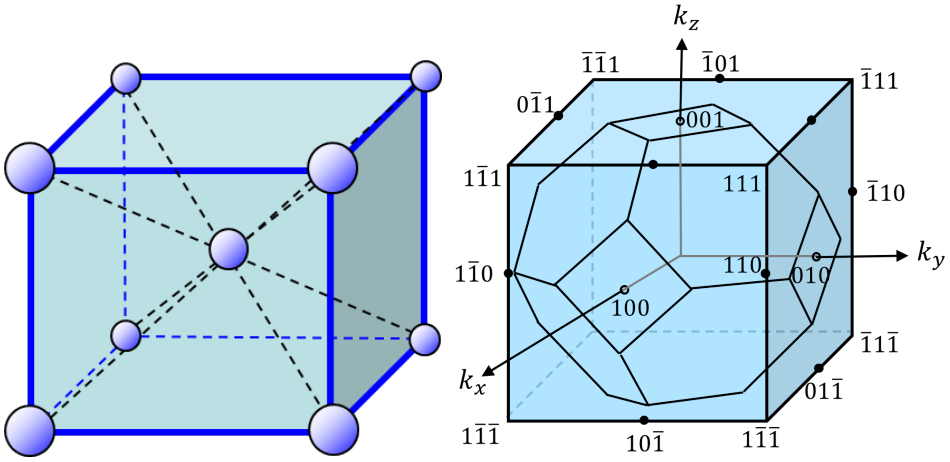


Figure 3.3: Illustration of the body centred cubic lattice (left) and its Wigner-Seitz cell which is a cropped octahedron (right).

As a result of the periodicity, we can describe the full dispersion relation within the first Brillouin zone. For example, we can evaluate it along the x-axis and as soon as k_x extends beyond $G/2$, we translate it back into the first Brillouin-zone by regarding the dispersion relation with a momentum equal to $k_x - G$. We can do the same at the other end of the first Brillouin zone, meaning that

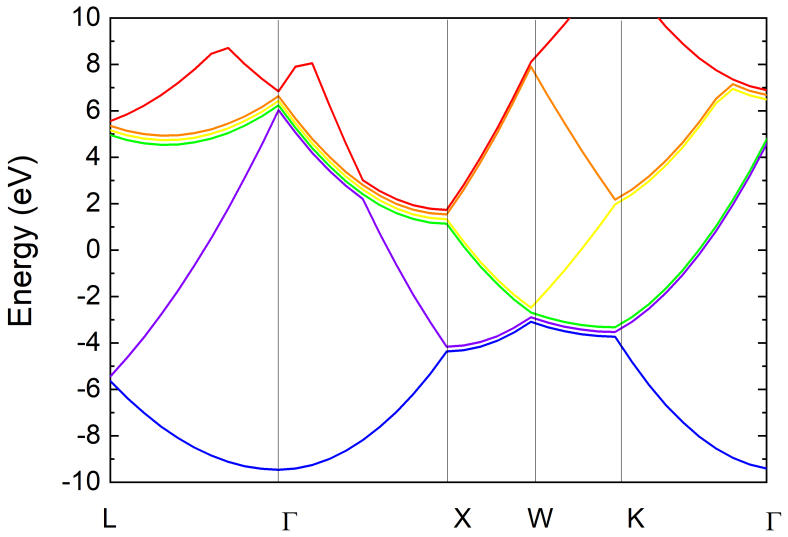


Figure 3.4: The band diagram of free electrons in the fcc lattice.

instead of regarding the dispersion relation beyond $-G/2$, we take shift it by taking $k_x + G$.

This is illustrated in figure 3.4 which plots the dispersion relation of a free electron in the fcc lattice. Between $\Gamma = \{0, 0, 0\}^T$ (the origin) and $X = \{G/2, 0, 0\}^T$, the dispersion relation is given by the blue parabola, and rather than continuing it beyond X, it flips and continues along the purple characteristic. The same happens if we follow the dispersion relation from X along the diagonal towards $L = \{G/2, G/2, G/2\}^T$. However, we need to keep track by how many times we applied the translation. We could do this by noting the colour, more formally this is done by the *band index* n that we already included in the definition of eq (3.4). For bands with high indices the processes of translation and folding gets ever more complicated and we notice that some bands get folded to the same location in the first BZ. We call these degenerate, and to avoid overloading of the diagram, only the first six bands are plotted in figure 3.4.

3.2 Bandgaps

Our understanding of the band structure of silicon and other semiconductors started with the interpretation of cyclotron resonance measurements using the $k \cdot p$ perturbation method and symmetry arguments to minimize the amount of calculations [15, 16]. Eventually, access to increasing amounts of computing power facilitated band structure calculations with the method of pseudo-potentials [17] which was also applied to calculate the band diagrams shown in figure 3.5.[1] Figure 3.5 shows the resulting energy-eigenvalues along various directions in k-space for the semiconductors Si and GaAs in a band-diagram. Looking at bands with with low index (blue and purple), it appears that

[1] The $k{\cdot}p$-method typically uses a small basis of 6 or 15 plane-waves and employs group-theoretical arguments to find a small number of non-zero matrix elements of the momentum operator which are subsequently related to measurable quantities like the curvature of the bands, i.e. their effective mass. The method of pseudo-potentials describes the periodic potential of elemental semiconductors by three Fourier coefficients (six for compound semiconductors such as GaAs), but the number of plane waves and therefore the number of matrix elements is much larger. For example, the band diagram shown in figure3.5 was calculated with a set of 124 waves. A Mathematica notebook is available from Aaron J. Danner, see: `https://www.ece.nus.edu.sg/stfpage/eleadj/pseudopotential.htm`.

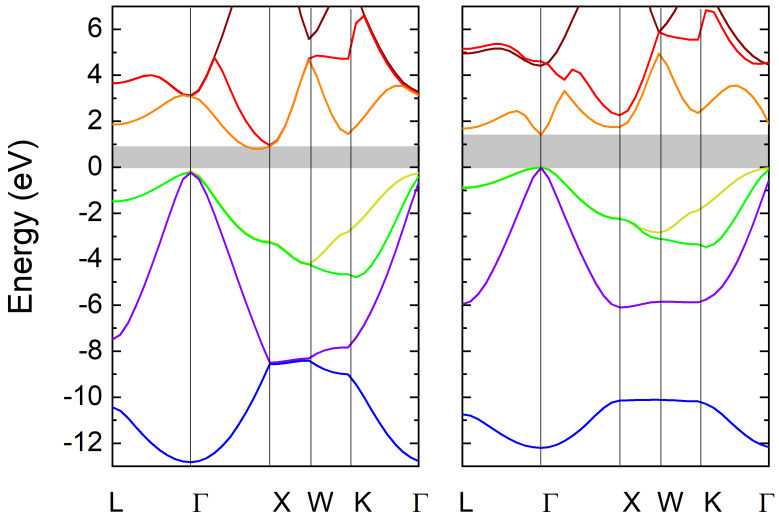


Figure 3.5: The band diagram of silicon (left) and GaAs (right) calculated with the pseudopoential method. The grey bars illustrate the bandgaps.

they propagate rather undisturbed and follow essentially the parabolic dispersion relation of free electrons. Elsewhere the characteristics can become considerably distorted, especially when the wave-vectors approach points of high symmetry in the reciprocal lattice. For example, at L, the blue branch does not connect continuously with the purple one, but it opens a gap in which no real energy-eigenvalue exists.

Further up the grey areas denote conditions where such an energy gap extends over all directions of k-space. Bands below this gap are called *valence bands* (VBs) because they are completely filled with valence electrons at 0 K. At temperatures above 0 K some electrons are excited into states above the gap which allows them to transport current, exactly like the electrons in a metal. For this reason the bands above the gap are called conduction bands (CBs). Note that only three of the CBs are shown for clarity, but above the red and the orange characteristics there are infinitely more bands at higher energies. As there are only a few thermally excited electrons in the CB, the conductivity is low and therefore we call these materials *semiconductors*.

Whenever electrons get excited into the CB, they leave behind unoccupied states in the VB which we call *holes*. Holes contribute to the charge transport in the semiconductor as well, but they are characterised by a positive charge and in most cases they have a lower mobility than electrons. For silicon, the bandgap extends from the valence band maximum (VBM) at the Γ-point to the conduction band minimum (CBM) which is found on the orange band at a wave-vector of $0.85X$. This behaviour is called *indirect bandgap* and the value of the energy gap is 1.12 eV at RT. For GaAs, the VBM and the CBM are both located at Γ, making it a *direct bandgap* with a value of 1.42 eV.

Most effects in electronics take place with charge carriers that are close to the bandgap. Looking at the CBM in figure 3.5, the dispersion relation does not look too different from a parabola. Therefore, we can apply a Taylor-expansion around the CBM and obtain a parabolic dispersion relation where the inverse of the curvature plays the role of an *effective mass*. More details on this approach are discussed in section A.2.

$$E_{k,\mathrm{CB}} = E_g + \frac{\hbar^2(\vec{k} - \vec{k}_{min})^2}{2m_e^*} \tag{3.5}$$

Applying the Taylor expansion to the VB yields negative curvature and thus negative effective mass. Whereas this appears strange at first sight, we can explain it as follows; when an electron gets excited from the VB, it leaves behind a hole. An electron from an adjacent state in the VB can easily jump into this hole, leaving behind a hole in the state it came from. Thus, we can effectively consider the hole like a moving entity, but its movement is opposite to the the one of an electron. Moreover, charge neutrality requires that a hole should have a positive charge which we can accommodate with a negative curvature. The dispersion relation of holes reads:

$$E_{k,\mathrm{VB}} = -\frac{\hbar^2\vec{k}^2}{2m_h^*} \tag{3.6}$$

For most of our work on solar cells, it is sufficient to know the values of the bandgap and the effective masses. Except for a description of optical absorption effects in chapter 4 the finer details of the band diagram are hardly needed anymore. Figure 3.6 illustrates once more the distinction between *metals*, *semiconductors* and *insulators*. For metals, there is usually no gap, or the VB is only *partially filled*. Thus, the electrons need only a small amount of energy to hop into a free state that allows them to move. For semiconductors and insulators there is a gap and the VB is entirely filled. Thus, the electrons need to cross the bandgap before they can find free states. For semiconductors this energy gap is in the order of 1 eV, thus the thermal energy which is $kT = 26\,\mathrm{meV}$ at room temperature, is sufficient to promote a few carriers into the conduction band. It also means that a semiconductor kept at very low temperature will behave like an insulator because the thermal energy is too small to excite electrons. In insulators the band gap is much larger (several eV) and thus it is very improbable for electrons to be excited across the gap at RT.

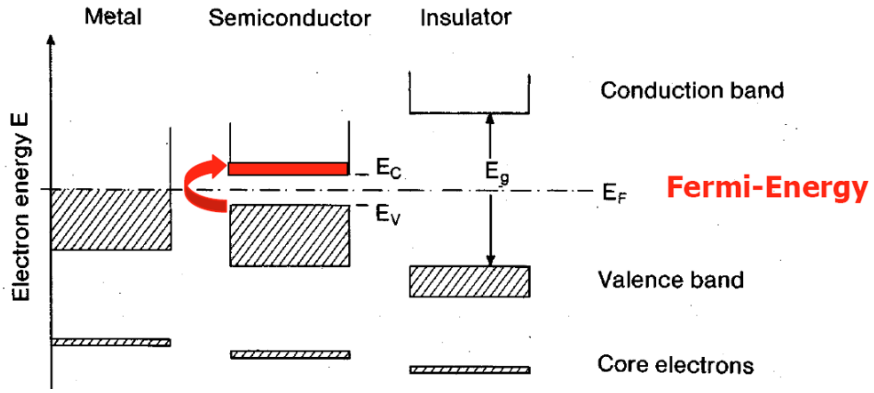


Figure 3.6: Behaviour of metals, semiconductors and insulators.

3.3 Fermi-Dirac statistics, density of states and concentration of carriers

As electrons and holes are *fermions* (half-integer spin particles) their occupation is governed by the exclusion principle. Thus, we can put only *one* electron into a state which is uniquely described by

a band index, the components of the wave vector and a spin.[2] The *occupation probability* of these states is described by *Fermi-Dirac statistics* which is given by:

$$f_{FD}(E, E_F) = \frac{1}{\exp\left(\frac{E-E_F}{kT}\right) + 1} \tag{3.7}$$

For electrons this means, that at $T = 0\,\mathrm{K}$ all the states below the *Fermi level (energy)* E_F are occupied and all the states above are empty.[3] At a finite temperature this distribution *smears out* and states a few kT away from the Fermi level can be occupied. The evolution of the Fermi-Dirac distribution function at different temperatures is shown in Figure 3.7.

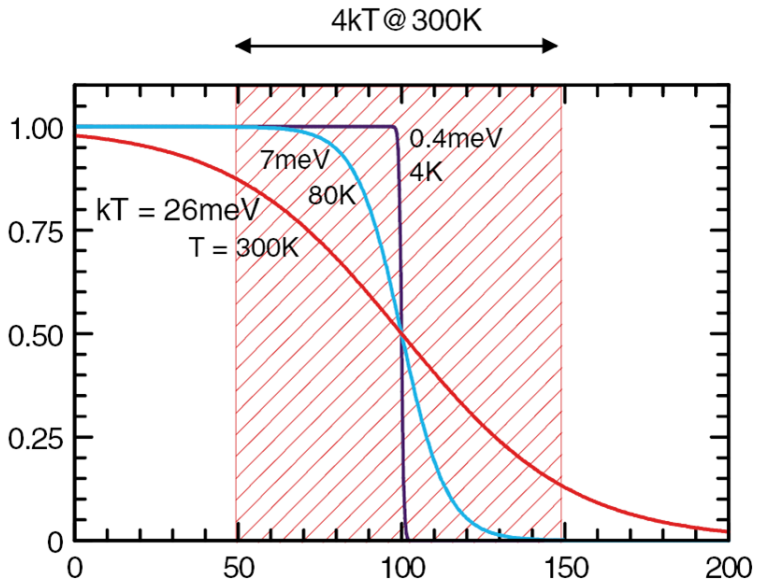


Figure 3.7: Fermi-Dirac distribution $f_{FD}(E, E_F)$ at different temperatures for a Fermi-energy of 100 meV.

To find out how many electrons are excited, we consider the occupation statistics in $\vec{k}$-space. Starting with the state $\vec{k} = (0, 0, 0)^T$, we go through all states of the conduction band and calculate their energy. Next, we use $f(E(\vec{k}))$ to decide whether the state is occupied or not, and we sum them up. Finally, we note that each state can accommodate two carriers of opposing spin, and we divide by the sample volume $V = L^3$ as we are normally interested in carrier densities.

$$n_0 = \frac{2}{L^3} \cdot \sum_{\vec{k}(CB)} f_{FD}(E(\vec{k})) = \frac{2}{L^3} \cdot \frac{L^3}{(2\pi)^3} \sum_{\vec{k}(CB)} f_{FD}(E(\vec{k}))\Delta\vec{k} \tag{3.8}$$

Here, we added an index "0" to denote thermal equilibrium. We also extended the sum by the volume element of $\vec{k}$-space, $\Delta\vec{k} = (2\pi)^3/L^3$. The extension is cancelled by the pre-factor, but it allows us replace the sum by a volume integral:

[2] Since we do not explicitly deal with magnetic fields that would express spin interactions, it is usually safe to work with *two* electrons per state.

[3] Strictly speaking the definition is the other way round, the Fermi level (or energy) is defined as the energy up to which electronic states are occupied at $T = 0\,\mathrm{K}$. We should more correctly use the term *chemical potential* μ which is defined also for positions within the bandgap, regardless whether there are allowed states or not.

$$n_0 \approx \frac{1}{4\pi^3} \int_{\vec{k}(CB)} f_{FD}(E(\vec{k}))d^3k \tag{3.9}$$

By extending the integral only over the distribution function $f(E)$, we obtain the density of excited electrons. More generally, we can use the same integral to calculate other averages that need to be weighted by the carrier density, simply by multiplying them to $f_{FD}(E)$ in the integrand. For example, the energy density of the electrons is determined by the following:

$$u_0 = \frac{U}{V} = \frac{1}{4\pi^3} \int_{\vec{k}(CB)} E(\vec{k}) \cdot f_{FD}(E(\vec{k}))d^3k \tag{3.10}$$

As the integral of eq. 3.9 is not dependent on any direction in $\vec{k}$-space but only on the energy,[4] we can carry out the integration in spherical coordinates, and we can substitute the wavevector modulus k by the energy E which appears already in $f(E)$.[5] Thus, the volume element of the integration is transformed as $\frac{1}{4\pi^3}d^3k = g(E)dE$. For the case of a spherical energy surface with one single with effective mass m^*, $g(E)$ is given by:

$$g(E) = \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2}\right)^{\frac{2}{3}} (E - E_C)^{\frac{1}{2}} \tag{3.11}$$

Since $g(E)$ describes the number of states per volume and per energy, we call it *density of states* (DOS). It allows us to calculate the electron density n_0 by the following integral:

$$n_0 = n(E_F) = \int_{E_C}^{\infty} g(E)f(E, E_F)dE \tag{3.12}$$

$$\approx \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2}\right)^{\frac{2}{3}} \underbrace{\int_0^{\infty} E^{\frac{1}{2}} e^{-\frac{E}{kT}} dE}_{\frac{1}{2}(kT)^{\frac{3}{2}}\sqrt{\pi}} \cdot e^{-\frac{E_c - E_F}{kT}} \tag{3.13}$$

Here, the rapid exponential decay of the Fermi-Dirac distribution function allowed us to replace the upper limit by ∞. In the second step we shifted the lower limit of the integral by E_C since . $g(E) = 0$ is for energies $E < E_C$. Finally we approximated the unwieldy Fermi-Dirac distribution with the Maxwell-Boltzmann distribution. This is only allowed when the Fermi-level is more than $3kT$ away from the band edges. Cases where we have to evaluate the full integral of eq. 3.12 are discussed in appendix B.

The same Fermi-Dirac distribution function also describes the states in the VB, but it is convenient to regard empty electron-states as filled hole-states. Thus, we can define the probability of filling a state in the VB by $f_{FD,p} = 1 - f_{FD,n}$ is filled To assess the properties of holes in the VB,

[4] A minor directional dependence arises when the energy surfaces at the VBM or the CBM are not spherical but elliptical as discussed in appendix A.1. The resulting differences in the effective masses can be incorporated into scaled spherical coordinates. Another example for a directional dependence is discussed in appendix H where we calculate the current density flowing in positive x-direction, i.e. the current of those carriers that have a positive velocity component v_x.

[5] In three dimensions, the Jacobian for spherical coordinates is $4\pi k^2 dk$ and we can replace k^2 and dk with the help of the dispersion relation $E(\vec{k})$. The resulting DOS varies with the square-root of E. For two dimensions we have to use polar coordinates and the Jacobian is given by kdk, resulting in a constant DOS. For one dimension the Jacobian is constant and $g(E)$ varies like $E^{-1/2}$.

we consider the same distribution function can be treated exactly the same way, we can obtain the following expressions for the *thermally excited carrier densities*.

$$n_0 = 2\left(\frac{m_e^* kT}{2\pi\hbar^2}\right)^{\frac{3}{2}} \cdot \exp\left(-\frac{E_C - E_F}{kT}\right) = N_C \cdot \exp\left(-\frac{E_C - E_F}{kT}\right) \tag{3.14}$$

$$p_0 = 2\left(\frac{m_h^* kT}{2\pi\hbar^2}\right)^{\frac{3}{2}} \cdot \exp\left(-\frac{E_F - E_V}{kT}\right) = N_V \cdot \exp\left(-\frac{E_F - E_V}{kT}\right) \tag{3.15}$$

The prefactors N_V and N_C are called *effective density of states* of the VB and of the CB, respectively. As every excited electron leaves behind exactly one hole, we have $n_0 = p_0$. In this case, the *intrinsic Fermi level* E_i is close to mid-gap with a small dependence on the temperature T:

$$E_i = \frac{E_C + E_V}{2} + \frac{kT}{2}\ln\left(\frac{N_V}{N_C}\right) \tag{3.16}$$

The product of n_0 and p_0 yields the square of the *intrinsic carrier concentration*:

$$n_0 \cdot p_0 = n_\mathrm{i}^2 = N_V N_C \exp\left(-\frac{E_g}{kT}\right) = \text{const.} \tag{3.17}$$

For silicon at 300 K, the values for N_C and N_V are $2.86 \times 10^{19}\,\mathrm{cm}^{-3}$ and $3.1 \times 10^{19}\,\mathrm{cm}^{-3}$, respectively, and $n_i \approx 1 \times 10^{10}\,\mathrm{cm}^{-3}$ is often used for the intrinsic carrier density.[6] Compared to the atomic density in silicon of $5 \times 10^{22}\,\mathrm{cm}^{-3}$ the intrinsic carrier density is not very high, i.e. about *one electron and one hole* for 10^{12} atoms. The values for Ge ($E_g = 0.8\,\mathrm{eV}$) and GaAs ($E_g = 1.42\,\mathrm{eV}$ are 1×10^{12} and $1 \times 10^{6}\,\mathrm{cm}^{-3}$, respectively. Obviously the higher the band gap, the lower the intrinsic carrier density.

3.4 Doping

It is possible to change the carrier concentration through *doping*, i.e. introducing different atoms into the host material. For silicon, the most common dopants are boron for p-doping and phosphorus for n-doping. For P occupying a substitutional site $\mathrm{P_{Si}}$, only four electrons are needed for bonding to the surrounding lattice. The fifth electron is not involved in the bonding and thus easily released to become a free carrier. For boron this is a little less intuitive because it lacks one electron to fully bond to its neighbours. Assuming that occasionally an electron is released from neighbouring bonds, the released electrons may then attach to the boron atom, leaving behind a hole somewhere else.

We can use the model of the hydrogen atom to estimate the Bohr radius and the binding energy. To this end, we have to use the permittivity of the semiconductor and the respective effective masses of the carriers.

$$a_0^* = a_0 \cdot \frac{\epsilon_{Si}}{m^*/m_0} \tag{3.18}$$

[6] In undoped silicon, $n_i = 9.65 \times 10^9\,\mathrm{cm}^{-3}$ at 300 K, however, the value increases with doping [18]. The value quoted above is still a good approximation for p-type wafers with resistivities of 1 - 2 Ω cm typically used for solar cells.

For the hydrogen atom, the Bohr radius of the electron is $a_0 = 0.3\,\text{nm}$, but for silicon with $\epsilon_{Si} = 11.7$ we obtain 5 - 10 nm which means that an electron is delocalised over several atoms. To estimate the binding energy, we can use the formula for ionisation from a state in orbital n:

$$E_{e,h}^{H} = \frac{m_{e,h}^{*} q^4}{2\left(4\pi\epsilon_0\epsilon_{\text{Si}}\hbar\right)^2} \cdot \frac{1}{n^2} \tag{3.19}$$

Instead of 13.2 eV which is the binding energy of the electron in the hydrogen atom, we obtain binding energies of 20 - 30 meV. Thus, the thermal energy kT at room temperature is sufficient to fully dissociate an electron from the donor. In a band diagram, we can depict donor states within the bandgap just below the CB. Likewise, holes will easily dissociate from their acceptors and we can depict acceptor states just above the VB. The probabilities of ionizing an acceptor or a donor are given by the following relations:

$$N_A^- = N_A \frac{1}{1 + 4\exp\left(\frac{E_A - E_F}{kT}\right)} \tag{3.20}$$

$$N_D^+ = N_D \frac{1}{1 + 2\exp\left(\frac{E_F - E_D}{kT}\right)} \tag{3.21}$$

Here, the different prefactors in the denominators are related to the number of bands contributing to the VB and the CB. Overall, the doped semiconductor is still neutral, and therefore we can formulate an equation for charge neutrality:

$$p_0 - n_0 - N_A^- + N_D^+ = 0 \tag{3.22}$$

Doping changes the equilibrium carrier densities n_0 and p_0, but their product $n_0 p_0 = n_i^2$ is still defined by eq. (3.17), similar to the mass-action law in chemistry. Also the Fermi-level does not stay at its intrinsic level, but it moves towards the CB or the VB for n-type and p-type doping, respectively. For example, if we are dealing with boron doping in silicon, virtually all dopant states are ionized at room temperature because of their low ionization energy, and therefore we can approximate as follows:

$$p_0 \approx N_{\text{A}}^- \approx N_{\text{A}} \tag{3.23}$$

$$n_0 = n_i^2 / N_{\text{A}} \ll p_0 \tag{3.24}$$

If the semiconductor is p-doped with $N_A = 1 \times 10^{16}\,\text{cm}^{-3}$ acceptors, then the equilibrium hole density can be approximated very well by $p_0 = 1 \times 10^{16}\,\text{cm}^{-3}$ as well. The equilibrium electron density is only $n_0 = 1 \times 10^{4}\,\text{cm}^{-3}$. Because of the enormous numerical differences, the holes are called *majority carriers* in a p-type semiconductor whereas the electrons are *minority carriers*. In n-type material it is the other way round.

Often we know the carrier density from measurement and we are interested in the position of the Fermi level. Therefore, we rearrange eqns. (3.14) and (3.15) as follows.

$$E_C - E_F = kT \ln\left(\frac{N_C}{n_0}\right) \tag{3.25}$$

$$E_F - E_V = kT \ln\left(\frac{N_V}{p_0}\right) \tag{3.26}$$

Literature sometimes defines an electron potential Ψ_n and a hole potential Ψ_p by referring the Fermi-levels to the intrinsic level. We can achieve this by subtracting eq. (3.25) from eq. (3.26) and identifying the intrinsic level as defined in eq. (3.16):

$$q\Psi_n = E_F - E_i = kT \ln\left(\frac{n_0}{n_i}\right) \tag{3.27}$$

$$q\Psi_p = E_i - E_F = kT \ln\left(\frac{p_0}{n_i}\right) \tag{3.28}$$

For the numeric example of the p-doped semiconductor we find $q\Psi_n = -0.359\,\mathrm{eV}$ and $q\Psi_p = 0.359\,\mathrm{eV}$.

3.5 Quasi Fermi levels

So far we assumed that there is a thermal equilibrium between electrons and holes. This is no longer the case when we create additional charge carriers, for example by absorption of light. This is discussed in more detail in chapter 4, but in essence the absorption process creates an additional pair of electrons and holes for every absorbed photon. Depending on the energy of the photons, these are excited into states that are far from their respective band edges and we call them *hot carriers* because it would need a temperatures of several thousand K to describe the occupation of such states with statistics. From there, they undergo multiple interactions with phonons which are on on a timescale of ps. Eventually they reach states close to the band edges. Ultimately, from there they will recombine through one of the processes discussed in chapter 5, but since at least two carriers are needed for recombination, it takes place on a much longer timescale which we call bulk lifetime τ_b. In most cases the lifetime is long enough to assume that the carriers are in thermal equilibrium within their respective bands. Since the two occupation statistics are governed by the temperature of the lattice, the whole process is called *thermalisation*.

The semiconductor is thus in a *steady state* where the illumination creates hot carriers which continuously thermalize to the pool of carriers at the band edges. This allows us to define two different (quasi-)Fermi levels $E_{F,n}$ and $E_{F,p}$ for the two carrier populations, both of them in terms of the lattice temperature.

The steady state carrier densities are called Δn and Δp, respectively, and since the absorption of photons creates pairs of electrons and holes we have $\Delta n = \Delta p$. Staying with the numerical example of a p-doped semiconductor with $N_A = 1 \times 10^{16}\,\mathrm{cm}^{-3}$, let us assume that illumination creates an additional density of electrons and holes equal to $\Delta n = \Delta p = 1 \times 10^{15}\,\mathrm{cm}^{-3}$. Thus, we have to drop the subscript "0" from the hole density since we are no longer dealing with thermal equilibrium, and we can find $p = p_0 + \Delta p = 1.1 \times 10^{16}\,\mathrm{cm}^{-3}$. As p is not very different from the majority density that we had in thermal equilibrium, we can use $p \approx p_0$ for the majority carriers. However, for the minority density there is a huge difference because it increases from $n_0 = 1 \times 10^4\,\mathrm{cm}^{-3}$ by 11 orders of magnitude to $n = n_0 + \Delta n = 1.000\,000\,000\,01 \times 10^{15}\,\mathrm{cm}^{-3}$. For the illuminated case it is therefore safe to approximate the minority carrier density by $n = \Delta n$.

We can describe this steady state situation by borrowing the definitions from the equilibrium, but we have to keep in mind that there are now two different Fermi levels for electrons and for holes. We define them the same way as in eqns. (3.25) and (3.26), but we have to insert the non-equilibrium carrier densities.

$$E_C - E_{F,n} = kT \ln\left(\frac{N_C}{n}\right) \tag{3.29}$$

$$E_{F,p} - E_V = kT \ln\left(\frac{N_V}{p}\right) \tag{3.30}$$

The splitting of the QFLs is obtained by $\Delta_{QFL} = E_{F,n} - E_{F,p}$. For the p-type semiconductor we can approximate as follows:

$$\Delta_{QFL} \approx kT \ln\left(\frac{\Delta n N_A}{n_i^2}\right) \tag{3.31}$$

It is important to note that under illumination the carrier densities n and p are no longer related via the mass action law. Instead, we define the pn-product as a new quantity for the steady state.

$$pn = (p_0 + \Delta n)(n_0 + \Delta n) = n_i^2 \exp\left(\frac{\Delta_{QFL}}{kT}\right) \tag{3.32}$$

In terms of the electron and hole potentials, the splitting of the QFLs is obtained by $\Delta_{QFL} = E_{F,n} - E_{F,p} = q\Psi_n + q\Psi_p$. For our numerical example from above, we obtained $q\Psi_n = -0.359\,\text{eV}$ and $q\Psi_p = 0.359\,\text{eV}$ without illumination and we obtain $q\Psi_n + q\Psi_p = 0$, expressing that there is a unique Fermi level. Under illumination, the potentials change to $q\Psi_n = 0.299\,\text{eV}$ and $q\Psi_p = 0.362\,\text{eV}$ and the Fermi levels split by $q\Psi_n + q\Psi_p = 0.661\,\text{eV}$.

In chapter 6 we will see that the splitting of the quasi-Fermi levels is related to the maximum value that we can ideally expect for the open circuit voltage of a solar cell. In a ideal case we could therefore hope to measure an open circuit voltage of 0.661 V at the electrodes.

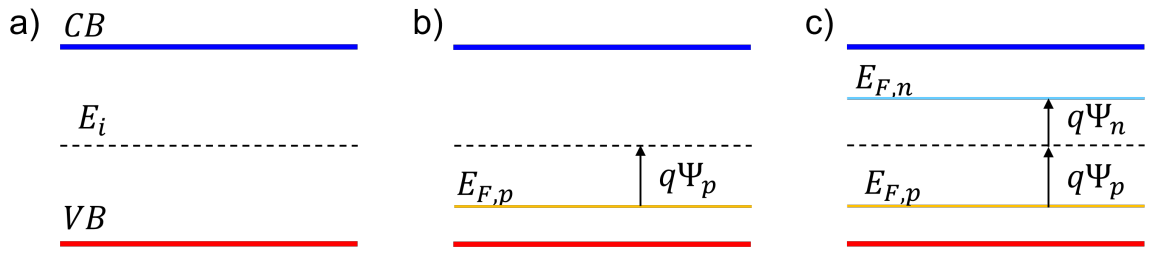


Figure 3.8: Illustration of a) the intrinsic Fermi level, b) a Fermi level close to the VB in a p-type semiconductor with the definition of $q\Psi_p$, and c) QFL splitting of a p-type semiconductor under illumination and the definition of $q\Psi_p$ and $q\Psi_n$.

3.6 Drift current

Paul Drude proposed a rather simple, but very successful model to describe the transport of electrons in metals [19, 20]. He assumed that the valence electrons are no longer bound to their atoms, but form a plasma that moves within the lattice of the ionised cores like an ideal gas of point-like particles. Occasionally the electrons undergo collisions which fully randomise the direction of their

velocity. The time between the collisions is called relaxation time τ_r. In the absence of an electric field, the average over all velocity vectors evaluates to zero. Nevertheless, the individual velocities are not zero. We can calculate the average thermal energy of an electron by dividing the energy density of all carriers from eq. (3.9) by the carrier density according to eq.(3.10), and identify this with the kinetic energy of the electron.

$$E = \frac{u_0}{n_0} = \frac{3}{2}kT = \frac{1}{2}m_e^* v^2 \tag{3.33}$$

This yields a relation for the average velocity of the electron which we call *thermal velocity*:

$$v_{th} = \sqrt{\frac{3kT}{m_e^*}} \tag{3.34}$$

We can obtain the same result from the equipartition theorem of thermodynamics. It states that an energy of $1/2 \cdot kT$ can be attributed to each degree of freedom, here referring to movement in three directions of space. At a temperature of 300 K, $v_{th} \approx 1 \times 10^7\,\mathrm{cm\,s^{-1}}$ with a small difference between electrons and holes because of their different effective masses as discussed in appendix A.2.

In the presence of an electric field, electrons still undergo collisions that fully randomise their directions. However, between the collisions the field accelerates them for a time τ_r, and thus the ensemble average moves with the *drift velocity* $\vec{v}_d$ which is directed parallel to the field. For the case of electrons and holes, we obtain:

$$\vec{v}_{d,n} = -\vec{a} \cdot \tau_{r,n} = -\frac{q\vec{E}}{m_e^*} \cdot \tau_{r,n}\vec{E} = -\mu_n \cdot \vec{E} \tag{3.35}$$

$$\vec{v}_{d,p} = \vec{a} \cdot \tau_{r,p} = \frac{q\vec{E}}{m_h^*} \cdot \tau_{r,p} = \mu_p \cdot \vec{E} \tag{3.36}$$

Here, we defined the charge carrier mobilities μ_e and μ_p in units of [$\mathrm{cm^2/Vs}$]. They describe the ability of charge carriers to move within an electric field $\vec{E}$. Sometimes we want to describe them in terms of the effective masses and and the relaxation times.

$$\mu_n = \frac{q\tau_{r,n}}{m_e^*} \tag{3.37}$$

$$\mu_p = \frac{q\tau_{r,p}}{m_h^*} \tag{3.38}$$

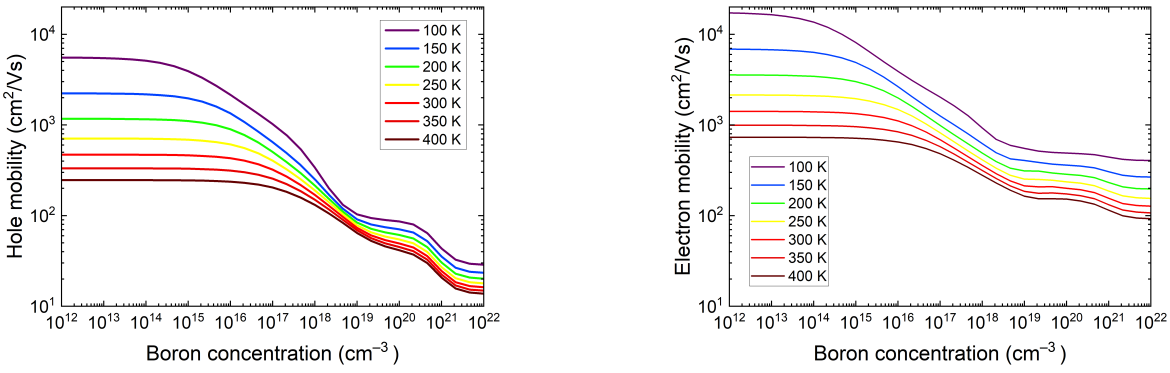
More often we will deal with the *drift currents* that are associated to the moving charges. They are defined in terms of the electric field and the respective conductivities σ_n and σ_p:

$$\vec{j}_n^{\,\mathrm{drift}} = -q \cdot n\vec{v}_{d,n} = \underbrace{q \cdot n \cdot \mu_n}_{\sigma_n} \cdot \vec{E} \tag{3.39}$$

$$\vec{j}_p^{\,\mathrm{drift}} = q \cdot p\vec{v}_{d,p} = \underbrace{q \cdot p \cdot \mu_p}_{\sigma_p} \cdot \vec{E} \tag{3.40}$$

There are two main mechanisms that limit the mobility in crystalline material. A fundamental and unavoidable scattering process is due to the thermal oscillations of the atoms around their

lattice sites. These are very well described by sound waves or phonons, and therefore we speak of *phonon scattering*. As we introduce dopants into the material, moving charge carries are additionally scattered by the Coulomb potential of the ionised cores. Accordingly, we speak of *ionised impurity scattering*. In multicrystalline material, there is additional scattering at grain boundaries [21]. For p-type c-Si, the variation of the hole mobility with doping concentration and temperature is illustrated in figure 3.9(a).



(a) (b)

Figure 3.9: Dependence of the mobility on dopant concentration and temperature in p-type silicon. Panel (a) shows mobility of holes (majority carriers), panel (b) the one of electrons (minority carriers); data from pvlighthouse.

Since electrons and holes contribute to the current, the total drift current is given by:

$$\vec{j}_{\text{total}}^{\,\text{drift}} = \underbrace{q\,(n\mu_n + p\mu_p)}_{\sigma}\,\vec{E} \tag{3.41}$$

Here σ_n and σ_p are the *conductivities* in units of $\mathrm{S\,cm^{-1}}$, and the subscripts n and p denote electrons and holes, respectively. Thus, $\sigma = \sigma_n + \sigma_p$ is the total conductivity. The inverse of the conductivity is the *resistivity* ρ in units of $\Omega\,\mathrm{cm}$. Generally, c-Si wafers are classified by the doping type and ρ. Other properties like the doping concentrations N_A and N_D or the mobilities μ_n, μ_p can then be inferred. Typical values are:

- **p-type**: $\rho = 2\,\Omega\,\mathrm{cm} \rightarrow \sigma = 0.5\,\mathrm{S\,cm^{-1}}$
  In this case $N_A = 7 \times 10^{15}\,\mathrm{cm^{-3}}$ and E_F is 0.23 eV above the VB edge.
- **n-type**: $\rho = 4\,\Omega\,\mathrm{cm} \rightarrow \sigma = 0.25\,\mathrm{S\,cm^{-1}}$
  In this case $N_D = 1.1 \times 10^{15}\,\mathrm{cm^{-3}}$ and E_F is 0.26 eV below the CB edge.

Besides the majority mobilites used here in the context of the conductivity, we will later find that the description of solar cells actually relies very strongly on the properties of the of the minority carriers. The figure 3.9(b) shows that the electron (minority) mobility in p-type c-Si is about three times higher than the hole (majority) mobility.[7]

[7] Whereas the very first solar cells were made with n-type wafers, application as power source for satellites quickly showed that p-type cells are more tolerant against bombardment with electrons and protons in space, especially in the altitude of the van Allen belt. The longer minority diffusion length in p-type Si was a additional asset [22]. The choice of p-type wafers was maintained for almost 60 years and only in the 2020s n-type wafers started gaining a larger market share because of high-efficiency cell designs such as IBC, heterojunction or TOPCon.

The electric field vector $\vec{E}$ can be expressed by the gradient of the electrostatic potential Φ_{el} which is a scalar with units of [V]. In the context of semiconductors we are dealing more often with band diagrams in terms of energies. We can reconcile the two descriptions by recognizing that the electrostatic potential Φ_{el} imparts a potential energy $q \cdot \Phi_{el}$ on a positive charge, and by keeping in mind that band diagrams are referred to the vacuum level $q\Phi_{vac}$ which is the negative of $q\Phi_{el}$. Figure 3.10 illustrates the different definitions for an n-type semiconductor under applied bias voltage,[8] and it shows how the CB of the semiconductor is constructed by a downwards shift of the vacuum level by an amount equivalent to the electron affinity χ. The VB is found at yet lower energy, separated from the conduction band by an energy equivalent to the bandgap E_g.

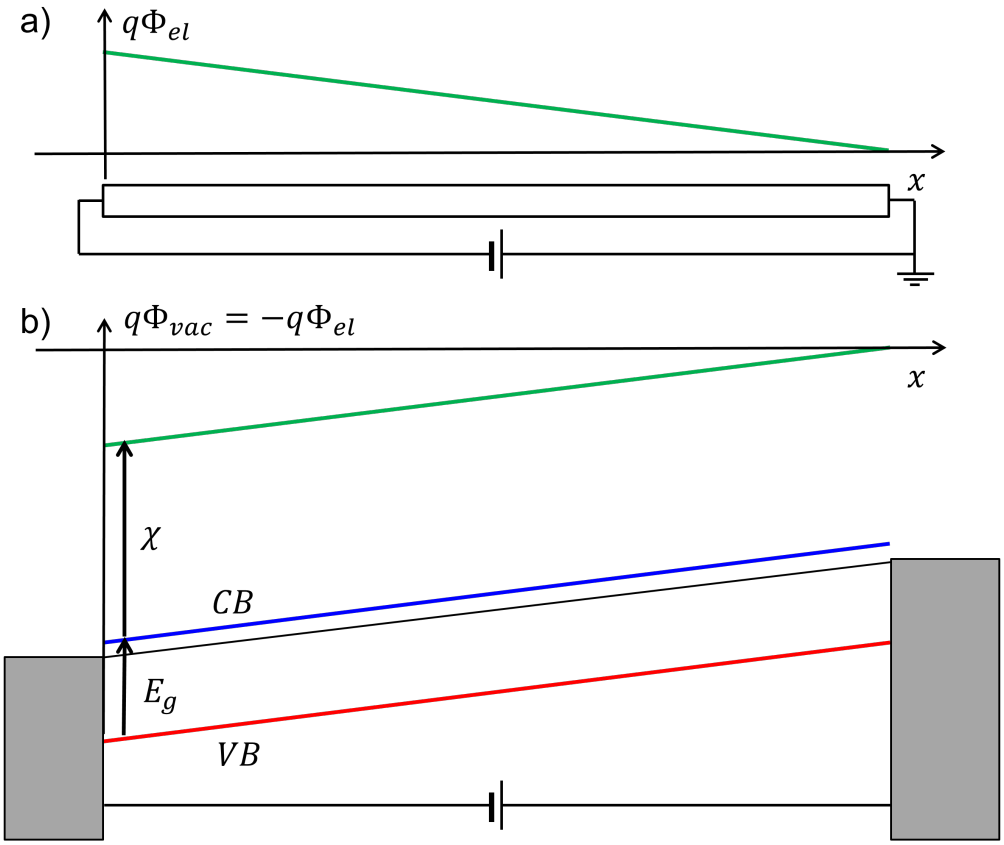


Figure 3.10: Panel a) shows the variation of the electrostatic potential along a uniform conductor under bias voltage and grounded negative electrode. Assuming that the conductor is an n-type semiconductor with Ohmic contacts, panel b) illustrates the corresponding band diagram.

The change of sign between the electrostatic potential and the vacuum level must also be kept in mind for the movement of charge carriers. In band diagrams, it is often depicted that electrons "roll down" the slope of the CB whereas positively charged holes "move up" along the slope of the VB like air-bubbles trapped under a sheet of ice. We have to be very careful though, after the next section we will have a closer look at the transport equations and we will find that this picture is a bit too simplistic.

[8] We assumed perfect Ohmic contacts where the Fermi-levels of the metal and the semiconductor align without depletion. The properties of this type of contact are discussed in more detail in appendix H.

3.7 Diffusion current

When the carrier density is spatially inhomogeneous, there is a tendency for carriers to diffuse from regions of high density towards regions of low density. This corresponds to a *diffusion current* which is defined by:[9]

$$\vec{j}_n^{\text{diff}} = q \cdot D_n \cdot \vec{\nabla} n \tag{3.42}$$

$$\vec{j}_p^{\text{diff}} = -q \cdot D_p \cdot \vec{\nabla} p \tag{3.43}$$

Here, D_n and D_p are the *diffusion coefficients* of electrons and holes, respectively. In chapter 6 it turns out that diffusion is of paramount importance for the treatment of most solar cells. The *Einstein relations* allow us to express the diffusion coefficients in terms of the temperature T and the mobilities μ_e and μ_h:

$$D_n = \frac{kT}{q}\mu_n \tag{3.44}$$

$$D_p = \frac{kT}{q}\mu_p \tag{3.45}$$

3.8 Total current

The total current is simply the sum of the *drift current* and the *diffusion current* for both, electrons and holes.

$$\vec{j} = \vec{j}_n + \vec{j}_p = q\left(n \cdot \mu_n + p \cdot \mu_p\right) \cdot \vec{E} + q\left(D_n \cdot \vec{\nabla} n - D_p \cdot \vec{\nabla} p\right) \tag{3.46}$$

We can rearrange the expressions for electron and hole currents a little by replacing the field by the gradient of the electrostatic potential and by applying the Einstein relations to express the diffusion coefficients in terms of the mobilities. Additionally, we resolve eqns. (3.29) and (3.30) for n and p to simplify the gradient operation in the second term. As the energy offsets between the vacuum level and the bands are constant, we find that their gradients cancel each other. The same rearrangement holds for the hole current and eventually we obtain the following:

$$\begin{aligned}
\vec{j}_n &= q\mu_n\vec{E} + qD_n\vec{\nabla} n \\
&= -\mu_n\vec{\nabla}(q\Phi_{el}) + kT\mu_n n\vec{\nabla}\left(N_C e^{(E_F-E_C)/kT}\right) \\
&= \mu_n\vec{\nabla}(q\Phi_{vac}) + kT\mu_n N_C e^{(E_F-E_C)/kT)} \cdot \vec{\nabla}\left((E_F - E_C)/kT\right) \\
&= \mu_n n\vec{\nabla} E_{F,n} \qquad (3.47) \\
\vec{j}_p &= \mu_p p\vec{\nabla} E_{F,p} \qquad (3.48)
\end{aligned}$$

This means that the true driving force for the electron and hole currents is not related to the gradients of the respective bands, but to the gradients of the respective quasi Fermi-levels.

[9] By definition, the gradient directs towards the place of higher concentration and therefore the direction of diffusing particles is opposite to the gradient. Thus, a "$-$" is needed in the definition of the hole current to express the direction of diffusion of positive charges. For the electron current, we have to multiply with another "$-$" to express the current direction opposite to the flow of the negative charges.

Chapter 4

Optical properties

4.1 Dielectric permittivity

The interaction of light with matter is described by the *dielectric permittivity* ϵ (sometimes also called dielectric function).

$$\epsilon(\hbar\omega) = \epsilon_1(\hbar\omega) + i\epsilon_2(\hbar\omega) \tag{4.1}$$

The response is always *causal*, i.e. has a certain delay to the cause, the excitation by the electromagnetic field. This is represented by a *phase shift* which in turn gives rise to the imaginary part. Additionally, the response is not the same for different frequencies ω of the incident light. The dependence of ϵ on ω (or the energy of the incident photons, $\hbar\omega$) is called *dispersion*.

We can imagine the interaction between light and matter very simplistically by assuming that an electron is bound to its core by a spring. If we describe the incident light by a sinusoidally varying electromagnetic field, it can interact with the dipole moment formed between the negatively charged electron and the positive core. Depending on damping and spring constant, the system will follow the excitation only little at low frequencies, very strongly at the resonance frequency, and again very little at very high frequencies. The sum of the individual dipole moments yields the macroscopic polarization P which is related to the incident field E through the susceptibility χ, resulting in a permittivity ϵ with Lorentzian shape:

$$\epsilon = 1 + \frac{e^2 N/V}{\epsilon_0 m_e} \cdot \frac{1}{\omega_0^2 - 2i\beta\omega - \omega^2} \tag{4.2}$$

Here, N/V is the volume density of oscillators, ϵ_0 is the dielectric constant, ω_0 is the eigenfrequency, β is the damping constant, and ω is the driving frequency. In case there are different oscillators in the material, the second term can be generalized to sum over all of them.

4.2 Refractive index

Alternatively to the permittivity, the refractive index n is frequently used to describe optical phenomena such as the refraction of lenses. Its imaginary part κ accounts for the decay of intensity as light travels through an absorbing medium. The refractive index and the permittivity are related as follows:

$$(n + i\kappa)^2 = \epsilon_1 + i\epsilon_2 \tag{4.3}$$

For an electromagnetic wave propagating in a medium, we can write the following relation for the field amplitude. It consists of two parts: the *attenuation term* and the *propagation term* ($\vec{E}_0$ is the initial amplitude of the electric field).

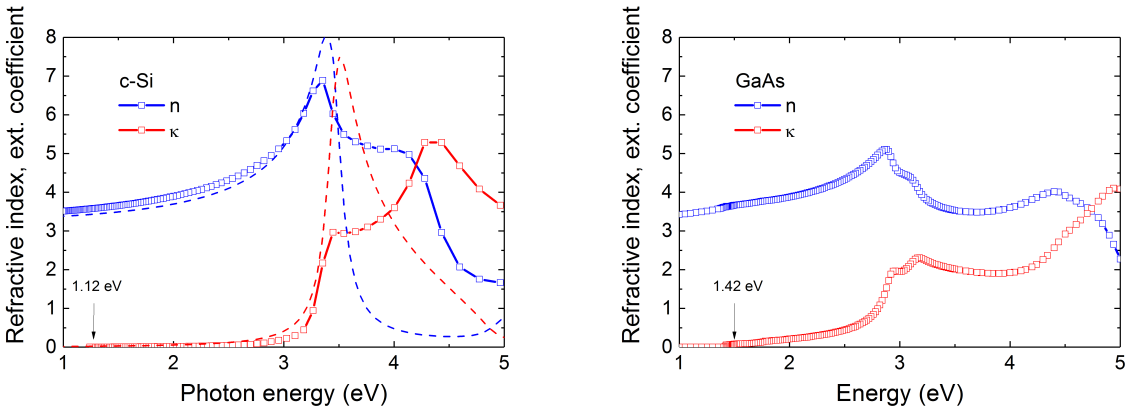
$$\begin{aligned} \vec{E} &= \vec{E}_0 \exp\left[i\left(\frac{2\pi(n+i\kappa)}{\lambda}x - \omega t\right)\right] \\ &= \vec{E}_0 \underbrace{\exp\left[-\frac{2\pi\kappa x}{\lambda}\right]}_{\text{attenuation term}} \underbrace{\exp\left[i\left(\frac{2\pi n}{\lambda}x - \omega t\right)\right]}_{\text{propagation term}} \end{aligned} \tag{4.4}$$

Given the definition of the electromagnetic wave, the decay of the intensity can easily be calculated with the law of *Beer-Lambert-Bougouer*:[1]

$$I(x) = \underbrace{I_0}_{=|\vec{E}_0|^2/2} \exp(-\alpha x) \tag{4.5}$$

Here, α is the material's *absorption coefficient* in units of cm^{-1}. It is related to the extinction coefficient κ in the following way:

$$\alpha = \frac{4\pi \cdot \kappa}{\lambda} \tag{4.6}$$



(a) (b)

Figure 4.1: (a) Complex refractive index $n + i\kappa$ of c-Si and (b) GaAs. The dashed curve overlaid in (a) represents a single harmonic oscillator.

The dispersion characteristics of the complex refractive index of silicon and GaAs are shown in figures 4.1(a) and 4.1(b), respectively. Overlaid to figure 4.1(a) is the refractive index n and the

[1]Note: For a sinusoidal variation in time, $I_0 = |\vec{E}_0|^2/2$ and $|\exp(ix)| = 1$.

extinction coefficient κ of the harmonic oscillator model of eq. (4.2). Over the range of energies from 1 - 3 eV, the simple model reproduces relatively well the increase of the refractive index of silicon from 3.5 to 6.5 and it also captures the increase of the extinction coefficient. For higher energies between 3 - 4 eV the behaviour is no longer well explained, and even a superposition of several oscillators does not improve the accuracy very much.

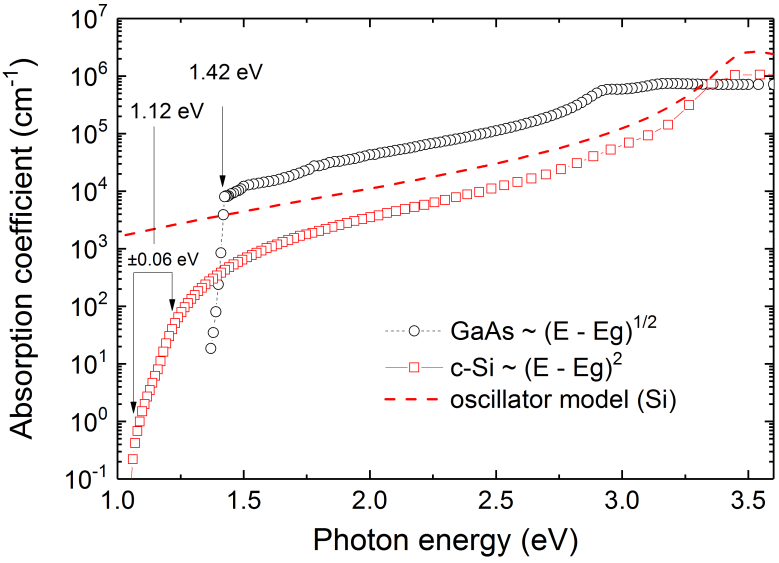


Figure 4.2: Absorption coefficients of c-Si and GaAs. The dashed line represents the oscillator model.

Since we want to understand the absorption of light in solar cells, we show the absorption coefficients in semi-logarithmic scale in figure 4.2. Over the visible range between 1.5 - 3.0 eV we notice that GaAs absorbs light much more strongly than silicon, but towards 3.3 eV the absorption in silicon goes up to the same level. The simple oscillator model falls between the two. It could probably be made to fit a little better by playing with the parameters, but it fails totally below 1.5 eV as the existence of a bandgap is a purely quantum-mechanic phenomenon that is not captured by the classical model of the harmonic oscillator.

There are also striking differences between the two semiconductors that were not visible in figures 4.1(a) and 4.1(b). For GaAs, the onset of the absorption is very steep at its bandgap energy of 1.42 eV whereas for silicon, there is a rather soft onset around its bandgap of 1.12 eV. Moreover, there are two small kinks to the left and to the right of the bandgap energy as indicated by the arrows.

The differences in the absorption coefficients can be explained with the direct bandgap of GaAs and the indirect bandgap of Si and the fact that the absorption process must conserve energy *and* momentum. Photons can carry large energies of several eV, but they have very small momentum. This means that a transition from a VB-state must go into a CB-state with very similar momentum. In the band diagrams of figure 3.5 this means that the transition across the bandgap should go almost vertically up.[2]

[2]Using the photon dispersion relation $E_\nu = \hbar\omega = \hbar c|\vec{k}_\nu|$, we find that a photon energy of 2 eV, i.e. a wavelength of 620 nm which is between orange and red, gives a wavevector modulus of $k_\nu = 1.1 \times 10^5\,\mathrm{cm}^{-1}$. For comparison, for the silicon in the fcc-lattice the the X-point is three orders of magnitude further out at $G/2 = 1.16 \times 10^8\,\mathrm{cm}^{-1}$.

Silicon, has an indirect bandgap as shown in figure 3.5(a). The VBM is at the origin of reciprocal space which is called Γ-point, but the CBM is at $0.85 \cdot X$. In that case, absorption is a second-order process that involves photons and phonons as illustrated in figure 4.3. The latter are the quanta of lattice vibrations that carry a lot of momentum but only very little energy.

There are two different possibilities to combine the two processes; when the photon energy is higher than the bandgap, the excess energy is spent by emitting a phonon into the lattice, heating it up a little. When the photon has less energy than the bandgap, the lattice can provide a phonon to assist the absorption process. In this case the lattice gets a little colder. The two processes are easily distinguished in measurements of the absorption coefficient at different temperatures. The first one is always possible, the second one becomes less and less likely at low temperature, it gets *frozen out*. More details on the bandgap and the temperature dependence of the absorption can be found in appendix A.1.

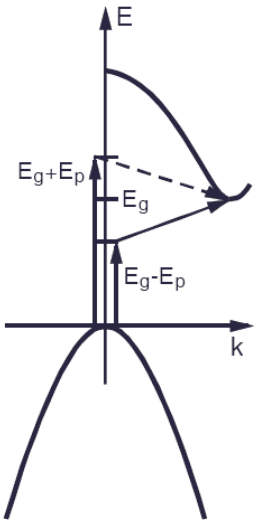


Figure 4.3: Indirect bandgap absorption.

To describe absorption phenomena in direct and indirect semiconductors, we have to use perturbation theory of first and second order, respectively, resulting in the following proportionalities:

$$\alpha \propto (E - E_\mathrm{g})^{1/2} \text{ (direct gap, e.g. GaAs, InP, CdTe, } CuInSe_2\text{)}$$
$$\alpha \propto (E - E_\mathrm{g})^{2} \text{ (indirect gap, (e.g. Si, Ge, AlAs)}$$

The derivation makes explicit use of the relation between E and $\vec{k}$, it does therefore not apply to disordered materials. Nevertheless, the absorption coefficient of amorphous silicon is reasonably well described by a the second case. Apparently, the near-range order over two to three neighbours is already sufficient to yield the behaviour of an indirect bandgap, but the higher absorption of a-Si compared to c-si suggests that the disorder relaxes the strict selection rules of perturbation theory and it is therefore often referred to as "non-direct" bandgap.

4.3 Reflection

The *reflection* R is governed by the refractive index n and the extinction coefficient κ. Figure 4.4(a) illustrates (almost) perpendicular incidence of light from a medium with refractive index n_1 on the flat surface of a medium with refractive index n_2. For this case, the reflection is given by the square of the modulus of the Fresnel reflection coefficient r_{12}:

$$r_{12} = \frac{n_2 - n_1}{n_2 + n_1} \tag{4.7}$$

Obviously, the incident medium is often air for which we can assume a refractive index $n_1 = 1$ in all practical aspects. For incidence on an absorbing medium, we have to use a complex refractive index and the reflected intensity is given by:

$$R_{12} = |r_{12}|^2 = \frac{(n_2 - n_1)^2 + \kappa_2^2}{(n_2 + n_1)^2 + \kappa_2^2} \tag{4.8}$$

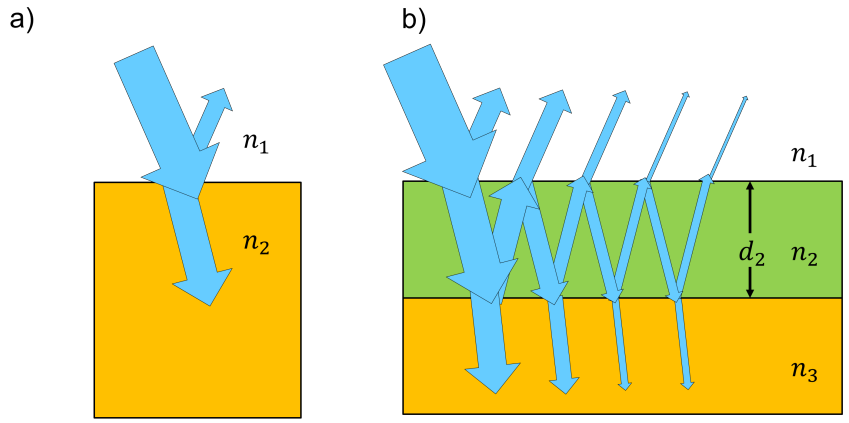


Figure 4.4: Reflection and transmission at a two layer system (a) and at a three-layer system (b).

For example, the refractive index of silicon is $n_{\mathrm{Si}} + i\kappa_{\mathrm{Si}} = 4.077 + 0.028i$ at $\lambda_0 = 550\,\mathrm{nm}$. Approximated to $n_{\mathrm{Si}} \approx 4$, we get a primary reflection of 36%. If we use the dispersion of the refractive index of silicon as shown in figure 4.1(a), we get the reflection characteristic shown by the black line in figure 4.5.

If we consider a slab or transparent material, the measured reflection also contains a contribution from the rear surface. To describe situation by ray-optics, we we have to add all the reflected intensities as illustrated by the arrows in figure 4.4(b). For the sake of generality, let us assume for the moment that $n_1 \neq n_3$.

$$\begin{aligned} R_{123} = R_{12} &+ T_{12}e^{-\alpha_2 d_2}R_{23}e^{-\alpha_2 d_2}\cdot T_{21} \\ &+ T_{12}e^{-\alpha_2 d_2}R_{23}e^{-\alpha_2 d_2}\underbrace{R_{21}e^{-\alpha_2 d_2}R_{23}e^{-\alpha_2 d_2}}_{\text{raised to power } m}\cdot T_{21} \\ &+ \ldots \\ = R_{12} &+ T_{12}T_{21}R_{23}e^{-\alpha_2 d_2}\sum_{m=0}^{\infty}\left(R_{21}R_{23}e^{-\alpha_2 d_2}\right)^m \\ = R_{12} &+ \frac{(1-R_{12})(1-R_{21})R_{23}e^{-\alpha_2 d_2}}{1-R_{21}R_{23}e^{-\alpha_2 d_2}} \end{aligned} \tag{4.9}$$

Obviously, we can simplify because $R_{12} = R_{21}$, but we kept ordering of the indices in the derivation for later use below. For a reflectivity measurement in air, we can further simplify because $R_{12} = R_{23}$. For silicon this is shown by the dashed line in figure 4.5, and we note that it fits the measured data very well.

If we made solar cells with bare silicon, more than a third of the light would be reflected without even getting a chance to be absorbed. Luckily we can do better than that. Consider again figure 4.4, but assume that medium two is a thin film and medium three is silicon. In this case, we enter the domain of wave-optics and can no longer use the intensities defined by eq. (4.8). Instead, we must in a first step sum up the amplitudes of the reflected waves according to eq. (4.7). Subsequently, the total reflected intensity is obtained in a second step by taking the squared modulus of that sum.

The sum is similar to the one leading to eq. (4.9), but there are two differences; first, we have to respect the ordering of the indices in the amplitude coefficient r_{12} defined in eq. (4.7). Second,

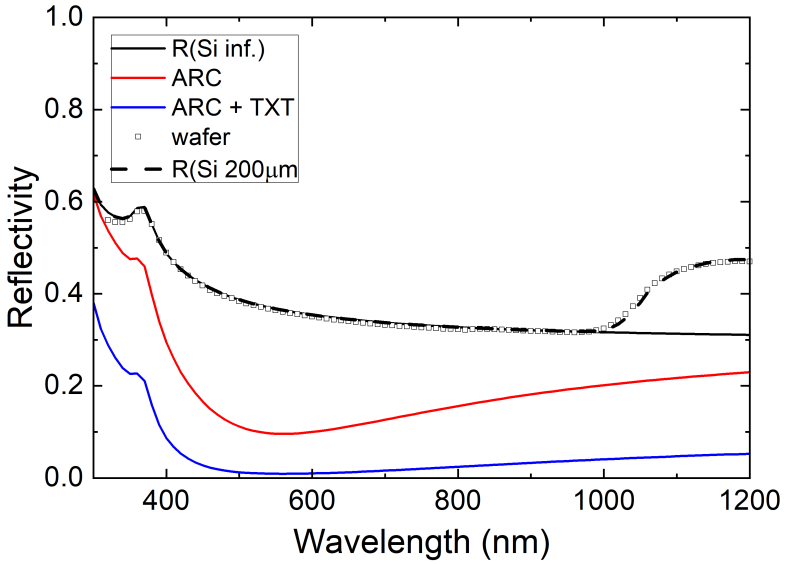


Figure 4.5: Modelled reflectivities for an infinite slab of silicon (black), the addition of of a 90 nm thick AR coating of $n_2 = 1.5$ (red), and the combination of the AR coating with a double rebound at a pyramidal surface texture (blue). Symbols show the measured reflectivity of a 270 µm thick wafer, the dashed black line is modelled with eq. (4.9) to take into account the rear surface.

we have to respecting coherency as we propagate waves across the film. Accordingly, we replace the exponents by $\exp(ik_2d_2)$ using the wavevector $k_2 = n_2k_0 = 2\pi n_2/\lambda_0$. The amplitude reflection coefficient r_{123} is thus given by:

$$\begin{aligned} r_{123} = r_{12} &+ t_{12}e^{ik_2d_2}r_{23}e^{ik_2d_2}\cdot t_{21} \\ &+ t_{12}e^{ik_2d_2}r_{23}e^{ik_2d_2}\underbrace{r_{21}e^{ik_2d_2}r_{23}e^{ik_2d_2}}_{\text{raised to power } m}\cdot t_{21} \\ &+ \ldots \\ = r_{12} &+ t_{12}t_{21}r_{23}e^{2ik_2d_2}\sum\nolimits_{m=0}^{\infty}\left(r_{21}r_{23}e^{2ik_2d_2}\right)^m \\ = r_{12} &+ \frac{(1-r_{12})(1-r_{21})r_{23}e^{2ik_2d_2}}{1-r_{21}r_{23}e^{2ik_2d_2}} \end{aligned} \tag{4.10}$$

Here, we used the Fresnel transmission coefficient which is defined by $t_{ij} = 1 - r_{ij}$. Different from the reflected intensity, we have $r_{ji} = -r_{ij}$ from the definition in eq. (4.7). Thus we can simplify r_{123} into a very compact form:

$$r_{123} = \frac{r_{12} + r_{23}\exp(2ik_2d_2)}{1 + r_{12}\cdot r_{23}\exp(2ik_2d_2)} \tag{4.11}$$

For absorbing media the complex refractive indices make eq. (4.11) rather unwieldy. For the sake of illustration, let us neglect once again the extinction coefficient. Doing so, r_{12} and r_{23} become purely real numbers and the modulus of the complex exponentials evaluate into cos-functions:

$$R = r_{123} r_{123}^* = \frac{r_{12}^2 + r_{23}^2 + 2 r_{12} r_{23} \cos(2 k_2 d_2)}{1 + r_{12}^2 r_{23}^2 + r_{12} r_{23} \cos(2 k_2 d_2)} \tag{4.12}$$

Approximating silicon at 550 nm by $n_3 = 4$ further assume that the coating is made from a non-absorbing material with a refractive index of $n_2 = 2$, we get $r_{12} = r_{23} = 1/9$ and and the reflectivity becomes:

$$R = \frac{1/9 + 1/9 + 2/9 \cos(2 k_2 d_2)}{1 + 1/9 + 1/9 \cos(2 k_2 d_2)} \tag{4.13}$$

The resulting reflectivity oscillates between 0 and 4/11, not exactly the value of 36% we had for the uncoated interface, but close. Minima are assumed whenever $\cos(2 n_2 k d) = -1$ and the first one occurs for $2 k_2 d_2 = \pi$. We can thus design an anti-reflection coating (ARC) with minimum reflection at $\lambda_0 = 560$ nm by applying a coating with $d_2 = 70$ nm and $n_2 = 2$. In this example the minimum is exactly zero because of a coincidence in the numerical values that we chose. In the general case, eq. (4.11) shows that zero reflection at the design-wavelength λ_0 is obtained under the following conditions for n_2 and d_2:

$$n_{\mathrm{ARC}} = (n_1 \cdot n_3)^{1/2} \tag{4.14}$$

$$d_{\mathrm{ARC}} = (\lambda_0 / 4) \cdot n_{\mathrm{ARC}} \tag{4.15}$$

The effect of a single-layer anti-reflection coating is illustrated in figure 4.5 for the case of SiO_2 which has a refractive index of $n_2 \approx 1.5$. To provide again an anti-reflection condition at $\lambda_0 = 560$ nm, we have to use a thickness of a bit more than 90 nm, but SiO_2 does not meet the condition of $n_2 = \sqrt{n_1 \cdot n_3}$. Thus, the reflection at the design-wavelength is clearly reduced with respect to the bare substrate, but it does not drop to zero.

Unfortunately, a single layer cannot reduce the reflection for all wavelengths, but a broader spectral range can be achieved by the use of multi-layers. Among the typical ARC materials are the following: MgF ($n = 1.38$), SiO_2 ($n = 1.52$), ZrO_2 ($n = 2.1$), TiO_2 ($n = 2.2$), CeF_3 ($n = 1.63$). For solar cells Si_3N_4 ($n \approx 2.2$) and Al_2O_3 ($n = 1.67$) are often used. Interference layers are also important for other purposes:

- Narrow pass filters for optical applications
- Selective reflectors for low emissivity glass and IR reflectors
- Antireflection coatings (ARC) for TV screen, scanners, lenses and prescription glasses, etc.

In solar cells, a second effect is used to reduce the reflection, namely surface texture. When 100-oriented silicon wafers are dipped in hot KOH or other caustic solutions, anisotropic etching exposes primarily 111-oriented surfaces.[3] This yields a texture of random pyramids with square base. The pyramids have a facet angle of 54° with respect to the plane. For perpendicular incidence, the primary beam arrives on a facet under 54° and reflects in forward direction. Consequently, there is generally one more incidence on the surface of an adjacent pyramid, and only after this second incidence the beam reflects backwards into the hemisphere of the source. To treat this case correctly,

[3] hint: (111) crystallographic planes have higher atomic density compared to (100)

we'd have to modify eqns. (4.7) and (4.11) to take into account the two polarisation directions and incidence under an angle. After going through all the trouble, it turns out that the polarisation-averaged reflection is reasonably close to the reflection of the flat surface. If this intensity hits the adjacent facet, another ratio of R gets reflected, and thus the intensity of the beam after its double rebound is reduced to R^2. This is illustrated by the blue line in figure 4.5 and the intensity transmitted into the textured wafer is given by $T \approx 1 - R^2$.

Overall, the combination of the AR coating and the texture reduces the reflection to less than 10% over the relevant part of the spectrum. For high-efficiency solar cells, there is yet a little advantage to gain by using a double ARC and working with inverted pyramids, but the fabrication of the latter requires lithography for the the opening of holes in an etch-mask. In case of multicrystalline silicon, anisotropic etching could not be used since there is no prevailing orientation of the crystallites. Thus, isotropic etching in acidic solution was used to etch hemispheric craters around the cracks that remained from the slurry-based wafer-sawing process. Generally, the acidic etch did not reduce the reflection as well as pyramidal textures and since current cell designs generally use monocrystalline silicon, isotropic acid etching is not used any more.

Note that in the processing of silicon solar cells we use a passivation layer of Si_3N_4 on the front side. As it happens, Si_3N_4 has an almost perfectly matched refractive index for silicon, and since we are relatively free to vary its thickness, we can get an AR coating for free! Adding an additional layer to obtain a broader AR effect adds cost to the process and that is why double AR-coatings are generally not used in cell manufacturing. Likewise, an etching process is anyway needed for the removal of sawing damage. The etching also create the surface textures that are used for prolonging the path within as discussed in the next section. Therefore also the AR effect of the texture comes for free! Many elaborate light scattering structures have been proposed for silicon, but few can compete with the built-in AR functionality, and none of them made it into production.

4.4 Absorption and generation

The properties of reflection discussed in the previous section were mostly governed by the real part of the refractive index. In the following, we will discuss how light is absorbed and therefore we need to look at the imaginary part. We must also sort out which part of the absorbed light is lost by parasitic absorption in supporting layers and which part actually creates electron-hole pairs. Even so, a part of the electron-hole pairs can be lost by by recombination as discussed in appendix D. The absorption characteristics discussed in the remainder of this section are therefore upper limits to the actual EQE that we could measure in a solar cell. We can correct the EQE for reflection losses, the resulting *internal quantum efficiency* (IQE) gives the probability that an absorbed photon will create an electron-hole pair. Thus, the IQE is useful to asses parasitic absorption effects within the cell.

Rather than using κ, we will work mostly with the absorption coefficient α. By taking the inverse of the absorption coefficient α, we obtain the *absorption length* which is the thickness that is needed to absorb 63% of the light at a given wavelength. This specific value comes from the Beer-Lambert law $I(x) = I_0 e^{-\alpha x}$ where for $x = 1/\alpha$, the initial intensity decreased to $\approx 37\%$. For silicon, the absorption length of IR light with energy close to the bandgap easily exceeds typical wafer thicknesses. In that case, *light trapping* must be used to enhance the absorption. This is achieved by textured surfaces that refract or scatter light into oblique angles and thus prolong the light path with respect to the device thickness.

The method of the infinite sum introduced in figure 4.4 can also be used to model the absorption in the active part of the solar cell. For simplicity, we start by assuming absorption over the full wafer

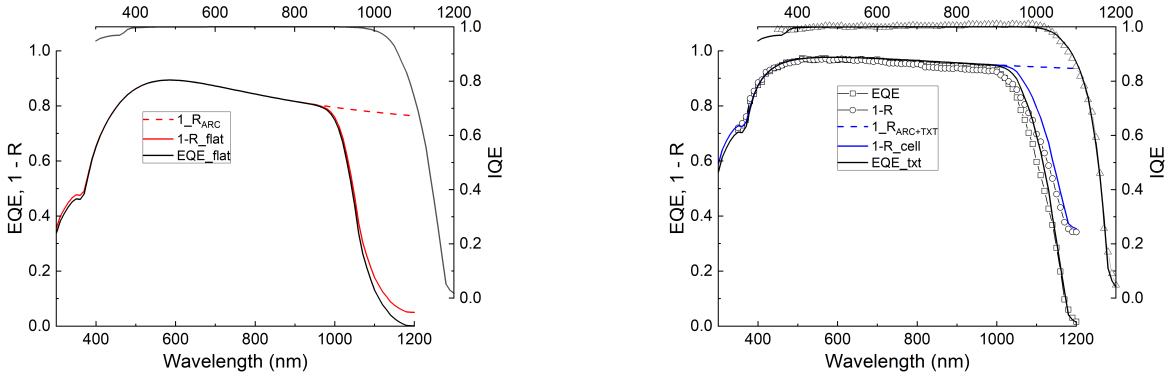
thickness d and we add up the exponentially decaying intensities along the zig-zag movement. Let us go through it step by step: Figure 4.5 showed that we have a primary reflection at the front interface, therefore we start with a term $(1-R_f)$ to describe the intensity with which the first forward beam starts propagating within the wafer. During the forward passage, its intensity decreases by a factor $\exp(-\alpha d)$ and thus contributes an absorption equal to $(1-\exp(-\alpha d))$. Let us call this term "fwd". Next, we have to determine the starting intensity for the backward beam. We already know how much the beam intensity decayed during the passage, after that we have to multiply with the reflectivity R_r of the rear surface. The backward beam thus starts with an intensity of $\exp\{-\alpha d\}R_r$ which we multiply with the term $(1-\exp(-\alpha d))$ to describee the contribution to the absorption during the backward propagation. Let us call this term "bwd". We are ready for the next term in the sum. The starting intensity of the next round trip is attenuated by the previous round-trip which we call "att", and starting from this intenstiy, the second trip contributes once again "fwd + bwd" and so forth. Note that the factor "att" contains the reflectivity R_f to describe the intensity that is bounced back into the wafer at the front side which is the same as the one going into the wafer. The starting intensity for every further round-trip simply piles up ever higher powers of the attenuation factor "att".

$$\begin{aligned} A &= (1-R_f)\left\{\underbrace{\left(1-e^{-\alpha d}\right)}_{\text{fwd}}+\underbrace{e^{-\alpha d}R_r\left(1-e^{-\alpha d}\right)}_{\text{bwd}}+\underbrace{e^{-2\alpha d}R_rR_f}_{\text{att}}\cdot(\text{fwd}+\text{bwd})+...\right\} \\ &= (1-R_f)\left\{(\text{fwd}+\text{bwd})\sum_{m=0}^{\infty}\text{att}^n\right\} \\ &= (1-R_f)\left\{\frac{\left(1-e^{-\alpha d}\right)+e^{-\alpha d}R_r\left(1-e^{-\alpha d}\right)}{1-e^{-2\alpha d}R_rR_f}\right\} \end{aligned} \tag{4.16}$$

For the flat wafer this was easy, textures are a bit more difficult. Let us tackle it with a few idealising assumptions that were first proposed by Deckman and Yablonovitch for the case of weakly absorbed light [23]. They proposed that that the incident light consists of many parallel beams. Upon scattering during the transmission of the textured front interface the directions of the beams get *fully randomised*. The resulting distribution of directions is called *Lambertian* and with respect to the wafer thickness d, a beam propagating under an angle δ gets prolonged by a factor of $1/\cos\delta$. Averaging all path prolongations yields an averaged path length of $2d$. Accordingly, in the formula, we just have to replace every instance of d by $2d$.[4] A second difference concerns the outcoupling at the front surface. Assuming that the beams returning to the front surface are also fully randomised, the majority of them will come in under an angle where they undergo total internal reflection and only a part of $1/n_{Si}^2$ leaves the cell through the so-called escape cone.[5] Thus, the factor R_f is maintained in the round brackets because it describes the primary incidence from air, but in the curly brackets it is replaced by $(1-1/n_{Si}^2)$ because there it describes the returning beams in the infinite sum.

[4] In reality, the Lambertian distribution is not maintained throughout the absorbing medium and thus the prolongation is less than two. A more rigorous treatment is discussed in appendix C.

[5] The original derivation assumed a nano-texture that is smaller than the wavelength, thus it made sense to define the incident angle of the backwards going beams with respect to the average plane of the textured interface. For the large pyramidal structures of crystalline silicon we would not expect that the assumption still holds, but surprisingly it does yield a fairly good description.



(a) (b)

Figure 4.6: Modelled absorption in the bulk for a cell thickness of 200 µm (black lines), assuming a flat cell with AR coating (a) and a cell with AR coating and texture (b). Coloured lines illustrate $1-R_{\text{primary}}$ (dashed lines, reproduced from fig. 4.5) and $1-R_{\text{total}}$ (full lines). Symbols in panel (b) refer to a 280 µm thick PERL solar cell [24].

We can use this result for projecting the efficiency limit of c-Si solar cells in appendix C. To check it against experimental data, we should still add parasitic absorption in the front of the cell. We can do this by multiplying a factor of $\exp(-\alpha d_f)$ to the terms fwd, bwd, and att in eq. (4.16) to describe an absorbing layer of thickness d_f whose attenuation is picked up by each of these terms. Figure 4.6(b) compares the result to experimental data of a 240 µm thick PERL solar cell with a 90 nm thick single-layer AR coating of SiO_2 [24]. The primary reflections of the flat and of the textured silicon surfaces from figure 4.5 are reproduced as $1-R_f$ in figure 4.6. For the textured case, the primary reflection yields a good agreement with the experimental data in the UV and in the visible range. To match the near IR and specifically at 1200 nm, a rear reflectivity of $R_r = 5\%$ was assumed.

The IQE suggests that there is very little parasitic absorption in the UV region because SiO_2 was used for its ARC. Later generations of the PERL cell used double ARC with ZnSe to improve the overall current at the cost of some parasitic absorption in the UV. Losses in the UV can also occur in cell designs with uniform phosphorus diffusion at the front whereas the the shown PERL cell uses a so-called selective emitter.

Both cells show a slow decay of the EQE in the the visible and near IR, even though the decay is much slower for the textured case because of a better ARC. Eventually, the flat cell shows a strong drop for wavelengths beyond 950 nm whereas the textured cell holds up to 1050 nm. This is related to the weak absorption by the indirect bandgap of c-Si and the enhanced absorption in the textured cell, underlining the importance of path length enhancement by scattering at the surface texture.

Light trapping is a combination of path length enhancement and of reduced out-coupling at the front surface. The latter relies on total internal reflection and above we already argued whether we can assume this for a textured cell, but we find in figure 4.6(b) that the blue characteristics fits astonishingly well to the measured data of the 280 µm thick PERL solar cell. Nevertheless, for less perfect devices than the PERL cell, we cannot assume that all of the photocurrent gets collected and consequently the EQE will differ from the calculated absorption characteristic. Similar to collection losses by interface recombination at the front, there is also surface recombination at the rear interface, and there can also be substantial collection losses in the bulk when the diffusion

length is less than the wafer thickness. For a better assessment of the collection, we have to know where the carriers are generated, and we have to describe the recombination effects.

The generation profile is the easier part and will be treated in the next section, but for recombination we first need to discuss different loss mechanisms in chapter 5 and then assess their impact on the charge carrier collection under operating conditions of the solar cell in the chapter 6.

4.5 Generation rate

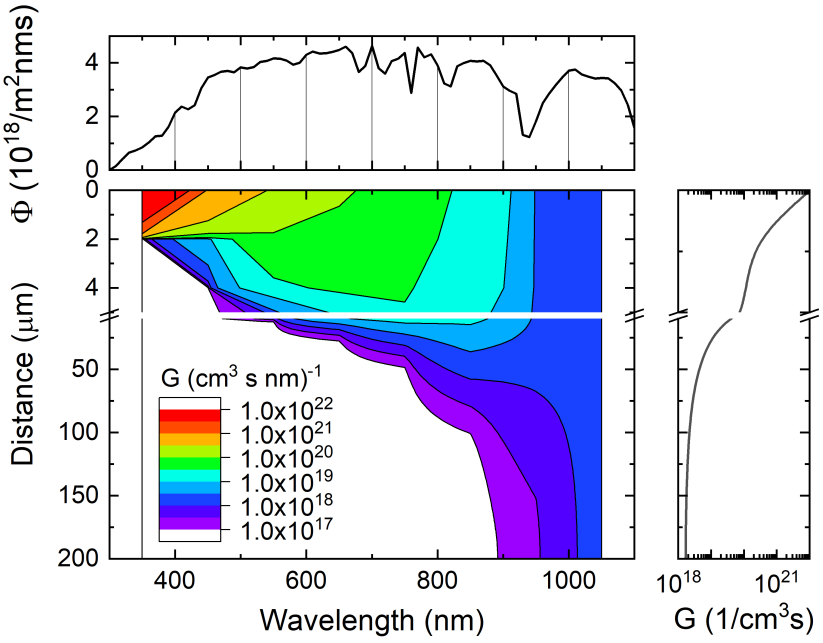


Figure 4.7: Absorption of the spectral photon flux in the AM1.5 spectrum (top panel) in a 200 µm thick solar cell (bottom left, illustrated for the seven wavelengths represented by the drop-lines). Integration over the wavelength yields the generation profile $G(x)$ (right panel).

A rough estimate of the average generation rate can be obtained by dividing the photocurrent by the device thickness. To do better than that, we start by calculating the incident flux of photons in the wavelength interval between λ and $\lambda + \Delta\lambda$. It is obtained by dividing the spectral power density shown in figure 1.1 by the energy that is carried by photons with wavelengths λ. The resulting photon flux will enter the cell at the front surface and decay exponentially according to the absorption coefficient corresponding to λ. To obtain the spectrally relsoved generation rate $G(x, \lambda)$, we calculate the difference $\Phi(x, \lambda) - \Phi(x + \Delta x, \lambda)$ at the positions x and $x + \Delta x$, and we divide this areal density by Δx to obtain a volume density. For an infinitesimal interval Δx, this is simply the negative of the derivative:

$$G(x, \lambda) = -\frac{d}{dx}\Phi(\lambda)e^{-\alpha x} = \Phi(\lambda) \cdot \alpha e^{-\alpha x} \tag{4.17}$$

This holds for a semi-infinite slab of material with perfect anti-reflection condition. After the considerations in the previous section, we can treat the decay more realistically by taking into account front- and rear reflection of a solar cell with finite thickness:

$$G(x,\lambda) = \Phi(\lambda)(1-R_f)\frac{\alpha e^{-\alpha x} + \alpha R_r e^{-\alpha(2d-x)}}{1 - R_f R_r e^{-2\alpha d}} \tag{4.18}$$

Figure 4.2 showed that α is large for high photon energies. Thus, highly energetic photons are absorbed within the first 10 - 100 nm. Photons with energy close to the bandgap are absorbed only weakly, they will transit the complete solar cell and reflect at the rear interface. Thus, more electron-hole pairs are generated at the front than at the rear of the device. The spectral generation profile $G(x,\lambda)$ is shown in by the contour plot in figure 4.7 for the case of a 200 µm thick solar cell. For the generation profile $G(x)$, we still have to integrate over the wavelength range of interest as shown by the lower right panel of figure 4.7. The part above the axis break shows that $G(x)$ is $1 \times 10^{22}\,\mathrm{cm^{-3}\,s^{-1}}$ at the front and decays to $1 \times 10^{20}\,\mathrm{cm^{-3}\,s^{-1}}$ within the first 5 µm. Over the remainder of the bulk it decays only very slowly to $2 \times 10^{18}\,\mathrm{cm^{-3}\,s^{-1}}$ at the rear.

In section 6.3 we will approximate $G(x)$ by a constant value because it yields a comparatively simple analytic solution, but we should keep in mind that the junction region at the front may not be described correctly. A better approximation with exponentially decreasing generation rate is discussed in appendix D.4.

Chapter 5

Generation and recombination

Let us assume a semiconductor in dark. For any finite temperature there is a probability to generate free carriers by thermal excitation across the bandgap. Since the generation of carriers does not accumulate infinitely, there must be an opposing process which we call recombination. In thermodynamic equilibrium between the two processes, the product of the electron density and the hole density is equal to the intrinsic carrier density n_i.

$$n_i^2 = n_0 p_0 \tag{5.1}$$

The zero in the subscripts indicates conditions of thermal equilibrium. In this case the rate of thermal generation G_{th} will be the same as the rate of recombination R_{th} .

$$G_{th} = R_{th} \tag{5.2}$$

The units of the rates are $\mathrm{cm}^{-3}\,\mathrm{s}^{-1}$. Besides the thermally generated charge carriers, there are various ways to create additional ones, for example by electric bias, by bombardment with energetic particles, or by illumination with photons whose energy is higher than the bandgap. Some of these generation processes create carriers in highly excited states, but in the following we will assume the carrier populations after thermalisation (c.f. section 3.5). Generation will yield carrier densities n and p that are higher than in equilibrium, but once again recombination will prevent them from accumulating infinitely. Thus, we find a *steady state situation* which we express with the *excess carrier densities* Δn and Δp.

$$\begin{aligned} n &= n_0 + \Delta n && (5.3)\\ p &= p_0 + \Delta p && (5.4)\\ & np > n_i^2 && (5.5) \end{aligned}$$

Optical generation is special because the absorption of photons always creates electrons and holes in pairs, and therefore $\Delta n = \Delta p$.

will return to the equilibrium value within a characteristic time τ that depends on the recombination rate R:

$$\tau = \Delta n / R \tag{5.6}$$
$$\tag{5.7}$$

In semiconductors there is a variety of bulk recombination phenomena such as **radiative recombination**, **Auger-recombination**, and **Shockley-Read-Hall recombination**. All of these recombination processes occur in parallel and the resulting recombination rates add up. Expressed in terms lifetimes associated to the different processes, this allows us to define the bulk lifetime τ_b:

$$\frac{1}{\tau_b} = \frac{1}{\tau_{\text{rad}}} + \frac{1}{\tau_{\text{Auger}}} + \frac{1}{\tau_{\text{SRH}}} \tag{5.8}$$

Lifetime measurement Let us look at a situation that is particularly important for silicon. Assume that we have a slab of material that is illuminated with a known intensity of weakly absorbed light, for example IR light with 800 - 1000 nm. From figure 4.7 we can conclude that this yields a uniform generation rate across the sample. If the illumination is long enough to establish a steady state condition, we will thus have additional electrons and holes of a concentration equal to Δn. They form a plasma of free carriers whose density can be measured, for example, through its reflectivity for radiation with low energy such as microwaves or radio frequencies at a few MHz. If we do not extract any carriers, $G = R$ and eq. (5.6) allows us to determine the minority carrier lifetime τ. Repeating this procedure for different illumination intensities, we can distinguish between the recombination phenomena as explained in the following sections.
We can also determine Δn for an illumination equivalent to one sun and insert the result in the definition of the QFLs of eq. (3.31). This allows us to project an upper limit for the V_{oc} that we can possibly reach, the so-called implied open circuit voltage iV_{oc}. We can do this at any time in the process without having to produce an actual device.

5.1 Continuity equations

So far we treated the recombination phenomena as if they were uniform across the full volume. This is obviously not the case, for example the generation rate is normally not constant. Other examples are non-uniformities at grain boundaries, or differences in the defect densities between the centre of a wafer and its borders. To describe these cases, we consider small independent volume elements within our material, each of them characterised by its own steady state condition. We do allow the transport of charges from one volume element to another, but on a timescale longer than thermalisation. Then, the integrated charge density in a given volume element changes over time by generation and by recombination, and by the flow of a current through its surface. Adding up all terms and applying a little mathematical manipulation,[1] we arrive at the *continuity equations*:

$$\frac{\partial n}{\partial t} = G - R_n + 1/q\vec{\nabla}\vec{j}_n \tag{5.9}$$

$$\frac{\partial p}{\partial t} = G - R_p - 1/q\vec{\nabla}\vec{j}_p \tag{5.10}$$

The most important application of these equations is the derivation of the current voltage characteristic of the *p-n* junction in chapter 6.

[1] For the derivation we have to extend a volume integral over the volume-related terms that describe the temporal change of the density, the generation, and the recombination. The change associated to the current to and from adjacent volumes is expressed by a surface integral. Finally, Gauss's theorem is applied to convert the surface integral over $\vec{j}$ into a volume integral over $\vec{\nabla}\vec{j}$. If we look only at the integrands, we find the continuity equations in their differential form.

5.2 Radiative recombination

The inverse process of absorption is *radiative recombination*, i.e. the emission of a photon after band-to-band recombination of an *e-h* pair. This process is much more probable for direct bandgaps than for indirect bandgaps as in the latter case the transition must once again be mediated by a phonon. The *recombination rate* R_{rad} is given by the following expression:

$$R_{\text{rad}} = B \cdot np \tag{5.11}$$

Here, B is the radiative recombination coefficient. For lowly doped silicon it is equal to B_{low} =4.73 $\times$ 10^{-15} $\text{cm}^3\,\text{s}^{-1}$ [25].

Let us assume the case of n-type silicon with a donor concentration N_D, such that we can approximate $n_0 = N_D$ and $p_0 = n_i^2/N_D$. We can identify two important limiting cases:

1. Under **low injection**, we have $n \approx n_0 \gg \Delta n$ and $p \approx \Delta n \gg p_0$. Then, eq. (5.11) simplifies to $R_{\text{rad}} = B \cdot N_D \Delta n$ and the minority lifetime of holes is given by:

$$\tau_p = \frac{1}{BN_D} \tag{5.12}$$

   Thus, τ_p becomes independent of the excess carrier density Δn.

2. Under **high injection**, $\Delta n > N_D$. In this case, both, n and p can be approximated by Δn, and eq. (5.11) simplifies to $R_e = B \cdot \Delta n^2$.

$$\tau_p = \frac{1}{r_{ec}\Delta n} \tag{5.13}$$

   Thus, the minority lifetime τ_p is inversely propotional to the excess carrier density.

5.3 Auger-Meitner recombination

The *Auger-Meitner effect* is a three particle process where the excess energy liberated by recombination is not used to emit a photon, but to excite a third charge carrier.[2] In a semiconductor, the third carrier is either a hole or an electron that gets excited into a state far from the band-edge. From there it *thermalises* until it reaches again a state at the band edge. Thus, the energy released by the *e-h* recombination heats the lattice. The recombination rate for the Auger process is given by:[3]

$$R_{\text{Auger}} = C_n n^2 p + C_p p^2 \tag{5.14}$$

[2]The effect was discovered independently by Lise Meitner in 1922 and by Pierre Auger in 1923 [26] for transitions between the inner shells of atoms.

[3]The formula applies only to highly doped material where a plasma of free carriers shields Coulomb attractions[27]. In lowly doped material, there is less shielding and consequently Coulomb attraction between the carriers can enhance the Auger effect with respect to eq. (5.14). A parametrisation for Auger recombination over a wide range of doping and excess carrier densities was proposed proposed by [28], later updated by Richter [29], end eventually further refined by Niewelt [30].

Here, C_n and C_p are the *Auger coefficients* and the fact that there is a n^2 in the electron term (and a p^2 in the hole term) indicates that two particles are involved. Numerical values for the Auger coefficients are $C_n = 2.8 \times 10^{-31}$ cm^{-6}s^{-1} and $C_p = 0.99 \times 10^{-31}$ cm^{-6}s^{-1}.

Just like for the case of radiative recombination, let us discuss the limiting cases of Auger recombination in n-type silicon for the two cases of low and high injection:

1. Under **low injection**, τ_p is given by:

$$\tau_p = \frac{1}{C_n N_D^2} \tag{5.15}$$

   Once again, the hole lifetime is independent of the the excess carrier density and assumes a constant value.

2. Under **high injection** we obtain:

$$\tau_n = \tau_p = \frac{1}{(C_n + C_p)\Delta n^2} \tag{5.16}$$

   The sum $C_n + C_p$ is also called *ambipolar Auger coefficent* and the lifetimes vary with the inverse square of the excess carrier density.

Figure 5.1 shows a log-log plot of the lifetime vs. Δn for different recombination processes. Auger recombination is easily distinguished from radiative recombination by its steeper slope at high excess carrier density. For the relatively low bulk doping chosen for the illustration, the Auger lifetime according to eq. 5.14 is reduced by Coulomb enhancement as shown by the dashed lines. The effect is discussed in more detail in appendix B.

5.4 Shockley-Read-Hall recombination (SRH)

The recombination via states in the bandgap is normally called SRH recombination because it was first described in two articles that appeared independently, one by Shockley and Read, the other one by Hall [31, 32]. States in the bandgap can be classified into dopant states that are typically closer than 0.1 eV to one of the band edges, trap states that are 0.1 - 0.2 eV away from the band edges, and recombination centres that are close to midgap. At room temperature we can assume that dopant states are always ionized, but When a charge carrier is captured by a trap state, there is still a chance that it gets excited back into the band. The main effect of trapping into such states is therefore to slow down the flow of carriers, i.e. to reduce the mobility, similar to the case of ionized impurity scattering in highly doped samples that is discussed in appendix B.

When the defect state is closer to midgap, thermal excitation is much less likely and it becomes more probable to capture a carrier of opposing type. Since this results in recombination and ultimately the loss of a minority carrier, these states are very detrimental to solar cells and the net recombination rate R_{SRH} is given by:

$$R_{\mathrm{SRH}} = \frac{N_{\mathrm{t}} v_{th} \cdot (np - n_i^2)}{(n + n_1)/\sigma_p + (p + p_1)/\sigma_n} \tag{5.17}$$

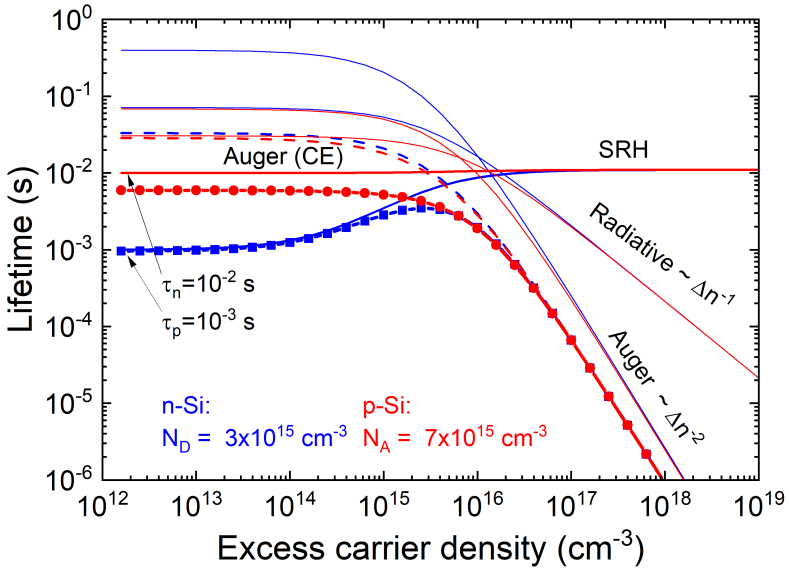


Figure 5.1: Dependence of the minority carrier lifetime τ on excess carrier density Δn for Si with bulk resistivity of $2\,\Omega\,$cm of n-type (blue) and p-type polarity (red). Radiative emission and Auger excitation are characterised by their typical slopes at high excess carrier density (the dashed lines illustrates the Auger effect with Coulomb enhancement that actually applies for the assumed doping concentrations). SRH recombination at a mid-gap defect is characterised by a transition between two flat regions. Symbols represent the resulting bulk lifetime.

Here, N_t is the density of the recombination centres[4] and v_{th} is the thermal velocity of the minority carriers which is typically $v_{th} = 10^7\,$cm/s. The energy level of the state in the bandgap is denoted by E_t and enters into the quantities p_1 and n_1 which are defined as follows:

$$p_1 = N_V \exp\left(-(E_t - E_V)/kT\right) = n_i \exp\left((E_F^i - E_t)/kT\right) \tag{5.18}$$

$$n_1 = N_C \exp\left(-(E_C - E_t)/kT\right) = n_i \exp\left((E_t - E_F^i)/kT\right) \tag{5.19}$$

Let us again discuss the limiting cases for an n-type semiconductor under low and high injection.

1. For **low injection** we approximate once again $n \approx N_\mathrm{D} \gg \Delta n$ and $p \approx \Delta n \gg n_i$, and we regard the case of a typical recombination centre. This means that E_t is close to midgap and that σ_n and that σ_p are in the same order of magnitude. Then, the denominator of eq. (5.17) simplifies massively because we can neglect n_1 with respect to N_D, and we can neglect the hole-term altogether:

$$\tau_\mathrm{SRH} = \frac{\Delta n}{R_\mathrm{SRH}} \approx \frac{1}{N_t v_{th} \sigma_p} \equiv \tau_p \tag{5.20}$$

2. For **high injection** we assume not only $\Delta n = \Delta p > n_0, p_0$, but also $\Delta n = \Delta p > n_1, p_1$, i.e. that the injected minority carriers split the quasi Fermi levels beyond the energy of the trap state. Then, the recombination rate becomes particularly simple:

[4]For historic reasons the index "t" is used for "trap" even though we are dealing with recombination centres.

$$\tau_{\mathrm{SRH}} = \tau_p + \tau_n \tag{5.21}$$

This means that for our assumption of a deep defect in n-type silicon the SRH lifetime makes a transition between a constant value of τ_p to another constant value of $\tau_p+\tau_n$ as shown in Figure 5.1. If assume p-type silicon with the same limiting lifetimes, the transition is from τ_n to $\tau_p+\tau_n$, which means that the lifetime starts out already from a much higher value and then improves only by a barely visible amount to the same sum of $\tau_p + \tau_n$ at high Δn. In appendix F we discuss SRH recombination in more detail and we will see that in case of shallow levels with strongly asymmetric cross sections the characteristic can even turn downwards.

To give an idea of the impact of defect recombination, let us assume a defect state close to midgap with a capture cross section of $\sigma = 10^{-15}\,\mathrm{cm}^2$ which is representative for typical metallic contaminants such as iron. For a defect density of $N_t = 10^{13}\,\mathrm{cm}^{-3}$, i.e. less than a ppb, the bulk lifetime is already reduced to less than 10 µs.

5.5 Surface recombination

The recombination phenomena discussed so far applied to the bulk. A different recombination mechanism is associated with the surface or, more generally, with an interface. In both cases the abrupt ending of the crystal lattice can lead to a high number of unsaturated bonds. In covalent semiconductors like silicon, these *dangling bonds* often form electronic states within the band gap that act as efficient recombination centres.

We define the *surface recombination velocity* S (SRV) to link the *net recombination rate* R_{s} at the surface to the *excess carrier density* Δn_{s} at the surface:

$$R_s = S\Delta n_s \tag{5.22}$$

This is similar to the definition of bulk recombination, but since R_{surf} has units of $[\mathrm{cm}^{-2}\,\mathrm{s}^{-1}]$ whereas Δn_{s} is measured in $[\mathrm{cm}^{-3}]$ in the semiconductor bordering on that surface, S must have the units of $[\mathrm{cm}\,\mathrm{s}^{-1}]$, i.e. of a velocity and not of an inverse time. Despite the unit, S is simply a measure of recombination at the surface and it should not be mistaken with the velocity of the carriers anywhere in the device or close to its surface. The recombination rate at the surface can be defined in analogy to the Shockley-Read-Hall theory for the bulk recombination (see section 5.4), the only difference is that the trap density N_t is replaced by an *interface trap density* D_{it} $[\mathrm{cm}^{-2}]$ and the carrier densities in the bulk are replaced by the densities n_s and p_s at the surface.

$$R_s = \frac{v_{th} D_{it} \left(n_s p_s - n_i^2\right)}{\left(n_s + n_1\right)/\sigma_p + \left(p_s + p_1\right)/\sigma_n} \tag{5.23}$$

There are special cases for which the recombination velocity can be approximated by a specific expression:

1. Assume an n-type semiconductor with a mid-gap defect ($E_t = E_F^i$) and a flat band condition such that the excess carrier density at the surface the same as the one in the bulk. Thus, $n_s \approx n$ and for a case of low injection $n > \Delta n = \Delta p \gg p$, we find the following for the SRV:

$$S_p = \sigma_p v_{th} D_{it} \tag{5.24}$$

2. Alternatively, in case of strong band bending, the carrier density at the interface is different from the one in the bulk. Such a situation can arise through fixed charges at or near the surface, and normally we don't know the values of n_s and n_p (c.f. appendix G). In that case, often an *effective surface recombination velocity* S_{eff} is defined for a position $x = d$ within the semiconductor for which can still assume $\Delta n(x = d) = \Delta p(x = d)$.

$$S_{eff} = \frac{R_s(n,p)}{\Delta n(x=d)} \tag{5.25}$$

The location of d is normally a position in the bulk just before the start of the band bending associated to surface effects. In a way this is similar to the description of the p-n junction in chapter 6 where we use the values of Δn just at the edges of the space charge region.

There are two main strategies to reduce the surface recombination. The first is called *chemical passivation* and works by reducing the density of defect states at the surface by covering it with a material that will saturate the dangling bonds. For silicon, the most effective passivation is a monolayer of hydrogen, but it is not very stable in ambient conditions. A more stable alternative is a carefully deposited layer of amorphous silicon whose hydrogen content reduces D_{it} to 1×10^9 - $1 \times 10^{10}\,\text{cm}^{-3}$. Another passivating material is a layer of silicon-oxide, especially with an additional hydrogenation treatment. The second method is called *field effect passivation* and works by repelling one type of charge carrier from the surface. Typically this is achieved by applying a *dielectric passivation layer* that contains *fixed charges*. For silicon, SiN_x and Al_2O_3 are frequently used; the former contains positive charge with density up to $Q \approx 1 \times 10^{12}\,\text{cm}^{-1}$ and is therefore used to passivate n-type emitters. The latter contains negative charge with even higher densities up to $Q \approx 1 \times 10^{13}\,\text{cm}^{-1}$ and is therefore very successfully used to passivate p-type regions. A drawback of dielectric passivation layers is that they are insulating. Thus, they require either localised contact openings, or they must be very thin for majority carriers to tunnel through.

With these passivation materials the values of S are 0.1 - $1000\,\text{cm}\,\text{s}^{-1}$ for the case of low doping. However, on the surface of an n-type emitter with typical surface doping concentration of $N_D \approx 1 \times 10^{20}\,\text{cm}^{-3}$, S_p of an oxide passivation layer is increased up to 1×10^4 - $1 \times 10^5\,\text{cm}\,\text{s}^{-1}$. For metallic contacts, typical values of S are 1×10^6 - $1 \times 10^7\,\text{cm}\,\text{s}^{-1}$, and in most cases they are simply assumed infinite.[5]

5.6 Effective lifetime

The joint effect of recombination phenomena in the bulk and at the surface can be described by an *effective lifetime*.

$$\frac{1}{\tau_{eff}} = \frac{1}{\tau_b} + \frac{1}{\tau_s} \tag{5.26}$$

$$= \frac{1}{\tau_{\text{rad}}} + \frac{1}{\tau_{\text{Auger}}} + \frac{1}{\tau_{\text{SRH}}} + \frac{1}{\tau_{\text{surf}}} \tag{5.27}$$

[5]The contact between silicon and aluminium is a special case since Al is also a dopant. When such a contact is annealed, Al intermixes with the wafer and creates a highly p-doped region. In this case, we revert to the definiton of an effective surface recombination velocity at the interface between the wafer and the highly doped region rather than SCR of the Schottky contact (c.f. appendix D.5).

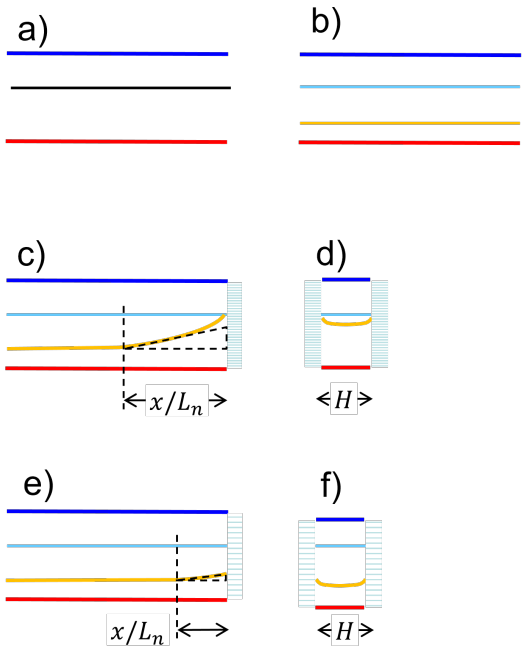


Figure 5.2: Illustration of a n-type semiconductor in dark with unique Fermi-level (a) and quasi Fermi-level splitting in case of a uniform generation rate (b). Strong surface recombination in a semi-infinite configuration will decay Δn to a reduced value at the surface (c), whereas strong recombination at both surfaces and $H < L_p$ yields low Δn across the full wafer. Panels (e) and (f) illustrate how surface passivation can maintain higher Δn at the surface and thus also in the bulk of a thin wafer.

To determine the contribution of the surface to the effective lifetime, we consider that surface recombination is fed by a current of minority carriers that flows from the bulk to the surface.

$$j_s = qS\Delta n \tag{5.28}$$

Thus, we have to use the continuity equations defined in section 5.1 and find a steady-state solution for the spatial variation of the carrier density and the resulting flow of current. For the example of a n-type wafer, figure 5.2 illustrates that the carrier density will globally decay exponentially with the diffusion lentgh L_p of the minority carriers like $\exp(-x/L_p)$ and without going into further detail,[6] two important limiting cases emerge:

1. For the case of a *badly passivated* surface with very high SRV, figure 5.2 illustrates that most of the minority carriers will reach one of the two surfaces and recombine there. In such a case, the surface lifetime τ_s can be expressed purely in terms of the sample thickness H and the minority diffusion coefficient of the minority carriers, in this case D_p.

$$\tau_s = \frac{1}{D_p}\left(\frac{H}{\pi}\right)^2 \tag{5.29}$$

[6]The relevant equations are derived by Kousik [33], limiting cases for practical use are described by Sproul [34]. A more critical analysis of the time-dependent equations and their applicability to the steady state can be found in the articles of Brody [35] and of Anfimov [36].

For example, assume an n-type wafer and $S = 1 \times 10^7\,\mathrm{cm\,s^{-1}}$ which is a typical value for the interface between silicon and a metal. For a typical doping concentration of $1 \times 10^{16}\,\mathrm{cm^{-1}}$ we have a minority diffusion coefficient $D_p \approx 10\,\mathrm{cm^2\,s^{-1}}$. Then, eq. (5.29) predicts $\tau_s \approx 10\,\mu\mathrm{s}$ for a 300 µm thick wafer whereas the bulk lifetime for the assumed doping concentration could be as high as 2.7 ms. This means that the effective lifetime is completely dominated by surface recombination, regardless of the bulk quality. Our only hope to retain higher effective lifetime would be to use a wafer thickness larger than the diffusion length, but for the assumed bulk quality this would more than 1.6 mm.

2. In solar cells we are hopefully dealing more often with the case of good surface passivation. In this case, we find that τ_s is defined in terms of the wafer thickness H and the SRVs at the front and and at the rear:

$$\tau_s = \frac{H}{S_{\mathrm{front}} + S_{\mathrm{rear}}} \tag{5.30}$$

Assume the same n-type wafer as above, this time passivated on both sides with SiN_x with a surface recombination velocity of $50\,\mathrm{cm\,s^{-1}}$. In this case, eq. (5.30) predicts $\tau_s \approx 300\,\mu\mathrm{s}$ for a 300 µm thick wafer. Applying this type of surface passivation to a multicrystalline wafer would be a bit of a waste as typical bulk lifetimes of this material are lower, but for our n-type wafer with bulk lifetime of 2.7 ms we'd still be limited by the surface and we'd better find a better surface passivation whose S_p is 10 times lower.

The definition of the surface recombination rate is essentially the same as the one of SRH recombination in the bulk and thus they cannot easily be distinguished by their dependence on the excess carrier density. However, if we apply the same surface passivation to wafers of different thickness, the dependence of τ_s on the wafer thickness H in eqns. (5.29) and (5.30) can be used to separate bulk and surface effects [37, 29].

5.7 Conclusion

Radiative recombination is the most fundamental recombination mechanism that cannot be avoided and it is at the heart of the Schockley-Queisser efficiency limit. Recent GaAs cells start getting close to this limit, for silicon solar cells it is usually not relevant.

Auger recombination is important at *high doping* and *high injection* as the lifetime is proportional to the inverse square of the *doping concentration* or the *excess carrier density*. It can become important for concentrator applications, and it limits the theoretical efficiency of silicon to values much below the Schockley-Queisser limit.

In contrast to the two intrinsic (unavoidable) recombination processes mentioned above, SRH recombination is extrinsic. This means it is avoidable e.g. by using highly purified material. It is less sensitive to the doping concentration or the excess carrier density, but mainly depends on the concentration of defect states, the ratio of capture cross-sections for for electrons and holes, and the position of the trap level within the bandgap.

Finally, surface recombination applies to all types of solar cell, but it is a particular issue for semiconductors with long diffusion length such as silicon. It gets more important as bulk quality and the diffusion lengths increase whereas cost considerations call for ever thinner wafers. It is also unavoidable as sooner or later we have to extract the current with metallic contacts. Since

metals have a unique Fermi-level with $\Delta n = 0$, the splitting of the quasi-Fermi levels must somehow collapse from its value in the semiconductor to $\Delta n = 0$ at the interface. In eq. (5.28) this means that S of the metallic contact tends to infinity and it is actually a major difficulty to accommodate this fact in our cell design.

In order to reduce surface recombination, we can passivate the surface and open only small areas to extract the current. Recombination at remaining contact areas is still unavoidable, but it can be reduced by modulating the doping concentration; beneath the passivated areas, doping can be reduced to the minimum that is needed for lateral transport to the contact fingers. Additionally, this reduces the parasitic absorption by the free carrier plasma and also Auger-recombination. Beneath the metallic contacts, doping can be increased to "shield" footnote The decay from the bulk value of Δn to $\Delta n = 0$ at the metallic contact typically happens over a diffusion length L. In the highly doped region underneath the contact, L is drastically reduced with respect to the bulk value, and thus the effect of $S = \infty$ is "shielded". the solar cell from recombination and also to establish a good electric contact with low contact resistivity. This concept was used in the "point contact solar cell" [38]. In appendix D.2 we discuss the concept in more detail and we illustrate it with some experimental data.

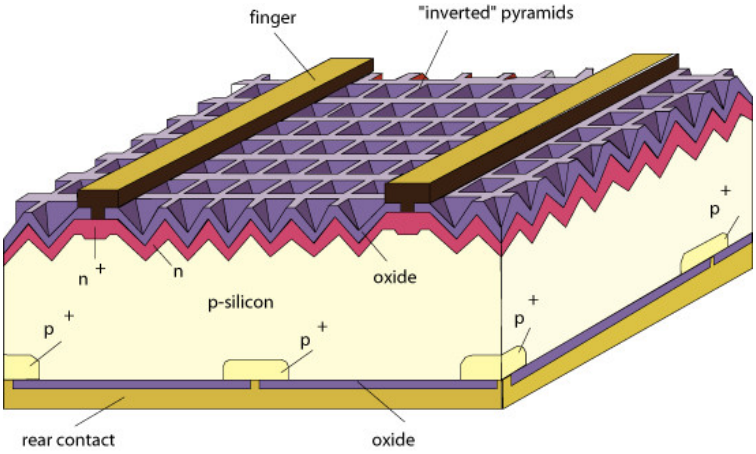


Figure 5.3: Design of a PERL cell with SiO_2 passivation, local openings for the metallic contacts, and localised diffusion of the p-contact (reference: `www.pveducation.org`, [24])

Localized front contacts were also used in a design called "passivated emitter solar cell" (PESC) [39]. The localisation was later extended to the rear in a design called "passivated emitter and rear cell" (PERC) [40] and eventually extended with enhanced doping concentrations under the contacts in a cell design called "passivated emitter with rear locally diffused" (PERL). Figure 5.3 illustrates that this cell type employs increased doping underneath the fingers of the metallisation and passivating oxide layers everywhere else. This concept is often called "selective emitter". Likewise, the rear contact is localized to small spots that connect to regions where the doping concentration is enhanced by locally diffused regions whereas the rest of the rear surface is covered by a stack of a passivating layer and the rear metallisation.The enhanced doping close to the rear contact is often called "back surface field" (BSF), regardless whether it is localized or applied to the full area.

We finish by mentioning that recent cells designs employ so-called "passivating contacts" where additional layers between the wafer and the metal ensure the passivation of the wafer surface while they establish the electric contact at the same time. These are discussed in chapter 10.

Chapter 6

The *p*-*n* junction

When a p-doped and an n-doped semiconductor are in contact, they form a p-n junction and the property of allowing or blocking the flow of current is a key to the operation of electronic devices. Generally a solar cell contains a p-n junction, and in dark it is nothing but a diode. In the following, we will develop a theoretical description of the p-n junctions in solar cells, going step by step from the idealised case of a junction between two semi-infinite semiconductors in dark towards a finite cell structure under operating conditions.

6.1 The infinite *p*-*n* junction in equilibrium

In this section we are dealing primarily with the properties of majority carriers. Let us assume we have two semi-infinite slabs of material with different doping, i.e. n-type material to the left and p-type material to the right. Upon joining the two materials together, free electrons in the n-type material will are free to diffuse across the junction. As the electron density in the p-material is much lower, fewer electrons will diffuse the other way. Effectively, there will be a flow along the the concentration gradient. Likewise, the holes from the p-type material will flow along the concentration gradient in the opposite direction. Generally the two types of charge-carrier will recombine when they meet at the interface. As the carriers move towards the interface, they leave behind the ionised cores of their dopant atoms. Thus, a field builds up between the positive cores in the n-doped region and the negative cores in the p-doped region. Consequently, free carriers do not only feel the diffusion gradient, but also the electric field. In steady state, the diffusion currents $j_{\text{diff}}^{n,p}$ associated to the concentration gradients exactly balance the drift currents $j_{\text{drift}}^{n,p}$ within this field. It turns out that not all carriers of the two semiconductors take part in this process, but only those from a small region which extends to both sides of the interface. This region is called *space charge region* (SCR) because of the charge of the ionised dopants. Beyond the space charge region, the charge of the dopants is compensated by the one of the free carriers, therefore these areas of the semiconductors are neutral.

In order to describe the junction, we restrict ourselves to a junction in 1D along the x-direction and we make use of the *depletion approximation* which is based on two assumptions.

1. Full ionisation: This means that $N_{D^+} = N_D$ and $N_{A^-} = N_A$. Denoting equilibrium conditions by a zero in the subscripts, the carrier densities are given by the following equations.

   n-region: $\quad n_{n,0} = N_D \quad p_{n,0} = \dfrac{n_i^2}{N_D}$

p-region: $\quad p_{p,0} = N_A \quad n_{p,0} = \dfrac{n_i^2}{N_A}$

2. Abrupt boundaries: This means that there is a sharp transition between the n-type and p-type regions and also sharp transitions from the fully depleted regions of the SCR and neutral regions beyond.

With n-material to the left and p-material to the right of the origin, the charge density of the SCR is approximated thus by a step-wise function:

$$\rho(x) = \begin{cases} 0 & x < w_n \\ +q \cdot N_D & w_n < x < 0 \\ -q \cdot N_A & 0 < x \ < w_p \\ 0 & w_p < x \end{cases}$$

The charge density is illustrated in panel a) of figure 6.1. We can insert this into the Poisson equation $d^2\Phi/dx^2 = -\rho/\epsilon\epsilon_0$ and integrate once to obtain a linear variation for the electric field $E(x)$.[1] The field is zero in the neutral part of the n-material up to $x = w_n$. From thereon, the field increases linearly to a maximum value value at $x = 0$. It must connect continuously[2] at the interface, and it decreases linearly in the p-material. Eventually, the field becomes zero at $x = w_p$. Charge neutrality requires $N_D \cdot w_n = N_A \cdot w_p$.

We can integrate once more to obtain a piecewise parabolic variation for the electrostatic potential $\Phi_{el}(x)$ which is illustrated in panel c) of figure 6.1. Note that in electrostatics a potential is defined by they way that a positive charge would move; here, it would move away from the positive charge of the ionised donors in the n-region towards the negative charge of the ionised acceptors in the p-region as illustrated by the direction of the arrows in panel b).

Next, we define the built-in voltage V_{bi} by the following potential difference:

$$V_{bi} = \Phi(w_n) - \Phi(w_p) = \frac{q}{2\epsilon\epsilon_0}(N_D \cdot w_n^2 + N_A \cdot w_p^2) \tag{6.1}$$

Here, w_n and w_p add up to the total width of the SCR w. If the two materials of the junction are differently doped, a proportionally larger part of w extends into the region with lower doping.[3]

$$w_n = \sqrt{\frac{2\epsilon\epsilon_0}{q}\frac{N_A}{N_D(N_A+N_D)} \cdot V_{bi}} \tag{6.2}$$

$$w_p = \sqrt{\frac{2\epsilon\epsilon_0}{q}\frac{N_D}{N_A(N_A+N_D)} \cdot V_{bi}} \tag{6.3}$$

$$w = \sqrt{\frac{2\epsilon\epsilon_0}{q}\frac{N_A+N_D}{N_A \cdot N_D} \cdot V_{bi}} \tag{6.4}$$

[1] The result is the x-component $E_x(x)$ of the electric field vector. Embedding this into 3D-space, we could write $\vec{E} = E_x(x) \cdot \vec{e}_x$.

[2] Continuity of the field can only be assumed for a homo-junction without interface states. Junctions between two different materials generally form interface states which can capture free carriers as discussed in appendix G.

[3] In silicon solar cells, the donor concentration of the n-region is usually much higher than the base doping of the wafer. Assuming a one-sided junction, we find $w_p \approx 3.5\,\mu\text{m}$ for $N_A = 1 \times 10^{14}\,\text{cm}^{-3}$ and $w_p \approx 350\,\text{nm}$ for $N_A = 1 \times 10^{16}\,\text{cm}^{-3}$.

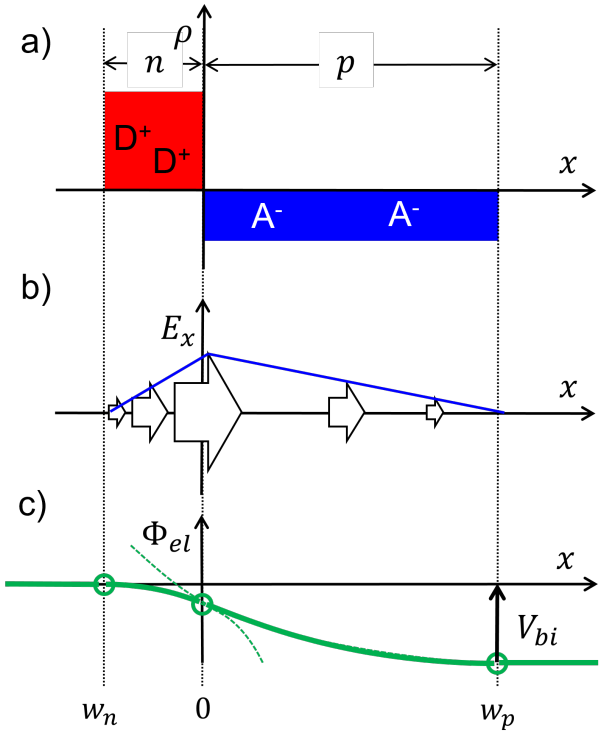


Figure 6.1: Spatial variation of the charge density (a), the electric field (b), and the electric potential (c) of a *p*-*n* junction in equilibrium (*n*-material to the left).

In semiconductors, the vacuum level is defined by $\Phi_{vac} = -\Phi_{el}$. Thus, the electrostatic potential of figure 6.1 c) has to be flipped vertically. Additionally, we have to multiply with q if we want to to relate it to the band bending in terms of energy as shown in figure 6.2 a).[4]

In equilibrium, the drift current associated to the band bending is opposed and exactly balanced by the diffusion current which is directed from high towards low concentrations. The bending of the bands is associated with V_{bi}, the *built in voltage*. In the derivation above it was just a value emerging as upper boundary of an integration, but now we can give it a value from the Fermi-level positions with respect to the intrinsic level E_i according to eqns. (3.27) and (3.28):[5]

$$qV_{bi} = q\Psi_n + q\Psi_p = kT \ln\left(\frac{N_A \cdot N_D}{n_i^2}\right) \tag{6.5}$$

For typical doping concentrations, we obtain:

[4] In a heterojunction, we join materials that have different values of χ and E_g, thus giving rise to a conduction band offset (CBO) and a valence band offset (VBO).

[5] Alternatively, we can use eqns. (3.14) and (3.15), using $E_C - E_V = E_g$ and $n_i^2 = N_V N_C \exp -(E_g/kT)$:

$$qV_{bi} = E_F^n - E_F^p = \left(E_C + kT \ln \frac{N_D}{N_C}\right) - \left(E_V + kT \ln \frac{N_A}{N_V}\right) = kT \ln\left(\frac{N_A \cdot N_D}{n_i^2}\right)$$

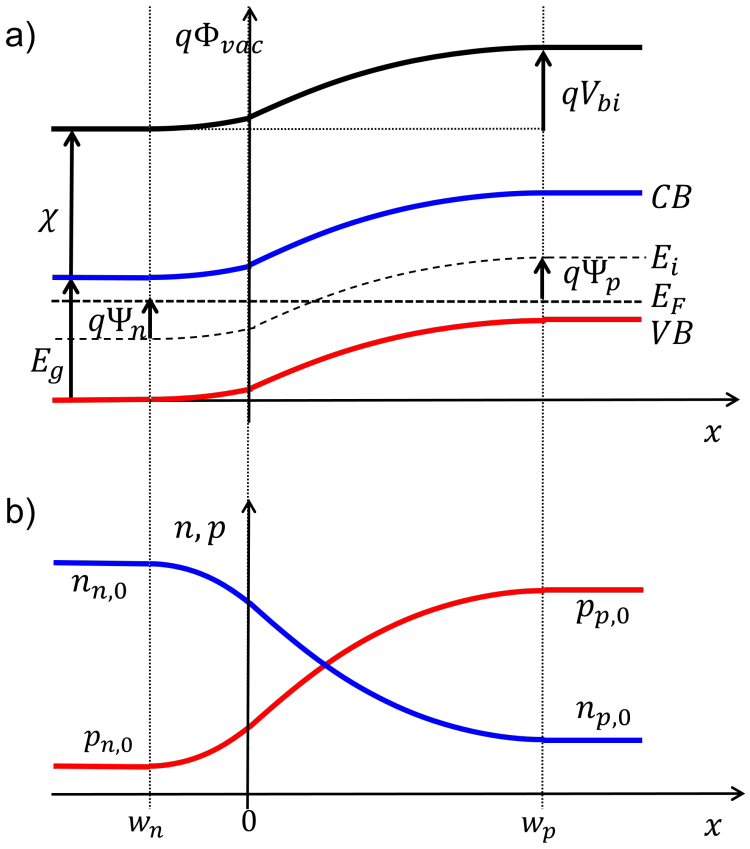


Figure 6.2: Spatial variation of band-bending (a), and carrier densities (b) of a p-n junction in equilibrium (n-material to the left).

$$\begin{aligned} N_D \approx N_A \approx 10^{16}\,\mathrm{cm}^3 &\Rightarrow V_{bi} = 0.730\,\mathrm{V} \\ N_D \approx 10^{18} \text{ and } N_A \approx 10^{16}\,\mathrm{cm}^3 &\Rightarrow V_{bi} = 0.810\,\mathrm{V} \end{aligned}$$

Figure 6.2 b) shows that the carrier densities vary by several orders of magnitude across the SCR. We can describe this variation by expressing the equilibrium between diffusion- and drift-currents. For electrons we have the following:

$$\begin{aligned} qn\mu_n E = -qn\mu_n \frac{d\Phi}{dx} &= -qD_n \frac{dn}{dx} \\ d\Phi &= \frac{kT}{q} \cdot \frac{dn}{n} \end{aligned}$$

Here, we used the Einstein-relation $D_n/\mu_n = kT/q$. The solution of the differential equation is easy, and we can use the borders of the SCR as limits of the integrals:

$$\Phi(w_p) - \Phi(-w_n) = \frac{kT}{q} \ln \frac{n_{p,0}}{n_{n,0}}$$
$$n_{p,0} = n_{n,0} \exp\left(-\frac{qV_{bi}}{kT}\right) \tag{6.6}$$

It tells us how much a given voltage difference reduces the majority carrier density along a region with band-bending, completely independent of the functional relationship of the band bending. In the example above, the band-bending extends over V_{bi} and the electron density varies from $n_{n,0}$ on the n-side into $n_{p,0}$ which is actually a minority carrier density since we crossed the junction. An analogous equation is obtained for holes, and these two simple equations will serve as entry-point to the description of the junction under electrical bias in the next section.

6.2 The infinite *p-n* junction in dark and under electric bias

The properties of the p-n junction become a little more interesting when we apply an electric bias voltage V. Let us take eq. (6.4) as starting point and simply replace V_{bi} by $V_{bi} - V$. Apart from changes in the band bending, the following will primarily impact the minority carriers. We distinguish three possibilities:

1. **rev,** $V < 0$: The case of reverse bias is shown on the left side of figure 6.3. It takes place when the p-side is polarised negatively and the n-side positively. Thus, the built-in potential is increased and according to eq. (6.4) the SCR is widened. Figure 6.3 shows that the minority carrier densities in the neutral regions actually drop below the equilibrium values. This means that we *extract* the minority carriers, but we can extract minority carriers only as fast as they are replenished by thermal generation, regardless of how much reverse bias we apply. The maximum we can draw is the *reverse saturation current density* j_0 and we call this the blocking direction.

2. **fwd** $0 < V < V_{bi}$: The case of forward bias is shown on the right side of figure 6.3. It refers to a situation where the p-side is polarised positively and the n-side negatively. This reduces the built-in potential and thus narrows the depletion region according to eq. (6.4). The term "forward" means that the p-n diode becomes conductive. Physically, this is related to an injection of majority carriers into the SCR (electrons into the n-side, holes into the p-side). As they propagate across SCR, their density gets reduced according to the reduced potential and at the junction at $x = 0$, they even become minority carriers. Nevertheless, at the far end of the SCR their density is sill above the value of thermal equilibrium. Thus, they get injected into the neutral region where they continue to propagate for some time until they recombine. The current that is flowing under forward bias is therefore a *recombination current*, but it is still related to the diffusion of minority carriers. Different from the previous case, we do not need to wait for thermal generation, but we can inject and recombine as many carriers as we want. Therefore, the current density does not saturate in forward direction, but grows exponentially with the applied bias.

3. **fwd** $V > V_{bi}$: In this case, the width of the depletion region according to eq. (6.5) would become imaginary which is physically not meaningful. The reason is that the depletion approximation cannot be used for this case any more, but we have to solve the Poisson equation for a situation of carrier accumulation, similar to the procedure outlined in appendix E.

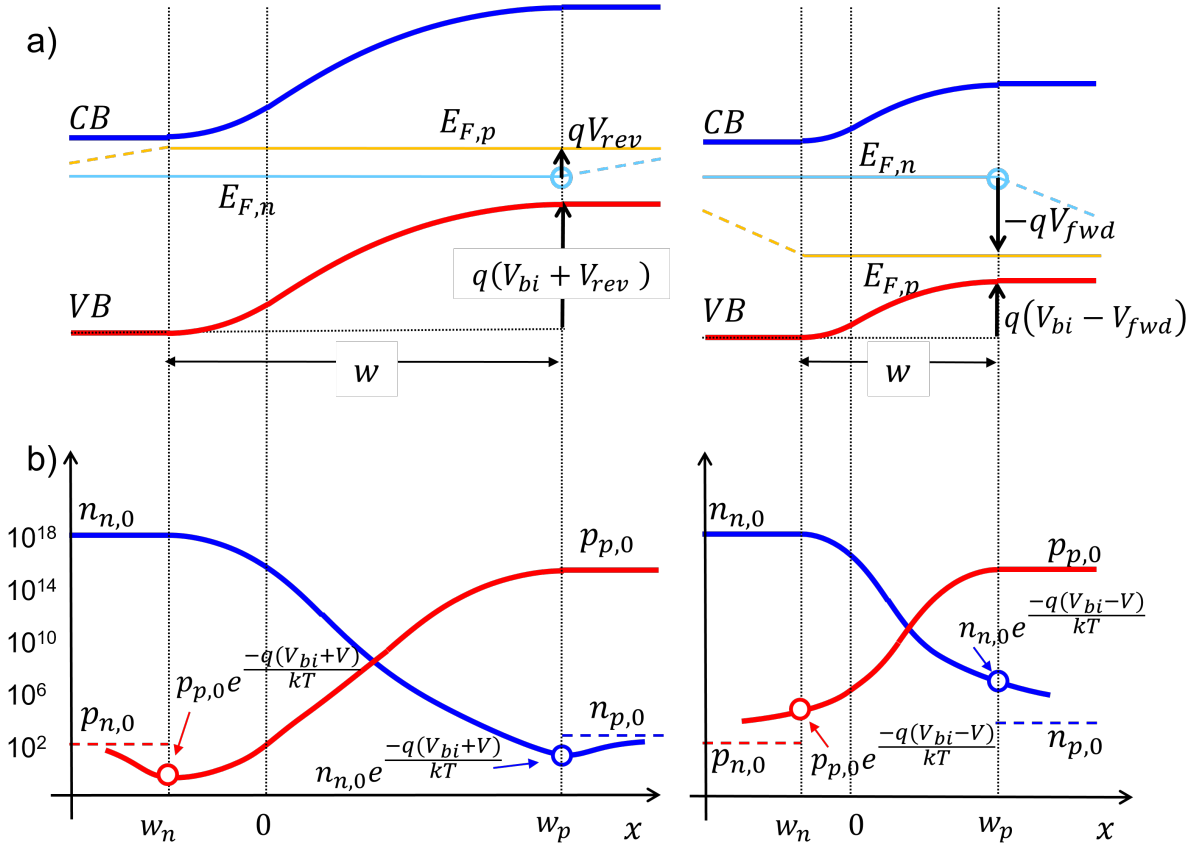


Figure 6.3: The p-n junction in dark under reverse bias (left) and under forward bias (right). The upper panels illustrate the band-bending and the splitting of the quasi Fermi levels, the lower panels represent the resulting carrier densities for typical doping concentrations in silicon (width of SCR not to scale).

Setting aside the third case, we will now describe the first two cases in terms of the minority carriers. To determine the *current-voltage characteristic* we calculate the minority carrier densities at the edges of the SCR with the help of eq. (6.6), but using $V_{bi} - V$. Then, we determine the spatial variation of the minority carrier density by solving the diffusion equation. Subsequently, we take the first derivative according to eq. (3.42) to obtain the current density. We can do this more easily by making the following assumptions:

1. **Depletion approximation:** We continue using the depletion approximation introduced in the previous section for the design of the band diagram of the unbiased case in figure 6.2. The biased case behaves similarly, reverse or forward bias simply widens or shrinks the SCR as shown in in figure 6.3.

2. **Low injection:** We assume that the injection at the edges of the SCR is negligible with respect to the majority densities at the external contacts far away from the junction. Thus, the Fermi level for electrons is defined by the donor concentration in the n-region and the Fermi level for holes is defined by the acceptor concentration in the p-region. We assume that the Fermi levels remain flat throughout the neutral regions which means that we carry the

external bias all the way to the edges of the SCR.[6] We have thus two different Fermi levels at the edges of the SCR. This means that there is no thermal equilibrium any more, but instead we have a steady state of flowing carriers. The deviation is expressed by defining quasi Fermi levels (QFLs).

3. **No recombination and no generation in the SCR:**. This assumption is justified by the fact that the diffusion length of the minority carriers is normally much larger than the width of the SCR. Consequently, they can transit the SCR without significant changes due to recombination. Similarly, we may also neglect the thermal generation.[7]

4. **The carrier densities n and p are described by the variation of the potential Φ.** This allows us to express the densities of the carries across the SCR as illustrated in figure 6.3. For example, let us look at the density of electrons. On the left side, of the SCR it is given by the equilibrium value of the neutral n-material, i.e. $n_{n,0}$. At the right side of the SCR, it gets reduced according to the drop of the potential across the SCR. The resulting densities are illustrated by the open circles. In the band-diagram, this is encoded by the simple fact that we kept the QFLs flat across the SCR.

   Depending on the applied bias voltage, these value are either lower (rev) or higher (fwd) than the equilibrium densities of the minority carriers that are denoted by the dashed lines. To maintain these non-equilibrium values of the carrier densities, we have to continuously extract (rev) or inject (fwd) minority carriers from or into the neutral region in a situation that we call steady state. Moving from the boundaries of the SCR into the neutral regions, the minority carrier densities will eventually approach the dashed lines, either from below (rev) or from above (fwd).

In the following, we consider electrons, but the same considerations hold for holes. We can easily express the density of electrons at $x = w_p$ with eq. (6.6) by replacing V_{bi} with $V_{bi} - V$. For extraction under revers bias, $V < 0$; for injection under forward bias $V > 0$. Thus, n denotes the electron density with applied bias V, and n_0 is the equilibrium value for $V = 0$. Usually we are interested in the density of injected (or extracted) carriers, i.e. $\Delta n = n - n_0$ which is given by:

$$\begin{aligned} \Delta n(w_p) &= n_{n,0} \exp\left(-\frac{q(V_{bi}-V)}{kT}\right) - n_{n,0} \exp\left(-\frac{qV_{bi}}{kT}\right) \\ &= \underbrace{n_{n,0} \exp\left(-\frac{qV_{bi}}{kT}\right)}_{n_{p,0}} \left(\exp\left(\frac{qV}{kT}\right) - 1\right) \end{aligned} \tag{6.7}$$

A similar relation holds for the holes in the n-region. Expredd in terms of flat QFLs across the SCR, the relevant Fermi level for electrons at w_p is the light-blue one which is higher by a value of qV_{fwd} (lower by qV_{rev}) than the orange one, i.e. the position of the equilibrium Fermi level without bias. Accordingly, the charge carrier densities are increased (decreased) by a factor of $\exp\{qV_{bias}/kT\}$. The non-equilibrium densities are indicated by the open circles in figure 6.3 b).

[6] This may not seem very intuitive, but we can illustrate it with a small numerical example: Let us take a typical current density, say $j_{ph} \approx 40\,\mathrm{mA\,cm^{-2}}$ of a silicon solar cell. The conductivity of the neutral bulk region is easily obtained from $\sigma = pe\mu_p$, using $p = N_A \approx 10^{16}\,\mathrm{cm^{-3}}$ and a hole mobility of ca. $400\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$. From $j = \sigma E$ we obtain a field of $0.06\,\mathrm{V\,cm^{-1}}$. Across a wafer thickness of 200 µm this is a potential difference of only 1.2 mV.

[7] Later in this section we will assume photogeneration which cannot be neglected.

Next, we calculate the spatial variation of the electron density in the p-region with the help of the continuity equation for electrons, eq. (5.9). For steady state and darkness, we find $\partial n/\partial t = 0$ and $G = 0$, and thus we to obtain:

$$qR_n = \frac{dj_n}{dx} \tag{6.8}$$

The recombination term R_n is replaced according to eq. (5.6), and we express j_n as a pure diffusion current according to eq. (3.42), assuming a constant diffusion coefficient D_n.

$$qD_n \frac{d^2n}{dx^2} = q\frac{(n - n_{p,0})}{\tau_n} \tag{6.9}$$

This is an inhomogeneous differential equation of second order. The homogeneous solution contains an exponentially increasing and an exponentially decreasing term. We can immediately discard the former one as it does not yield a physically meaningful solution for $x \to \infty$. Finding a particular solution is easy, it is simply the inhomogeneity itself. Applying the injected carrier density of eq. (6.7) as boundary condition at $x = w_p$, we obtain:

$$n(V, x) = n_{p,0}\left(\exp\left(\frac{qV}{kT}\right) - 1\right) \cdot \exp\left(\frac{w_p - x}{L_n}\right) + n_{p,0} \tag{6.10}$$

The solution approaches the equilibrium density $n_{p,0}$ either from above (injection under forward bias), or from below (extraction under reverse bias).[8] In both cases, the decay constant is governed by $L_n = (D_n\tau_n)^{1/2}$, the *diffusion length*.[9]

The figure 6.3 is helpful to clarify a common misconception about the function of the field. Based on the upper panels it would be inviting to claim that the field across the SCR pushes holes to the right and electrons to the left. Looking at the lower panels, it becomes immediately clear that a movement of holes from left to right is opposed by a strong concentration gradient, and the same is true for a movement of electrons from right to left. Across the SCR, the drift current in the field is exactly balanced by the diffusion current due to the concentration gradient. As a result, net flow of charge carriers must be expressed through both, the gradient of the potential and the gradient of the concentration. The two can be combined into the gradient of the Fermi-levels, more precicely the quasi Fermi-levels. Consequently, the only regions were the minority currents flow are those with gradients of the QFLs in the neutral part of the semiconductor close the SCR. In case of reverse bias there is flow of holes from the n-region towards the SCR, alas only very few carriers are involved. In case of forward bias, there is a flow of holes from the SCR into the n-region. To find an expression for the diffusion current of the electrons in the p-region, we apply eq. (3.42):

$$j_n(V, x) = -qD_n\frac{dn(V, x)}{dx} = \frac{qD_n n_{p,0}}{L_n}\left(\exp\left(\frac{qV}{kT}\right) - 1\right) \cdot \exp\left(\frac{w_p - x}{L_n}\right) \tag{6.11}$$

[8] Note that there is a small inconsistency here: Outside of the SCR, we assumed neutral regions where the majority concentrations are unchanged, but now we inject (or extract) minority carriers. Even if can neglect the minorities against the majorities, we should still refer more correctly to *quasi neutral* regions.

[9] The diffusion length depends massively on the doping through both, the lifetime τ and the diffusion coefficient through its dependence on the mobility $D = \mu kT/q$. The impact of doping on the minority-lifetimes is discussed in chapter 5, the dependence of the minority carrier mobilities on doping is discussed in Appendix B. As example, for a majority acceptor concentration of $N_A = 10^{16}\,\mathrm{cm}^{-3}$, D_n and D_p are 28 and 11 $\mathrm{cm}^2\mathrm{s}^{-1}$, respectively.

A similar expression describes the hole-current in the n-region. The spatial variation of $j_n(x, V)$ and $j_p(x, V)$ is shown in figure 6.4. The diagram is constructed as follows:

1. With our approach above, we calculated the the minority carrier densities in the quasi neutral regions adjacent to the SCR. These are are represented by the thin lines.
2. Since the drift- and the diffusion current of the equilibrium conditions are compensating each other, we have only the injected (or extracted) current densities $j_n(x)$ and $j_p(x)$ across the SCR. Both of them are constant between w_n and $+w_p$ because we assumed that there is no generation and no recombination in the SCR. The constant levels must connect continuously with the minority current densities at the respective ends of the SCR.
3. The total current density does not change throughout the whole of the solar cell, thus $j = j_n(x) + j_p(x) = const$ for all x. Since the current densities are constant across the SCR, we can simply use the values that we calculated for the left and the right end. Thus, $j = j_p(w_n) + j_n(+w_p)$.
4. Finally, majority current densities never came into play, but we could calculate them now by the difference between the total current density and the minority current density (denoted by the thick lines that mirror the thin lines of the minority current densities).

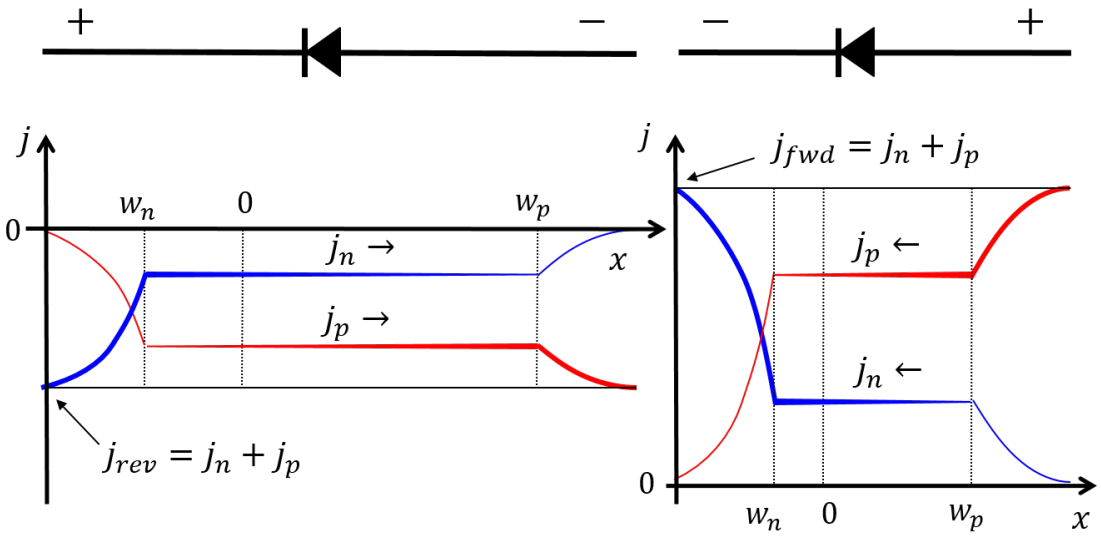


Figure 6.4: The contributions of hole- and electron-current under reverse (left) and forward bias (right). They add up to the to the total current which is constant across the junction.

Coming back to the total current density, we find:

$$\begin{aligned} j(V) = j_p(-w_n) + j_n(+w_p) &= \left(\frac{qD_p p_{n,0}}{L_p} + \frac{qD_n n_{p,0}}{L_n}\right)\left(\exp\left(\frac{qV}{kT}\right) - 1\right) \\ &= \underbrace{(j_{0,n} + j_{0,p})}_{j_0}\left(\exp\left(\frac{qV}{kT}\right) - 1\right) \end{aligned} \tag{6.12}$$

Here, we defined $j_{0,n}$ and $j_{0,p}$ which we call *saturation current densities*[10]. They describe the transport of the minority carriers, i.e. the one for the n-region contains the properties of holes, namely the *equilibrium hole density* $p_{n,0}$, the *hole diffusion constant* D_p and the *hole diffusion length* L_p. Likewise, $j_{0,p}$ for the p-region contains properties of electrons.

As the doping concentrations are normally better known than the equilibrium values of the minority carrier densities, j_0 is most often expressed as shown below:

$$j_{0,n} = \frac{qD_p p_{n,0}}{L_p} = \frac{qD_p n_i^2}{L_p N_D} = \frac{qn_i^2}{N_D}\sqrt{\frac{D_p}{\tau_p}} \tag{6.13}$$

$$j_{0,p} = \frac{qD_n n_{p,0}}{L_n} = \frac{qD_n n_i^2}{L_n N_A} = \frac{qn_i^2}{N_A}\sqrt{\frac{D_n}{\tau_n}} \tag{6.14}$$

We saw in the introduction that the saturation current densities should be as small as possible in order to obtain high V_{oc}. This can in principle be achieved through each of the individual parameters in eqns. (6.13) and (6.14), except q.

The most obvious parameter to vary is the doping concentration in the denominator. Figure 6.5 shows that this works well for low doping concentrations as $j_{0,n}$ and $j_{0,p}$ decrease with a slope equal to -1 in the double-logarithmic plot. However, the trends stops at dopant concentrations of about $1 \times 10^{16}\,\mathrm{cm}^{-3}$. An explanation for this behaviour is the onset of Auger recombination. If the doping is increased further, eq. (5.14) states that the lifetime τ is inversely proportional to the square of the doping concentrations.[11] As the expression for j_0 contains the lifetime within a square-root, the dependencies should cancel. This is illustrated by the dashed characteristics which are (almost) flat. In reality they continue to decrease slightly since the diffusion coefficients decrease with increasing doping.

Unfortunately, this is not the end of the story because high doping concentrations reduce the value of the bandgap and thus increase n_i^2. With the inclusion of this effect, the j_0 characteristics do not flatten at high doping densities, but they actually increase. In figure 6.5, the slope to the right indicates a proportionality to the square-root of the doping density. Details of bandgap narrowing are discussed in Appendix B.

Another handle for low j_0 would be to lower the mobility, but obviously this is not a strategy we would like to employ because we have to transport the minority carriers over the thickness of the wafer. A more successful strategy is the use of crystalline materials with high quality and thus high lifetime.

We discussed the impact of doping for the case of silicon, but the underlying principles apply also to other semiconductors. This brings us the remaining term, the intrinsic carrier density n_i^2. As it depends exponentially on the bandgap, it has a significant impact on j_0. For example, in Ge, Si, and GaAs with their respective bandgaps of 0.8, 1.1 and 1.45eV, $n_i \approx 10^{13}$, 10^{10} and 10^6 cm^{-1}. Besides the dependence on the bandgap, n_i^2 also varies exponentially with the temperature. This translates into a linear variation of the V_{oc} with temperature, suggesting that we should avoid high operation temperatures.

[10] For historical reasons, the p-type wafer of silicon solar cells is often called base in analogy to transistors. Accordingly, the n-type region at the front of is often called *emitter* and its saturation current is often denoted by $j_{0,e}$ instead of $j_{0,n}$.

[11] The Auger effect would yield very high lifetimes at low doping concentration, but even very pure material is generally limited by a bulk recombination effect; here we assumed an upper limit of 50 ms [29].

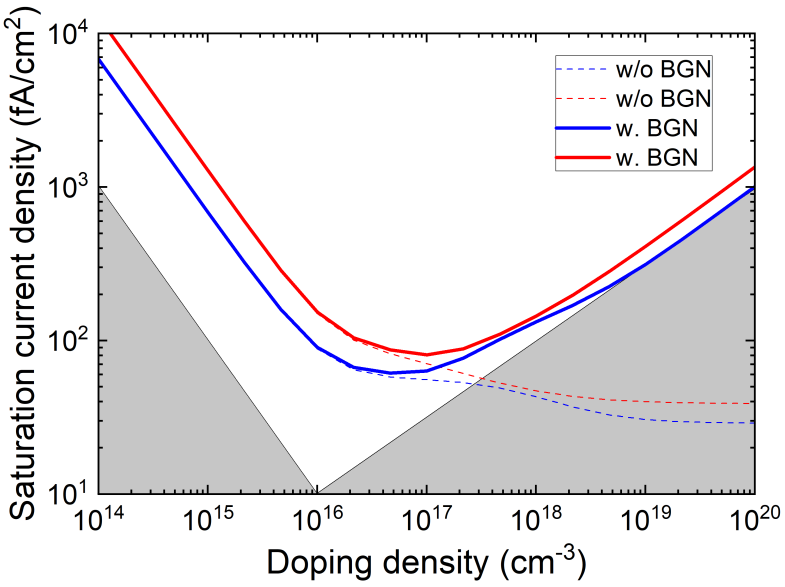


Figure 6.5: Dependence of the saturation current density on doping density. Red and blue curves refer to p-type and n-type material respectively. Dashed lines illustrate the behaviour with Auger recombination, but without band gap narrowing.

6.3 The infinite *p-n* junction cell with uniform generation

In case the junction is illuminated, photons with sufficient energy can excite electrons from the valence-band to the conduction-band, i.e. absorption creates additional electrons and holes. To keep the treatment simple, we start with a generation profile G that is constant throughout the full device. Under steady-state conditions this will yield an excess carrier density $\Delta n = G\tau$ for which we assume low-injection, i.e. that Δn is less than the doping concentration. Thus, it is sufficient to keep track of the minority carriers even though electrons and holes are generated in pairs. After generation somewhere in the neutral region, the minority carriers move around by diffusion until one of two things happens:

1. After a duration equal to the lifetime, they recombine with a majority carrier. This is not exactly what we want.
2. They diffuse across the junction and become majority carriers on the other side. In that case, they are less prone to recombination and there is a good chance for them to be collected (we assume an infinite cell with ideal contacts for the time being).

The current density under illumination is calculated the same way as the one in dark: For electrons in the p-region we start once again with the continuity equation, eq. (5.9), but now with a generation term $G = const.$ Then we apply the boundary conditions to determine $n(x)$ in the p-region, and finally we determine the diffusion-current density by taking the first derivative. Let us start with the continuity equation under steady state with recombination and generation:

$$\begin{aligned} D_n \frac{d^2 n}{dx^2} &= R_n - G \\ \frac{d^2 n}{dx^2} - \frac{n}{D_n \tau_n} &= \frac{n_{p,0}}{D_n \tau_n} - \frac{G}{D_n} \end{aligned} \tag{6.15}$$

The constant generation rate G can easily be integrated into the particular solution. As boundary condition at $x = w_p$ we use once more the density of electrons that gets injected by the bias voltage according to eq. (6.7).[12] The second boundary condition at $x \to \infty$ is $n = n_{p,0} + G\tau$. Thus we obtain the following profile for the carrier density:

$$n(x) = \left[n_{p,0} \left(\exp\left(\frac{qV}{kT}\right) - 1 \right) - G\tau_n \right] \cdot \exp\left(\frac{w_p - x}{L_n}\right) + n_{p,0} + G\tau_n \tag{6.16}$$

We can easily carry out the derivative to find the current density for electrons in the p-region at $x = w_p$:

$$j_n(V) = j_{0,p} \left(\exp\left(\frac{qV}{kT}\right) - 1 \right) - qGL_n \tag{6.17}$$

With respect to the cell in dark, there is an additional term of $-qGL_n$ which represents the photocurrent. Looking more closely, we see that we do not collect all carriers, but only those that are within a diffusion length of the SCR. For a real solar cell, we can immediately conclude that the minority diffusion length should be larger than the absorption length of light. In case of silicon where the absorption length of IR light is larger than the wafer thickness, we find that the diffusion length should also be larger than wafer thickness. We can achieve this by using single-crystalline material with a low concentration of dopants and impurities. For poor material quality such as in the case of some thin film solar cells, we can use an *electric field* across the active layer to drive the carriers towards the contacts by drift, but we can no longer solve the equations analytically for this case; see chapter 11 on thin film solar cells.

The total current across the junction is once again determined by adding the values of the minority currents at the edges of the SCR with one little difference; according to our assumption $G = const$, there is now generation also within the SCR. We assume that all of these carriers are collected, thus the SCR contributes a current $j_{SCR} = qG \cdot w$. Finally we obtain a diode equation very similar to eq. (6.12), but with a shift of j_L along the axis of the current-density:

$$j(V) \quad = \quad j_0 \left(\exp\left(\frac{qV}{kT}\right) - 1 \right) - j_L \tag{6.18}$$

The difference j_L is given by the sum of three contributions:

$$j_L \quad = \quad q\,G(L_n + L_p + w) \tag{6.19}$$

The fact that eqns. (6.12) and (6.19) differ only by the photocurrent j_L, is called *superposition principle*, i.e. our ideal solar cell can be described by a diode in parallel with a current source. As a consequence, the $j(V)$-characteristic is simply shifted along the j-axis. From the measurement of a $j(V)$ curve the following cell parameters can be extracted (in case there are no parallel- or series resistances):

1. We obtain the short circuit current by setting the voltage to zero:

$$j_{sc} = j(V = 0) = j_L \tag{6.20}$$

[12] Remember that the electron density at the position w_p is obtained from the potential Φ with respect to the flat continuation of the quasi Fermi-level throughout the junction. Formally, we are describing majority carriers all the way to $n(w_p)$. Sticking (half-heartedly) to the assumption of low-injection, we do not add the generation term at $n(w_p)$, but only when we are dealing with minority carriers at $x \to \infty$.

2. For the open circuit voltage, we set the current density to zero as we don't extract any current ($j(V_{oc}) = 0$). Thus, we obtain:

$$V_{oc} \approx \frac{kT}{q} \ln\left(\frac{j_L}{j_0}\right) \tag{6.21}$$

3. The fill factor *FF* is the ratio between the maximum power P_{max} and the product of j_{sc} and V_{oc}. As discussed already in section 1.5.1, it can be obtained from Lambert's W-function or from the following approximation [2].

$$FF_0 \approx \frac{qV_{oc}}{qV_{oc} + 4.7nkT} \tag{6.22}$$

The subscript "0" refers to the case with zero series-resistance R_s and infinite parallel-resistance R_p. The formula includes a quantity n which is not part of our derivation. This so-called diode quality factor is an empiric extension and for the time being, we can simply ignore it by setting $n = 1$. The effects the diode quality factor as well as of series- and parallel-resistance will be treated in more detail in section 6.5.

Before we continue to effects of finite cell thickness, we point out a relation between the V_{oc} and the quasi Fermi level splitting that we discussed in the context of eqns. (3.27) and (3.28). Note that this is not a rigorous proof, but more a useful heuristic. Let us assume for the moment that the behaviour of the cell is dominated by the p-type bulk, Then, we have $j_L \approx qGL_n$ and $j_0 \approx j_{0,p}$, and thus we can write:

$$\begin{aligned} V_{oc} &\approx \frac{kT}{q} \ln\left(\frac{qGL_n}{qD_n n_i^2/(L_n N_A)}\right) \\ &= \frac{kT}{q} \ln\left(\frac{G\tau_n \cdot N_A}{n_i^2}\right) \end{aligned} \tag{6.23}$$

Since $G\tau_n = \Delta n$, we reproduce eq. (3.31) and all our considerations of injection and diffusion are no longer part of the result. In other words, we can predict an upper limit of the V_{oc} already on the basis the wafer doping and the lifetime, no other input is needed. Since no actual cell structure is involved, we do not call this quantity V_{oc}, but *implied open circuit voltage* iV_{oc}.

6.4 Finite solar cell in dark

In contrast to the idealised device discussed in the previous sections, the differently doped regions in a real solar cell have *finite thicknesses*, and recombination effects at the interfaces must be considered. This is done by taking into account the *surface recombination velocity* S. For the bulk region of a p-type wafer with constant doping concentration, we find once again an expression with a saturation current density $j_{0,p}$, but it is multiplied with a term that we call *geometry factor* $G_{F,p}$ (c.f. appendix D):

$$j_{0,p} = \frac{qD_n n_i^2}{N_A L_n} \cdot G_F \tag{6.24}$$

$$G_{F,p} = \frac{\cosh(H/L_n) + S_\infty/S_n \sinh(H/L_n)}{S_\infty/S_n \cdot \cosh(H/L_n) + \sinh(H/L_n)} \tag{6.25}$$

Here, S_n is the surface recombination velocity of electrons at the rear of the p-type bulk and $S_\infty = D_n/L_n$ is a recombination parameter defined for the minority carriers in order to simplify the expression. We can distinguish three regimes of device operation:

- $S_n = S_\infty \Rightarrow G_{F,p} = 1$. We retrieve the formula of the *infinite cell* and $j_{0,p}$ depends only on the properties of the semiconductor. This case is primarily of mathematical interest as it would be very unlikely that S_n can be engineered to exactly match the properties of the semiconductor.
- $S_n > S_\infty \Rightarrow G_{F,p} > 1$. In this case, j_0 is increased with respect to the infinite cell and thus η is decreased. For very large S_n, we can approximate $G_{F,p} \approx \coth(H/L_n)$, and if additionally $L_n \gg H$, we find $G_{F,p} \approx L_n/H$. In the expression of $j_{0,n}$ this looks as if the diffusion length were replaced by the wafer thickness.
- $S_n < S_\infty \Rightarrow G_{F,p} < 1$. In this case, j_0 is reduced and η is increased. This means that we can fabricate a good cell from a qualitatively bad material (bad bulk lifetime) by applying a good surface passivation. With thinner wafers the ratio between device thickness and diffusion length decreases and this case becomes increasingly important.

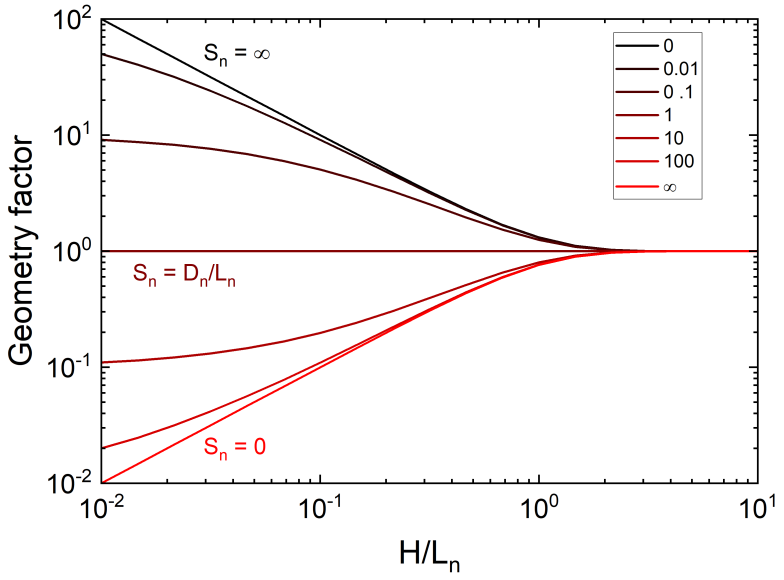


Figure 6.6: Geometry factor as a function of the ratio between base thickness and minority diffusion length, for different values of S_∞/S_n.

The relation for $j_{0,p}$ with the geometry factor of eq. (6.25) holds well enough for low doping and a constant profile such as it is the case for the p-region. In case of high doping concentrations and spatially varying concentrations such as in the diffused n-region, the situation becomes more complex and the following aspects have to be considered:

1. There is an *internal field* due to the variation of the doping concentration.
2. The high doping yields *band gap narrowing* and changes n_i^2 in the prefactor.
3. *Auger recombination* can become the dominant limitation for the lifetime.

4. The *mobility* is reduced due to ionised impurity scattering at the high density of dopants.

5. As combination of the previous two items, the diffusion length $L_{n,p}$ is reduced.

Unfortunately we need high doping concentrations at the front in order to ensure lateral collection to the metallisation grid, and also to achieve a low contact resistance. Besides their negative impact on recombination, high doping concentrations also yield *free carrier absorption* (FCA) through which a substantial part of the light can be lost. In modern cell designs these loss mechanisms are mitigated with a low surface concentration and a shallow diffusion profile. In turn, current extraction requires closely spaced fingers of the metallisation and a low $j_{0,n}$ can only be achieved with good surface passivation.

A compromise is the the *selective emitter* that we already discussed in section 5.7. Since recombination at the metal-semiconductor interface is high and necessarily yields high $j_{0,n}$, high doping is applied below the contacts in order to shield minority carriers. Since the metal contacts are opaque, FCA is not an issue there. Between the contacts, a lower doping concentration is applied and a surface passivation is applied. Early silicon solar cells with n-type region at the front were passivated with SiO_2 as this material saturates very effectively the dangling bonds of the silicon surface, an effect called chemical passivation. In modern cells Si_3N_4 is favoured as it contains a positive fixed charge that repels holes, an effect called field-effect passivation.

To summarize, we determined a formula for the $j(V)$-characteristic of a finite solar cell on the basis of the depletion approximation and by restricting ourselves to diffusion transport.

$$j(V) = j_0 \left(\exp\left(\frac{qV}{kT}\right) - 1 \right) - j_{\mathrm{L}} \tag{6.26}$$

The dark saturation current density j_0 is determined by the contributions from the two differently doped regions of the junction, i.e. $j_{0,n}$ and $j_{0,p}$. The respective doping concentrations enter directly in both terms, and then again indirectly through their impact on the minority lifetime and the minority diffusion coefficient. Yet another indirect mechanism is the effect on band gap narrowing and thus on n_i. Effects of the surface in finite cells are described by the geometry factor G_F which contains the surface recombination velocities and the ratio between the diffusion length and the thickness of the respective doped regions.

$$j_0 = j_{0,n} + j_{0,p} \tag{6.27}$$

$$j_{0,n} = \frac{qD_p n_i^2}{N_D L_p} G_{F,n} \tag{6.28}$$

$$j_{0,p} = \frac{qD_n n_i^2}{N_A L_n} G_{F,p} \tag{6.29}$$

We should keep in mind that the derivation still relies on several simplifications. In reality, lifetime, diffusion coefficient and n_i depend on the excess carrier density Δn which primarly varies with illumination level and applied voltage. The latter may not be obvious, but we should consider that only under short circuit conditions all the carriers are rapidly extracted from the cell. As we approach open circuit conditions, they build up higher excess carrier density. For illumination with sunlight, most historic c-Si solar cells worked at low excess carrier density of $\approx 1\times10^{13}\,\mathrm{cm}^{-1}$. In high efficiency solar cells, recombination in the bulk and at the surfaces is reduced, and the excess carrier density can reach up to $1 \times 10^{16}\,\mathrm{cm}^{-1}$. Such cells can enter the domain of Auger-recombination which ultimately defines the upper limit for the efficiency of silicon solar cells.

6.5 Two-diode model

In measurements of solar cells, the $j(V)$ curve is often modified by a series resistance R_s, a shunt resistances R_p, and additional carrier recombination processes, e.g. in the space charge region. Therefore, our analytical model generally fails to reproduce measured $j(V)$-characteristics. Consequently, the simple one-diode model is usually replaced by a so-called two-diode model as illustrated in the equivalent circuit of figure 6.7.

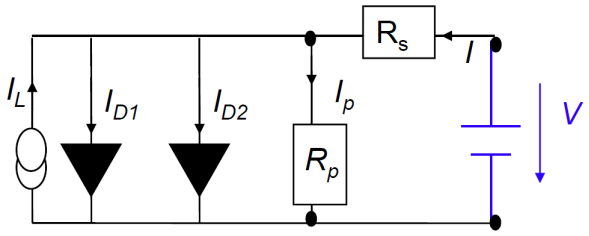


Figure 6.7: Equivalent circuit for the two-diode model.

The series resistance R_s can be found approximately from the illuminated $j(V)$ curve by taking the inverse of the slope at V_{oc}.

$$R_s \approx R_{oc} = \left(\left. \frac{dj}{dV} \right|_{V=V_{oc}} \right)^{-1} \tag{6.30}$$

The parallel resistance R_p, creates a shunt path around the diode(s). It reduces the V_{oc} as it is no longer possible to build up a high voltage. In first approximation it is the inverse of the slope of the $j(V)$ curve at $V = 0$.

$$R_p \approx R_{\mathrm{sc}} = \left(\left. \frac{dj}{dV} \right|_{V=0} \right)^{-1} \tag{6.31}$$

Usually the series resistance R_s and the parallel resistance of R_p are expressed as a product of resistance and area. Considering the two resistances and also the additional diode, we get the following expression for the current-voltage characteristic:

$$\begin{aligned} j(V) &= j_{01}\left(\exp\left(\frac{q(V-jR_s)}{n_1 kT}\right)-1\right) \\ &\quad + j_{02}\left(\exp\left(\frac{q(V-jR_s)}{n_2 kT}\right)-1\right) \\ &\quad + \frac{V-jR_s}{R_p} - j_L \end{aligned} \tag{6.32}$$

Here, n_1 and n_2 are called *diode quality factors*. Remember that we did not have a diode quality factor for the ideal junction, thus $n_1 = 1$ is considered the ideal behaviour. We can try to relate the diode quality factors to recombination phenomena in the bulk by using a relation for the j-V characteristic in terms of wafer thickness w and recombination rate R as proposed by Tiedje [41]:

$$j = qwR - j_L \tag{6.33}$$

To assess the dark characteristic, we set $j_L = 0$ and replace $R = \Delta n/\tau$. Next, we relate Δn to the applied voltage through the the pn product of eq. (3.32). As long as the applied voltage is low enough to keep Δn below the wafer doping, figure 5.1 suggests that τ is constant for radiative recombination and Auger recombination, and almost constant for SRH recombination. Assuming a p-type wafer with $p_0 = N_A$, we can approximate pn by $p_0\Delta n$ and find that Δn is proportional to $\exp(qV/kT)$. Thus, we obtain a diode quality factor equal to 1.

If the applied voltage is high enough for Δn to exceed the wafer doping, figure 5.1 suggests that τ is again constant for SRH recombination, decaying proportionally to Δn^{-1} for radiative recombination, and decaying proportionally to Δn^{-2} for Auger recombination. At the same time, we should approximate the pn product by Δn^2 which means that Δn is proportional to $\exp(qV/2kT)$. Together with the lifetime this yields diode quality factors equal to 2 for SRH recombination, equal to 1 for radiative recombination, and equal to 2/3 for Auger recombination.

Figure 6.8 reproduces temperature dependent I-V measurements of PESC solar cells, a device configuration with passivated emitter and a highly doped region at the rear that is covered with a full-area metallisation [42]. Generally the temperature dependence is well reproduced with the two diode model, but the ideal diode characteristic with slope equal to one is found only in a small region of 0.45 - 0.6 V. At lower bias voltages we find a slope of 1/3, suggesting a diode quality factor of 3. This is not explained by our considerations above. According to Breitenstein, it is indicative of cascaded recombination processes close to the contact regions that go beyond the theory of SRH recombination [43].

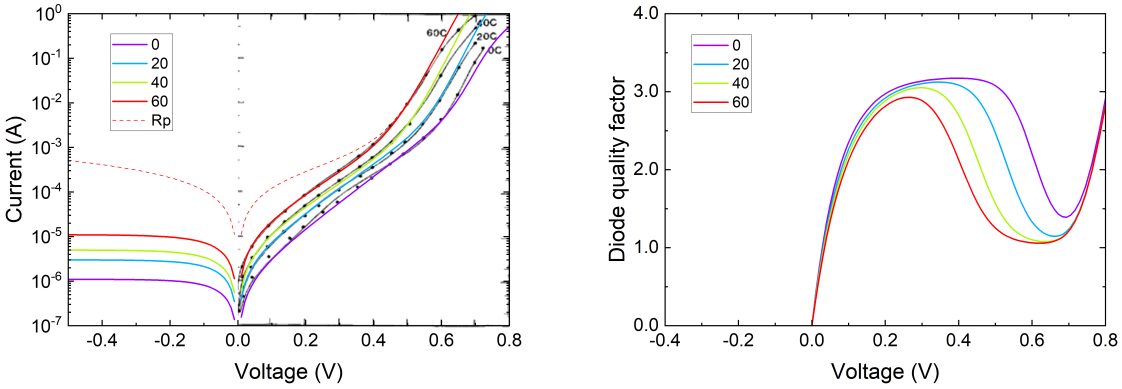


Figure 6.8: Illustration of $j(V)$ characteristics measured at different temperatures (left, symbols from [42]). The series resistance R_s bends the characteristics at $V \approx 0.7$ V, the effect of a parallel resistance R_p is schematically shown by the dashed line. The diode quality factor is obtained by differentiation and suggests $n_1 \approx 3$, and $n_2 \approx 1$ (right).

The series resistance and the parallel resistance have very distinct effects on the $j(V)$ characteristic. In figure 6.8 the former bends the characteristics at high forward bias whereas the latter would normally show up under reverse bias. In that region, the $j(V)$ characteristics normally flattens out asymptotically to the constant value of the reverse saturation current but the presence of a parallel resistance R_p yields a curvature. The parallel resistance can extend to small forward bias as illustrated by the dashed line in figure 6.8. In that case it would complicate the identification of the diode with high quality factor, but unfortunately we cannot assess R_p of the shown device as the measurements did not include reverse bias.

The presence of parallel and series resistances also has distinct effects on the illuminated $j(V)$ characteristics as illustrated in figure 6.9. A parallel resistance does not change the j_{sc}, but it does

not allow the cell to build up a high V_{oc}; ideally it is as high as possible. The series resistance does not change the V_{oc}, but in bad cases it can reduce the j_{sc}; ideally it is as small as possible.

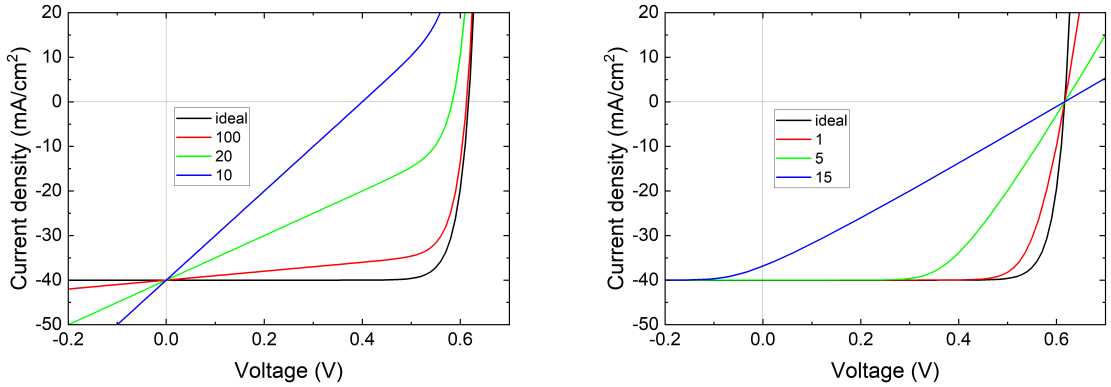


Figure 6.9: Impact of parallel resistances varying in the range of for $R_p = \infty$ - $10\,\Omega\,\mathrm{cm}^2$ (left) and of series resistances in the range of $R_s = 0$ - $15\,\Omega\,\mathrm{cm}^2$ (right) on the illuminated $j(V)$ characteristics.

Figure 6.9 shows that in both cases the FF is massively impacted. Starting from the ideal FF_0 given approximately by eq. (6.22), we can describe the effects as follows [2]:

$$FF_p = FF_0\left(1 - \frac{v_{oc} + 0.7}{v_{oc}} \cdot \frac{FF_0}{r_p}\right) \tag{6.34}$$

$$FF_s = FF_0(1 - 1.1 \cdot r_s) + \frac{r_s^2}{5.4} \tag{6.35}$$

The equations use the so-called reduced open circuit voltage which is quantity without units defined by $v_{oc} = qV_{oc}/nkT$. Similarly, the resistance area products R_p and R_s in units of $\Omega\,\mathrm{cm}^2$ are scaled to the unit-less quantities $r_p = R_p j_{sc}/V_{oc}$ and $r_s = R_s j_{sc}/V_{oc}$.

Chapter 7

Wafer processing

In the early years of PV, single-crystal silicon of very high purity was produced for microelectronics and left-overs of low quality were used for PV. As the massive growth of PV in the 90s and 2000s required larger and larger quantities, the PV industry reverted to multicrystalline silicon as it was found that the lower initial quality of this material could be improved during the manufacturing process by impurity gettering and hydrogenation. In the years after 2010 the fabrication cost of single crystals could be significantly reduced by better understanding of the Cz process and multiple use of crucibles. At the same time, kerf loss of the wafer sawing process was massively reduced by the introduction of diamond wire sawing. Thus, multicrystalline material was replaced almost completely by single-crystal Si by the year 2020. Let us look in detail at the individual steps that it takes to get a wafer:

7.1 Metallurgical grade silicon

Silicon is generally bound to oxygen in the form of quartz (SiO_2) or other minerals. To obtain it in elemental form, it is extracted by carbothermal reduction along the following chemical reaction:

$$SiO_2 + C \rightarrow Si + CO_2$$

The reaction is carried out in an arc-furnace as shown schematically in figure 7.1. In there, the global chemical reaction proceeds through a hierarchy of intermediate reactions that prevail in the different regions of the furnace. The uppermost part of the furnace is a relatively cold zone where gaseous SiO and CO rise from hotter regions below. Unless driven back into the reaction zone by an adequate feed of reactants, the SiO vapour either condenses or reacts with CO to fine particles of SiO_2 and C which are filtered from the exhaust. In a hotter zone below, the SiO_2 from the feed melts, and SiO reacts with C to SiC and CO. The reaction that actually produces elemental Si requires temperatures above 1800 °C [44] which are reached in the region below by an arc discharge between graphite electrodes. There, SiC and SiO_2 react to Si, SiO, and CO. Overall, the process consumes carbon from the feed and from the electrodes whereas elemental Si collects as melt below the arc from where it can be tapped. The intermediate reaction products SiO and CO are gaseous and rise in the furnace to feed the reactions in the colder zones above.

The carbon source is usually coal or coke. Additions of charcoal or wood-chips are used to enhance the reactivity in the upper part of the furnace and to reduce the loss of SiO [45]. The Si-producing reaction in the lowest region can be enhanced catalytically by additions of iron or calcium [46], resulting in Si with a purity of 98 - 99%. Most of this material is used in alloys with aluminium or magnesium, therefore it is called metallurgical grade silicon (MG-Si). In 2020, the

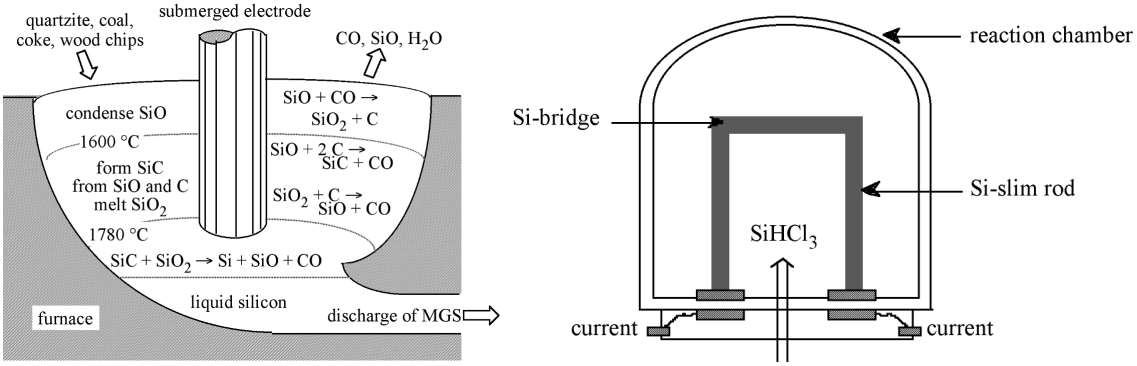


Figure 7.1: Schematic drawing of an arc furnace with different temperature zones (left) and of a Siemens reactor for the growth of poly-silicon (right). Images from `https://cnx.org/contents/bGvrPT9L@3/Semiconductor-Grade-Silicon`

global production amounted to about 4 Mio tons of which about 20% went into photovoltaics and electronics.

Besides the carbon consumption of the chemical reaction, the process requires electric energy of approximately 11 - 14 kW h per kilogram of raw silicon to sustain the arc discharge [45].

7.2 Purification

As MG-Si contains 1-2% of impurities, it must be further purified. To this end, it is ground to powder and mixed with hot HCl at 300 - 400 °C to produce H_2 and $SiHCl_3$ (Tri-Chloro Silane, TCS) which is liquid at RT. Subsequently, TCS is separated from metallic impurities like Fe, Ni, Cu etc. by repeated distillation in 3 - 6 stages.

After that, the liquid is processed into purified poly-silicon.

1. The majority of silicon is processed with the so-called **Siemens process** as shown in figure 7.1. It uses a vacuum vessel where where U-shaped rods of high-purity Si are heated electrically to about 1150 °C.[1] Then, TCS and hydrogen are introduced into the vessel and silicon grows by thermal decomposition on the rods. Byproducts like HCl are evacuated from the reaction zone. This process is energetically costly as the rods have to be heated, whereas the reactor vessel must be cooled to avoid deposition on the walls. Overall, this process adds another 45 - 50 kW h per kilogram of silicon.

2. Alternatively, processing in a **fluidized bed reactor (FBR)** works by releasing purified silicon granules and heated silane (SiH_4) into a reaction chamber. Silane dissociates and deposits pure Si on the granules which grow heavier and exit via the bottom. Despite its higher energy efficiency and the advantage of continuous operation, this process is not (yet) widely used.

[1] The legs of the rods are often rectangular bars of 10 - 15 mm sawed from a 2 - 3 m long ingot with a wire saw. The deposition rate is low at the begin of the process because of the small surface area, eventually it increases as the rods grow thicker. Alternatively to the thin rods, hollow silicon tubes are increasingly used. As they start already from a larger diameter, the process starts directly with high deposition rate.

7.3 Ingot fabrication

There are different possibilities how to produce an ingot, a large piece of crystalline or multi-crystalline silicon.

1. **Czochralski (CZ) process**: Poly-Si is molten in a crucible and a seed crystal is inserted from the top. When the crystal is slowly pulled up, crystalline silicon grows at the interface between seed and melt. The challenges are contamination due to contact of the melt with the crucible and incorporation of oxygen into the upper part of the growing crystal since it is exposed to the ambient during the whole pulling process. If boron doping is used, this is the source for the B-O complex described in section F.2.2. The process yields single crystals with diameters of up to 30 cm which can be controlled via the pulling- and rotation-speed. Typical seed crystals support and ingots with weight up to 400 kg.

2. **Float-zone (FZ) process**: For this process a seed crystal is soldered to a poly-Si rod. Then, the rod is melted with an RF coil and the molten zone is advanced from the seed-end along the rod. On the cooling side of the molten zone, the melt crystallises with the orientation of the seed crystal. Impurities are collected in the melt because of a higher solubility in molten silicon. Thus, they travel along with it, eventually accumulating in a region at the end which is discarded. The method yields material of highest purity because there is no contact with walls and because it can be performed under inert gas atmosphere. Thus it is possible to obtain higher lifetimes than with the CZ process. The drawback is the price.

3. **Cast mc-silicon**: This method starts by pouring molten silicon into a crucible. Then, the bottom is cooled and the zone of solidification eventually moves upwards. Nucleation at random sites in the bottom of the crucible yields multicrystalline material. The main problems are the impurities that are introduced by the atmosphere (O and C) but also the impurities that diffuse into the ingot from the crucible (O, N, Fe, Cr, Ni and Cu). Thus, the quality within an ingot varies a lot, especially at the edges.

4. **Quasi mono process**: This method is similar to the growth of mc-Si, but it uses mono-crystalline seeds at the bottom and melting of the seed-crystals at the bottom melt should be avoided. Upon solidfication from the bottom to the top, quasi monocrystalline material starts to grow epitaxially. Even though the electrical properties of this material are better than mc-Si, ingots do not crystallise homogeneously, there are still multi-crystalline parts and dislocations.

7.4 Wafering

Wafer sawing is a key step of manufacturing. A first sawing step is applied after solidifaction in order to cut the ingots or the blocks into (pseudo-)square bricks. In the past, this was done with circular or band saws (fast, but large kerf loss). Modern processes usually use wire saws (slow but little kerf loss),

After sawing, the bricks are glued onto a holder and passed to a second sawing step where wafers are produced. Historically, a thin steel wire was used in combination with a so-called slurry, a liquid that contains small abrasive particles. The movement of the wire carries the slurry into the groove where the particles propagate in a rolling movement between the silicon and the wire. During rolling, occasionally a sharp edge of a particle impacts into silicon with sufficient force to initiate a crack. Eventually, silicon chips break away from the crack and get carried out of the grove by the

slurry. Slurry sawing was replaced by diamond wire sawing where abrasive diamond particles are bonded to a metallic wire. Since the particles move with the speed of the wire, abrasion is faster and proceeds in a ductile mode without the formation of cracks. No slurry is needed, only a pure lubricant is added to cool the wire and to extract cut silicon-particles. For both methods there are advantages and disadvantages. The slurry method works with a cheaper wire, but the wire wears and can only be used once. As the wire wears, the diameter of the cut groove shrinks, resulting in wedge-shaped wafers. Recycling of silicon from the abrasive particles in the slurry is difficult and further complicated by accumulation of metallic particles from the wearing wire. The diamond coated wire is more expensive, but it can be used several times and the grove diameter remains constant. Moreover, the silicon particles can be recycled more easily from the lubricant.

After sawing, the finished wafers have to be detached from the support, cleaned and characterized. The state-of-the-art wafer thickness is 150 - 200 µm but wafering is improving continuously. With the current sawing and processing techniques the practical limit is probably around 100 µm. Thus, up to 6000 wafers can be produced from an ingot.

In case of sawing with slurry, the cutting leaves behind a surface that is full of cracks. In case of multi-crystalline material, these cracks are used as nucleation sites for the etching of the surface texture in an acidic solution. In case of monocrystalline material, the cracks are not desired and have to be etched off in the so-called *saw-damage etch*, only after that a caustic etch is applied to obtain a texture that consists of square pyramids.

7.5 Alternative wafer preparation

Besides the classical wafering techniques there are alternative processes as described below:

1. Edge defined film fed growth (Schott). Instead of ingots, an octagon is pulled from the melt. The speed of pulling and rotation can be used to control the thickness.

2. Wafering with *heat-resistant wires* that are pulled through the Si melt. The liquid Si spans a membrane between the wires and solidifies between two wires (Evergreen Solar).

3. Cleaving wafers directly from a Si brick, which involves the implantation of a "crack seed" and subsequent cleavage. Wafers with thicknesses of 20 - 150 µm can be obtained without any kerf loss (SIGEN).

4. Stress-assisted cleaving works by depositing a metal on a heated crystalline substrate. Upon cooling, the mismatch of the thermal expansion coefficients introduces a stress field below the surface. After initiating a crack at the side, the metal can be peeled off with a thin layer of silicon. Thereby wafer thicknesses of around 50 µm can be obtained.

5. Epitaxial silicon wafers can be grown by chemical vapour deposition on a silicon seed wafer with detachment zone. For example, the detachment zone can be obtained by etching a region of porous silicon into the surface. Alternatively, annealing of deep trenches etched by RIE can coalesce into a membrane on top of a buried void.

6. Growth of silicon sheets on a bath of molten Si. Like ice on water, solid silicon is less dense that the liquid and therefore floats on top. The company 1366 Inc. developed a process that grows sheets of solid Si on the surface of a melt. Introduction of doping gases during the growth can be used to created arbitrary doping profiles in the growing material.

Chapter 8

Standard silicon solar cell processing

In 1984, the team at the University of New South Wales (UNSW) in Sydney proposed a cell design based on p-type wafers and phosphorus diffusion to create the n-region at the front (the "emitter"). They used a full-area metallisation on the rear side and a thermally grown oxide to passivate the front. Subsequently they defined small openings in the oxide by lithography for the metal-contacts to the n-layer. They demonstrated 19% with this device design and called it *Passivated Emitter Solar Cell* (PESC) [47].

Also in 1984, the team at the University of Stanford presented a cell design with both contact polarities on the rear side. Compared to an earlier cell type with interdigitated back contact (IBCs) that covered almost the full rear side [48], their device used SiO_2 passivation on both sides whereas the metallic contacts were confined to very small regions, so-called *point contacts*, using local diffusion of the two polarities through openings in the oxide passivation. With this design they achieved a very high V_{oc} of 685 mV under one sun [38]. The result clearly underlined the fundamental importance of surface passivation, and the need to minimize the area of the metallic contacts, but the team noted the difficulty of connecting the small point contacts to an external electrode.

Two years later in 1986, UNSW added a light trapping texture at the front and a back surface field (BSF) to their device [40]. Eventually they adopted localized contacts also for the rear in a cell design called PERC for *Passivated emitter and rear cell* [49], and they improved it further with locally diffused regions in the PERL cell [24].

Even though limited by recombination at its full-area rear metallisation, the BSF design was applied very successfully to low-cost, multicrystalline material and eventually became the standard for industrial production. In 2014, it was still applied in about 90% of all silicon solar cells and thus in about 80% of the global photovoltaic production. In a parallel development the cost for Cz-wafers was reduced by advances in ingot pulling and wafer sawing, such that by 2020 most cells were manufactured with mono-crystalline Si. Accordingly, recombination at grain boundaries was no longer the main limitation, but recombination at the full-area rear contact became a major bottleneck. Thanks to advances in laser-technology for the opening of local contact areas in the rear passivation, by 2021 the PERC[1] architecture with its localized rear contacts had become the new standard in industrial production, 35 years after its first presentation.

Even though the basic design did not change, there were nevertheless countless refinements to the process. The following sections outline the major fabrication steps that are used in manufacturing lines.

[1] Due to the formation of a highly doped region by Al-alloying in the openings, it was actually a PERL cell, but industy adopted the name PERC.

Starting from wafers, the fabrication of standard solar cells involves several steps:

1: saw damage etch (SDE)

2: wafer cleaning

3: emitter diffusion, at e.g. 900 °C

4: edge isolation

5: removal of the silicate glass that contains the dopants

6: anti-reflection coating (ARC) by PECVD

6': inserted here for PERC cells: rear passivation and laser opening

6": inserted here for TOPCon cells: oxidation and n-layer deposition at the rear, 2nd annealing

7: screen printing of the contacts and drying

8: Co-firing of the contacts at 800 - 900 °C

8': inserted here for the latest generation of PERC cells and as standard in TOPCon cells: laser enhanced contact optimization (LECO)

9: measuring and sorting

The individual steps are explained in more detail in the following sections:

8.1 Base material

Standard c-Si solar cells used p-type wafers until the early 2020ies. To some extent the choice of p-type material was historic because of the thicker cells used in early PV production. Since electrons are minority carriers in p-type material, their higher mobility could better ensure transport across thick wafers. For silicon with typical doping concentrations of $1 \times 10^{16}\,\mathrm{cm}^{-3}$ $D_n \approx 36\,\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ whereas $D_p \approx 12\,\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$. Another historic reason is that most of those cells were used to power the first satellites and p-type material turned out to be more tolerant against particle radiation in earth orbit as discussed on page 36. Additionally, phosphorus diffusion is easier and works at lower temperatures than boron diffusion.

As boron in p-doped wafers forms a variety of defect complexes such as B-Fe and B-O (c.f. the examples discussed in section 5.4), by 2020 its was increasingly replaced by Ga. With the change from PERC to TOPCon solar cells shortly thereafter, the most common base material has become phosphorus doped n-type silicon

8.2 Etching and cleaning

The purpose of this production step is to remove saw damage and to create a light-scattering surface texture. Saw damage is usually removed by etching 10 - 15 µm of the material. The texture etch depends on the wafer material.

1. **alkaline etching**:

- very directional, i.e. anisotropic.[2]
- NaOH or KOH (high concentration) for saw-damage removal
- (111) random pyramids are etched from (100) wafers with NaOH or KOH, or alternatively K_2CO_3 or Na_2CO_3
- Owing to its anisotropy, alkaline etching is primarily used for mono-crystalline wafers

2. **acidic etching**
   - isotropic etch that yields hemispherical pits, usually starting from defects such as a cracks of the sawing process
   - uniform etch of mc-Si material, i.e. the grain boundaries "disappear"
   - hydrofluoric acid (HF); etching of silicon-oxide, not silicon (deep etching only in combination with oxide-forming chemicals)
3. **plasma etching**: dry etching method thus less cost intensive (but not common)

With different etching agents different textures can be obtained, which is clearly visible in *reflectance measurements*. The etching process happens either by batch or by in-line processing which means that either a certain number of wafers is loaded in a carrier and immersed in a bath at the same time, or that the wafers are transported through the tool on a conveyor belt and they are processed while they are moving.

After the etching steps the surface has to be cleaned once more. This is usually done with hydrochloric (HCl) and hydrofluoric acid (HF), followed by a rinsing step in deionized water. After drying, the wafers are ready for diffusion.

8.3 Diffusion of the *p*-*n*-junction

The main goal of the diffusion step is to form the junction and a highly doped region for lateral current collection between the contact-fingers. Some key parameters are:

1. Ideally, the surface donor concentration does not exceed the solubility limit of phosphorus in silicon which is $N_s \approx 1$ - $3 \times 10^{20}\,\mathrm{cm}^{-3}$ at typical diffusion temperatures 800 - 850 °C [50]. However, the diffusion process can force a shallow region at the surface into supersaturation which means that the actual phosphorus concentration exceeds the solubility limit. Only the soluble part of the phosphorus concentration can form dopants on substitutional sites whereas excess phosphorus atoms form clusters that act as recombination centres.

2. Junction depth: $> 200\,\mathrm{nm}$. This is an important parameter since the silver metallisation spikes to a certain depth into the $n-$doped region to make contact, but it should not short-circuit the junction.

3. The emitter sheet resistance should be $R_{sheet} \approx 65 - 130\,\Omega_{\mathrm{sq}}$.[3] Historically, lower sheet-resistances and higher doping concentrations were used for the n-doped layer. In recent cell designs the trend goes towards shallower doping with higher sheet resistance in order to reduce free carrier absorption and recombination by the Auger-effect within the doped region as well as at the interface to the passivating dielectric.

[2] This means that certain crystallographic planes are etched more strongly than others.

[3] Can be measured via the transfer length method (TLM) since the underlying junction shields the contribution from the wafer.

Just as for the etching steps, for the diffusion there is either *batch processing* or *in-line processing*. In batch mode, several thousand wafers can be processed at the same time by exposure to an atmosphere of $POCl_3$ and O_2 in a tube furnace. Typical in-line processes load the wafers on a conveyor belt and spray them with phosphoric acid (H_3PO_4) or some other P-containing compound. Subsequently, the belt runs through an IR heating zone. For this step it is crucial to obtain a homogeneous sheet resistance over the full wafer area to ensure uniform contact resistance to the metallisation.

The doping takes place through the growth of a phosphorus silicate glass (PSG) at the surface. Since the PSG can be seen as inexhaustible reservoir of phosphorus, simple diffusion theory predicts that the diffusion profile follows a complementary error function cerf(x). In reality, the diffusion coefficient of phosphorus depends on the density of vacancies and interstitials which are injected from the surface as silicon is oxidised into PSG, and additionally on the doping concentration, i.e. it is self-enhancing. The result is a flat plateau close to the surface, a steep decay called kink, and once more a flat profile called tail [51]. Phosphorus overcompensates the boron-doping of the wafer and eventually the tail of the diffusion profile forms the buried p-n-junction.

Apart from the formation of the junction, there is a positive side effect, since metallic impurities are gettered into the doped region or into the PSG at the surface. This is related to the fact that impurities can diffuse faster in highly doped regions and that they can pair to vacancies, resulting in a different saturation concentration from the lowly doped bulk. By choosing an adequate temperature profile, it is possible to segregate impurities at the surface and possibly to remove them by combining the PSG removal with a short etch of the wafer surface. Gettering is especially important for mc-Si, but the diffusion-temperature should not be too high, otherwise impurities may remain dissolved in the wafer rather than segregate to the front.

Typically, the PSG and the diffused regions form on all sides of the wafer. The PSG is usually removed by dipping into 5% HF for 2 minutes at RT. Afterwards, the wafer is rinsed with deionized water and dried. Depending on the technique, the P-doping of the rear side is later overcompensated by doping with Al, or a single-side etching has to be applied to the rear.

At any rate, the n- and p-regions have to be separated by applying an edge isolation.

8.4 Anti-reflective coating

The most widely used material in c-Si solar cells is SiN_x. Its purpose is not only to provide an anti-reflective coating (ARC) but also to passivate the wafer surface and thus to reduce the effective surface recombination velocity at the front surface. SiN_x can saturate some of the dangling bonds at the silicon surface, additionally it contains a positive fixed charge that repels holes, the minority carriers in the underlying n-layer. The third purpose is to create a hydrogen reservoir that releases hydrogen during the firing process to passivate defects at the interface and in the bulk of the wafer.

Silicon nitride is deposited either by *PECVD* in a parallel-plate configuration, or by *microwave PECVD* using a remote plasma. The key parameters of the SiN_x are its thickness and its refractive index n. Typical values are 70 nm and $n = 2.05$ - 2.1.

8.5 Rear passivation

The PERC design requires two additional process step at this point in the process sequence. The first step consists of applying a passivating layer stack of Al_2O_3 and SiN_x at the rear. Owing to the negative fixed charge in Al_2O_3, electrons which are the minority carriers in the p-type bulk

are repelled from the rear interface. The second step is the opening of small contact holes in the insulating layer stack, such that electric contacts can be established in the next step.

8.6 Contacting and screen printing

Electric contacts are achieved by screen-printing of different metallisation pastes: For the front, the paste contains silver-particles, glass-frit and an organic binder. The paste for the rear contains aluminium particles and the binder, possibly some glass-frit and sometimes some boron to increase the surface doping concentration for a lower contact resistance and the print extends only only over the contact openings. Usually, a pure silver-paste is printed locally after the aluminium to ease soldering of the busbars. Recent cell designs don't use full-area printing anymore, resulting in bi-facial cells.

The screen printing technique used for solar cells is basically the same as the one used for printing pictures on T-shirts. It uses a metallic mesh which is coated with a polymer except for the areas that should be printed. During printing, a squeegee is pushed over the mesh and squeezes the paste through the openings.

One challenge of screen printing is to obtain fingers with a high aspect ratio, i.e. as high as possible (typically 10 - 20 µm) and as narrow as possible (typically 40 - 70 µm). This is done to ensure both, low surface coverage (low shadowing losses) and low line resistance. To this end, the rheology of the screen-printing pastes and the screen design are essential.

Note that the optimum number of fingers and their width is not necessarily decided by best electric performance but by cost. A producer may opt to use finer fingers and to compensate the loss in fill factor by lower module costs due to lower silver usage.

8.7 Co-firing process

The temperature profile for the firing is not straightforward and has to be designed very carefully since this step combines more than one purpose:

8.7.1 Formation of the front contact

During the firing process, the silver particles in the paste sinter together and yield conductive lines (approximately a factor of three below bulk conductivity). At the same time, the glass-frit melts through the SiN_x anti-reflection coating. The melt eventually dissolves some of the silver contained in the paste and some of the silicon of the wafer surface. In the first phase of cooling, excess silicon and silver are expelled from the melt. Silicon solidifies onto the n-doped region and silver nucleates at random sites. During further cooling, the glass-frit solidifies, leaving behind a thin insulating layer of glass between the epitaxially regrown silicon and the silver crystallites at the wafer surface, and the sintered silver line.

The dimensions of the silver crystallites are in the order of 100 - 500 nm. As they crystallize in an epitaxial relation with the silicon wafer, they form a perfect metal-semiconductor interface.[4] Thus, the contact from the n-doped region to the Ag-crystallites may be perfect, but only about 1 out of 10'000 crystallites extends to the sintered line and contributes to the current flow into the contact finger [52].

[4]Even at a perfect interface there are electronic defects since a continuous density of states is in contact with a bandgap.

The doping profile of the n-doped region, the quantity of glass frit, and the firing temperature must be carefully adapted. Ideally, the crystallites should penetrate to a depth of about 100 nm, but if the n-doped region is too thin, the Ag contacts might spike through and form shunts with the p-doped wafer underneath. This can drastically decrease the V_{oc} by creating a shunt-path with low parallel resistance.

8.7.2 Formation of the back-surface field

At the rear, no glass-frit is required since the Al-paste is in direct contact with the p-doped wafer and because Al melts already at 660 °C. At the liquid/solid interface, Si from the wafer dissolves in liquid Al until the solubility limit of Si in Al is reached. Depending on on the thickness of the screen-printed Al-paste and the kinetics of the process, this can produce a liquid region of 3 - 10 µm. Upon cooling, the solubility of Si in Al is reduced and Si gets expelled from the melt. The expelled Si grows epitaxially to the wafer and Al atoms get incorporated into the growing layer up to the saturation concentration of Al in solid silicon. Since Al is a p-type dopant, a thick p^+-region with doping concentration of 3 - 5 cm^{-3} is formed. It is often claimed that the band-bending between the wafer and the highly doped region forms an electrostatic field (the back surface field, BSF) that repels electrons away from the rear contact. In reality, this electrostatic field is exactly compensated by the concentration gradient, resulting in a flat Fermi-level. Under operating conditions, the QFLs split, and they remain essentially flat between the bulk and the highly doped BSF region. Consequently, n_{p^+}, the minority-density in the highly doped regions, is lower than n_p, the minority-density in the bulk. Thus, fewer minority carriers are available for recombination at the rear contact than in a design without highly doped BSF region. The effect of the BSF is discussed in more detail in appendix D.5.

8.7.3 Hydrogenation

Hydrogenation is achieved by release of hydrogen from the SiN_x:H layer during firing. About half of the hydrogen leaves into the air, the other half is transported to the interfaces and into the silicon wafer where it improves passivation.

8.8 Measuring and sorting

After edge isolation the cells' $j(V)$-curve is measured (usually by a flasher) to extract parameters like: J_{sc}, V_{oc}, FF, series resistance, V_{mpp}, j_{mpp}. With this information the cells are sorted according to their j_{MPP} values. This last step is crucial for the cell interconnection into modules. In order to maximize the power output of a module, the currents within a string have to be matched as well as possible. The reason is that the cell with the lowest current will limit the entire module as the current through all the cells has to be the same in every cell.

Chapter 9

Advanced cell processing

After introducing the PESC design and equipping it with light trapping and a BSF, the team at the UNSW further improved the efficiency of crystalline silicon solar cells. In 1989 they presented the passivated emitter and rear cell (PERC) which had a localised doping profile at the front (in modern parlance a "selective emitter") and an oxide-passivated rear with localised metal contacts. Finally, the in the PERL design presented in 1990 also the doping profile at the rear was localised (local BSF).

The transfer of localised rear contacts to large scale manufacturing was not trivial and required many developments in the early 2000s. By 2014, PERC cells had a market share of ca. 1%, but they were gaining rapidly; industrial versions of the PERC and the PERL concepts replaced the BSF solar cell by 2020. Let us look how the the main limitations of the BSF cell were addressed:

9.1 Optical improvements

Optically, the standard cell offered many routes for improvement.

- **Primary reflection** at the front may be as high as 15%. Alkaline etching of multicrystalline material gives only a marginal reduction of the reflection and results in clearly visible grains. The development of acidic etching reduced the reflection and gave the cells a more uniform appearance. For monocrystalline material, the primary reflection is massively reduced by alkaline etching since it produces random pyramids that reduce the reflection by a double rebound at the steep facets.

- **Shadowing losses** by broad grid lines can amount to as much as 8% at the front. Improved designs use narrow lines with high aspect ratio. High-aspect lines with steep sides can yield forward reflection, patterning of the line itself can be used to scatter the reflected light such that there is a chance for total internal reflection at the glass cover.

- **Parasitic absorption** in the doped layers at the front and at the rear is due to a high density of free carriers with low mobility. This is generally mitigated by lower doping concentrations. At the front, this can be achieved by a phosphorus diffusion with subsequent drive-in that reduces the surface concentration, or by reducing the doping concentrations selectively between the metallisation ("selective emitter").

- **Poor reflection of the rear contact**. The reflection of Al in direct contact to Si is only $\approx 80\%$; it can be improved by inserting a dielectric layer with lower refractive index. In the

original PERC concept, this was a layer of SiO_2, modern designs use a stack of Al_2O_3 and SiN_x. Even better reflection is obtained by replacing Al with Ag.

Further information about losses in real cells can be found in the literature[1].

9.2 Reduced recombination losses

The simplified process of the standard cell has many sources of recombination:

- **Recombination at the interface with the front dielectric** is due to interface states. For SiO_2, hydrogenation is successfully used to reduce the interface state density, alternative materials such as SiN_x contain positive fixed charge that repels minority carriers (holes in the n-doped region)

- **Recombination in the "dead layer"** of the phosphorus diffusion. When the phosphorus concentration exceeds the solubility in silicon, it forms inactive clusters that are very active recombination centres. Concentrations below the solubility limit are still critical because of a high level of Auger-recombination. The solution are shallower diffusion profiles.

- **Recombination at grain boundaries** of multicrystalline material can be mitigated by hydrogenation during the firing. It is also possible to reduce the number of detrimental grain boundaries with so-called "high-performance" (HP) material which is obtained by putting seed crystals into the crucible.

- **Recombination at bulk defect states** such as the B-Fe and the B-O complex can be mitigated by working with purer starting material. Fe-impurities can be gettered into the highly doped regions at the surfaces. O-impurities can be avoided by advanced Cz processes, or they can be deactivated with dedicated annealing procedures or temperature profiles. Alternatively, the boron doping concentration of the wafer can be reduced, or it can be replaced by gallium.

- **Recombination losses in the p^+ region of the BSF** follow the same physical principles as those in the n^+ region of the front, except that the BSF region is normally less doped and much deeper. In a pure Al-BSF process, doping is fixed by the saturation concentration and recombination can only be mitigated by reducing the depth of the Al-doped region. In case a very shallow BSF is used, it may be desirable to further *increase* the doping concentration. This is possible by adding B to the metallisation paste since it has a higher saturation concentration than Al.

- **Recombination losses at the Al-contact** can be mitigated by using a dielectric passivation layer with local openings (PERC or PERL concepts). In modern PERC designs, Al_2O_3 is used since it contains a negaive fixed charge that repels minority carriers (electrons in the adjacent p-type material).

The optical improvements obviously yield higher j_{sc}, locally enhanced doping close to the contacts improves *FF* because of lower series resistance losses. Reduced recombination losses lower the impact of the j_{02} and thus primarily improve the V_{oc}. Typical values are:

[1] Smith, "Toward the practical limits of silicon solar cells", IEEE Journal of Photovoltaics, vol. 4, no.6, pp. 1465–1469, 2014

$$J_{01} \sim 8.5 \cdot 10^{-13}\,\mathrm{A/cm^2} \rightarrow \sim 7.0 \cdot 10^{-13}\,\mathrm{A/cm^2} \Rightarrow +\,4\,\mathrm{mV}$$
$$J_{02} \sim 4.5 \cdot 10^{-11}\,\mathrm{A/cm^2} \rightarrow \sim 2.0 \cdot 10^{-11}\,\mathrm{A/cm^2} \Rightarrow +\,4\,\mathrm{mV}$$

Note that an improvement in one part of the cell may express a limitation elsewhere. Improving the front-side passivation only makes sense if the rear-side is good. Progress with the contacts of the cell won't yield an improvement if bulk recombination is too high.

9.3 Selective emitter

The term *selective emitter* refers to a specific emitter design with an increased doping beneath the metallic contacts (as shown in figure 9.1). Thus, majority carriers are collected with a low contact resistance, minority carriers are repelled, and shunting is prevented as the probability of silver penetrating through a more deeply doped region is reduced. In the remaining front-region, the doping can be reduced which will decrease $j_{0,e}$ due to lower Auger recombination. Figure D.3 in the appendix shows that this concept requires a low surface recombination velocity, thus the need for an excellent passivation layer such as SiN_x.

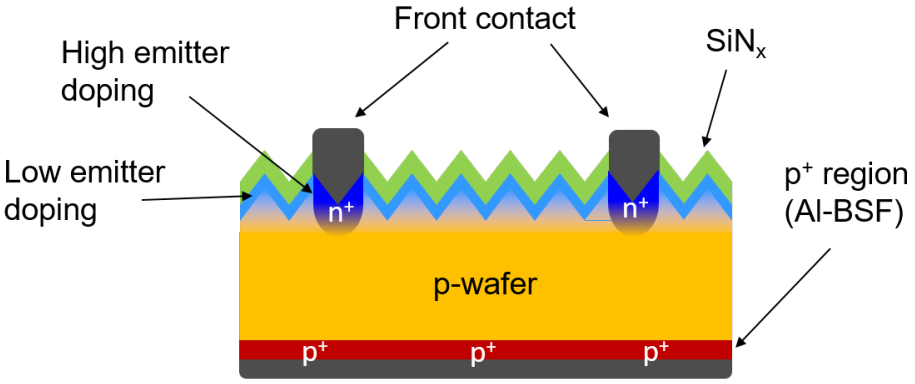


Figure 9.1: Sketch of a solar cell with selective emitters.

A simplified way to look at the emitter saturation current is to divide it into two contributions: a *metallised* and a *passivated* part (with A_m being fraction of the metallised area). The total $j_{0,e}$ is then simply given by the area-weighted average:

$$j_{0,e} = A_m j_{0,met} + (1 - A_m) j_{0,pass} \tag{9.1}$$

Obviously it is important to minimise the metallised area as much as possible in order to obtain low j_0 and thus high V_{oc}. At the same time, this improves j_{sc} over all of the spectrum by reduced shading and reduced Auger recombination, and additionally in the red region because of reduced free carrier absorption.

We are not free to reduce the doping concentration as much as we want because of a trade-off with the lateral sheet resistance in the lowly doped regions. For low sheet resistance the spacing of the finger-lines can be large, which will decrease the shadowing losses. High sheet resistance of a selective emitter requires smaller finger spacing. The Ohmic losses in the finger-lines scale with the square of the spacing a.[2] The *relative power loss* is given by:

[2]Note that increasing the spacing will increase the current that is flowing to the fingers. From the equation for the power $P = U \cdot I = R \cdot I^2$ it is obvious that increasing the finger spacing, and thus the current, will lead to an increase in dissipated power.

$$\Delta P/P \quad = \quad R_{\text{sheet}}\, a^2\, j_{\max}/12\, V_{\max} \tag{9.2}$$

One possibility to implement a selective emitter without increasing the complexity of the process uses a laser to remelt small areas of silicon and the covering phosphorus silicate glass at the front of the wafer [53]. This dissolves additional phosphorus atoms into the molten region and subsequently the phosphorus glass is etched as usual. The process requires only one additional step and it is easy to implement into a production line.

Alternatively, the a thin dielectric buffer layer can be deposited on the full front surface and patterned with a laser. During the subsequent emitter diffusion, the dielectric layer retards the diffusion of phosphorus into the c-Si bulk except for those regions where the dielectric layer has been opened locally by the laser.

9.4 Passivated rear surface with local contacts

Various industrial methods have been proposed to manufacture localised rear contacts. The most common covers the rear-side with a passivation layer and a laser is used to open this layer locally. Then, Al-paste is screen-printed over the openings. Finally, the Al is driven into the wafer with a regular firing step. At first sight one would expect that melting and ablation deteriorate the interface, but the process is surprisingly insensitive to laser damage since the formation of the local BSF melts the region beneath the opening once more to a depth of several µm and then regrows it epitaxially [54].

Figure 9.2 illustrates the gains of a PERC cell with respect to a BSF cell when all other process steps are the same. Due to the insertion of Al_2O_3 and Sin_x between silicon and Al, parasitic absorption is massively reduced. At the same time, reduced recombination at the rear yields a better collection, resulting in a gain of $1.9\,\text{mA}\,\text{cm}^{-2}$. As a result of the improved passivation and the higher excess carrier density, the V_{oc} improves by almost 20 mV [55].

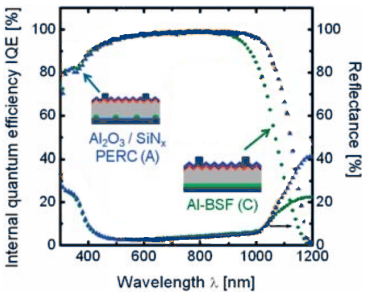


Figure 9.2: EQE and reflection of BSF and PERC solar cells, after Gatz, PSS (2011).

Usually the different approaches are summarised under the name "PERC", even though comparison with the original cell designations would make them "PERL" devices because of their localized doping profiles [56]. Note that it has become common to cover only the laser-opened regions with the rear metallization, thus resulting in bi-facial cells that collect stray light reflected from the ground or adjacent modules. Additionally, the issue of wafer bowing is considerably reduced. Bowing results from strain that builds up by different thermal expansion coefficients of Al and Si when thin cells with full-area Al printing are cooled down from the firing.

9.5 Metallization processes

An important part of each cell are the contacts that are used to extract the current from the solar cell. For the standard c-Si technology a metal grid is created at the front surface, which can be soldered when assembling cells into a module. There are two methods that are widely used at the moment: *screen printing* and *plating* (electro or electro-less, e.g. light induced plating).

Screen printing uses printing pastes that consist of metallic particles and organic binders. The front metallisation requires an additional ingredient to etch through the SiN_x layer. This is achieved by adding glass frit, finely ground particles of glass with low melting point such as lead-glasses or bismuth-glasses. Printing is carried out by pressing the paste through openings is a screen. After that, the pastes are dried and cured or fired, typically by passing through a belt furnace. Typical dimensions for printed contact finger lines are 80 - 120 µm.

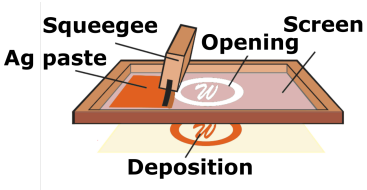


Figure 9.3: Screen printing method

- The use of silver represents a significant part of the total costs. Note that it can be economically interesting to sacrifice efficiency by reducing the use of silver, thus ending up with a lower cost/W ratio.
- Low temperature Ag-pastes are normally still porous after curing, thus their resistivity is a factor of 3 to 5 higher than bulk Ag [57]. Fire-through Ag-pastes contain varying amounts of glass-frit, after firing their resistivity is 5 to 10 times higher than bulk Ag [58].
- As the Al-paste is directly applied to the *p*-layer, it does not need any glass-frit. Since Al and Si do not mix in their solid form, the fired films consist of phase-separated conglomerates. Normally Al cannot be soldered because of its surface oxide, therefore a second printing process with a solderable material is required.
- Instead of glass-frit, the Al-paste may contain Si powder to prevent void-formation in the local contacts of PERC cells. The voids are formed as follows: When a Si-free Al-paste becomes liquid, it dissolves silicon until it reaches saturation, and all of this silicon has to flow through the small contact opening. This is not an issue at the peak temperature, but during cool-down the silicon is grown epitaxially from the liquid and Al gets expelled through the contact opening. If Si cannot flow fast enough from the liquid into the opening, epitaxial growth stops and a void is formed. If the paste contains Si powder, the Si-saturation in the liquid Al is reached earlier and less Si is dissolved from the wafer and the BSF in the contact opening is shallower. Consequently, less Si has to flow back into the opening.

Plating is an electrochemical process where the oxidation state of a metal in solution is reduced by flowing a current through the solution. It requires a conductive seed layer which may typically be 30 - 40 nm. Plating can enlarge the dimensions to 20 - 80 µm.

- Plating yields pure and dense finger material with conductivity close to bulk characteristics.
- Possible with Ni and Cu. Note that Cu is a low-cost and abundant alternative to Ag,[3] but it is a *fast diffusor* even at room temperature and an active recombination centre. Thus, a Cu diffusion barrier is needed underneath the metallisation.

The advantages of the *plating method*:

- narrower lines → less shading (5-7% down to 2-3%)
- seed layer can be optimized to yield low contact resistance

[3]The price for copper is in the order of 2 CHF/kg, whereas silver is 400 times as expensive 800 CHF/kg (2012).

- pure, dense material for the top layer / fingers → lower series resistance
- copper could be used ⇒ reduction of Ag consumption ⇒ cheaper

Plating can reduce the combined losses of shadowing and series resistance from 10 - 11% to 7 - 9%. Nevertheless, the lower conductivity of printed lines is often accepted because of lower manufacturing-cost, and there is still a steep learning curve for screen printing towards printing thinner fingers with a higher aspect ratio.[4]

Besides screen printing and plating, there are advanced metallization techniques like:

- aerosol printing (finest)
- inkjet
- pad printing

9.6 Other concepts

- **Laser grooved, buried contact solar cell**: Laser scribing is used to create deep gooves which are filled by plating. This yields very narrow lines with high aspect ratio. Shadowing is thus reduced whereas the deep grooves can still contain a sufficient volume of the metal.
- **Metallisation wrap through (MWT) cell**: By drilling holes through the wafer at several points with a laser and guiding the contact from the front to the rear, the shadowing can be reduced by 2 - 3%. However there are still issues with *shunting* at the rear-side where the p-doped and n-doped regions should not be connected. Compared to standard technology the estimated maximal gain is $\approx 0.5\%$.
- **Emitter wrap through (EWT) cell**: For this structure no metal grid is required at the front and thus no shadowing losses are present. The estimated gain is $> 1\%$. But just like for the MWT cell, there are shunting issues at the rear. Another aspect is the need for several thousand holes that have to be drilled to feed the emitter through to the rear.

9.7 Conclusion

It is imaginable to combine various technologies and concepts (e.g. cell geometries, advanced passivation schemes and metallization etc.), but the question remains whether it is feasible and economically viable. It is interesting to show high performance in the lab, but if it is not possible to transfer a technology to industry in a cost-competitive way, it might be better to invest in other technologies.

Research and development of c-Si solar cells is ongoing and production costs will continue to go down. The practical limit for mono c-Si cells is estimated around 27%, the latest demonstrated record efficiency is at 27.1% (held by Longi), the theoretical limit is at 29.5% [5, 6].

[4] For example, finger width of $\sim 80\,\mu m$ with ASADA meshes.

Chapter 10

Very high efficiency silicon cells

In order to obtain efficiencies of 20% and higher for c-Si cells, various advanced concepts have to be combined. Ideally, the starting material is single-crystal silicon with high lifetime, the front contact is made in the form of a selective emitter, and surface passivation is applied to the rear. The contacts on the front are narrow and with high aspect ratio, often obtained by plating, or both contacts are located at the rear. High efficiency cells are not only pursued for scientific curiosity, but they do have a market because they require less mounting space for a desired system power, and because usually they maintain good performance also at high temperatures.[1]

There are the following three different approaches in this segment.

1. SunPower commercialised cells with *Interdigitated Back-Contacts* (IBC) where both contacts are located on the rear. This means that the front surface can be optimised fully for optical performance and the resulting cells yield very high current density. The contact strategy is not disclosed; likely they use passivating contacts (c.f. item 3 of this list).

2. The Japanese company Sanyo (now merged into Panasonic) developed *Silicon Heterojunction Cells* under the brand name HIT. They were able to claim the 25 year old world record efficiency of the PERL cell in 2015 [59]. In 2022, the Chinese company Longi claimed the world record efficiency for Si solar cells with 26.8%, also using heterojunction technology.

3. The tunnel contact briefly mentioned at the begin of chapter 8 was investigated by various research groups because it can mitigate recombination of the metallic contact. Currently, the highest efficiencies are held by the Fraunhofer ISE with 26.0% in a both-side contacted device and by the ISFH with 26.1% in a IBC cell.

With the exception of the very last one, all of these cells are based on n-type wafers.

10.1 Interdigitated back-contacted cells

The concept of contacting the cells at the rear-side seems obvious since the shadowing losses are completely eliminated and the serial connection of cells to form a module is simplified as there is no need to wrap the interconnection ribbons from the front of one cell to the rear of the next one. However, there are also drawbacks like the complex production process where a high precision is needed to align the wafers during all steps of the fabrication. Accidentally contacting both the

[1] The temperature coefficient decreases with increasing V_{oc} and as high efficiency cells usually have high open circuit voltage values, they perform better even if the temperature is increased.

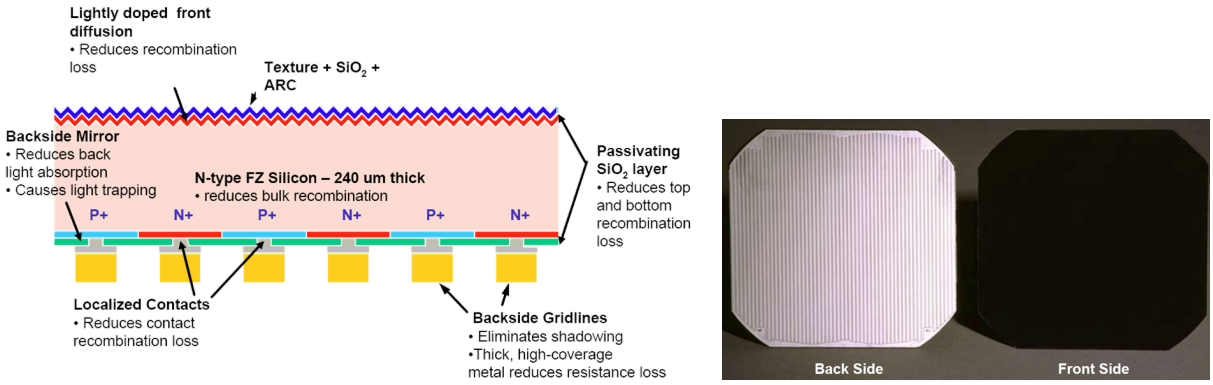


Figure 10.1: IBC cell schematic and pictures of a cell showing the absence of front metallization

n- and the p-doped parts due to misalignment leads to short circuits. Furthermore the lifetime ($\sim 1\,\mathrm{ms}$) and diffusion length ($\sim 1\,\mathrm{mm}$) of the minority carriers have to be as good as possible as most of them are created at the front of the cell whereas they are all collected at the rear-side. This leads to resistive losses. Furthermore the carriers have to be able to diffuse laterally, given the fact that they are not collected on the entire rear surface.

The front of the cell consists of an n^+ doped layer introduced by diffusion ("front surface field" (FSF) or "floating emitter") and e.g. a SiO_x/SiN_x layer stack simultaneously acting as an anti-reflective coating and a passivation layer. Note that the doping level of the n^+ layer should not be too high as to avoid Auger recombination. At the rear-side, p-doped and n-doped regions are applied in the form of two combs that stick into each other without touching. For an n-type wafer, electrons can flow as majority carriers in the wafer and require thus only small contacts. The p-doped regions have to be larger in order to collect the minority carriers more efficiently.

An interesting feature of the IBC cells is the fact that the p- and n-regions on the rear-side form a pn-junction. This junction can act as integrated by-pass diode. A by-pass can save the actual cell or module when one of the cells is shadowed. In this case it acts as resistive load to the other cells in the string. Since it dissipates a lot of energy, the shadowed cell can heat until failure. The breakdown voltage of the by-pass is lower than the one of the actual cell and thus the current will be deviated through the by-pass, sparing the cell.

10.2 Silicon heterojunction solar cells

There is a common issue with the cell designs in the previous two chapters and also the one discussed in the previous section; in all of them the presence of a metal on the surface forces a unique Fermi-level. i.e. the surface recombination velocity is virtually infinte. The large diffusion length of silicon transports this problem into the interior of the solar cell and thus limits the V_{oc}.

Silicon heterojunction (SHJ) solar cells follow a radically different philosophy; they move the metallic contact away from the wafer by inserting layers of doped amorphous silicon at front and rear. The low mobility in doped amorphous silicon is just sufficient to extract majority carriers, but it allows the minority Fermi-level to split. Researchers at Sanyo in Japan pioneered this technology in the mid 80ies but they noticed that the interface between doped amorphous silicon and the wafer is very defective. By 1991 they found that the insertion of an undoped layer between the wafer and the p-doped layer at the front boosted their efficiencies by reducing recombination losses at the

wafer surface, thus the name "HIT" for Heterojunction with Intrinsic Thin layer.

The use of amorphous silicon restricts the process to temperatures below 250 °C. This prevents the use of all the work-arounds applied to low-qualtiy wafers such as gettering during diffusion, hydrogenation during firing, or deactivation of thermal donors. Consequently, heterojunction technology was always applied to n-doped wafers and it is not applicable to multicrystalline material. The manufacturing processes for a cell as shown in figure 10.2 is described below.

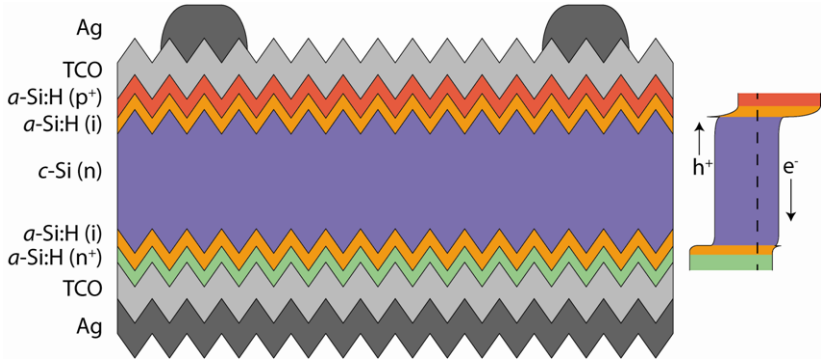


Figure 10.2: Sketch of the layers (note: not drawn to scale!) and the band diagram of a SHJ solar cell.

10.2.1 Wafer cleaning and texturing

Typically the wafer surface is cleaned in three steps. The first step is (1) a wet *chemical oxidation* (e.g. with sulfuric acid), i.e. a thin layer of the wafer oxidizes and contaminants at the surface are encapsulated. This step is followed by (2) a *rinsing* step with de-ionized water to remove remnants of the acid. The final step (3) *removes the oxide including the contaminants* created by step 1 through an etch with hydrofluoric acid (HF).

After cleaning the wafers are textured. For this e.g. a 200 - 300 µm, $< 100 >$ wafer is etched with alkaline solution (typically KOH) which etches selectively, i.e. it attacks the different crystallographic planes at different rates. This creates *random pyramids* with $< 111 >$ facets. These pyramids help to efficiently confine the light in the bulk of the wafer.

10.2.2 Deposition of amorphous silicon layers

In order to passivate the wafer and to create the doped regions at the front and at the rear, a method called Plasma Enhanced Chemical Vapour Deposition (PECVD) is used. In contrast to standard CVD this process dissociates gas molecules with the help of a plasma rather than by high temperatures. The reactors typically used for this process consist of two parallel capacitor plates in between which the plasma is ignited by applying an AC power of 10 - 100 MHz (see figure 10.3). In the alternating field electrons are accelerated to high enough energies to be able to crack the gas molecules by collision. The resulting gas radicals reach the substrate's surface where they adsorb and start forming the layer. Thus, double-layer stacks of *ip* and *in* are deposited at the front- and on the rear-side of the wafer, respectively. Each layer has typically a thickness around 10 nm measured on a flat wafer, corresponding to 6 nm on the side of a pyramid.

The precursor gases that are used for the deposition of the hydrogenated intrinsic amorphous silicon passivation layer ((i) a-Si:H) are silane (SiH_4) and hydrogen (H_2). For the doped layers

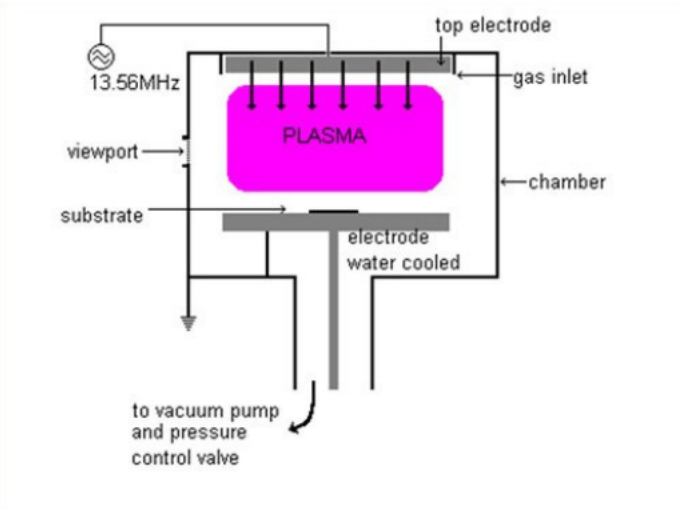


Figure 10.3: Sketch of a PECVD reactor.

either phosphine (PH_3, for n-type) or tetramethylboron (TMB, for p-type) are added to the gas mixture. The whole process is done at temperatures around 200 °C and deposition rates of up to several nm/s can be reached. For the deposition in a single chamber reactor, cross contamination is a big issue and has to be prevented. A cluster tool with dedicated chambers for intrinsic, n-type and p-type layers avoids this problem almost completely.

10.2.3 Heterojunction contacts

The metal-semiconductor contacts are critical as they are prone to be very defectuous. In the SHJ structure the metallic contacts are completely decoupled from the c-Si wafer, and the insertion of a passivating i-layer reduces the recombination even further. This means that j_0 does not depend on the metallization, only on the passivation quality of the amorphous layers. Another advantage is the fact, that a-Si:H can easily be doped either n- or p-type in order to form the carrier-selectivity of the contacts.

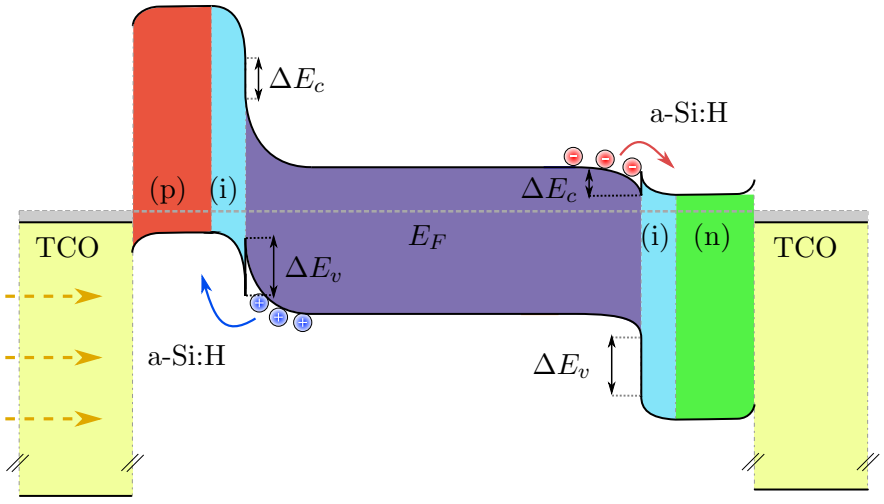


Figure 10.4: Schematic band diagram of a standard silicon heterojunction (arbitrary scale).

Figure 10.4 shows that there are band-offsets ΔE_v and ΔE_c between the a-Si and c-Si due to the different band gaps of a-Si ($\sim$1.7 eV) and c-Si (1.1 eV). The size of these barriers depends on the electron affinities and the bandgaps[2] of the two materials that make up the heterojunction.

Normally the height of barriers would impede the transport of all carriers from the wafer to the amorphous layers, but the width of the barriers can still be manipulated by adequate doping. Figure 10.4 illustrates that holes can still pass the narrow barrier on the left and electrons the one on right. In contrast, electrons have a high barrier on the left and holes have a high barrier on the right.

10.2.4 Surface passivation

The insertion of an undoped intrinsic layer is a crucial part of the device since intrinsic amorphous silicon is not as defective as the doped layers.[3] It has thus the ability to enhance the passivation of the a-Si - c-Si interface. The thickness of the i-layer has to be chosen very carefully:

- **thicker i-layer**: better passivation **but** more absorption and blocked transport
- **thinner i-layer**: less absorption and better transport **but** worse passivation

The trade-off is made more complicated by parasitic absorption in the amorphous layes at the front. Due to a high absorption coefficient α of a-Si, light with short wavelengths is absorbed in the front a-Si layers, but it does not contribute to the photo-current. Hence it is important to keep these layers as thin or as transparent as possible. This can be achieved e.g. by using materials with a wider band gap (a-SiOx, a-SiC) or with lower absorption coefficient (μc-Si).

10.2.5 Loss contributions

We distinguish *optical* and *electric losses*. The former include the reflection at the front side, shading by the front contacts as well as parasitic absorption by the a-Si layers, the TCO and the back-contact (concerning IR). Losses due to recombination and series resistance on the other hand are referred to as electric losses.

So in order to improve the J_{sc} the optical losses have to be reduced e.g. by:

- improved wafer texture
- reduced shadowing but maintain good collection properties (compromise)
- lower parasitic absorption (in a-Si, TCO and rear-reflector)

Improving the electrical losses will especially help to increase the V_{oc}. The possibilities are the following:

- Higher wafer quality (fewer impurities)
- cleaner c-Si surface

[2]The work function Φ_{sc} is the energy that is needed to extract an electron from the Fermi level E_{F} of the semiconductor and put it to the vacuum level E_{∞}, the electron affinity χ is the energy difference of the conduction-band edge and the vacuum level. A third quantity is the ionisation potential (IP) which is the energy-difference between the valence-band edge and the vacuum level.

[3]Defects are an intrinsic and unavoidable property of amorphous silicon. Since the formation enthalpy of defect-states depends on the Fermi-level position, doped material contains more defects than intrinsic material.

- best passivation
- reduction of plasma or thermal damage introduced during fabrication (a-Si, TCO sputtering and contacts)
- optimization of band offsets to improve carrier-transport

10.2.6 Advantages of silicon heterojunction solar cells

Some advantages of SHJ solar cells are listed below.

1. **High T performance**: In terms of *temperature coefficient* (-%/°C) SHJ cells, due to their high V_{oc} are comparable to a-Si cells.

   a-Si: -0.2%/°C
   std. c-Si: -0.5%/°C
   SHJ: -0.3 - -0.2%/°C

2. **Towards thinner wafers**: Decreasing the thickness of the device will *increase the* V_{oc} [4] but *decrease the* j_{sc}. When a good light management is applied the two effects balance over wide range of wafer thicknesses. Since all processes are carried out at low temperature, wafer-bowing is not an issue and it is possible to have thin, high efficiency cells that cost less due to lower material consumption.

3. **Bifacial cells and modules**: The SHJ structure allows for bifacial cells, i.e. cells that can be illuminated from both sides.

10.2.7 Temperature behaviour of heterojunctions

Heterojunctions generally exhibit potential barriers caused by *band offsets* at the interface between two materials of different bandgaps. The charges must be transferred through these barriers in order to be collected, and the crossing is done by tunnelling or thermionic emission. These two phenomena are both enhanced at high temperature meaning that the charge transport is strongly temperature dependant. The main effect is on *FF*.

At high temperature, a heterojunction cell will behave similarly to a homojunction, but with a lower temperature coefficient due to its high V_{oc}. However at low temperature, the transport is inhibited at the heterointerfaces and the *FF* drops, pulling the efficiency down as shown in figure 10.5.

Finally, the overall temperature coefficient is dominated by the V_{oc} at high temperatures (above 25 °C) and by the *FF* at low temperatures (below 10 °C). Note that the maximum of *FF* and efficiency can vary to higher or lower temperature according to the cell design.

10.2.8 Sputtering of the transparent conductive oxide

The conductivity of amorphous silicon generally very low, even with doping.[5] Consequently, lateral transport to the contact fingers must be ensured by a transparent conductive oxide (TCO). The

[4] This only true if the surface passivation is good enough to transform the generation rate G into a high excess carrier density.

[5] As mentioned in footnote 3, doping creates additional defects that capture free carriers. Thus, doping becomes self-limiting. The room-temperature conductivity of n- and p-type material hardly exceeds 10^{-2} S/cm, even for dopant concentrations in the range of several at%.

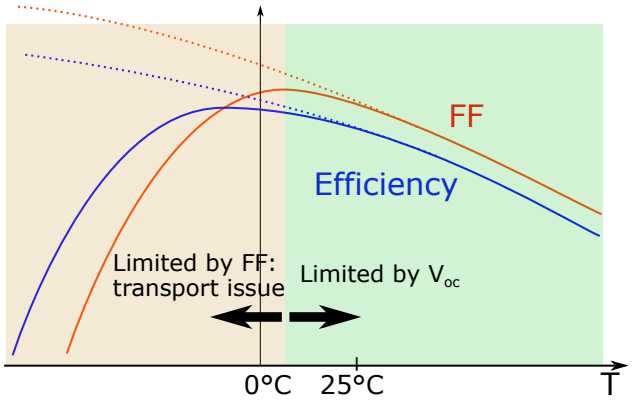


Figure 10.5: Temperature dependence of a typical silicon heterojunction solar cell (solid line) and a standard homojunction (dotted line.)

most commonly used TCO for SHJ solar cells is indium tin oxide (ITO), despite years of intensive research on alternative TCOs. TCOs are deposited onto both the front and back of the cell. At the front, it improves lateral conduction for the carriers to be able to reach the contact fingers and also acts (with an optimised thickness for interference) as an anti-reflective coating. At the rear of the SHJ cells, TCOs serve to minimize parasitic absorption from the rear metal reflector and as electrical contact.

The method used to deposit these layers is called *sputter deposition* (see figure 10.6) and uses an RF or DC plasma. One of the electrodes is the TCO target, the other one is the plate that holds the substrate. Typically the chamber is pumped down to low pressures and then partially refilled with a mixture of argon and oxygen. The the plasma is sustained by impact ionisation between electrons and gas particles. Ionised particles are accelerated towards the target, sputtering particles from the surface. Generally they rest neutral and traverse the plasma towards the substrate. The properties of the TCO are manipulated by the deposition temperature, gas pressure, gas flow, gas composition (typically the oxygen content in Ar) and deposition time.

As the name suggests, TCOs are transparent *and* conductive. Compared to metal, TCOS are 10 - 100 times less conductive, i.e. their sheet resistances are usually in the order of 40 - 100 $\Omega_\square$ for a 100 nm layer. There is a trade-off between conductivity and transparency; σ depends on the free carrier density, electron mobility and electron charge:

$$\sigma = e \cdot n \cdot \mu \tag{10.1}$$

Increasing the free carrier density leads to higher conductivity (lower sheet resistance) but it also generates higher free carrier absorption (FCA), which renders the material less transparent in the IR. In this regard, it is better to increase the electron mobility rather than the carrier density. To increase the mobility, several approaches are possible, some of them are: passivating trap states or defects states, structural relaxation of the atomic networks by curing, increasing the size of the TCO crystallites, etc. For example, cerium-doped indium oxide (ICO) can reach mobilities around 120 cm^2/Vs compared to around 30 cm^2/Vs for ITO.

10.2.9 Metallisation

Silver is used for the metallization of the cell because it has high conductivity. The back contact is usually created by covering the entire surface with about 100 nm sputtered silver. This layer

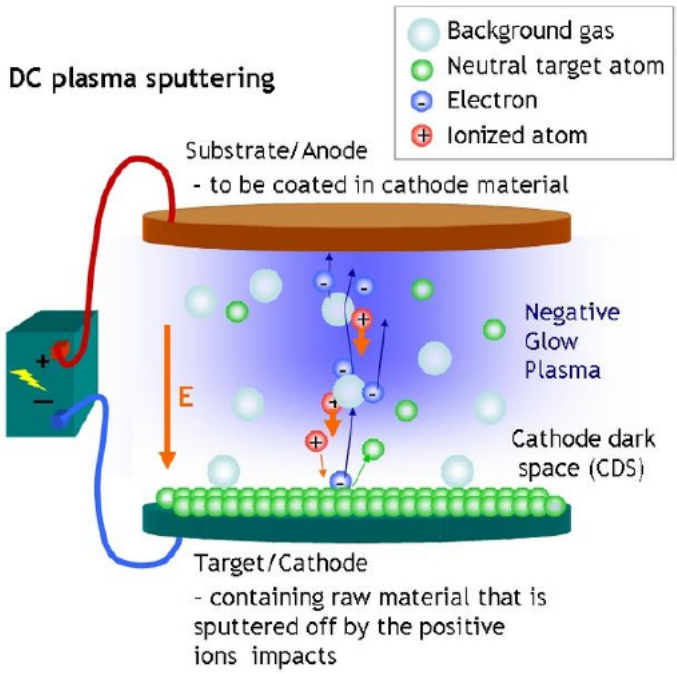


Figure 10.6: Sketch of the sputtering deposition method.

not only provides the rear contact but it also acts as rear reflector for the IR radiation which is reflected back into the c-Si. In terms of device-optics, the rear TCO works by reducing the parasitic absorption of the interface between c-Si to the metal reflector. This is similar to optical effec of the dielectric passivation layer in the PERC design.

The front contacts are created by *screen printing*. Different from the screen-printing pastes used for the metallisation of standard cells, SHJ cells require low temperature silver pastes which are cured at temperatures below 200 °C. Note that it is also possible to create a back contact by screen printing.

After this final step, the cell is ready to be characterised by measuring e.g. the $J-V$ curve to extract the information about the cell performance.

10.3 Tunnel-oxide passivating contacts

Silicon oxide provides excellent surface-passivation, but it is also an excellent insulator. It can still be used in contacts if it is thin enough for tunnelling. The beneficial effects of interfacial oxide layers were investigated for bipolar transistors because often a thin oxide film was formed before the growth of the n-type emitter and it was found that this oxide actually reduced the leakage current of minority carriers (holes) into the emitter during transistor operation [60]. Unfortunately current transport is also somewhat impeded for electrons, especially for for the current densities in transistor applications and eventually the interest of the transistor community moved to MOS technology which is still in use today. Nevertheless, current transport was acceptable for the lower current densitites of solar cells and there the effect of surface passivation was noticeable though high V_{oc}. For example, UNSW demonstrated $V_{oc} \approx 680\,\mathrm{mV}$ as early as 1981 for their MINP device architecture which separated the diffused emitter from the metal contact by a thin tunnel oxide [61].

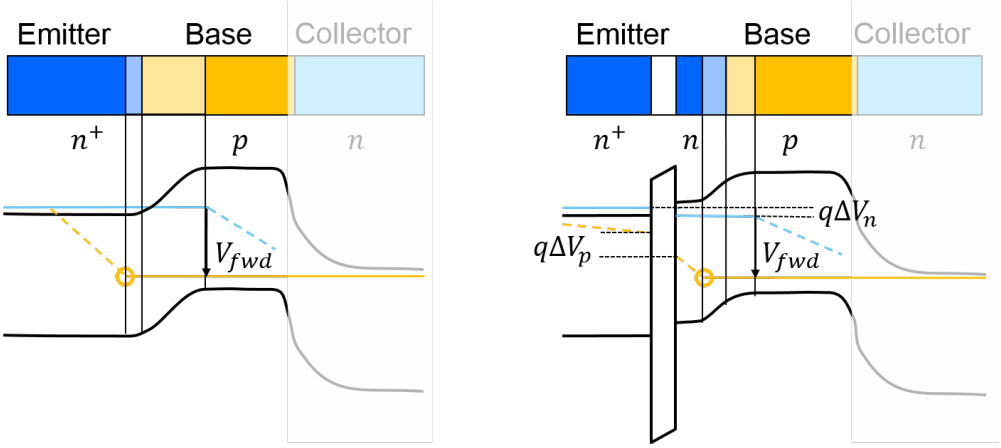


Figure 10.7: Emitter region of a bi-polar transistor without (left) and with interfacial oxide layer (right) under forward bias. The collector is shaded since it is not present in a solar cell.

The design of bipolar transistors was typically based on an n-type wafers which also served as collector. Next, an oxide mask was applied to the full surface, areas for the p-type base were opened by lithography, and the base was created by a deep boron diffusion. At this stage the oxide was introduced, either unintentionally by native oxidation in ambient air, or under somewhat more controlled conditions as part of the mask removal and the subsequent cleaning process, or even intentionally by thermal oxidation. Finally, small pads of a polycrystalline n^+ emitter were deposited on top of the base, often followed by an annealing step to activate the dopants in the layer and to create a shallow diffusion into the base.

If we ignore the collector and assume a very deep base, we have a junction that resembles very much the front of a standard p-type solar cell. This is the reason why the n-region of solar cells is often called emitter, even though in an illuminated solar cell it does not emit electrons into the base as in transistor operation, but it collects photo-generated electrons from there!

Figure 10.7 illustrates a bipolar transistors without and with interfacial oxide and and the corresponding band diagrams under forward bias. Current transport through thin barrier layers depends on a variety of parameters such as the width of the oxide, its barrier height, the carrier types and the densities that are available on both sides of the oxide, etc. For n^+-emitters it was favourable to diffuse a shallow n-region that defined electrons as majority carriers also at the surface of the wafer.[6]

Under the assumption of flat QFLs, the excess hole densities at $x = -w_n$ are given by $p = p_{n^+,0} \exp(qV_{fwd}/kT)$ and $p = p_{n,0} \exp(q(V_{fwd} - \Delta V)/kT)$ for the cases without and with tunnel oxide (circles in figure 10.7). In the latter case the oxide prevents or reduces the flow of holes into the emitter, exactly what we want in solar cells. This is illustrated by the fact that the voltage drop across the oxide is higher for holes ($q\Delta V_p$) than the one for electrons ($q\Delta V_n$).

In case of thick oxides tunnelling should no longer be possible, but researchers at Stanford University noted that the annealing can be used to break up the oxide at randomly distributed sites, thus permitting local current flow through pinholes rather than relying on tunnelling [62]. Eventually, these researchers funded the company SunPower which developed into a leading supplier of high-efficiency solar cells.

Public research on solar cells with thin interfacial oxides picked up again in 2014 when researchers

[6]Contacts without diffused region are discussed in appendix E.

at the Fraunhofer-ISE demonstrated efficiencies of 24.4% with tunnel-contacts [63]. Researchers at the ISFH used an IBC design based on the thermal breakup of thicker oxides and demonstrated even 26.1% in 2018 [64].

The contact strategy with silicon oxide is appealing since it is compatible with processes at high temperature such as diffusion and the firing of standard screen-printing pastes. Since this includes the temperature range of gettering and hydrogenation, these contacts could potentially be applied to p-type Cz wafers which is not possible for heterojunctions with amorphous silicon.

V_{oc} and FF loss

Let us look at the contact resistivity a bit more closely. In Si/SiO_2 tunnel contacts the thickness of the oxide is typically 1.5 nm and the theoretical barrier heights between Si and SiO_2 are around 3 eV for electrons and around 4.5 eV for holes. Tunnel contacts should thus work much better for electrons. In reality this is not the case, likely because very thin films result in effective barrier heights that are different from bulk values. Thus, resistivities for n-type as well as for p-type contacts were reported in a very similar range of 1 - 100 $m\Omega\,cm^2$ [65, 66]. For a full-area contacts this means that even in the worst case a voltage drop of 4 mV is sufficient to transport a current density of 40 $mA\,cm^{-2}$ which is a typical value for silicon solar cells.

To assess the impact on the FF, we use eq. (6.35) with typical values of $V_{oc} \approx 0.7\,V$ and once again $j_{sc} \approx 40\,mA\,cm^{-2}$. For the upper value of $R_s = 100\,m\Omega\,cm^2$ this yields a relative FF reduction of 0.4%, i.e. a FF of 80% would be reduced to 79.7%, illustrating why higher values of the series resistance should be avoided.

Chapter 11

Thin film technologies

In the early days of photovoltaics, c-Si solar cells were manufactured from electronic-grade silicon. The high cost of this approach triggered research into thin film materials. The idea of thin film solar cells is to deposit the active layers directly on inexpensive large area substrates and to use monolithic interconnection schemes to form modules between the different deposition steps. In a final step the raw module is contacted, encapsulated and a junction box is added.

Early on there were three technologies namely CIGS, CdTe and thin film silicon, and all of them were developed to the level of mass manufacturing. As regards thin film silicon, despite the demonstration of very low manufacturing cost of 30 - 50 cents per Watt, its aesthetically pleasing appearance, and even the option of semitransparent modules, this technology could not convince the customer due to limited efficiencies.

The term "thin film solar cells" is generally extended to emerging technologies and novel concepts such as *dye sensitized cells*, *organic solar cells*, *perovskite cells* or *quantum dot cells*.

The main advantages of thin films are first of all the low consumption of material (in the order of a few µm of semiconductor) and thus their potential for low cost production on both rigid and flexible substrates (e.g. glass or steel foil). Compared to standard c-Si cells, the energy intensive steps of ingot fabrication are not needed. Thus, thin film solar cells typically much less energy for their production. Besides, no special substrate treatment is needed, e.g. no wafering, and in total there are fewer production steps. Another point is the fact that the thin film development can benefit from other sectors like e.g. the flat panel industry.

Besides these advantages there are also drawbacks and challenges. Compared again to standard c-Si technology, until now they exhibited somewhat lower efficiencies, there is less field experience, i.e. it is not clear yet how stable the modules really are on the long term. Furthermore the production process is much more delicate as the thin layers are prone to shunting and the processing environment has to be as clean as possible. For this reason system development is expensive, i.e. a lot of money has to be invested before cells can actually be produced (high capital investment but then low manufacturing costs). One big challenge that remains is the up-scaling to industrial standards which is not trivial. Homogeneity in thickness, doping and material structure across large substrates has to be assured for high quality devices.

Inherent to their fabrication, the deposited films are *amorphous* or *polycrystalline*. Some of them have typical grain sizes in the nm range, some are between 0.1 to a few µm. Grain boundaries often suffer from high defect densities, leading to high recombination rate and thus to low cell performance. Possibilities to avoid or mitigate this effect are e.g. to try to drain impurities from the bulk and collect them at the surfaces where they can be removed, to passivate grain-boundary defects e.g. with hydrogen, or to implement drift-devices where electric fields across the active layer

assist collection.

11.1 Monolithic serial interconnection

In order to limit the current and thus the ohmic losses, the active layers are "cut" into individual cells by *monolithic serial interconnection*. This step is easily integrated in the deposition process using *laser scribing techniques* which ablate parts of the already deposited layers. Series connection of the resulting cells is achieved by depositing the following over grooves offsetting the next scribe-line laterally. Thus, the region between the first and the last laser-scribe serves for series-connection. Since it does not contribute to the output power, this region is called *dead area* and laser lines have to be optimised to minimised it. The width of the cell-stripes between the dead-area must be optimised with respect to the series resistance of the contacting-layers; typically it is in the order of cm.

Normally, pulsed lasers are used for scribing. Thus, the overlap of the spots and the imaging optics have to be optimised for overlap. Shunts are easily created if the layers not cleanly ablated. Infrared lock-in thermography was developed as powerful tool for shunt detection, but additional analysis is needed to find the root-cause, e.g. peeling of layers from the substrate, partial melting, sputtering of molten droplets, ablating too deep or to shallow, etc.

11.2 Comparison of different thin film technologies

Table 11.1: Comparison of thin film technologies

	CdTe	CIGS	Si, SiGe (a-Si and μc-Si)
record η (lab)	21.0% (1 cm^2)	21.0% (1 cm^2)	12.7% micromorph on glass 11.8% mini module of 14 cm^2 10.2% a-Si
record module[1] η	18.6%	17.5%	12.3%
prod. quant. 2010 prod. quant. 2020	1.4 GW 1.9 GW	430 MW 1.1 GW	1350 MW -
strengths	fast and easy process for absorber deposition	η close to c-Si (on cell level)	synergy with FPD resources unlimited
weaknesses	availability of Te Cd toxic	availability of In sensitive process	medium/low η

Note that cells with a direct band gap like CdTe and CIGS can in principle be much thinner than devices with an indirect (μc-Si) or non-direct band gap (a-Si). Fabrication issues may nevertheless require thicker film than needed for absorption.

11.3 CdTe based solar cells

This cell type is fabricated in *superstrate configuration*, i.e. the cell is illuminated through the glass substrate on which the cell layers are grown.

1. TCO: usually 500 nm of (n^+) SnO_2:F (3.3 eV) with $R_{\text{sheet}} \approx 10\,\Omega/\square$
2. 100 nm CdS (2.4 eV) as n-type buffer / window layer from a chemical bath (heterojunction to next layer)
3. p-type CdTe (1.45 eV) absorber, thickness $\sim$ 6 - 10 µm
4. sputtered ZnTe with an additional metallic back-contact
5. chlorine post treatment

The fact that the bond between cadmium (Cd) and tellurium (Te) does not break upon evaporation renders the deposition very simple (CdTe evaporated congruently). It is usually performed by *close-space sublimation* (CSS), however *electrodeposition* can also be used. For close-space sublimation the challenge is the glass handling, the back contact formation and diffusion of impurities along the grain boundaries, as high temperatures are reached during the evaporation ($\sim$550 - 600 °C, close to the glass melting temperature). The process itself extremely fast: a few seconds for the absorber layer.

11.4 CIGS or chalcopyrite solar cells

In contrast to CdTe cells, CIGS cells are fabricated in *substrate configuration*, i.e. the layers are deposited onto a substrate that is coated with a non-transparent metal or can start directly with a metal plate or foil. Then the following layers are deposited:

1. 600 nm sputtered molybdenum (Mo), opaque
2. 1.5 µm of CIGS (naturally p-type): graded, i.e. increased Ga content during the growth (graded doping). This leads to a gradient in band gap from. The available range of bandgaps varies between 1.04 eV ($CuInSe_2$) to 1.67 eV ($CuGaSe_2$), but due to defect creation for high Ga-content, 1.3 eV are normally not exceeded. The CIGS absorber thickness is typically 1 - 2 µm.
3. 70 nm p-type CdS (2.4 eV) (buffer layer) from chemical bath or alternatives
4. ZnO:i
5. ZnO:Al

The challenges for this technology are the delicate control of the composition of the absorber material; up to four materials have to be deposited simultaneously to obtain a correct composition. Furthermore, the interfaces between the layers can be an issue, as well as the high processing temperatures (up to 600 °C). Another aspect is the availability of In which could have a drastic impact on the applicability of this technology on very large scale. This is why several groups are working on alternatives, e.g. copper, zinc, tin, sulphur and selenium (CZTS).

11.5 Thin film silicon

For the deposition of thin film silicon several alternatives have been tested. These are listed below:

1. **CVD epitaxy** where thin films are deposited on a Si seed layer at high T (800 - 1000 °C). The deposited layers need to be rather thick, but they can have an efficiency potential close to c-Si. Issues are impurity diffusion (typical at high temperature), finding a suitable low-cost substrate for epitaxy, and up-scaling.

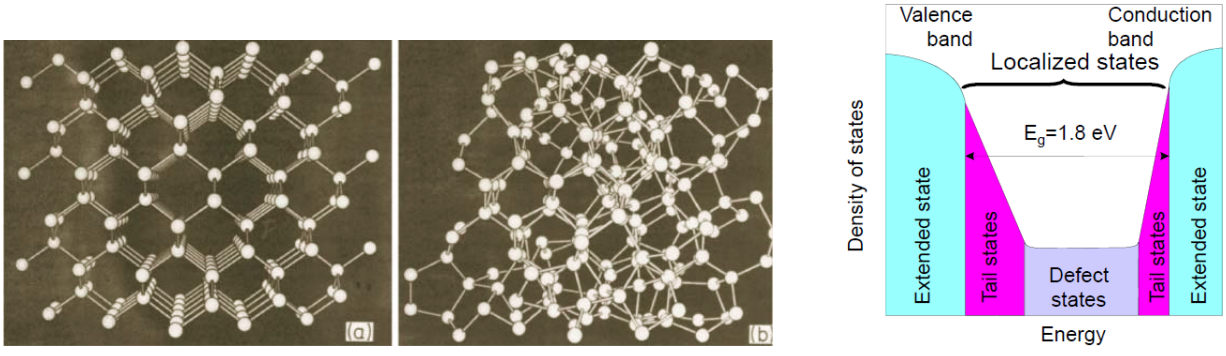
2. **Recrystallization of a-Si** is done at lower T (600 - 900 °C) compared to CVD epitaxy. There are attempts to make this work. An alternative would be to find a method to peel off thin layers directly form a c-Si ingot.

3. **Low temperature deposition** for the deposition of a-Si, μc-Si or SiGe is performed at 180 - 250 °C. Up to now this is **the only** method that is used in industry. It has the potential to drive down the costs to 0.8 - 1.0 €/W_p.

Among the low temperature deposition techniques there is *Plasma Enhanced Chemical Vapour Deposition*, (PECVD) *Hot Wire CVD* (HWCVD) and *Microwave Plasma Deposition* (MWCVD). Some critical parameters for the deposition of silicon films are the reactor geometry, inter-electrode distance, temperature, pressure, plasma power and gas concentrations. The material structure can be altered by varying the deposition conditions, e.g. from a-Si:H to μc-Si:H. This transition depends on the silane depletion fraction (the amount of silane molecules that are dissociated) and the silane concentration in the plasma. Increasing the hydrogen dilution, the pressure or the power will drive the layer properties from a-Si:H to μc-Si:H.

11.5.1 Amorphous silicon

In contrast to crystalline silicon, a-Si:H does not exhibit long range order, i.e. the matrix is strongly disordered (see figure 11.1(a)). This fact is clearly visible in the high density of strained and *dangling bonds* (broken Si-Si bonds), both forming electronic states within the band gap. Strained bonds form states at the band edges, rendering them no longer sharp as in crystalline material but leading to so-called *band tails* (see figure 11.1(b)). The dangling bonds form states close to mid gap, therefore they are recombination-active. They can passivated to some extent with hydrogen, but generally not below $1 \times 10^{15}\,\mathrm{cm}^{-3}$ because of a thermodynamic equilibrium between defect formation and defect curing.



(a) (b)

Figure 11.1: (a) Crystalline (left) and amorphous (right) silicon structure. (b) Band gap of amorphous silicon.

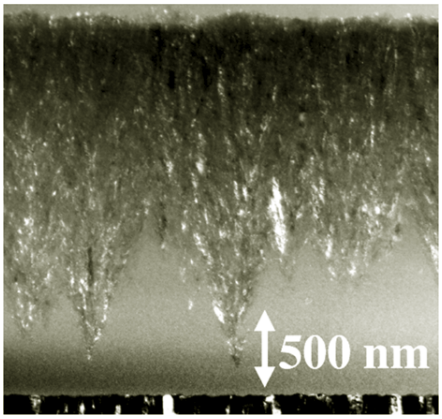
As can be seen in Fig 11.1(b), the band gap of a-Si is very different from its crystalline counterpart. Due to the states within the band gap, which act as carrier traps or recombination centres,

one sometimes refers to it as a *mobility gap*. Compared to c-Si the carrier mobility in a-Si is very low, 0.1 - 1 $\mathrm{cm\,V^{-1}\,s^{-1}}$ and the carrier lifetimes are of the order $\sim 1\,\mu s$. This results in short diffusion lengths $\sim 0.15\,\mu m$ [2]. Upon doping the electronic properties degrade even more which is related to the fact, that the Fermi level deviation out of mid-gap enters into the thermodynamic equilibrium of defect formation. Essentially doping makes it easier to transfor states in the *band tails* into dangling bonds.

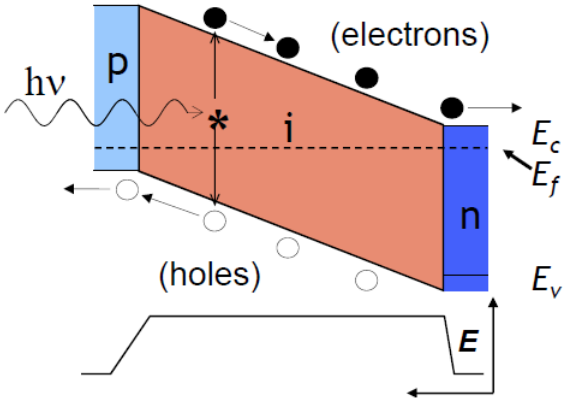
It is also important to note that due to its structure and composition, amorphous silicon is a *meta-stable* material, i.e. it is subject to *light induced degradation* which is described by the *Staebler-Wronski effect (SWE)*. Upon illumination the material will degrade up to an equilibrium state.

11.5.2 Micro- or nano-crystalline silicon

This material consists of *nano-scale particles* ($\approx 20\,nm$) embedded in an a-Si matrix. The crystallites are usually interconnected and their grain boundaries are passivated by the a-Si phase. But apart from this they show the same properties as c-Si, same band gap 1.12 eV and same activation energy $E_{act} \sim 0.5\,eV$. The deposition of a μc-Si layer is usually preceded by an a-Si *incubation layer*, which, depending on the deposition parameters, can be in the order of several 100 nm (see figure 11.2(a)). The incubation layer can be avoided by treating the interface in CO_2 Plasma before μc-Si deposition.



(a)



(b)

Figure 11.2: (a) Transmission electron micrograph of a microcrystalline layer. The incubation layer can clearly be seen. (b) Doped layers create an electric field across the entire intrinsic layer to efficiently separate and collect the photo-generated charge carriers.

11.5.3 Drift and diffusion length

Both lengths describe the same distance, namely the distance a carrier can travel before it recombines. The difference is that in the case of the drift length an electrical field is present that drives the carriers.

$$L_{\text{drift}} = \tau v_E = \tau \mu |\vec{E}| \quad (11.1)$$

$$L_{\text{diff.}} = \sqrt{D\tau} \quad (11.2)$$

[2] The mobility in a-Si:H is approximately 500 times lower than the one of c-Si. This and the fact that the lifetime is as well very low, is the reason why **drift devices** are necessary.

Depending on the field that is applied, the drift length can be much larger than the diffusion length. Applied to a solar cell, this means that the collection probability increases if an electric field is present, for example in a p-i-n junction.[3] cells (see figure 11.2(b)). The concept here is to use the doped layers, which are *electronically dead*, to create an electric field across the intrinsic layer and efficiently separate the charge carriers. The thickness of the doped layers has to be optimized, i.e. strongest field but lowest parasitic absorption.

Amorphous and microcrystalline silicon were combined in tandem cells and developed up to the point of multi-MW production with excellent production cost below 0.5€ /W around 2014. Since the efficiency of commercial modules stagnated around 10% more and more customers turned towards other module types.

11.6 Perovskite solar cells

The organic-inorganic metal halide perovskite solar cells emerged in 2009 with a first publication showing 3.8% efficiency. Impressive improvements increased the efficiencies to more than 20% within less than 8 years, a speed that was never seen with any other photovoltaic technology. Nowadays, there is almost no photovoltaic research lab in the world which does not have perovskite activities. Let's see briefly see what makes these materials so interesting, what are the perspectives and the challenges:

11.6.1 Material properties

Perovskite solar cells take their name from the crystal structure with the general formula ABX_3 (e.g. calcium titanate). A and B are for cations, X an halide anion. The most commonly used composition in solar cells is methylammonium lead tri-iodide ($CG_3NH_3PbI_3$ or MALI), but increasingly more cations[4] such as formamidinium and even alkali-metals such as Cs and Rb are added for stability. In parallel, Sn is investigated to replace the problematic Pb, and different halides such as Br and Cl are studied to manipulate the bandgap.

Perovskite materials have been shown to have exceptional optoelectronic properties, such as a strong optical absorption coefficient, high electron and hole mobility and diffusion length, high defect tolerance, low surface recombination velocity, low influence of grain boundaries, low sub-bandgap absorption and steep absorption edge, tuneable bandgap from 1.1 eV to >2 eV, photon recycling, direct/indirect bandgap character, high internal photoluminescence quantum efficiency, high voltage/low V_{oc} losses. These properties bring them close to single-crystal III-V materials.

11.6.2 Cell architectures

Historically, perovskite solar cells are derived from dye sensitized solar cells which featured a mesoscopic scaffold layer of TiO_2 impregnated with dye molecules. The perovskite solar cells started by replacing these dye molecules with a MALI perovskite material. The most used hole transport layer (p-contact) is still spiro-OMeTAD, which was first developed for solid state dye sensitized solar cells. This layer is usually spin coated onto the annealed perovskite layer. The cells are then finished with a evaporated layer of gold. Such cell is therefore in the n-i-p polarity and measured in superstrate configuration (illumination through glass substrate).

[3] Thin film silicon solar cells are classified as p-i-n or n-i-p devices which refers to the growth sequence. As the mobility of holes is much lower, the p-layer is always facing the sun. Thus, the deposition on opaque substrates (e.g. metal foils) has to start with the n-layer.

[4] DOI: 10.1039/C5EE03874J, 10.1126/science.aah5557

It was quickly demonstrated that the perovskite material can have sufficiently good electrical properties to render the mesoscopic scaffold layer unnecessary for charge extraction. Therefore, inspiration could be found in organic solar cells, providing a huge reservoir of research on n- an p-contacts, e.g. pedot-PSS, poly-TPD, PTAA, fullerenes (PCBM, C60)... Planar perovskite cells can be made with either p-i-n or n-i-p polarities and either substrate or superstrate configurations. Hence, it gives a larger flexibility in the substrate choice, deposition techniques and final applications.

The record perovskite cells usually maintain a mesoscopic structure, likely due to the much larger number of groups working on this type of cells compared to planar cells.

11.6.3 Deposition techniques

Many different techniques have been demonstrated:

- Hole and electron contacts: spin-coating, sputtering, ALD, evaporation, CVD...
- Perovskite absorber layer: spin coating, co-evaporation, hybrid techniques mixing evaporation and solution processes, die coating, printing techniques, blade coating, CVD...

11.6.4 Perovskite-based multijunction solar cells

Although single-junction perovskite solar cells have very rapidly shown high efficiencies, their first commercial appearance will likely not be on their own. Entering the photovoltaic market in competition with wafer-based crystalline silicon solar cells will be extremely difficult. However, with their steep absorption edge, low sub-bandgap absorption, tuneable bandgap and various processing techniques, perovskite cells can be excellent candidate to make tandem solar cells, combined with silicon or CIGS bottom cells.

Perovskite/Silicon[5] tandem devices have been demonstrated in both mechanically stacked 4-terminal tandem (up to 25.2%) and monolithically integrated 2-terminal tandem configurations (up to 23.6%). The challenges are the process compatibility (temperature, texture), cell size (upscalability of the perovskite cell), parasitic absorption (minimize absorption in non-active layers where photons are lost, e.g. TCO, HTL, ETL), light management (reflections on flat surfaces, interfaces reflections, refractive index engineering, surface texture) and reliability (perovskite cells must prove at least the same long-term reliability as the silicon cells.

Perovskite/CIGS or CIS[6] tandem devices have also been demonstrated in the last two years as a promising all-thin film tandem technology.

However, the ultimate thin film tandem is obviously a perovskite/perovskite[7] monolithic tandem cell, with high efficiency potential $> 30\%$ and low costs. This was made possible with the development of low band gap perovskite cells, which are still in their infancy.

11.6.5 Challenges

The perovskite solar cells have great potential. However they have many challenges to overcome before reaching commercialisation:

- Accurate characterization: Often Perovskite cells showed hysteresis with respect to the direction and rate of the voltage scan during the $j(V)$-measurement. This issue led to difficulties

[5]DOIs: 10.1021/acs.jpclett.5b02686, 10.1021/acsenergylett.6b00254
[6]DOIs: 10.1038/nenergy.2016.190, 10.1002/aenm.201500799
[7]DOIs: 10.1002/AENM.201602121, 10.1126/science.aaf9717

in comparing results between labs and misinterpretation of results. Hysteresis was a very hot topic in 2014-2015, with hundreds of different explanations, e.g. ion migrations, transient capacitive current, ferroelectric polarization... Currently, it is still not fully understood, but solutions to avoid it have been successfully demonstrated and the best reported cells have now negligible hysteresis.

- Showing stabilized efficiencies was an important step towards reliable performance reporting. This can be done by either fixing a voltage or measuring the current output for some time or by actively tracking the maximum power point with a proper algorithm.
- Moisture/humidity: perovskite materials tend to degrade with moisture exposure, e.g. MALI usually degrades to form a PbI_2 phase that is not electricall active. Newer more stable compositions and proper encapsulation should prevent this issue in the future.
- Temperature stability: the original perovskite compositions were severely affected by temperatures around 60 - 80 °C. With the recent advances on new compositions and new contacts, it becomes less of an issue.
- Light stability: light soaking effects are often observed, but rarely fully understood. Their origin lies either in the contacts which get doped during light exposure, or degradation of the perovskite material .
- up-scalability with industrial techniques: fabricate large scale modules with high throughput and high yield. From OLED and OPV industries, printing technologies such as roll-to-roll are investigated. But also vacuum deposition processes such as were used in silicon thin-film technologies and flat-panel industry.
- Toxicity: the lead content is a discussed controversially in the photovoltaic community. Research is ongoing towards a lead-free compound, but performance is still far behind the mainstream lead-containing perovskites.
- etc

Chapter 12

Module technology

PV installations are divided into three subassemblies capable of energy generation: the cell, the module and the array. The cells are composed of a delicate chemistry that cannot be directly exposed to environmental elements, such as moisture, thermal cycles, mechanical stress... For this reason, PV cells are electrically connected into a string, and a series of strings are packaged into a module composed of multiple layers of glass and polymers.

12.1 Module design

Poor packaging is the predominant cause of system failure. Thus it is crucial to find the best solution in module technology and design to ensure system stability, i.e. long lifetime and reliability while keeping the losses due to encapsulation at a low level. The main stress factors that cause failure of a PV module are:

- Irradiance: sun, sky
- Humidity
- Moisture: rain, dew, frost
- Temperature: heat, frost, night-day cycles
- Mechanical stress: wind, snow, hail
- Atmosphere: salt mist, dust, sand, pollution

12.1.1 Module materials

In a PV bulk silicon modules the active component is sandwiched between two polymeric encapsulants, protected with a transparent top surface, a rear layer and a frame around the outer edges. In most modules, the top surface is glass, the encapsulant is EVA (ethyl vinyl acetate) and the rear layer is Tedlar or glass, as shown below.

Front surface material

The front surface material must have high transmission in the wavelengths which can be used by the solar cells encapsulated in the PV module. Moreover, it should also present a low reflection property. Besides the optical properties, the top surface material should be impervious to water, should have good impact resistance, stable under prolonged UV exposure and should have a low thermal resistivity. In most modules the front surface is used to provide the mechanical strength

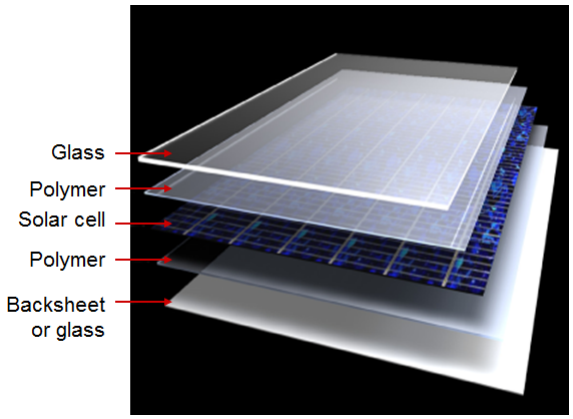


Figure 12.1: Typical lay-up stack for c-Si PV module.

and rigidity, therefore either the top surface or the rear surface must be mechanically rigid in order to support the solar cells and the wiring. Tempered, low iron-content glass is most commonly used as it is low cost, strong, stable, highly transparent, impervious to water and gases and has good self-cleaning properties.

Encapsulant

The encapsulants are polymers composed of repeating structural units (monomers) typically connected by covalent chemical bonds. Depending on their characteristics properties we can have:

- Thermoplastics: polymer which softens with temperature and can be made to flow. It can be viscous / elastic / plastic... (e.g. polyvynyl butyral (PVB), Thermoplastic Polyurethane (TPU), Iononers...)

- Thermoset: polymer which undergoes a chemical change to produce a networked polymer (after cross-linking, vulcanization, ets.) (e.g. Ethylene Vinyl Acetate (EVA), polyurathane casr resins (TPU), Polyacrylate cast resins...)

The encapsulant is responsible to provide adhesion between the solar cells, the top surface and the rear surface of the PV module. Moreover, the encapsulant should be optically transparent and should be stable at high temperatures and high UV exposure. EVA is the most commonly used encapsulant material because it is cheap, stable and reliable under long time exposure.

Rear surface

The backsheet of a PV module is responsible to provide physical and mechanical stability and also prevent the ingress of water or vapour water. The backsheet should present a good adhesion with encapsulant and the J-box adhesive, good weatherability, high internal reflectivity, resistant to abrasion or scratch and pass successfully fire tests. In most modules, a thin polymer sheet, typically Tedlar, is used as the rear surface because it has excellent UV and chemical resistance. However, there are on the market other backsheet materials such as polyester (PET), multistack materials (PVF-PET-PVF (TPT) and also backsheet which include metals (PVF-Al-PET). Special modules can be found with glass at the back. For example, the use of bifacial solar cells required light trapping from either the front or the rear of the module.

Frame

After assembly all the components of a module, the module can receive an edging (no frame) or be protected by an aluminium frame. The frame structure should be free of projections which could result in the lodgement of water, dust or other matter.

12.1.2 Module manufacture

The most common PV module manufacture process is the Vacuum Laminator Line. The overall back-end module manufacturing process is illustrated in figure 12.2.

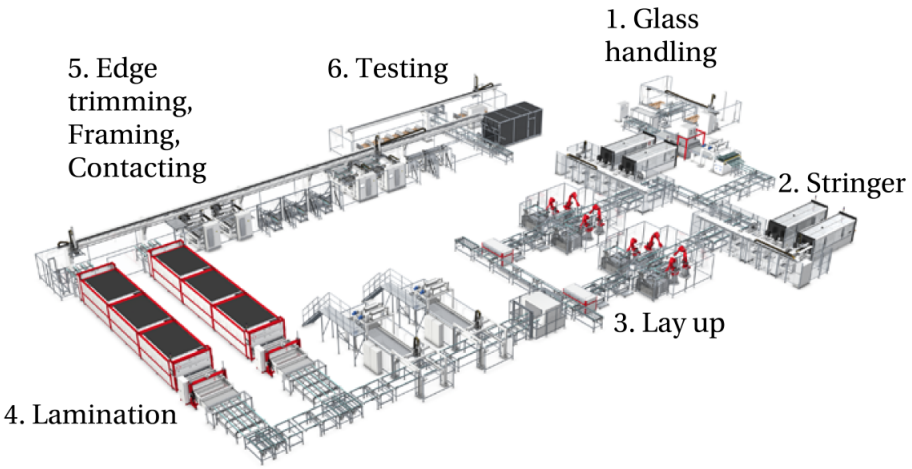


Figure 12.2: An overview of the process flow of PV module manufacturing. [1]

The lamination process is composed of 6 steps:

1. Glass handling
2. Stringer: in this section dedicated solar cells are stringed (soldering process) to form the full module array.
3. Lay up: all materials are assembled together based on the structure presented in figure 12.1.
4. Lamination: module layup is introduced in the laminator and a typical temperature and pressure profile (T-P-t profile) is applied. The module layup first undergoes a preheating step (lasting 100 - 500 s), with the aim of easily removed the air entrapped in the module layup to minimize the risk of voids formation. After the pre-heating step, the softening temperature of the encapsulant should be reached and the pressure on the module is applied. The curing step starts taking place by typically 300 - 900 s. The hot press is required to suppress any bubble or void remaining in the encapsulant. An overall encapsulation cycle can take 10 to 20 minutes.
5. Post-lamination: it includes an optional active cooling process, where PV module cool down rapidly to room temperature under 1 atm pressure. The post-lamination also includes the edge trimming, the placements of the junction box and the framing.

[1] H. Li, Open the Black Box: Understanding the Encapsulation Process of Photovoltaic Modules PhD thesis, École polytechnique fédérale de Lausanne, 2013.

6. Testing

12.2 Characterization

There are different test methods that can be used to characterize materials and modules.

12.2.1 Material Characterization

There are many different encapsulants that can be used during the module lamination. The selection of the best encapsulant is based on the properties required for the different case studies. These properties can be quantified based on:

1. **Tensile test**

   The tensile test is a fundamental materials science test in which a sample is subjected to a controlled tension until failure. From the obtained stress-strain curve, the Young modulus is given by its slope. From this test we can differentiate 4 different classes of materials: a) strong fibre; b) brittle material; c) tough plastic and d) elastic / rubber.

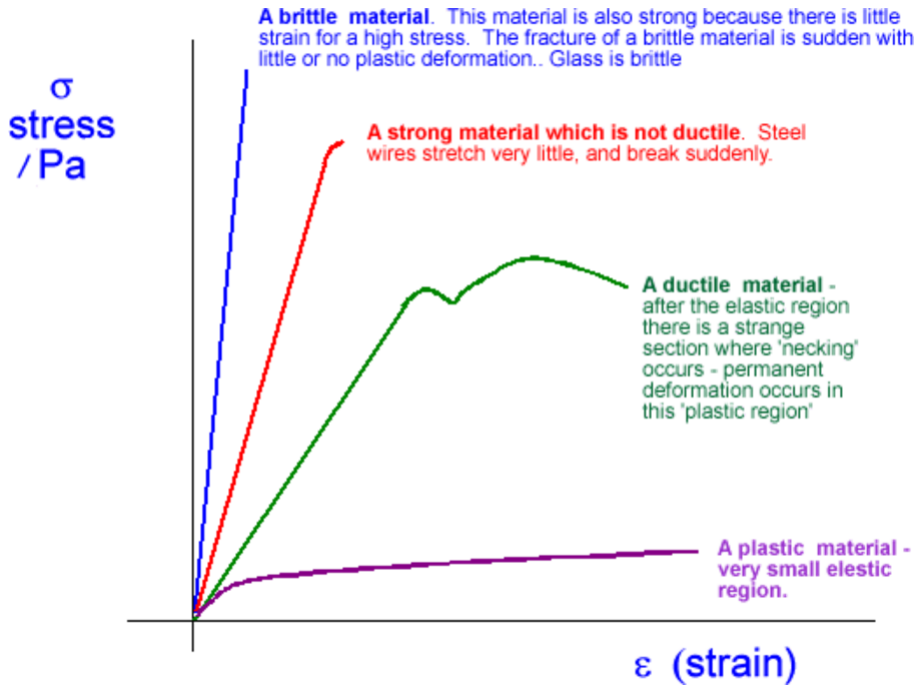


Figure 12.3: Stress-stain curves for typical material classifications.

2. **Dynamic Mechanical Analysis (DMA)**

   DMA is a technique widely used to characterize material's properties as a function of temperature, time, frequency, stress, atmosphere or a combination of these parameters. With this test we can quantify the glass and melting transitions over a wide range of temperatures. At low temperatures polymer is its brittle state. By increasing the temperature the Young modulus is reduced of 3 orders of magnitude over typically 10 - 20 °C. The material reaches

an elastic behaviour. If temperature is further increased polymer will reach its melting temperature passing a viscoelastic state and finally, a complete viscous state. With this study we can quantify the ideal lamination temperature for different types of encapsulants.

The glass transition temperature can be changed by adding molecules, such as plasticizers, to our polymer formulation. Plasticizer is a material that enhances the processability or flexibility of the polymer. These compounds may reach 50% in some polymer formulations.

The peroxides are other type of compounds which are added to the polymer formulation. These elements promote the creation of a network polymer which has new covalent bonds between macromolecules, the called cross-linked effect.

3. **Adhesion test**

There are different methods used to measure adhesion: lap shear, compressive shear and peel test. The selection of the best test is based on the type of materials we want to study (rigid material, flexible material...). In all this types of tests, the maximum force until the test specimen breaks / delaminates is a measure of the adhesion of the polymer to the substrate. The value for adhesion can be expressed in $\mathrm{N\,mm^{-2}}$ or $\mathrm{N\,mm^{-1}}$.

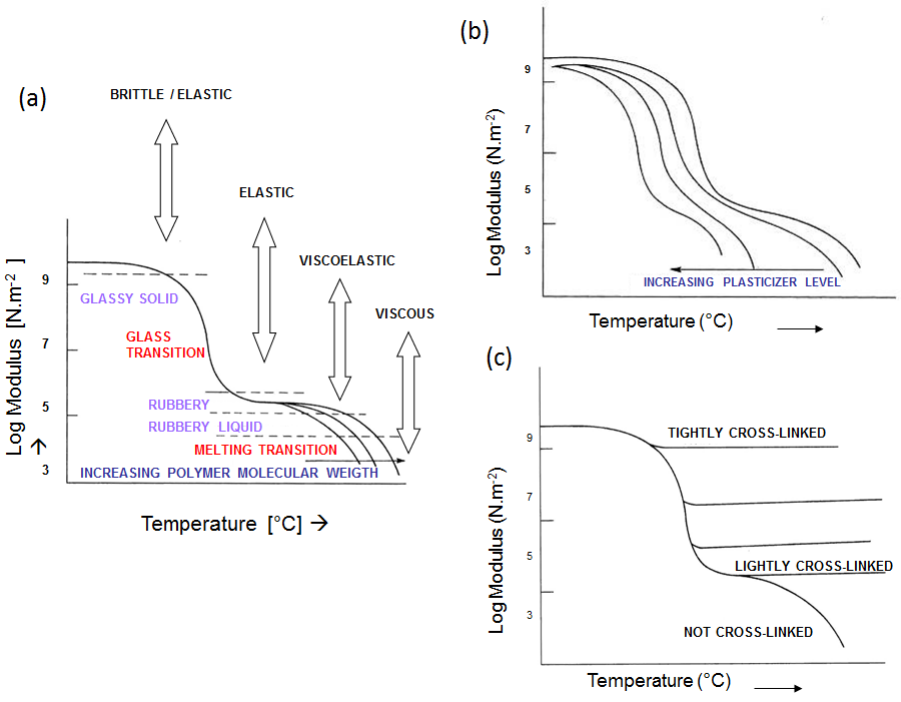


Figure 12.4: (a) Modulus values change with temperature and transitions in materials. (b) Effect of plasticizers and (c) effect of peroxides on the modulus properties

4. **Moisture vapour transmission rate (WVTR)**

The WVTR is a measure of the passage of water vapour through a material. The moisture ingress in polymers obeys a diffusion law. The important parameters are:

- the diffusivity of the water vapour in the polymer, D
- the concentration of water vapour in the polymer at saturation C_{sat}

12.3 Reliability and Norms

Commercial modules are generally sold with a product guarantee (2-10 years) and a performance guarantee (typically 90% of nominal power at 10 years and 80% after 25 years). The guarantee of a PV module is given by the company based on a typical certification for PV modules: IEC 61215 for c-Si PV modules. By applying Accelerated lifetime tests (high stress factors) we can predict how a module will perform in an outdoor condition. Within this test method we have:

- Thermal-Cycling test: determine the ability of the module to withstand thermal mismatch, fatigue and other stress caused by repeated changes in temperature. During this test modules are subjected to cycles in temperature ranging from $-40\,°C$ and $85\,°C$ with current injection equal to the STC peak power current of the module.
- Damp-heat test: determine the ability of the module to withstand the effects of long-term penetration of humidity. Modules are subjected to with high temperature ($85\,°C$) and high humidity (85%) during 1000 h.
- Mechanical load test: determine the ability of the module to survive wind, snow, static or ice loads. Module is mounted in a support and load of 2400 Pa is gradually applied during 1 h in sections of load and suction.
- Hail test: determine the capability of the module to withstand the impact of hailstones. For this test, module is installed vertically and ice balls with a diameter of 25 mm are shoot at $23\,m\,s^{-1}$ at 11 specific locations.

The figure 12.5 shows all test sequence which has to be performed for a PV module certification.

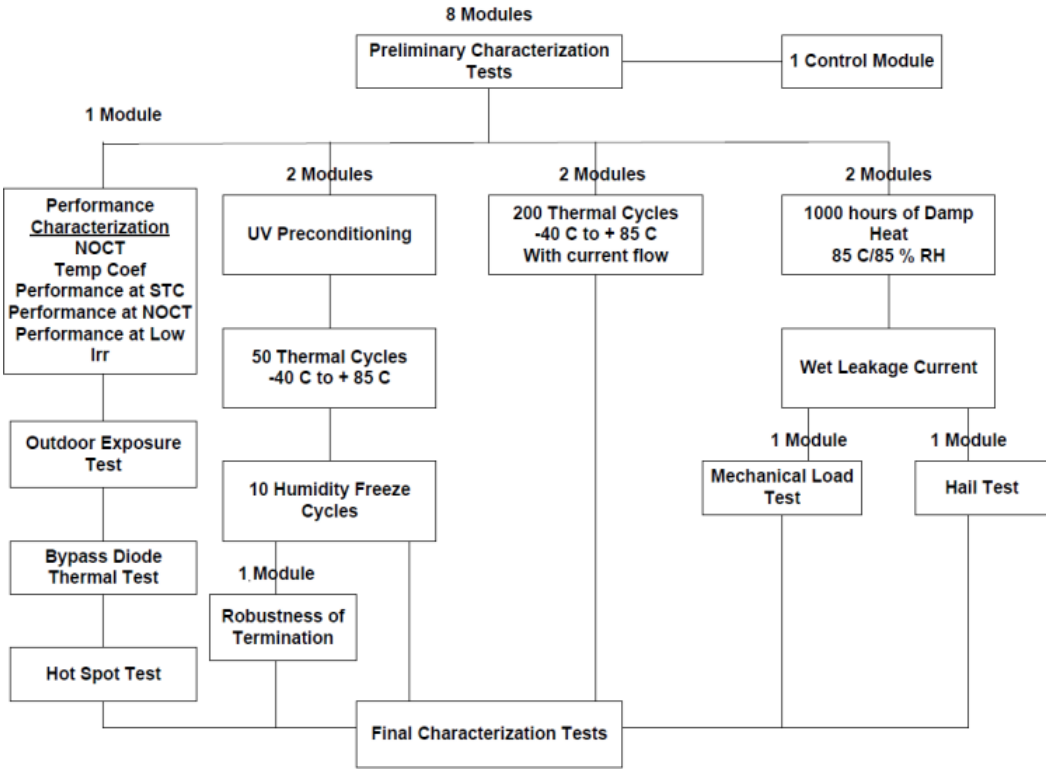


Figure 12.5: Diagram of test sequence based on the IEC 61215 qualification tests

The main failure modes observed in PV modules in outdoor installations are:

- interconnect failures: it can happen due to the temperature cycles in between day and night
- hot spots: it can cause visual defects on the front and rear side of the module(>300 °C) and it can induce glass breakage, EVA melting or backsheet foil melting
- cracked cells
- broken glass
- encapsulant yellowness: can be avoided by choosing compatible materials
- moisture ingress: avoid "contaminants" getting into the module by choosing a good packaging material
- delamination: EVA + H_2O + UV = CH_3COOH (acetic acid) which produces yellowing of the polymer and loss of adhesion. It can be reduced by adding stabilized materials to the EVA formulation.
- corrosion / degradation (TCO): due to the ΔV, ion migration and water, TCO start delaminating. We must choose "clean" materials which retards / avoid the ion migration from glass.

12.4 Cost

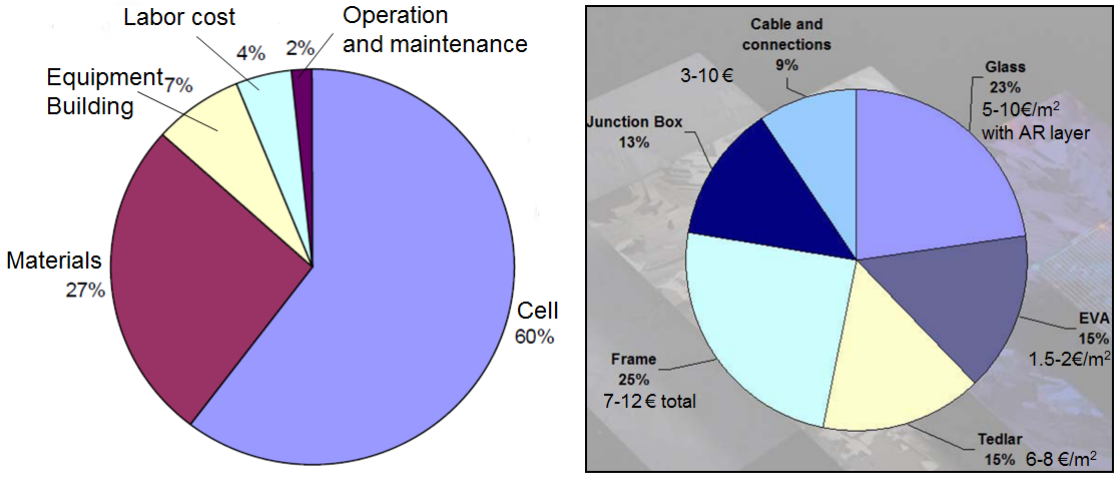


Figure 12.6: Main costs at module and material level

With this in mind, the goals are:

- long lifetime (chemical stability, electrically immune, stable and good optical properties, thermal stability)
- security (electrically, non-toxic, fireproof, mechanically robust, weight)

- low cost (abundant and safe materials, recyclable, fast and automatable processes, low energy consumption processes)
- minimization of grey energy and energy payback time
- aesthetics
- easy building integration

Appendix A

Band gap and effective masses

A.1 Bandgap

Properties of the bandgap are most directly probed by measurements of the absorption coefficient. Figure A.1 shows data for silicon in the temperature range between 4 - 415 K [67, 25]. At the lower temperatures there two clearly distinguishable changes of the absorption coefficient that occur at 1.08 and at 1.23 eV. Both of them shift to lower energy as the temperature increases, additionally the one at higher temperature is no longer well distinguished towards room temperature. These abrupt changes are interpreted as onset of indirect absorption processes with phonon assistance. For the one at lower energy, the photon actually has less energy than the bandgap, but it can still excite an electron across the bandgap when the energy of the assisting phonon is added. As the phonon is provided by the lattice, this process becomes less and less probable at low temperature. For the absorption process at higher energy the assisting phonon is emitted into the lattice. As this process can always take place, it is only little impacted by the temperature.

The right panel of figure A.1 shows the energies of the transitions with respect to temperature. The strongest phonon assistance takes place with an energy of 60 meV, but at low temperature the right panel in figure A.1 shows that there are also contributions of phonons with energy of 20 meV. The energy of the bandgap is subsequently found in the middle between these transitions.[1]. To find the value of the bandgap we must still take into account that the absorption process does not immediately create a free electron-hole pair but a bound entity called exiton. The true bandgap is only obtained after adding the binding energy of the exciton which is 14 meV in case of silicon.

The right panel in figure A.1 shows that the bandgap varies little at low temperature, but eventually it bends into linear variation around room temperature and above. The transition between these regimes is rather generic for a variety of semiconductors and can be described empirically as follows [68]:

$$E_g = E_g(0) - \alpha T^2/(T + \beta) \tag{A.1}$$

For silicon, the parameters are $E_g(0) = 1.169\,\mathrm{eV}$, $\alpha = 4.9 \times 10^{-4}\,\mathrm{eV\,K^{-1}}$ and $\beta = 655\,\mathrm{K}$ [68, 69, 70]. Alternatively, Pässler suggested a more detailed model to capture the two major physical phenomena of the temperature dependence, namely electron-phonon interactions and thermal lattice expansion [71].

[1] The diagram includes energies of additional phonon-assisted processes at energies below 1.05 eV, so-called replicas [25]. These are explained by transitions between the equivalent minima of the conduction band, so-called f- and g-processes that are mediated by similar phonon energies.

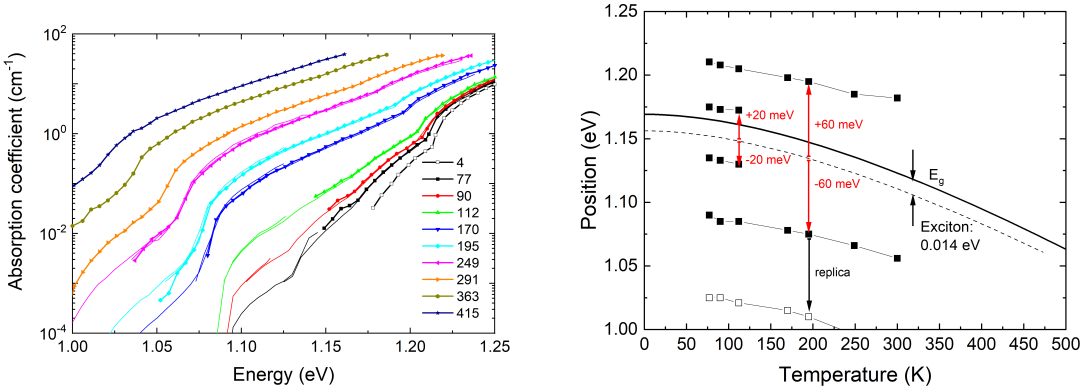


Figure A.1: Temperature dependence of the absorption coefficient in silicon (left, [67, 25]) and energies of the phonon signatures (right).

Besides the indirect transitions in silicon, Figure 4.3 shows that there are states at an energy of 3.4 eV directly above the VBM. Incidentally, this is exactly the energy where the absorption coefficient of silicon in figure 4.2 has a kink and reaches the same level as the one of the direct semiconductor GaAs.

To identify the type of the contributing phonons, figure A.2 reproduces the phonon dispersion diagram of silicon [72]. Different from electronic waves, phonons start at the origin with two linear dispersion relations according to $\hbar\omega = c\vec{k}$. These are called acoustic branches since they describe the propagation of sound in the solid by means of transverse and longitudinal waves. Towards the edges of the first BZ the branches start bending and eventually they fold back into the optical branches. Light excites more easily with optical vibrations than acoustic ones, and among the two preferably with the transverse ones since its electric and magnetic field amplitudes are perpendicular to the propagation direction. The transitions with energy of ± 60 meV can thus be identified with transverse optical phonons whereas the weaker ones with ± 20 meV are due to transverse acoustic phonons.

Mathematically, the indirect absorption process can be described in terms of a quantum-mechanical transition probability from a state in the valence band $|v\rangle$ to a state in the conduction band $|c\rangle$ using Fermi's Golden Rule of second order.

$$w_{v\to c} = \left| \sum_m \frac{\langle c|H_{er} + H_{el}|m\rangle \langle m|H_{er} + H_{el}|v\rangle}{E_m - E_v} \right|^2 \cdot \frac{2\pi}{\hbar} \delta(E_c - E_v) \tag{A.2}$$

Here, H_{er} and H_{el} are the Hamiltonians that couple the electronic states to the radiation field and to the lattice vibrations, respectively. The δ-function selects among the infinitely many intermediate states $|m\rangle$ those that comply with energy conservation. Thus, emission or absorption of a phonon takes an electron from the VB almost horizontally into a virtual state within the bandgap, followed by the absorption of a photon whose energy takes it vertically up into a free state in the CB. The reverse order can also happen, but the term $E_m - E_v$ in the denominator yields a larger number and thus it contributes much less to the overall transition probability.[2] The absorbed power is obtained by weighting the transition probability with $\hbar\omega$ and summing up all the contributions of possible transitions. The Poynting theorem subsequently provides a link to the imaginary part of the permittivity:

[2]Details on the evaluation of the integrals can be found in the textbook of Hamaguchi [73].

$$P_{abs} = \hbar\omega \sum_{\vec{k}_c,\vec{k}_v} w_{v\to c} = \frac{1}{2}\omega\epsilon_0\epsilon" E_0 E_0^* \tag{A.3}$$

Alternatively to this heuristic link between the transition probability and the imaginary part of the permittivity, a more rigorous treatment requires the use of the density operator [74]. Once the imaginary part is obtained, the real part can be obtained by a Kramers-Krönig transformation.

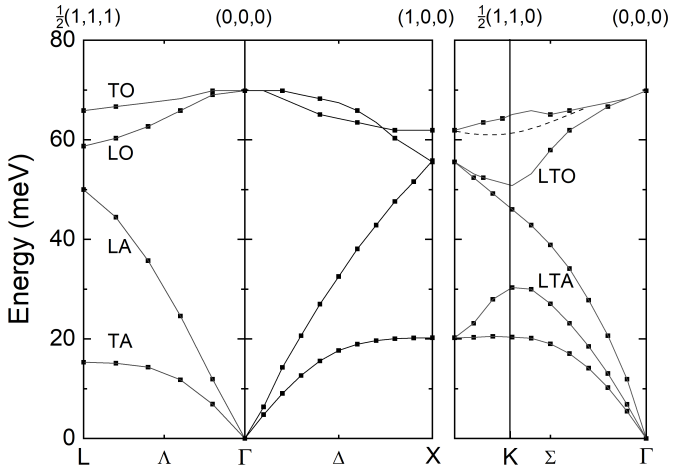


Figure A.2: Dispersion relation of phonons in c-Si (digitized from [72]).

A.2 Effective mass

In the main text, effective masses were defined in the context of eqns. (3.5) and (3.6). There, we argued that the 1D dispersion relations resemble parabolas close to the CBM and the VBM. More generally, the dispersion relation is defined in 3D $\vec{k}$-space where a second order Taylor expansion around an extremum of the energy ideally yields ellipsoids or saddle-points which can have up to three different radii of curvature. Accordingly, there may be up to three effective masses for a given band. Even though this yields already a lot of complexity, in the next section we will encounter situations where the approximation with ellipsoids is still not sufficient.

A.2.1 Valence band

In silicon, the VB is composed of three bands shown in figure 3.5. There is a single band at low energy (purple) followed further up by two more bands that are identical from L over Γ to X (green and yellow). A closer look to the right of the diagrams reveals that one of them is flatter between Γ and K (yellow). The properties of these three bands can be described by $k \cdot p$ perturbation theory taking into account spin-orbit coupling. The first band is thus associated with an angular momentum of $j = 1/2$ and its energy surface can be described as follows [15]:

$$E_{so} = -\Delta_0 - A\frac{\hbar^2 k^2}{2m} \tag{A.4}$$

In silicon the curvature is characterized by the single parameter $A = 4.28$ which corresponds to an effective mass of $m^*_{so} = 0.24\ m_0$ and a spherical shape as shown in figure A.3. We call this band

the split-off band because the parameter $\Delta_0 = 0.044\,\mathrm{eV}$ locates it below the common maxima of the two other bands which are associated to $j = 3/2$:

$$E_{lh/hh} = -A\frac{\hbar^2 k^2}{2m} \pm \sqrt{B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)} \tag{A.5}$$

In silicon the parameters of these two bands are $B = -0.68$ and $C^2 = 24$. This yields energy surfaces resembling an octaeder with rounded corners and a cube with rounded corners and dented sides. Owing to its large curvature the octaeder-shaped band is called light-hole band whereas the cube-shaped one is called heavy-hole band. As the shapes are clearly not elliptical, effective masses can be defined only locally in different directions or in terms of averages, either over specific directions in k-space, or over all possible directions. The latter procedure yields effective masses of $m^*_{lh} = 0.15\ m_0$ for the light hole and $m^*_{hh} = 0.5\ m_0$ for the heavy hole [15, 75].

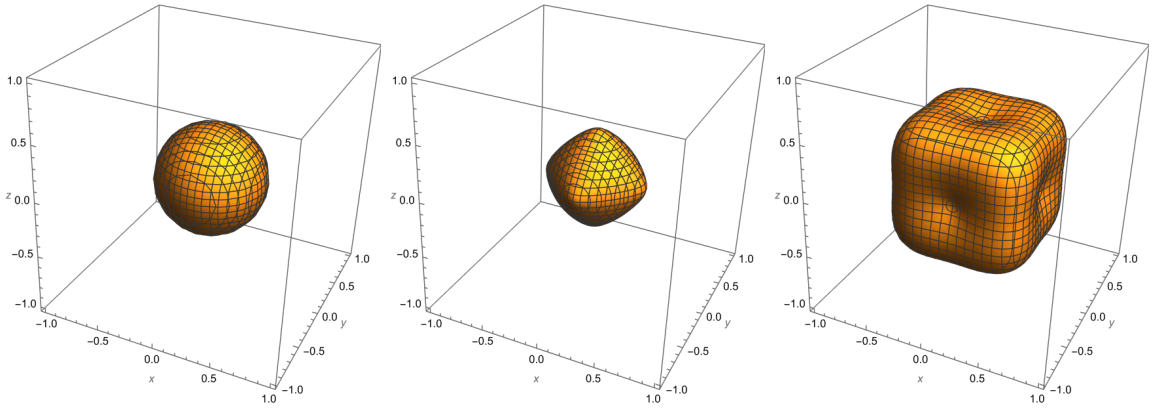


Figure A.3: Hole energy surfaces of silicon for the split-off band (left), the light-hole band (middle) and the heavy-hole band (right).

For the calculation of the DOS we should take into account the different curvatures of the bands when doing the coordinate transform of the Jaobian. Moreover, the split-off band does not contain holes at low temperatures, it is occupied only at room temperature and above. Accordingly, the following expression was suggested for the DOS effective mass [76]:

$$m^*_{VB,DOS} = \left((m^*_{lh})^{3/2} + (m^*_{hh})^{3/2} + (m^*_{so}e^{-\frac{-\Delta_0}{kT}})^{3/2}\right)^{2/3} \tag{A.6}$$

For the case of silicon, this yields an effective DOS hole mass of $m^*_{VB,DOS} = 1.15\ m_0$ at room temperature.

The electric conductivity for holes is obtained by $\sigma = qp\mu$ where $\mu = q\tau/m^*_h$. Therefore, we should use the harmonic mean of the effective masses:

$$m^*_{VB,\sigma} = 3 \cdot \left(\frac{1}{m^*_{so}} + \frac{1}{m^*_{lh}} + \frac{1}{m^*_{hh}}\right) \tag{A.7}$$

This yields a value of $m^*_{VB,\sigma} = 0.21\ m_0$. For the effective mass appearing in the thermal velocity, yet a different averaging was proposed to take into account the warped shape of the upper two bands, resulting $m^*_{VB,v} = 0.41\ m_0$ [77].

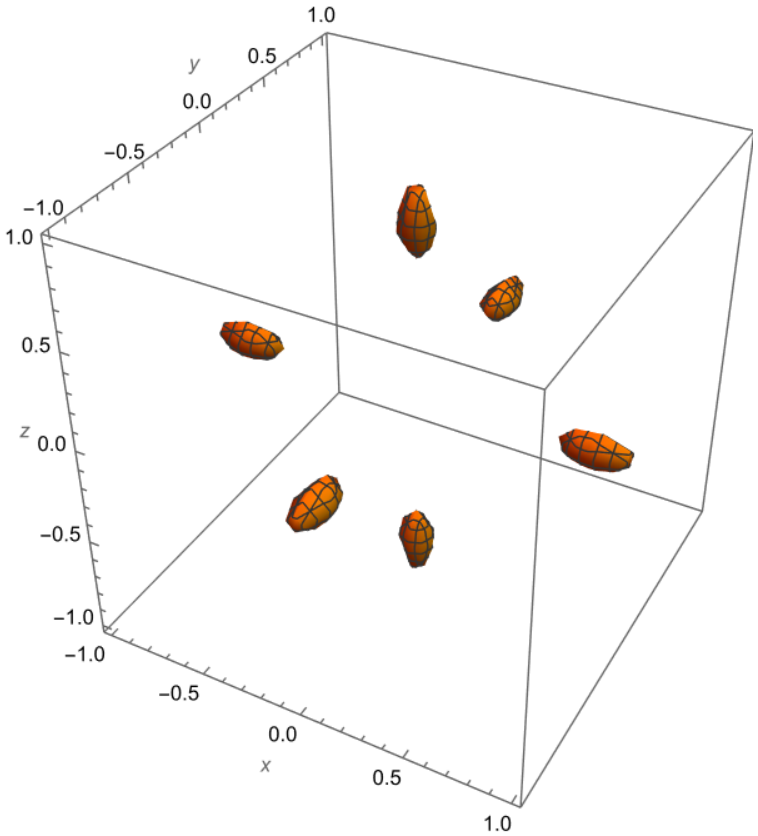


Figure A.4: The energy surface of the conduction band in silicon consists of 6 equivalent ellipsoids located at $0.85X$ along the main axes.

A.2.2 Conduction band

Different from the VB, the CB is not degenerate in the sense that different bands have the same energy for a range of k, but figure A.4 illustrates that there are six equivalent ellipsoids at a distance of $0.85X$ along the main axes, all of them having the same energy. The low curvature parallel to the axes gives rise to a large longitudinal effective mass of $m_l^* = 0.91\ m_0$ whereas the identical curvatures in the two directions perpendicular to the axes yield a transverse mass of $m_t^* = 0.19\ m_0$ [15, 78]. Entering the six ellipsoids into the coordinate transformation of the Jacobian, we obtain the following relation:

$$m^*_{CB,DOS} = 6^{2/3} \left(m_t^* (m_l^*)^{1/2} \right)^{2/3} \tag{A.8}$$

Thus, we obtain $m^*_{CB,DOS} = 1.09\ m_0$ for the calculation of the DOS effective mass. For the conductivity we should again use the harmonic mean:

$$m^*_{CB,\sigma} = 3 \cdot \left(\frac{2}{m_t^*} + \frac{1}{m_l^*} \right)^{-1} \tag{A.9}$$

This yields $m^*_{CB,\sigma} = 0.26\ m_0$. The averaging proposed for the effective mass appearing in the thermal velocity yields a slightly higher value of $m^*_{CB,v} = 0.28\ m_0$ [77].

Appendix B

Degenerate semiconductors

Effects of high doping concentrations semiconductors become noticeable when the carrier densities approach the effective density of states. If carrier densities become yet higher, we call the semiconductor degenerate. Typical cases are the highly doped layers close to the contacts, or a semiconductor illuminated with high light intensity, or a junction under very strong forward bias. In the first case, we imagine a plasma of free carriers that floats between the immobile cores of the ionised dopants, the other two cases are characterized by a neutral plasma of electrons and holes.

B.1 Band gap narrowing (BGN)

A very fundamental effect of high carrier densities is the modification of the bandgap itself. To understand this, let us draw an analogy to the electrons of the hydrogen molecule; in that case, the overlap of the electronic wave functions between the cores yields the binding energy of the molecule. Similarly, the correlation energy of the free carrier plasma can become strong enough to alter the binding energy of the atoms within the semiconductor. In case of high doping we have the two types of doping as the correlation energies between the majority carrier plasma and the ionised cores are different for for the case of electrons and of holes. In case of high injection, the correlation energy of the neutral plasma due to free electrons and free holes is different again. All cases result in *band gap narrowing* (BGN), but the amount differs for every one of the three cases as in the first two cases only one of the majority Fermi-levels approaches the band edge, whereas in the third case the quasi Fermi-levels split and both approach their respective band-edges.

The effect of BGN is important for all expressions that involve n_i^2 since this quantity scales exponentially with the bandgap energy. For solar cells, the most prominent example is the saturation current density. Some of the experimental data on BGN was even determined from studies of $j_{0,e}$ in highly doped emitters. The literature was critically reviewed by Altermatt and related to a rigorous quantum-mechanical treatment of BGN [18] as older literature did not always properly account for the various effects. Figure B.1 illustrates this for the two polarities of silicon. Thick full lines represent the quantum mechanical calculation whereas dashed lines in illustrate a simple empiric model for BGN that predicts the same extent for both [79]. Once the positions of the bands are determined, we can use them as reference to overlay E_F.

B.2 Fermi-level position

For non-degenerate doping and as long as the Fermi-level is more than than 3 kT away from the band-edge, we can use the Maxwell-Boltzmann approximation which resulted in eqns. (3.14) and

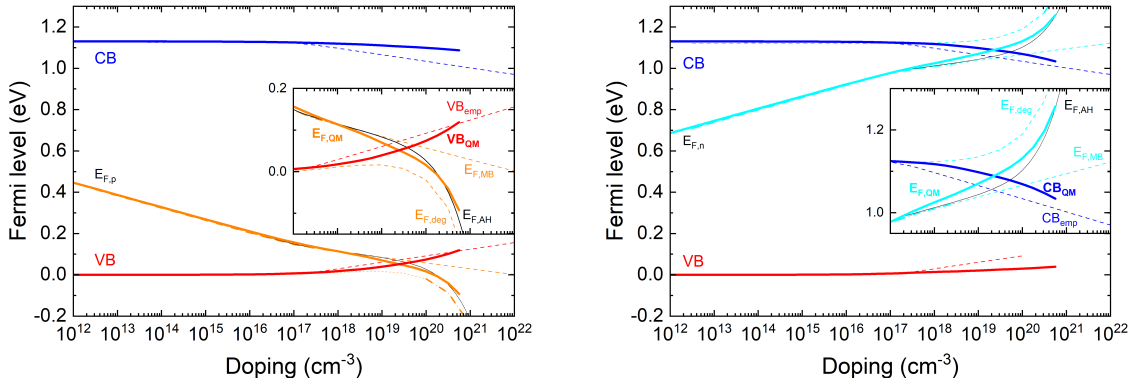


Figure B.1: Fermi-level and band structure with respect to carrier density of p-type (left) and n-type silicon (right). Thick continuous lines illustrate results of a quantum mechanical calculation. Thin dashed lines represent empiric models for BGN and the approximations for non-degenerate and degenerate statistics. For the latter two, the inset illustrates how they deviate from the rigorous results.

(3.15). We reproduce the result for the electron density n, the one for the hole density p reads accordingly:

$$n = N_C \exp\left(\frac{-(E_C - E_{F,n})}{kT}\right) \tag{B.1}$$

For degenerate doping, we can once more use an approximate treatment when the Fermi-level is more than 3 kT from the band edge. In this case, we use the 0 K approximation which is commonly applied to metals. This means that the Fermi-Dirac statistics simplifies into a step function. Starting from the state with the lowest energy at the origin of k-space, we fill all states below the Fermi-energy with two carriers of opposing spin until the ensemble of the filled states occupies the so-called Fermi-sphere around the origin. Then, the carrier density n corresponds to the number of unit cells that we can fill into this sphere. The radius of the sphere is the Fermi wave-vector which is given by $k_F^3 = 3\pi^2 n$. Accordingly, the Fermi energy is $E_F = (\hbar k_F)^2/2m^*$. This yields:

$$n = \frac{1}{3\pi^2}\left(\frac{2m^*(E_{F,n} - E_C)}{\hbar^2}\right)^{\frac{3}{2}} = \frac{4}{3\sqrt{\pi}} N_C \left(\frac{(E_{F,n} - E_C)}{kT}\right)^{\frac{3}{2}} \tag{B.2}$$

Both relations can easily be inverted to give $E_{F,n}$ as function of n. In fig. B.1 they are plotted by the dashed lines with respect to the empiric relation for BGN. The insets illustrate their respective ranges of validity at $\pm 3kT$ away from the band edges. For the region in between we have to either carry out the correct integration, or use a somewhat more difficult approximation as proposed by Aymerich and Humet [80]:

$$n \approx N_C \left(e^{-E_F/kT} + \frac{3\sqrt{\pi}}{4}\left[E_F/kT + 2.13 + \left(|E_F/kT - 2.13|^{2.2} + 2.6^{2.2}\right)^{\frac{1}{2.2}}\right]^{-\frac{3}{2}}\right)^{-1} \tag{B.3}$$

The latter relation is shown by the black line connecting the non-degenerate and the degenerate approximation. The accuracy of eq. (B.3) is better than 0.5% in the critical region around the

band edge and it gives the correct asymptotic behaviour, meaning that it contains the exponential relationship of the non-degenerate case as well as the power law of the degenerate case.

B.3 Ionized impurity scattering

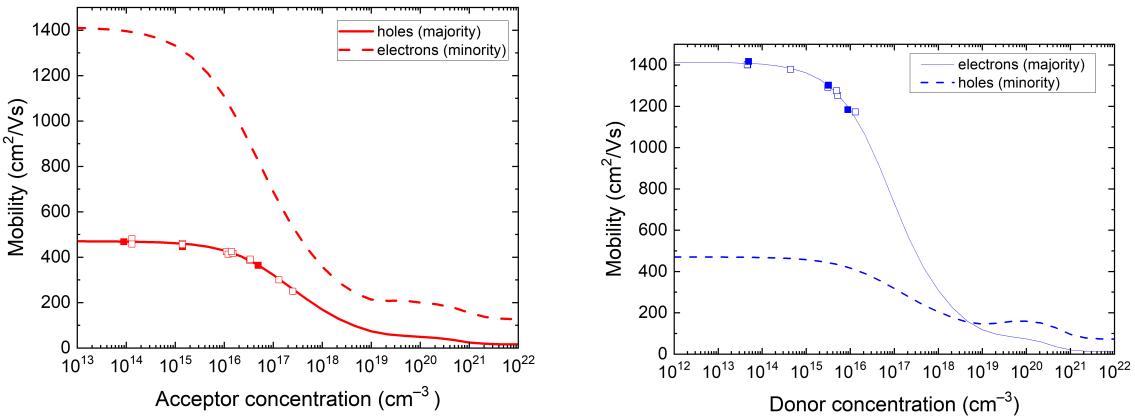


Figure B.2: Mobilities p-Si (left) and n-Si (right). Full and dashed lines illustrate the majority and minority mobility, respectively. Data from PV-Lighthouse.

High carrier densities have a profound impact on the charge carrier mobilities. At low doping concentration, the mobility is limited mostly by phonon scattering, i.e. by lattice vibrations. At high doping concentrations, charge carriers are scattered in the potential of the ionised impurities. Carrier transport in the presence of scattering is described by the Boltzmann transport equation after inserting an adequate expression for the differential scattering cross section σ in the collision term. The Coulomb interaction between charged scattering partners such as a free charge carrier and an ionised core is described by the Rutherford scattering cross section. The expression is identical for attractive and repulsive potentials, therefore we would not expect a difference for majority or minority carriers. Nevertheless, in highly doped material the Coulomb potential becomes increasingly shielded by the plasma of the free carriers, turning it eventually into a Yukawa potential. We distinguish Debye shielding for low doping and Thomas-Fermi shielding for the degenerate case. The hump in the minority mobility of fig. B.2 is explained by the fact that other than the Coulomb force, the shielding effects are not the same for attractive and repulsive interactions. The former apply to majority carriers and their ionised cores, the latter to minority carriers and the ionised cores, and the difference becomes especially important for high doping concentrations.

B.4 Free carrier absorption

When we describe the charge carriers in the solid by electronic wave functions, semiconductors are no different from metals. Thus, we should be able to describe the optical properties of free charge carriers similarly to the description of metals, except with a much lower density. Under the assumption of a solid filled by a plasma of free carriers, Drude obtained the following frequency dependence of the permittivity [19, 20]:

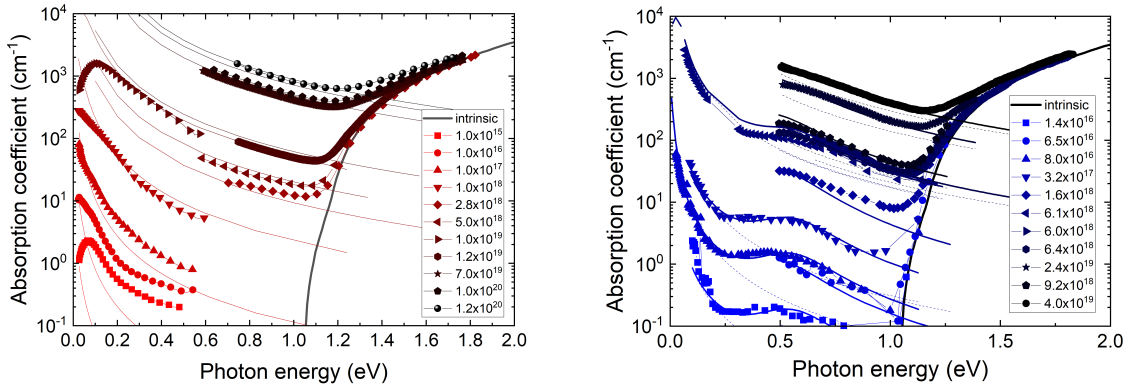


Figure B.3: Spectral absorption coefficient of p-Si (left, B-doped) and n-Si (right, As and Sb doped). The free carrier part at low energies varies like $(\hbar\omega)^{-2}$ in accordance to Drude theory (lines). In n-Si, the humps at 0.6 eV are due to an absorption within the CBs. Data digitized from [81, 82, 83].

$$\epsilon(\omega) = \underbrace{1 - \frac{\omega_p^2\tau^2}{1+\omega^2\tau^2}}_{\epsilon_1} + i\underbrace{\frac{1}{\omega\tau}\cdot\frac{\omega_p^2\tau^2}{1+\omega^2\tau^2}}_{\epsilon_2} \tag{B.4}$$

Here, ω_p is the plasma frequency given by $\omega_p = \sqrt{nq^2/(\epsilon_0 m^*)}$ where n is the electron density (replaced by p in case of holes) and τ is the relaxation time.[1] If the frequency of the incident light is less than ω_p, the free carrier plasma is able to follow the excitation and prevent the electromagnetic field from entering the medium. Consequently, the incident light is reflected in this range of frequencies. In case the frequency of the incident light exceeds ω_p, the plasma can no longer follow the excitation, and consequently the medium becomes transparent. In between, there is the region of free carrier absorption (FCA). The absorption coefficient associated to FCA is illustrated by the lines in figure B.3. Note that in case of n-Si there is an additional absorption effect 0.6 eV due to intra-band transitions [81].

The excitation of plasma oscillations does not create electron-hole pairs but plasma oscillations that ultimately dissipate the absorbed energy into heat. The FCA is therefore considered a parasitic absorption effect. In solar cells we can reduce absorption losses in the n-doped region at the front by using a selective emitter where the doping concentration between the fingers of the metal grid is reduced.

B.5 Burstein-Moss shift

Besides absorption in the IR, the results in the previous section suggest that also the fundamental absorption should change with high doping concentration. Normally a semiconductor absorbs light whenever the photon energy exceeds the band gap energy and the onset is exactly at the bandgap energy when an electron from the VBM gets excited to the CBM. However, in a degenerately n-doped semiconductor the lowest states of the CB are occupied by a electrons. Likewise, a degenerate hole density means that the states at the VBM are empty (filled by holes). In both cases, transitions

[1] The relaxation time τ depends on the effective mass and the mobility. For the former we should use conductivity effective masses as discussed in appendix A.1. For the latter, he dashed lines in fig. B.3 were modelled with the theoretical mobilites of fig. B.2. In a few cases lower values would be needed to better fit the data.

involving states directly at the band-edges are forbidden, either because there are no electrons in the initial states at the VBM, or because the final states at the CBM are occupied. The absorption onset is thus shifted to an energy that can excite electrons from the highest VB state filled by electrons into the lowest CB state that is free. The widening of the optical gap with respect to the bandgap is called Burstein-Moss effect.

Figure B.1 suggests that we can expect the Burstein-Moss effect in n-doped silicon for carrier densities higher than 10^{19} cm^{-3}. For lower doping, BGN prevails. In fig. B.3 neither of the effects is directly visible since the weak absorption of silicon close to the bandgap is easily hidden below an absorption tail of FCA.

B.6 Coulomb-enhanced recombination

Historically, highly doped samples were used to investigate the Auger effect in silicon and to determine the Auger coefficients in eq. (5.14). Later measurements with lowly doped material found larger coefficients, suggesting a dependence on the bulk doping [37]. The enhancement was eventually explained by Coulomb attraction between electrons and holes which can form a bound state called exciton [27]. In highly doped samples the plasma of free-carriers increasingly shields the Coulomb attraction and thus the enhancement is no longer observed. The resulting lifetime with Coulomb-enhanced Auger recombination is illustrated figure B.4 by the blue and red lines for n-type and p-type material, respectively.

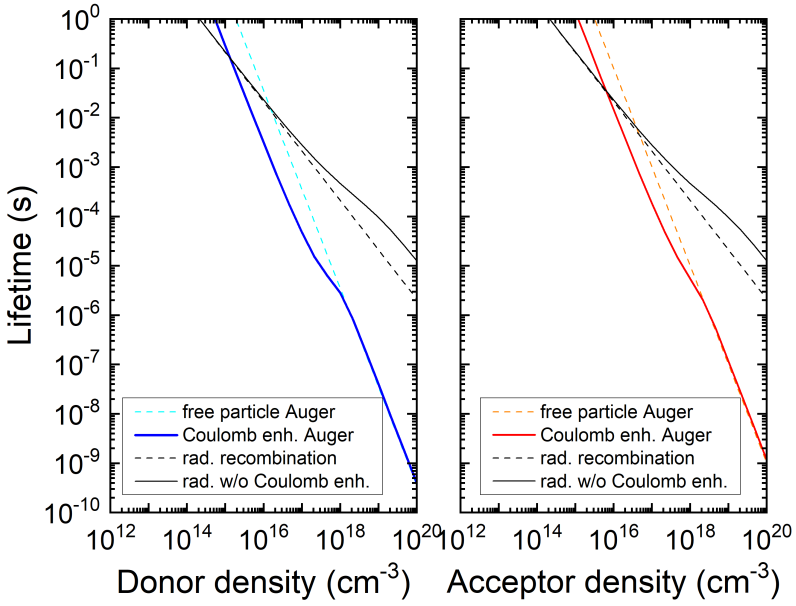


Figure B.4: Impact of Coulomb-enhancement on the lifetimes associated to radiative recombination (black) and Auger recombination (blue and red).

For the case of radiative recombination it seems that there was a different sequence of events. The generally accepted value for the radiative recombination coefficient was measured on very lowly doped silicon [25]. Later it was concluded that this yields a coefficient that is already enhanced by Coulomb attraction; consequently it should be called B_{low} and at high doping concentration it should be multiplied by a factor B_{rel} as discussed by Altermatt [84]. The resulting radiative

lifetimes are illustrated in figure B.4 by the full black lines.

Appendix C

The efficiency limit of c-Si solar cells

In the main text, we wrote at several places that the efficiency of c-Si solar cells is limited by Auger recombination. In this chapter we will discuss the main elements that are needed to come to this conclusion. The line of thought is as follows: Looking only at the value of the bandgap, c-Si would be able to generate a photocurrent of $46\,\mathrm{mA\,cm^{-2}}$ under the AM1.5g spectrum. Since silicon has a low absorption coefficient, we need a thick cell to get close to the potential photocurrent. However, by spreading the generation over a large cell thickness we reduce the resulting excess carrier density. According to eq. (6.23) this results in a low V_{oc}. On the other hand, in a very thin cell the current would be lower, but the generation rate and thus the excess carrier density would be higher, leading to a higher V_{oc}. Somewhere in between there is an optimum thickness, and it turns out that the corresponding generation rate falls into the region where Auger recombination dominates over radiative effects.

We follow an argument of Tiedje [41] who first pointed out the limitation of c-Si solar cells by Auger recombination. According to the superposition principle, the $j(V)$ characteristic is simply the sum of the recombination current and the photocurrent where the former is expressed with the recombination rate R and the cell thickness d:

$$j(V) = qdR - j_L = qd\Delta n/\tau_{eff} - j_L \tag{C.1}$$

The voltage V enters through the excess carrier density Δn, and indirectly two more times through the dependence of R and j_L on Δn. To relate V and Δn, we make use the np-product.

$$np = (n_0 + \Delta n)(p_0 + \Delta n) = n_i^2 \exp\left(\frac{E_{F,n} - E_{F,p}}{kT}\right) \tag{C.2}$$

To identify the voltage in this expression, we have to apply ideal selective contacts to the cell. In that case, the potential between the two external electrodes is the same as the splitting between the QFLs inside of the semiconductor. To construct the $j(V)$ curve, we start by setting an operating voltage V and resolve the np-product for Δn. The solution is slightly complicated by the fact that also n_i depends on Δn. This is due to band gap narrowing in case of high excess carrier densities [18].

C.1 Photocurrent

To calculate the photocurrent, we follow again the derivation of eq. (4.16), i.e assuming zero reflection at the front, and perfect reflection at the rear. To account for light scattering, we modify

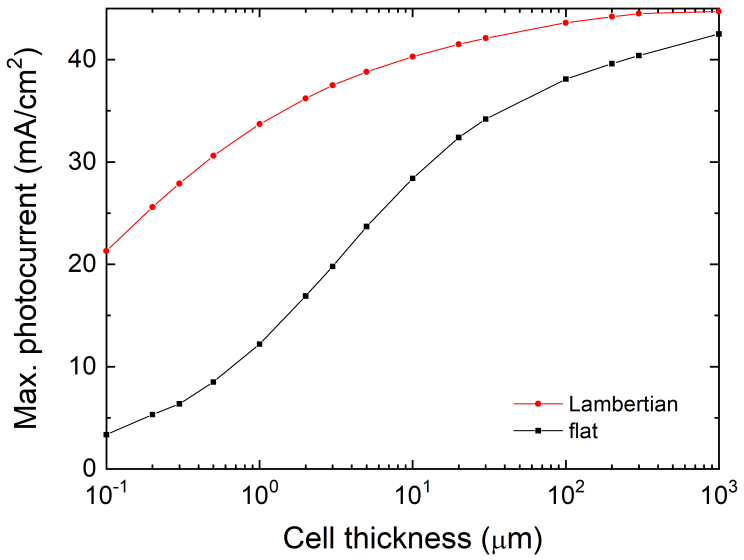


Figure C.1: Maximum achievable photocurrent in c-Si vs. the absorber thickness, assuming pure band-to-band absorption, zero front reflection and perfect reflection at the rear. Circles and squares denote flat interfaces and Lambertian scattering, respectively.

the outcoupling at the front as discussed in the section after eq. (4.16), and to account correctly for the transmitted intensity after each of the passages through the absorbing medium, we replace each occurrence of $\exp(-\alpha d)$ by a more general expression $T(\alpha d)$.

$$A_{bb} = \frac{\alpha_{bb}}{\alpha_{bb} + \alpha_{FCA}} \frac{(1 - T(\alpha d)) + T(\alpha d)\,(1 - T(\alpha d))}{1 - T^2(\alpha d)(1 - 1/n_{Si}^2)} \tag{C.3}$$

For Lambertian light scattering and a weakly absorbing medium, the average path length enhancement is equal to two. i.e. $T(\alpha d) = \exp(-2\alpha d)$. In all generality the approximation of weak absorption is not fulfilled, and then the angular distribution of Lambertian scattering is no longer maintained within one passage through the medium. If we can still assume that the angular distribution is re-established by the scattering event at the next interface, $T(\alpha d) = \exp(-\alpha d) \cdot (1 - \alpha d) + (\alpha d)^2 \cdot \mathrm{E_i}(\alpha d)$ where $\mathrm{E_i}(\alpha d)$ is the exponential integral [6].

In eq. (C.3), the absorption coefficient α is the sum of two parts, $\alpha = \alpha_{bb} + \alpha_{FCA}$ where α_{bb} is the regular band-to-band absorption of silicon which yields electron-hole pairs and ultimately the photocurrent. The term α_{FCA} is due to free carrier absorption as discussed in appendix B. Alternatively to deriving the absorption coefficient from eq. (B.4), a parametrisation for n- and p-doped silicon was proposed by by Rüdiger [85].[1] Finally, eq. (C.3) includes a weighting factor which ensures that the photocurrent is calculated only with that part of the absorbance that actually creates electron-hole pairs.

To calculate the photocurrent, we weight the absorbance A_{bb} of eq. C.3 with the incident photon flux, and we integrate over the spectral range of interest. The photon flux is obtained from the spectral illumination intensity after dividing by the photon energy $E = hc/\lambda$, and for the ideal cell it is assumed that every absorbed photon creates an electron-hole pair.

[1] Note that our treatment is slightly inconsistent. To do it correctly, we should consider the different susceptibilities, compose the permittivity, and calculate the refractive index and the absorption coefficient from there.

$$j_L = q \int_{300}^{1200} A_{bb}(\lambda) \cdot \Phi(\lambda)/E \cdot d\lambda \tag{C.4}$$

Figure C.1 shows the dependence of j_L on the cell thickness d for flat interfaces and for ideal Lambertian scattering (in both cases for short circuit, i.e. without the effect of free carrier absorption). In the former case, a device thickness of more than 1000 µm would be needed for a photocurrent above 40 mA cm^{-2} whereas 10 µm would be sufficient with ideal light scattering.

C.2 Recombination rate

The recombination rate R is the sum of R_{rad} and R_{Auger} which are inherent mechanisms. Assuming an ideal cell in all other aspects, we will stick to these two in the following, SRH recombination and surface effects will be discussed at the end of the next section.

C.2.1 Radiative recombination

Let us start with the radiative recombination rate R_{rad}. It is expressed in terms of the radiative recombination coefficient B_{low} which is obtained by integrating over the spectrally resolved rate of spontaneous emission [25].

$$B_{low} = \int B(E) \cdot dE = \int \frac{n_{Si}^2 E^2 \alpha(E)}{\pi^2 c^2 \hbar^3 n_i^2} \cdot \frac{1}{\exp(E/kT) - 1} dE \tag{C.5}$$

For lowly doped material and low excess carrier density we can make a Boltzmann approximation, resulting in $B_{low} = 4.73 \times 10^{-15}\,\mathrm{cm^3\,s^{-1}}$ for silicon. At high excess carrier density, a correction factor B_{rel} should be used [86]. The net radiative recombination rate is thus given by:

$$R_{rad} = \left(np - n_i^2\right) B_{rel} B_{low} \tag{C.6}$$

The photons emitted by a radiative recombination event can either leave the cell through the escape cone, or they can be re-absorbed by the semiconductor in a process called photon recycling. In the latter case, electron-hole pairs are generated and subject to the recombination rate once more, just as if the preceding events of radiative recombination and re-absorption had not taken place. Consequently, we can use a reduced recombination rate $R'_{rad} = (1 - P_{PR})R_{rad}$ where P_{PR} is the probability of photon recycling [5]. To find P_{PR}, we need the spectrally resolved radiative recombination coefficient $B(E)$ [25]:

As emission and absorption are reciprocal phenomena, the probability of photon recycling can be expressed in terms of the cell absorbance $A(\lambda)$ according to the following expression [5]:

$$P_{PR} = \frac{\int A(\lambda) \cdot B(E) \cdot dE}{\int B(E) \cdot dE} \tag{C.7}$$

To evaluate the integrals, we use $E = hc/\lambda$ and replace the differentials according to $dE = (dE/d\lambda)d\lambda$. The absorbance $A(\lambda)$ was discussed in the previous section and the integral in the denominator yields once again the radiative recombination coefficient B_{low}.

C.2.2 Auger recombination

The first estimate of the limiting efficiency by Tiedje was based on the recombination rate as stated in eq. (5.14) [41]. As discussed in appendix B, it was later found that Coulomb attraction enhances the Auger effect in lowly doped silicon [28]. Based on additional experimental data, Richter adjusted the parametrisation for the recombination rate as follows [29]:

$$R_{Auger} = \left(np - n_i^2\right)\left(2.5 \times 10^{-31} g_{eeh} n_0 + 8.5 \times 10^{-32} g_{ehh} p_0 + 3.0 \times 10^{-29} \Delta n^{0.92}\right) \tag{C.8}$$

The terms g_{eeh} and g_{ehh} describe the Coulomb enhancement at low excess carrier density. They are given by $g_{eeh} = 1 + 13 \cdot (1 - \tanh\left[(n_0/N_{0,eeh})^{0.66}\right])$ with $N_{0,eeh} = 3.3 \times 10^{17}\,\mathrm{cm}^{-3}$ and $g_{ehh} = 1 + 7.5 \cdot (1 - \tanh\left[(p_0/N_{0,ehh})^{0.63}\right])$ with $N_{0,ehh} = 7.0 \times 10^{17}\,\mathrm{cm}^{-3}$.

C.3 Efficiency limit

We are now in a position to evaluate R and j_L for different voltages and thus to assemble a $j(V)$ characteristic from which we can determine the MPP and the efficiency. To find the limiting efficiency, we repeat the procedure for different cell thicknesses. The result is shown in figure C.2. We find a maximum of 29.6% for a thickness of ca. 100 µm which is in agreement with the most recent calculation of the limiting efficiency by of Schäfer [6]. Earlier results by Kerr and by Richter are included for reference.

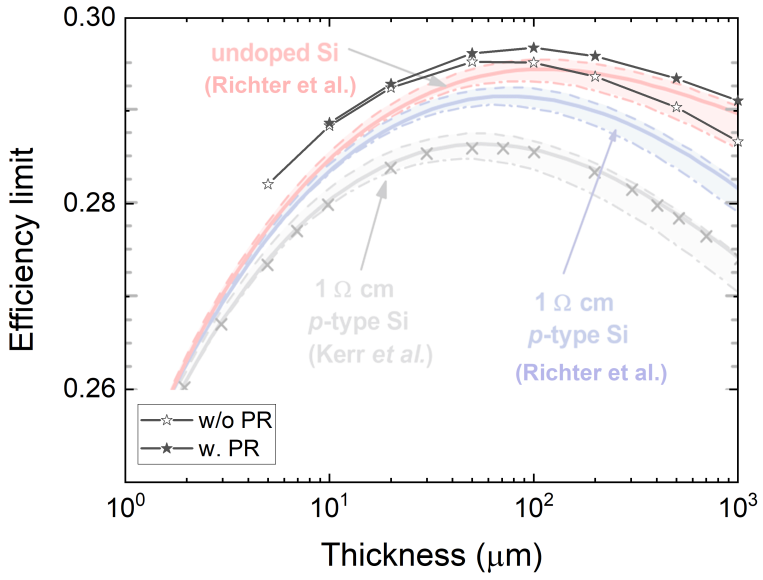


Figure C.2: Efficiency limit for c-Si solar cells with ideal Lambertian light scattering, w/o and with the effect of photon recycling (stars). Shaded characteristics refer to earlier calculations by Kerr and by Richter [28, 5].

Figure C.2 also shows the importance of photon recycling. Looking at eq. (C.5), the emission of photons with high energy is very unlikely because of the exponential term. Towards the other end of the spectrum, the appearance of α prevents emission of photons with energy below the bandgap except for those that are mediated by phonon assistance and some more in the IR that

originate from α_{FCA}. Taking both effects together, the energy of the photons emitted by radiative recombination peaks very closely to the bandgap energy. However, the absorption coefficient of c-Si is very low for these energies, giving the photons a fair chance to be emitted through the front interface. Thus, we can only hope to recycle an appreciable amount of photons if we use cells with thickness greater than 200 µm.

We can modify the calculation of the efficiency limit to gain insight into some experimental limitations. First of all, we can add recombination effects in the bulk and at the two surfaces. The former can be achieved by simply a term with an empiric value for the lifetime into the recombination rate, using $R = \Delta n/\tau$. This is also used in the parametrization of the Auger effect [29]. The latter can be implemented similarly by using empiric values for the SRVs at the two surfaces according to eq. (5.30). In this case we have to keep in mind that the contribution to the effective lifetime is different for every modelled cell thickness. Figure C.3 illustrates this for lifetimes of 1 and 10 ms and SRVs of 1 and 10 $\mathrm{cm\,s^{-1}}$. Regarding first the V_{oc}, the high lifetime and the low SRV do not show up appreciably, but choosing a low lifetime yields low V_{oc} for large device thickness whereas high SRV is most noticeable at low thickness.

Next, let's check the FF. We already mentioned in the main text that it is intrinsically related to the V_{oc} through the Lambert W-function, or in a good approximation by $FF_0 \approx qV_{oc}/(qV_{oc} + 4.7nkT)$ [2]. In figure C.3, the FF values modelled for pure Auger recombination are rather high since in the case the diode quality factor is $n = 0.7$. When we choose $\tau = 1$ ms and $S = 10\,\mathrm{cm\,s^{-1}}$, the FF is massively reduced and approaches the characteristic for $n = 2$ for large and small device thickness, respectively. Assuming less severe recombination by $\tau = 10$ ms and $S = 1\,\mathrm{cm\,s^{-1}}$, the loss in FF is not as drastic and dashed line illustrates that the combined effect of the two assumptions complies well with the values reported by AIST for their pFF, i.e. the FF without effects of the series resistance [87]. The pFF as well as the FF values reported by Longi for their record cells suggest bulk lifetimes above 10 ms and surface recombination velocity below 1 $\mathrm{cm\,s^{-1}}$ [88].

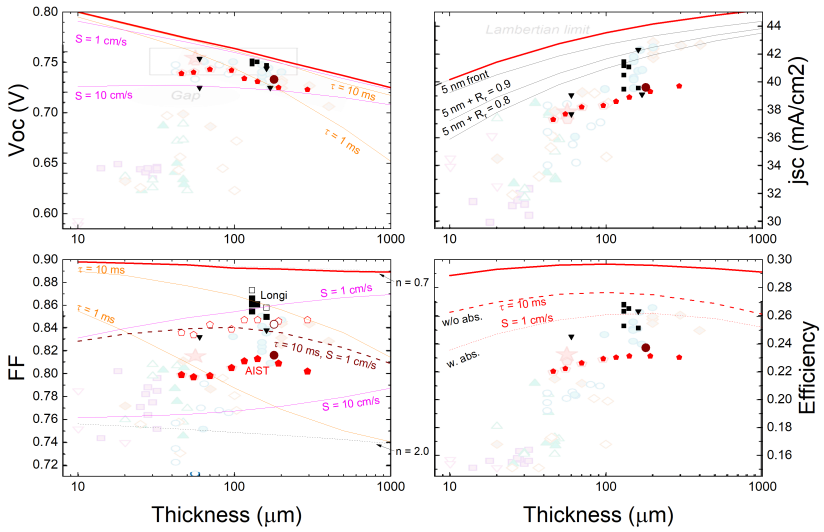


Figure C.3: Efficiency limit for c-Si solar cells after [89], additional data from [87, 88].

We can also include optical losses into the model of the efficiency limit. The eq. (4.16) already includes the option to describe primary reflection at the front as well as parasitic absorption losses

in the rear reflector. In addition we can include parasitic absorption in a layer at the front by slightly modifying the terms entering the infinite sum. For the panel of the j_{sc} in figure C.3, we kept perfect anti-reflection, but we represent the front contact with a layer that absorbs like silicon and has a thickness of 5 nm. This yields a uniform reduction of about $1\,\mathrm{mA\,cm^{-2}}$. Adding a less than perfect rear-reflector further reduces the current density predominantly for the thinner cells.

Appendix D

Finite cell

This chapter discusses the impact of the surface in case of finite wafer thickness. Figure D.1 illustrates this for a solar cell with a metallic electrodes that form Schottky junctions with the semiconductor. We will discuss the effect of the surface for the case of electrons in a p-region of finite thickness H, using a procedure similar to the development of section 6.2.

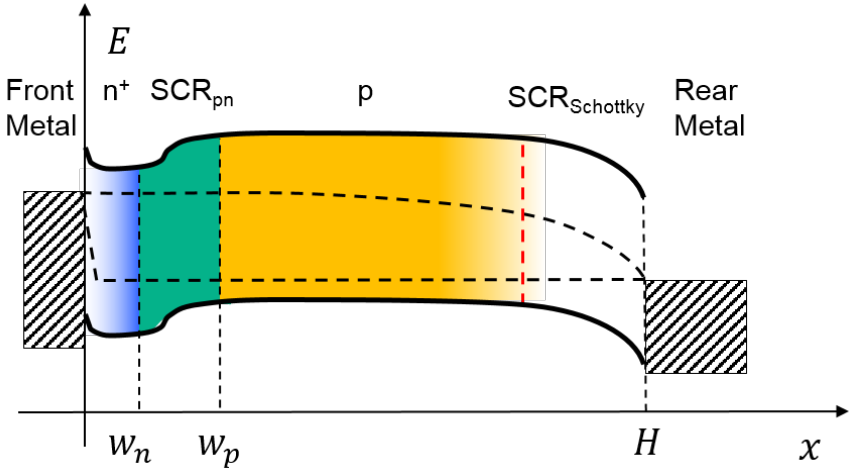


Figure D.1: Schematic band diagram for a solar cell with metallic rear contact under forward bias in dark. The red dashed line illustrates where the effective surface recombination velocity S_{eff} is at a position where the band bending of the Schottky junction has not yet started (the width of the space charge region is exaggerated for illustration).

D.1 Saturation current density of the finite bulk in dark

For the description we start once more with eq. (6.9):

$$\frac{d^2n}{dx^2} - \frac{n}{D_n\tau_n} = -\frac{n_{p,0}}{D_n\tau_n} \tag{D.1}$$

For the finite case we must consider the decaying as well as the increasing part of the homogeneous solution, and since the inhomogeneity is constant, we have once more a very simple particular solution:

$$\begin{aligned} n_{hom} &= c_1 e^{\frac{x}{L_n}} + c_2 e^{-\frac{x}{L_n}} \\ n_{part} &= n_{p,0} \end{aligned} \tag{D.2}$$

We apply the following boundary conditions:

1. At $x = w_p$: We request that the minority carrier density at the edge of the SCR is defined by injection (or extraction) according to eq. (6.6). Since the p-doped bulk is generally much thicker than the SCR of the pn-junction, we may neglect w_p and evaluate for $x = 0$ instead:

$$n(0) = c_1 + c_2 + n_{p,0} \stackrel{!}{=} n_{p,0} e^{\frac{qV}{kT}}$$

2. At $x = H$: We do not know the carrier density at the rear surface, but we do know the current that flows through it. Thus, we request that the diffusion current density according to eq. (3.42) equals a surface current density as defined by eq. (5.28).

$$j(x = H) = qD_n \left.\frac{dn}{dx}\right|_{x=H} \stackrel{!}{=} qS\left(n(H) - n_0(H)\right) \tag{D.3}$$

Thus, our second boundary condition becomes:[1]

$$qD_n \left[\frac{c_1}{L_n} e^{\frac{H}{L_n}} - \frac{c_2}{L_n} e^{-\frac{H}{L_n}}\right] \stackrel{!}{=} qS_{eff}\left(c_1 e^{\frac{H}{L_n}} + c_2 e^{-\frac{H}{L_n}}\right) \tag{D.4}$$

The two boundary conditions in eqns. (D.3) and (D.4) allow us to determine the constants c_1 and c_2, and to derive an analytic expression for the concentration of the electron density in the p-region:

$$n(x) = \frac{\left[n_{p,0}\left(e^{\frac{qV}{kT}} - 1\right)\right] \cdot \left[S \sinh(\frac{H-x}{L_n}) + \frac{D_n}{L_n} \cosh(\frac{H-x}{L_n})\right]}{S \sinh\frac{H}{L_n} + \frac{D_n}{L_n} \cosh\frac{H}{L_n}} + n_{p,0} \tag{D.5}$$

Finally, the current density is once again obtained by applying eq. (3.42) and evaluating the result at $x = w_p \approx 0$. After rearranging a little, we obtain:

$$j_n = \underbrace{\frac{qD_n n_i^2}{N_A L_n} \left\{\frac{S \cosh\frac{H}{L_n} + \frac{D_n}{L_n} \sinh\frac{H}{L_n}}{S \sinh\frac{H}{L_n} + \frac{D_n}{L_n} \cosh\frac{H}{L_n}}\right\}}_{j_{0,p}} \left(e^{\frac{qV}{kT}} - 1\right) \tag{D.6}$$

The term in the round brackets is once again the dependence on the applied voltage and the whole term before can be defined as saturation current density $j_{0,p}$. The difference to the infinite cell of eq. (6.14) is given by the term in curly brackets which is called geometry factor $G_{F,p}$. If we set $G_{f,p} =$

[1] In principle we should write S_{eff} because we do not refer to the real interface between silicon and the rear contact, but to a virtual surface within the wafer for which the condition of purely diffusive transport still holds. For the example of a metallic contact as shown in figure D.1, we would define this virtual interface at the edge of the SCR where the band bending of the Schottky contact starts.

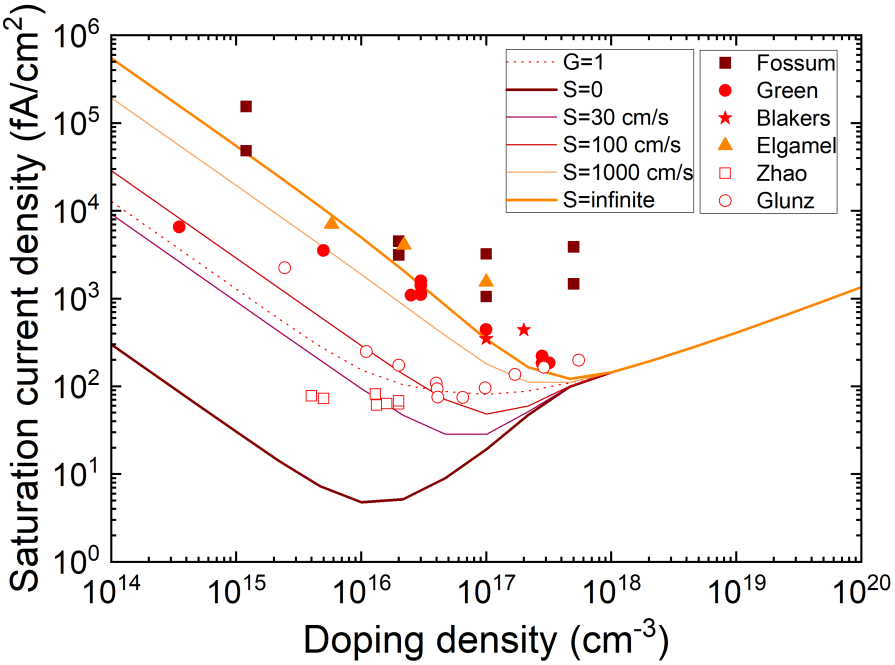


Figure D.2: Saturation current density of a 200 µm thick bulk region with respect to the acceptor density N_A. Full symbols refer to experimental results with metallisation on the full rear area [90, 42, 40, 91], open symbols refer to PERC and PERL cells [92, 93].

1, we reproduce the result of an infinite bulk region that was discussed in figure 6.5. Figure D.2 illustrates the impact of G_F for a 200 µm thick bulk. For carrier densities above $5 \times 10^{17}\,\mathrm{cm}^{-3}$, the surface recombination velocity has no effect because doping levels above this value reduce the diffusion length to values below 200 µm and surface effects are simply not noticeable at the junction. For lower doping, $j_{0,p}$ shows a strong dependence on the surface. For S_{eff} in the range of 0 - $10\,\mathrm{cm\,s}^{-1}$, $j_{0,p}$ can be reduced by almost two orders of magnitude with respect to the infinite case, and the optimum bulk doping concentration is found at $1 \times 10^{16}\,\mathrm{cm}^{-3}$. As eq. (1.7) tells us that the V_{oc} increases logarithmically with decreasing j_0, such a gain could improve the V_{oc} by almost 120 mV.

Figure D.2 includes some experimental data of solar cells with passivated emitters at the front. At the rear they had either full-area metallic back contacts without highly doped region, or localized contacts with passivation. We can thus assume that j_0 is dominated by the rear. The former fall close to the characteristic of infinite S_n, the latter are closer to $S = 100\,\mathrm{cm\,s}^{-1}$. Overall, the agreement suggests that the geometry factor of our rather simple treatment gives a powerful description of the dependence on the wafer doping concentration.

D.2 Saturation current density of electron contacts

In the majority of silicon solar cells, the electron contact at the front is obtained by diffusion in an atmosphere of $POCl_3$. Thus, the doping profile is not constant and the calculation of the geometry factor has to take into account a drift component. Consequently the continuity equation can no longer be solved analytically. In case of phosphorus diffusion, the concentration profile usually assumes a comparatively flat plateau close to the surface before it decays steeply into the

kink. We can therefore try to apply the simple model to the n-region as long as we are aware of its limitations. Figure D.3 shows experimental data of different diffusion conditions. The overlaid curves were determined under the assumption that the surface-concentration remains constant over the full n-region. Accordingly, n_i^2, D_p and L_p were evaluated for these concentrations and kept constant over the width of the n-layer. Finally, a thickness of $H = 150\,\mathrm{nm}$ was chosen to reach a "reasonable" correspondence with the data. Note that this is *not* the junction depth given by the authors for their different diffusion profiles. Let us look at the salient features first, a few words of caution are given at end of the section.

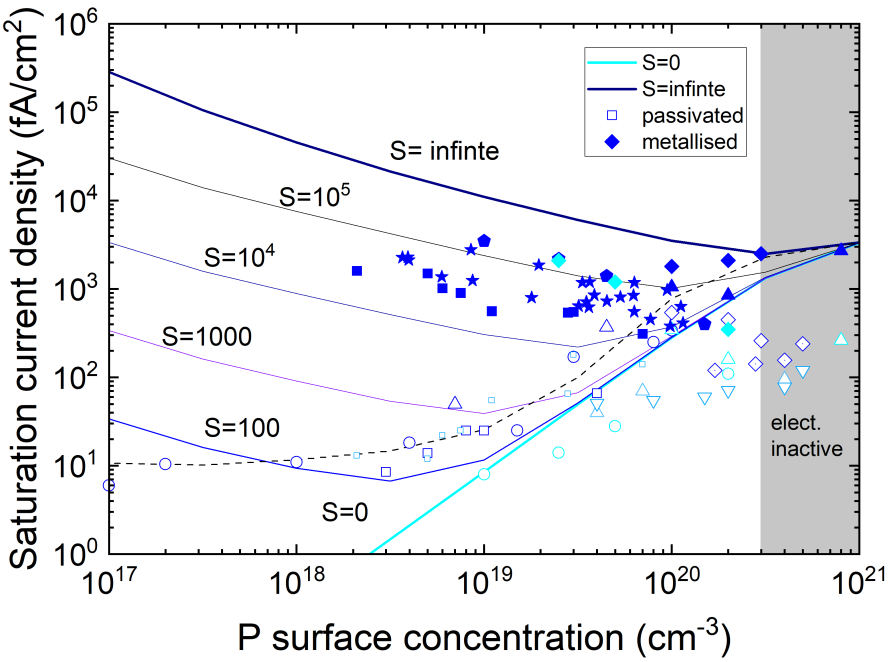


Figure D.3: Saturation current density of diffused n^+-regions with respect to surface concentration. Full symbols refer to samples without surface passivation, open symbols denote samples with SiO_2 passivation. Full lines correspond to a G_F model with constant doping equal to $N_{D,s}$ and fixed values of S_{eff}, the dashed line takes into account the variation of the S_{eff} with $N_{D,s}$ for SiO_2 passivation.

Similar to the case of bulk doping in figure D.2, the influence on S_{eff} disappears for high doping concentrations at the right side of figure D.3. As before, this happens when the diffusion length drops below the thickness of the n-region, here for a dopant concentration of $2 \times 10^{20}\,\mathrm{cm}^{-3}$. This suggests that even a 150 nm thick region can shield the interior of the cell from surface effects as long as it doped highly enough. Unfortunately, at these doping concentrations j_0 is already strongly increased due to BGN.

A detailed investigation of surface passivation effects revealed that S_{eff} of the Si/SiO_2 interface is not constant, but actually it varies with the surface concentration $N_{D,s}$ [94, 95], see also figure G.5. The dashed line in figure D.3 takes into account this variation of S_{eff} with $N_{D,s}$ for SiO_2 passivated samples. In the region of low doping it corresponds extremely well with the experimental data. Appendix G discusses the effect in more detail for the interface between silicon and silicon oxide.

In order to estimate the saturation current density of the front contact in a typical solar cell, we can proceed as follows: Let us assume that we have a surface concentration of $1 \times 10^{20}\,\mathrm{cm}^{-3}$

after the diffusion of phosphorus. For the metallised area the surface recombination velocity S_{eff} is around $1 \times 10^5\,\mathrm{cm\,s^{-1}}$ and we read off $j_{0,met} \approx 1000\,\mathrm{fA\,cm^{-2}}$. Between the fingers we can assume that the surface is well passivated by a dielectric, resulting in $S_{eff} = 100\,\mathrm{cm\,s^{-1}}$ and we read off $j_{0,pass} \approx 400\,\mathrm{fA\,cm^{-2}}$. The total saturation current can now be calculated by weighting these values with their respective area fractions:

$$j_{0,tot} \approx f \cdot j_{0,met} + (1 - f) \cdot j_{0,pass} \tag{D.7}$$

Here, f is the fraction of the metallised area and the example suggests $j_{0,tot} \approx 460\,\mathrm{fA\,cm^{-2}}$. As a second example, we can consider a process where the surface concentration of phosphorus was reduced to $1 \times 10^{19}\,\mathrm{cm^{-3}}$ by a drive-in diffusion. For the same fraction of metallised area, we can estimate $j_{0,tot} \approx (0.1 \cdot 2000 + 0.9 \cdot 20)\ \mathrm{fA\,cm^{-2}} = 218\,\mathrm{fA\,cm^{-2}}$. Plugging these values into eq. (1.7) with a typical photocurrent of $j_L = 40\,\mathrm{mA\,cm^{-2}}$ and forgetting for a brief moment the contribution of $j_{0,p}$, we estimate that these front contacts could potentially give V_{oc} of 655 and 674 mV, respectively. The difference of 20 mV is not as spectacular as the one in the previous section, but it is still significant and worth the effort of the drive-in annealing.

In advanced cells the metallisation fraction is obviously much smaller and a so-called *selective emitter* is used. In this case, the doping concentration is enhanced only in those regions where eventually a metal contact with high S_{eff} is added. In the regions in between where a passivating oxide or nitride is applied, the doping concentration can be reduced to the minimum that is needed for lateral current transport between the contact fingers. For the optimum values and $f = 0.05$, we can project $j_{0,tot} \approx (0.05 \cdot 1000 + 0.95 \cdot 10)\ \mathrm{fA\,cm^{-2}} = 59.5\,\mathrm{fA\,cm^{-2}}$.

Before we proceed, we should add a few words of caution to the interpretation of experimental data in figure D.3.

- The theoretical characteristics are based on Boltzmann statistics whereas for the dopant densities in the right part of the figure clearly Fermi-Dirac statistics should be used [95].

- The equation for G_F was derived by assuming a pure diffusion current and constant dopant concentration. Even if we use Boltzmann statistics, the presence of a diffusion profile yields band-bending and we have to take into account the drift current. Analytic solutions are more difficult and generally rely on additional assumptions [96, 97], or we have to revert to numerical simulation.

- The electron density n was approximated with the donor concentration. This is accurate only for low doping, but as the Fermi-level rises close to the CB, the ionisation probability of the dopant atoms drops to 75% for dopant concentrations around $2 \times 10^{18}\,\mathrm{cm^{-3}}$. For higher doping, it approaches unity again [98].

- For concentrations above $2 \times 10^{20}\,\mathrm{cm^{-3}}$, phosphorus is actually no longer soluble in silicon at the temperatures that are typically used for phosphorus diffusion [50]. The dopant source can obviously have higher concentration and force more phosphorus into a region close to the surface, but the excess above the solubiltiy limit is not incorporated on substitutional lattice sites and generally does not act as dopant. Instead, phosphorus forms inactive clusters that contribute to recombination [99]. The total phosphorus concentration is typically measured with SIMS (secondary ion mass spectroscopy) whereas the electrically active concentration can be measured with ECV (electrochemical capacitance-voltage profiling). Note that the data in figure D.3 was taken from a variety of sources, partially based on SIMS and partially on ECV.

A correct treatment of $j_{0,n}$ is discussed in ref. [95] and integrated into *EDNA*, a numerical calculator that is freely accessible on the website PVLighthouse. It offers a variety of predefined dopant profiles and it accepts also uploaded data.

D.3 Finite bulk with uniform generation

Let us continue with the finite bulk, but this time with a constant generation profile G. We can directly use the differential equation for the transport of minority carriers as discussed in section 6.3. The presence of another constant term changes only slightly the particular solution which is now given by $n_p = n_{p,0} + G$. The boundary conditions remain the same as in section D.1, i.e. injection according to the applied bias voltage at $x \approx 0$ and a current density according to S at $x = H$. The determination of the constants c_1 and c_2 becomes a bit more lengthy and yields the following equation for the electron density:

$$n(x) = \frac{\left[n_{p,0}\left(e^{\frac{qV}{kT}} - 1\right) - G\tau\right] \cdot \left[S\sinh(\frac{H-x}{L}) + \frac{D}{L}\cosh(\frac{H-x}{L})\right] - G\tau S \sinh\frac{x}{L}}{S\sinh\frac{H}{L} + \frac{D}{L}\cosh\frac{H}{L}} + n_{p,0} + G\tau \quad \text{(D.8)}$$

Here, the index n is once again dropped on S, D, and τ for simplicity. We can easily carry out the derivative to find the minority current density of the electrons in the p-region at $x \approx 0$:

$$j_n(V) = \frac{qD}{L}\frac{\left[n_{p,0}\left(e^{\frac{qV}{kT}} - 1\right) - G\tau\right] \cdot \left[S\cosh\frac{H}{L} + \frac{D}{L}\sinh\frac{H}{L}\right] + G\tau S}{S\sinh\frac{H}{L} + \frac{D}{L}\cosh\frac{H}{L}} \quad \text{(D.9)}$$

With can obtain the usual of form of a $j(V)$ characteristic, i.e. superposition of a dark current density that is composed of a saturation current density j_0 and an exponential variation with the voltage, and a photocurrent density j_L.

$$j_n(V) = \underbrace{\frac{qDn_i^2}{N_A L}G_F}_{j_0}\left(e^{\frac{qV}{kT}} - 1\right) - \underbrace{\frac{qDG\tau}{L} \cdot \frac{S\cosh\frac{H}{L} + \frac{D}{L}\sinh\frac{H}{L} - S}{S\sinh\frac{H}{L} + \frac{D}{L}\cosh\frac{H}{L}}}_{j_L} \quad \text{(D.10)}$$

To check consistency, we evaluate two extreme cases for the photocurrent:

- $S \to 0$: In this case, the expression for j_L simplifies to:

$$j_L = \frac{qDG\tau}{L}\tanh\frac{H}{L} \quad \text{(D.11)}$$

  For the case of a diffusion length L that is (much) less than the wafer thickness H, $\tanh(H/L) \to 1$ and because of $D\tau = L^2$ we find $j_L \approx qGL$. This means that we do not collect all carriers generated in the wafer, but only those within a distance of L from the junction. This resembles the case of the infinite cell and corresponds to the right part of figure D.4.

  A technologically more relevant case emerges for a diffusion length much larger than the wafer thickness. In that case, $\tanh(H/L) \approx H/L$ and $j_L \approx qDH$. This means that wafers with high quality allow us to collect carriers from the full volume, provided we also fulfil the overarching condition of low surface recombination velocity.

- $S \to \infty$: In this case, the expression for j_L simplifies to:

$$j_L = \frac{qDG\tau}{L} \frac{\cosh(H/L) - 1}{\sinh(H/L)} \tag{D.12}$$

Let us start by looking at the limit of a short diffusion length L. In that case, $\cosh(H/L) >> 1$ and the second fraction simplifies to $\coth(H/L) \to 1$. Once again, we collect only those carriers generated within a distance L of the junction.

If the diffusion length is much larger than the wafer thickness, the argument H/L becomes very small. Expanding the hyperbolic functions to second order, we end up with $j_L \approx qDH/2$ which means that we collect carriers only from the front half of the wafer thickness. The carriers generated in the rear half are lost because they recombine at the rear interface. Figure D.4 shows that the cells for which we concluded $S \approx \infty$ in figure D.2 appear to fall into this category, even though here we find a lesser value of $S \approx 1000\,\mathrm{cm\,s^{-1}}$. Moreover, these are historic devices and we cannot exclude additional effects of bulk recombination.

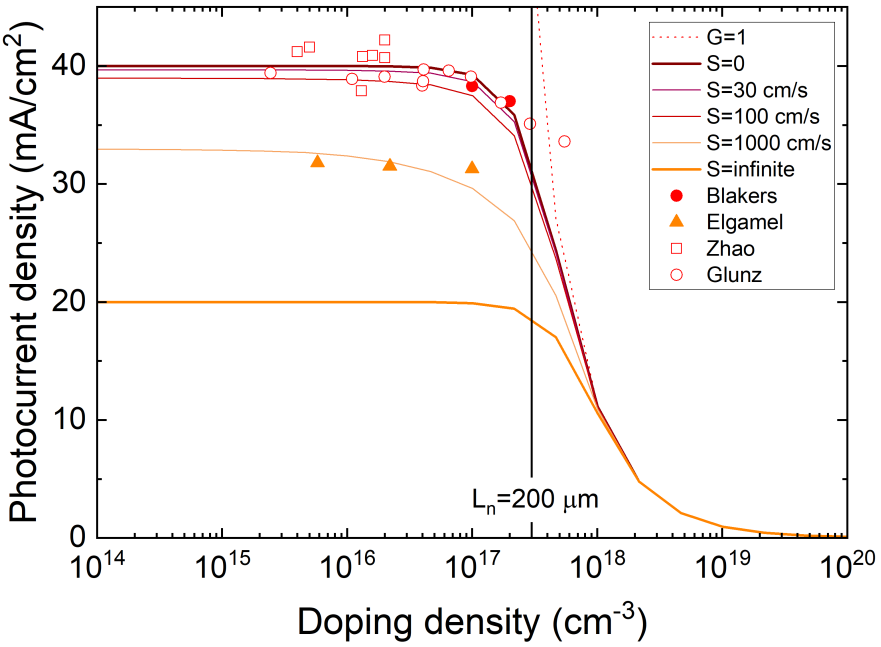


Figure D.4: Short circuit current density of a 200 µm thick bulk region with respect to the acceptor density N_A. The verticla line illustrates the doping concentration for with the diffusion length becomes 200 µm. Lines illustrate different values of the surface recombination velocity S, full symbols refer to experimental results of cells with passivated emitter and full-area metallic rear contact [40, 91], open symbols refer to cells to PERC cells with reduced rear recombination [92, 93].

D.4 Finite cell with exponential generation profile

The differential equations can also be solved for an exponential variation of the generation profile such as the one given by eq. (4.17) [100], but unfortunately not for the more realistic profile

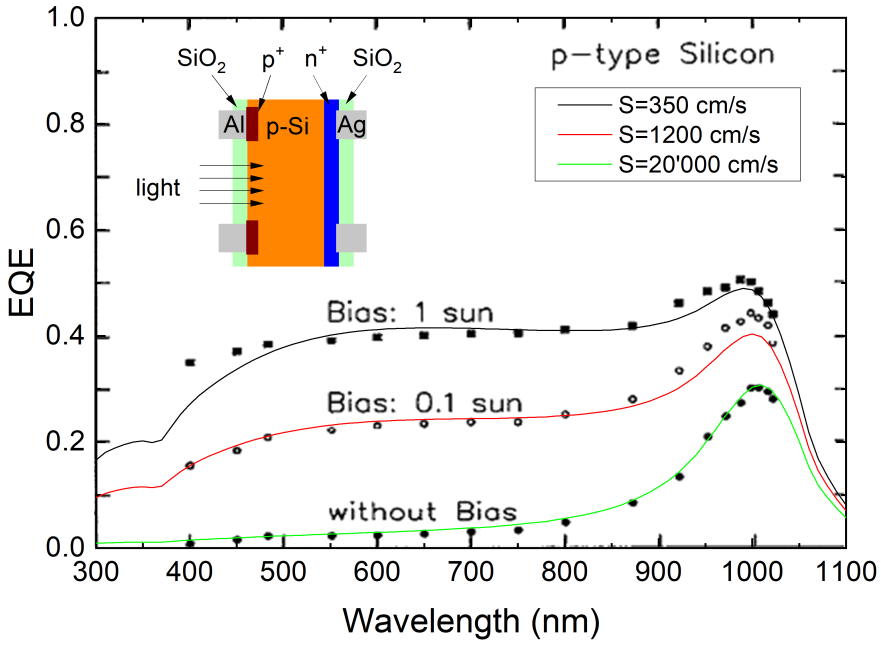


Figure D.5: EQE of a 250 µm thick bifacial cell under rear illumination. A bias light is added to define the excess carrier density Δn (after [102]).

of eq. (4.18). Nevertheless, both types of generation rate give a gradient in the density of the photogenerated carriers from which we would expect a diffusion towards the rear. In a standard cell configuration with n-layer at the front, this is fine for the hole current $j_{L,p}$ generated in the n-layer because we want it to flow towards the junction and from there onward to the rear. The situation is different for the electron current $j_{L,n}$ in the p-region because the gradient of the generation rate would still promote the diffusion of electrons to the rear, but we want to collect them at the front. Accordingly, there are two different expressions which are generally given for short circuit conditions of $V = 0$ to describe the photocurrent density j_L in a measurement of the EQE [101]:

$$j_{L,p} = \frac{q(1-R)\alpha L_p \Phi}{\alpha^2 L_p^2 - 1}\left(\frac{S_p + \alpha D_p - \left(S_p \cosh\frac{H'}{L_p} + \frac{D_p}{L_p}\sinh\frac{H'}{L_p}\right)e^{-\alpha H'}}{S_p \sinh\frac{H'}{L_p} + \frac{D_p}{L_p}\cosh\frac{H'}{L_p}} - \alpha L_p e^{-\alpha H'}\right) \quad \text{(D.13)}$$

$$j_{L,n} = \frac{q(1-R)\alpha L_n \Phi}{\alpha^2 L_n^2 - 1}e^{-\alpha H'}\left(\alpha L_n - \frac{S_n \cosh\frac{H}{L_n} + \frac{D_n}{L_n}\sinh\frac{H}{L_n} + (\alpha D_n - S_n)\,e^{-\alpha H}}{S_n \sinh\frac{H}{L_n} + \frac{D_n}{L_n}\cosh\frac{H}{L_n}}\right) \quad \text{(D.14)}$$

Here, R is the primary reflectivity at the front, the thickness of the n-region at the front is denoted by H', and the thickness of the p-region is approximated by the wafer thickness H.

We finish by an application from the days when EQE measurements were one of the few ways to assess the surface recombination velocity [102]. Figure D.5 shows EQE characteristics of bifacial cells under rear illumination as illustrated in the inset. In this case, eq. (D.13) should be used to describe current generated between the illuminated surface and the junction. Note that this is an electron current and that we have to insert the full wafer thickness for H'.

So far, we treated only the photogeneration by a single passage of light across the wafer. We can gain a little in accuracy by adding the contribution of the carriers that are generated by the

light that is reflected by the surface close to the junction. This contribution to the generation rate decays from the junction towards the illuminated surface whereas the electrons still flow towards the junction. Accordingly, we should use eq. (D.14) just as we would do for regular illumination through the n-side, but we have to insert the wafer thickness H into the leading factor $\exp(-\alpha H')$ and we have to multiply with the reflectivity R_r of the n-Si/SiO_2/air interface.

The result is shown by the lines in Figure D.5. They fit the data acceptably and the experiment showed that the SRV of the interface between silicon and SiO_2 is not constant, but varies with illumination intensity. A variation with surface concentration was already hinted at in the discussion of figure D.3. The dependence of the SRV on doping and excess carrier concentration was investigated in great detail because of its importance for transistor technology and a broader discussion is given in appendix G.

D.5 The back surface field

Let us come back to the numerical example of the electron-contact in section D.2 where we projected values for the $j_{0,n}$ between 240 - 460 $\mathrm{fA\,cm^{-2}}$ but neglected the contribution of the rear contact. Looking at figure D.2, a full area metal contact on the rear of a 200 µm thick wafer with doping concentration of $1 \times 10^{16}\,\mathrm{cm^{-3}}$ would contribute $1 \times 10^{4}\,\mathrm{fA\,cm^{-2}}$ and obviously it would not make any sense to think about improvements of the front side. Luckily we can do better than the simple full-area metallisation of the wafer by introducing the so-called *back surface field* (BSF).

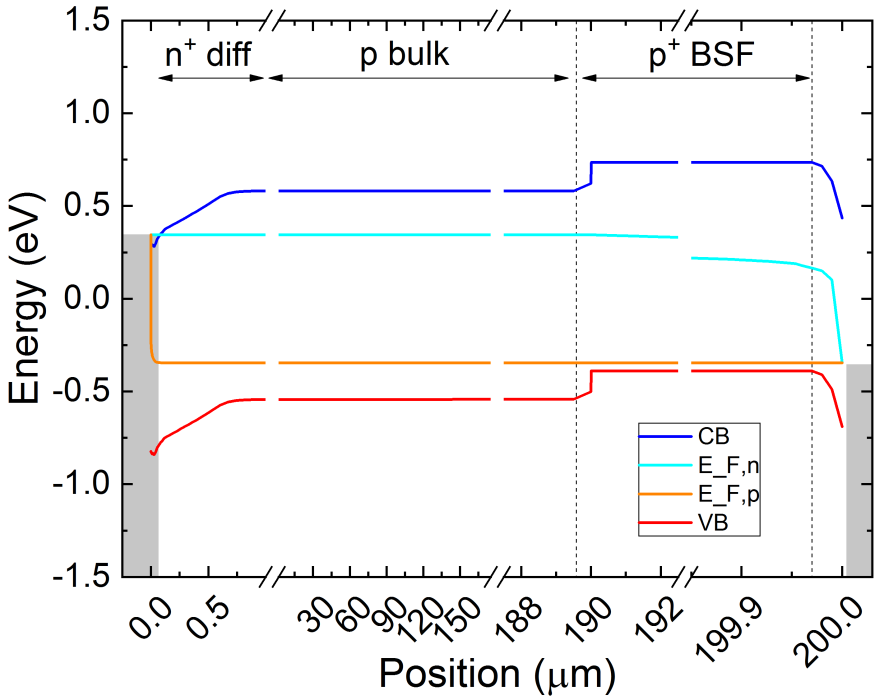


Figure D.6: Band diagram of a solar cell with p-type BSF under V_{oc} conditions. Note the different scales between the axis-breaks.

The name refers to a highly doped region at the rear of the wafer which serves two functions. First, it reduces the contact resistance as discussed further in appendix H. Second, it shields the

minority carriers against recombination at the metallic contact. In the previous chapter we concluded that the shielding is related to a reduction of the diffusion length as a consequence of high doping concentrations, but we should not extend high doping throughout the full wafer because this would entail a lot of Auger recombination.

There is a variety of processes to create a highly doped region at the rear. A very simple and cost-effective method is the rapid annealing of a screen-printed Al-paste as discussed in section 8.7.2. Alternatively, doping can be obtained by diffusion from B-doped layers at the rear or by annealing in an atmosphere of BBr_3.[2]

Figure D.6 shows the band diagram of an Al-BSF cell under V_{oc} conditions. The thickness of the Al-doped region is $H_{BSF} = 10\,\mu m$. At the rear, the band diagram is characterised by an upwards bump at the interface between the p-type bulk and the p^+ region, and an SCR with downwards band bending towards the Schottky-contact with the metal. In order to relate the properties of the bulk with those of the highly doped region and its contact with the metal, we start by deriving three important limiting cases for the geometry factor of eq (6.25). For zero and infinite S_n we obtain:

$$G_{F,p} \overset{S_n \to 0}{\approx} \tanh H/L_n \overset{H << L_n}{\approx} H/L_n \tag{D.15}$$

$$G_{F,p} \overset{S_n \to \infty}{\approx} \coth H/L_n \overset{H << L_n}{\approx} (H/L_n)^{-1} \tag{D.16}$$

In some cases $S \to 0$ may be a bit too restrictive, but if S is "small" and the minority diffusion length L_n is much larger than the wafer thickness H, we can approximate as follows:

$$G_{F,p} \overset{L_n >> H}{\approx} \frac{S_n + \frac{D_n}{L_n}\frac{H}{L_n}}{D_n/L_n} \approx \frac{S_n L_n}{D_n} \tag{D.17}$$

$$j_{0,p} \approx \frac{q n_i^2}{N_A} \cdot S_n \Big|_{x = H - H_{BSF}} \tag{D.18}$$

This gives a particularly simple expression for j_0 in terms of the surface recombination velocity at the interface between the bulk and the p^+ region. To determine the S_n, we use eq. (5.22) to describe the current density flowing through the surface at $x = H - H_{BSF}$. Since this is a current density that gets injected into the highly doped region, we can express it by eqns. 6.24 and 6.25, using the voltage V' that reflects the QFL splitting just after the interface and the properties of the highly doped region:

$$\underbrace{q S_n \left(n(H) - n_{p,0}\right)}_{\text{bulk values}} = \underbrace{\frac{q n_i^2 D_{n,BSF}}{N_{BSF} L_{n,BSF}} \cdot G_{F,BSF} \left(e^{\frac{qV'}{kT}} - 1\right)}_{\text{BSF values}}$$

Assuming that the QFLs in the bulk are flat, we can express the electron density at the virtual interface by the applied voltage via $n(H) = n_{p,0} \exp\{qV/kT\}$. If we assume further that the QFLs remain flat across the region with the induced band bending, we have $V' = V$ [103]:

[2] Common to all methods is that they must either remove or compensate the n-type region that the preceding P-diffusion normally creates also at the rear side of the wafer. The Al-Si eutectic of the Al-BSF process normally melts through several µm of the rear surface and easily compensates the P-doping. Alternatively, the region must be removed by one-sided etching processes.

$$S_n = \frac{N_{bulk}}{N_{BSF}} \frac{D_{n,BSF}}{L_{n,BSF}} G_{F,BSF} \tag{D.19}$$

Here, we cancelled the intrinsic carrier density of the bulk against the one of the BSF region. In reality, the latter is higher because of bandgap narrowing and thus it would be more appropriate to keep their ratio in the equation. Putting this back into eq. D.18, we obtain:

$$j_{0,p} \approx \frac{q n_i^2 D_{n_{BSF}}}{N_{BSF} L_{n_{BSF}}} G_{F,BSF} \tag{D.20}$$

Thus, when we measure a high-quality wafer with p^+ region, $j_{0,p}$ is entirely governed by the properties of that region, the parameters of the wafer don't come into play. Figure D.7 shows the results of such an experiment with different thicknesses of the highly doped BSF region and two types of surface finish.

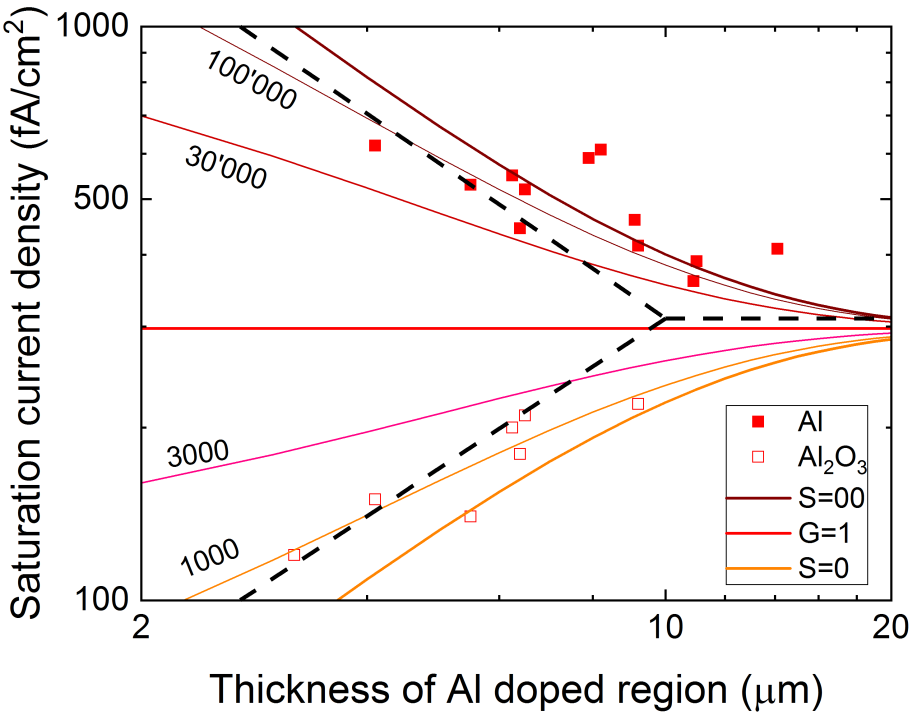


Figure D.7: Saturation current density of p^+ rear contact regions with various thicknesses. Full and open symbols refer to coverage with Al and Al_2O_3, respectively (data from ref. [104]). Full lines correspond to different values of the surface recombination velocity S, dashed lines illustrate the linearised approximations with slopes of $+1$ and -1, respectively.

We can now compare the measured data to our simple model. The superimposed lines in figure D.7 illustrate $j_{0,p}$ characteristics calculated by setting H equal to the measured thicknesses of the BSF and approximating the doping concentration by a constant value[3] of $N_{BSF} = 1.5 \times 10^{18}\,\mathrm{cm}^{-3}$.

[3]The authors measure a slowly varying Al-concentration which peaks around $3 \times 10^{18}\,\mathrm{cm}^{-3}$, but they report that only about 50% of the Al-dopants are ionised.

Thus, n_i is increased to $1.8 \times 10^{10}\,\mathrm{cm}^{-3}$ and ionised impurity scattering reduces the electron mobility to $310\,\mathrm{V\,s\,cm}^{-2}$ which yields an electron diffusion coefficient of $D_{n,BSF} = 8.1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$. We used a diffusion length of $L_{n,BSF} = 10.4\,\mu\mathrm{m}$ to obtain agreement between the measured data and the simple formula. This would suggest a lifetime of $\tau = 135\,\mathrm{ns}$ in the BSF region, half of the value of 270 ns that the authors obtain from SRH theory. Note that both values are much below the Auger limit which is $\tau = 2.2\,\mu\mathrm{s}$ for this doping concentration, presumably because Al creates a wealth of recombination centres in addition to the acceptor state.

The calculated characteristics suggest $S_n \approx 1 \times 10^5\,\mathrm{cm\,s}^{-1}$ for the Al-covered surface, but for most practical cases we can assume $S_n = \infty$. We note that the lowest j_0 are obtained when the thickness of the doped region is similar to the diffusion length.

For the case of the Al_2O_3 covered surface, the open symbols roughly fall on the characteristic corresponding to $S_n \approx 1000\,\mathrm{cm\,s}^{-1}$. Current state-of-the-art passivation to p-type surfaces more often uses a shallow B-diffusion and Al_2O_3 passivation; the combined effect of the thinner doped region and field-effect passivation by the negative fix charge in Al_2O_3 yields a surface recombination velocity S_n that is about four times lower [105].

How about such a rear contact in a solar cell? The majority of the data for the Al-covered surface scatters around $500\,\mathrm{fA\,cm}^{-2}$. We can now use eq. D.18 and estimate $S_{eff,n}$ for the interface between the wafer and the p^+-region. This yields $S_{eff,n} \approx 200\,\mathrm{cm\,s}^{-1}$ for a typical wafer doping of $1 \times 10^{16}\,\mathrm{cm}^{-3}$ and thus much less than values around $1 \times 10^5\,\mathrm{cm\,s}^{-1}$ which we obtained for a metal-covered surface. Values between 200 - $500\,\mathrm{cm\,s}^{-1}$ were reported by Rose [106]. The agreement is encouraging, but it is not very satisfactory that $S_{eff,n}$ depends on the doping of the wafer whereas by name it is a supposed to be a surface quantity. This is one of the reasons why it has become more common to work with j_0.

Appendix E

Isotype junctions

The depletion approximation may appear a bit crude, but it gives a very powerful and accurate description of the *pn*-junction. Unfortunately, it does not apply to *isotype junctions*, i.e. junctions between materials of the same doping type. A well-known example is the junction between the p-type bulk and the p^+ BSF. As the two different materials are in direct contact, holes can move from the highly doped region to accumulate in the lowly doped region until a unique Fermi level is formed. The built-in voltage of the iso-type junction is still given by a very familiar expression:

$$qV_{bi} = kT \ln\left(\frac{N_{BSF}}{N_{bulk}}\right)$$

For typical doping concentrations of $N_{bulk} = 1 \times 10^{16}\,\mathrm{cm}^{-3}$ and $N_{BSF} = 1 \times 10^{18}\,\mathrm{cm}^{-3}$, we obtain $V_{bi} = 0.120\,\mathrm{eV}$.[1] Let us assume that the wafer is in the negative half-space and that its surface is at $x = 0$. To find the electrostatic potential $\Phi(x)$ in the accumulation region we have to solve the Poisson equation for $x < 0$. The charge density is given by $\rho = q(-N_{bulk} + p)$ where the first term is due to the negative charge N_{bulk}^- of the ionised dopants and the second term is due to the total density of positive holes, i.e. the sum of those that are present in the p-doped bulk in equilibrium and of the additional ones that are injected. The relation between the hole density and the potential is given by $p = p_0 \exp(-q\Phi/kT)$. With this definition, accumulation is obtained if the potential assumes negative values. This yields the following Poisson equation for $x < 0$:

$$\frac{d^2\Phi}{dx^2} = -\frac{\rho}{\epsilon_{Si}\epsilon_0} = -\frac{q}{\epsilon_{Si}\epsilon_0}\left(p_0 e^{-\frac{\Phi}{kT}} - N_{bulk}\right) \tag{E.1}$$

Since the differential equation for Φ contains the charge density that once again contains the potential $\Phi(x)$ which we are looking for, the solution is called *self consistent*. It is safe to assume that the accumulation takes place close to the interface and that the bulk is neutral far away from the interface. Thus, the field E will be zero at $x \to -\infty$. If we take this as reference for the potential by assuming that it is zero there as well, we find $p_0 = N_{bulk}$.

We cannot solve this equation analytically and most numeric solvers have difficulties to accept boundary conditions at $x \to -\infty$. Fortunately, we can solve the equation partially up to a point that allows us to evaluate the boundary condition manually. To this end, we multiply both sides with $2 \cdot d\Phi/dx = -2E$. After recognising that $2E \cdot dE = d(E^2)$, we can drop the dx and carry out

[1] For this case, the Fermi level of the highly doped region is located at $E_F - E_{VB} = 0.06\,\mathrm{eV}$. This is closer than $3kT \approx 0.077\,\mathrm{eV}$ and we are entering the domain where the Maxwell-Boltzmann approximation is no longer applicable.

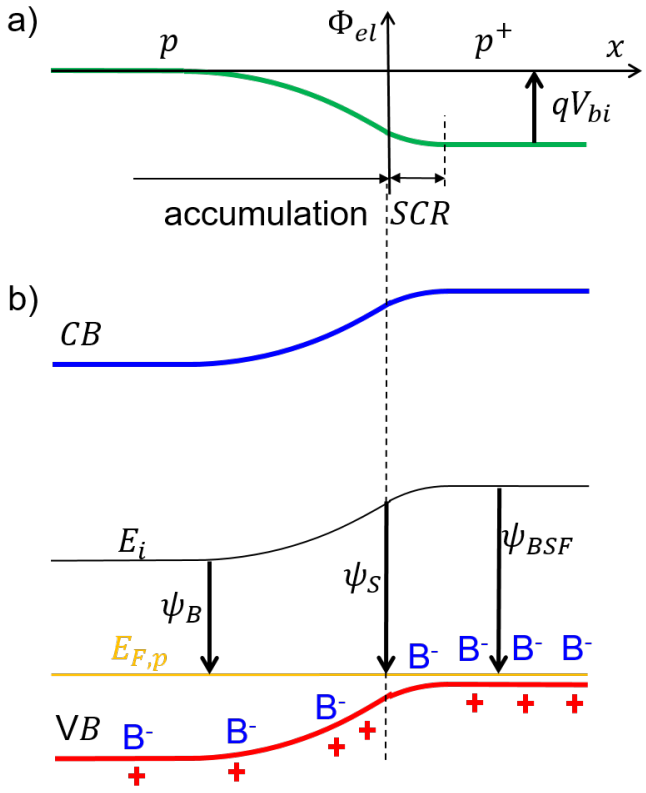


Figure E.1: Electrostatic potential (a) and band diagram (b) of an isotype pp^+ junction.

the integration. On the left side, the boundary values of the integral are the fields at $x = -\infty$ and at $x = 0$, on the right side the boundaries are potentials at these positions:

$$\begin{aligned} 2\frac{d\Phi}{dx}\cdot\frac{d^2\Phi}{dx^2} = 2E\cdot\frac{dE}{dx} = \frac{d(E^2)}{dx} &= -\frac{2qN_{bulk}}{\epsilon_{Si}\epsilon_0}\left(e^{-\frac{\Phi}{kT}}-1\right)\frac{d\Phi}{dx} \\ E^2(0^-) - \underbrace{E^2(-\infty)}_{=0} &= -\frac{2qN_{bulk}}{\epsilon_{Si}\epsilon_0}\left[-kT\left(e^{-\frac{\Phi(0)}{kT}}-1\right)-\Phi(0^-)\right] \end{aligned} \tag{E.2}$$

Thus, we find a relation between the field $E(0^-)$ and the potential $\Phi(0^-)$ where we use the superscript "$-$" to denote that both quantities are evaluated to the left side of the interface. Next, we use Gauss' law to relate the field with σ_s,the areal density of charges in the accumulation region of the semiconductor.[2]

$$D(0^-) = \epsilon_{Si}\epsilon_0 E(0^-) = \sigma_s = qQ_s \tag{E.3}$$

[2]Here, we assumed the the Gaussian pillbox is deep enough to contain all of the accumulation region and for consistency with the treatment in the next chapter, we write σ_s to denote an actual charge density in $\mathrm{A\,s\,cm^{-2}}$. We use Q_s when we we refer to the number of charges per area in $\mathrm{cm^{-2}}$.

If we select the positive root because we are dealing with the charge of the accumulated holes, we obtain.

$$qQ_s = \sqrt{2q\epsilon_{Si}\epsilon_0 N_{bulk}\left[kT\left(e^{-\frac{\Phi(0)}{kT}}-1\right)+\Phi(0^-)\right]} \tag{E.4}$$

We can turn now to the highly doped region. Here, we may apply the depletion approximation which means that we drop the exponential term in eq. (E.1). We denote a superscript "+" to the right side of the interface and integrate over the width of the SCR from $x = 0^+$ to $x = d_{SCR}$:

$$\underbrace{E^2(d_{SCR})}_{=0} - E^2(0^+) = -\frac{2qN_{BSF}}{\epsilon_{Si}\epsilon_0}(\underbrace{-\Phi(d_{SCR})}_{=-qV_{bi}} + \Phi(0^+))$$

We use the fact that the field will disappear at the end of the SCR, and since we assumed that $\Phi(-\infty) = 0$, we find $\Phi(d_{SCR}) = -qV_{bi}$ because the potential difference over the full junction must sum up to qV_{bi}. Applying once more Gauss' law, we obtain the areal charge density of the depletion region. This time, we select the negative root as the SCR contains the negative charge of the ionized dopants:

$$\sigma_{BSF} = qQ_{BSF} = \epsilon_{Si}\epsilon_0 E(0^+) = -\sqrt{2q\epsilon_{Si}\epsilon_0 N_{BSF}(\Phi(0^+) - \Phi(d_{SCR}))} \tag{E.5}$$

Finally, we invoke charge neutrality by requesting that the sum of eqns. (E.4) and (E.5) equals zero and we are ready to solve numerically for $\Phi(0)$. For our dopant concentrations from above, this yields $\Phi(0) = -0.106\,\text{eV}$, i.e. most of V_{bi} drops over the accumulation region, the band banding of the SCR is only $-0.014\,\text{eV}$. Having determined a numerical value for $\Phi(0)$, we can easily calculate $E(0)$ with eq. (E.4) or eq. (E.5). These values can be processed more easily as boundary conditions by a numerical solver.

The resulting band diagram is schematically shown in figure E.1 with an alternative definition of the potentials in terms of the midgap energy E_i. We can identify $\Phi(0) = \psi_S - \psi_B$ and the round bracket in eq. (E.5) yields the band bending as positive value which it should be for the term under the root.

Note that here it was not strictly necessary to invoke Gauss' law. Even before bothering about choosing the correct signs for the charge we could have claimed that the squares of the field amplitudes must be the same. However, when we are dealing with a hetero-junction where the permittivities are different on the two sides of the junction, the electric field is discontinuous and we have to request equality for the electric displacement, i.e. $D^- = D^+$ or $\epsilon^- E^- = \epsilon^+ E^+$. Another example where we need to go through the more general case of charge neutrality, is when we have charged interface states. In that case, even the electric displacement is discontinuous and the potential acquires a kink. We'll come across interface charges in appendix G.

Appendix F

Bulk recombination at defects

Different from the recombination rate of a single SRH defect state described by eq. (5.17), there are often two or more defect states whose recombination rates add up. Accordingly, the resulting lifetime is obtained by dividing the excess carrier density Δn by the sum of all recombination rates.

$$R_{tot} = R_{SRH,1} + R_{SRH,2} + ... \quad \text{(F.1)}$$

$$\tau = \Delta n / R_{tot} \quad \text{(F.2)}$$

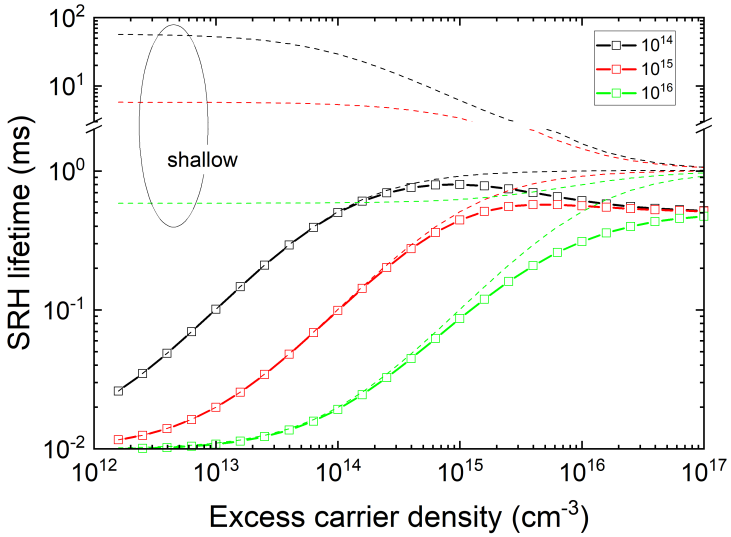


Figure F.1: Minority carrier lifetime τ for the combination of two defect states (symbols), assuming assuming p-type Si with the indicated bulk doping. The dashed lines illustrate the components of a shallow and a deep defect located 0.1 and 0.4 eV above the VB, respectively.

Figure F.1 illustrates the resulting lifetime for a deep and a shallow defect in p-type silicon. For clarity the contributions of radiative recombination and of Auger processes are not included. The deep defect state is assumed to be at 0.4 eV above the VB and its limiting lifetimes τ_p and τ_n are assumed to be 0.01 and 1 ms, respectively. Thus, there is the typical transition from τ_p at low Δn towards $\tau_p + \tau_n$ at high Δn. For the lowest bulk doping the onset of the transition starts already at

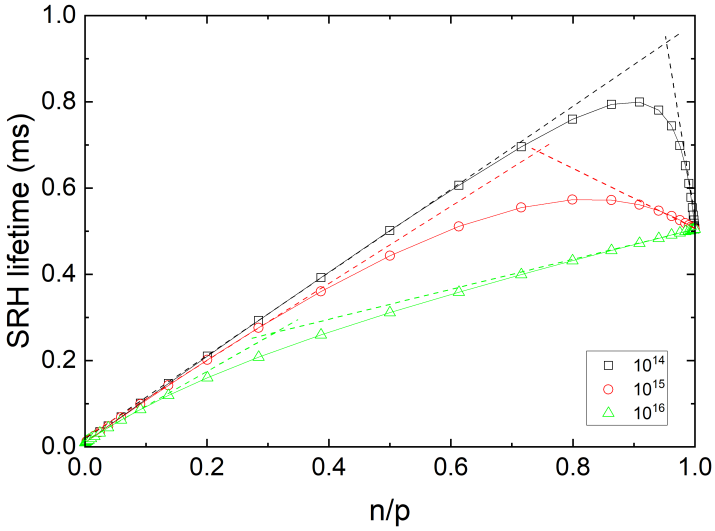


Figure F.2: Plot of the minority carrier lifetimes from figure F.1 over n/p. The dashed lines illustrated the linear components of the two defects.

low excess carrier density whereas highly doped material requires rather high Δn to transit towards high lifetime. This suggests that highly doped material is more sensitive to the presence of defects.

For the second defect state in figure F.1 we assumed the same values for τ_p and τ_n, but a shallow energetic position at 0.1 eV above the VB. In this case, a transition from low to high lifetime is observed only for the highest bulk doping whereas the less doped materials show an inverted transition. In case of shallow defects the high lifetime values at low Δn are not directly related to the assumed values of τ_p and τ_n, only the lifetime at high Δn approaches again the value of $\tau_p + \tau_n$.

F.1 Defect analysis

To disentangle contributions of different defects, Murphy proposed a linearization procedure [107]. Defining k as the ratio of the capture cross sections by $k = \sigma_n/\sigma_p$ and x as the ratio between minority and majority carrier densities which becomes $x = n/p = \Delta n/(p_0+\Delta n)$ for p-type material, eq. (5.17) can be approximated to a much simpler expression for the lifetime:[1]

$$\begin{aligned} \tau &= \frac{1}{\sigma_n v_{th} N_t}\left[1 + \frac{k n_1}{p_0} + \frac{p_1}{p_0} + x \cdot \left(k - \frac{k n_1}{p_0} - \frac{p_1}{p_0}\right)\right] \\ &\equiv a + b \cdot x \end{aligned} \tag{F.3}$$

The strength of the linearized form in eq. (F.3) becomes immediately clear when the the modelled lifetime data of figure F.1 is plotted over n/p in figure F.2. The linear relations associated to the two different defects are now clearly expressed. Except for the highly doped case, deep and shallow defects are characterized by raising and falling characteristics, respectively.

The two fitting parameters a and b in eq. (F.3) are not sufficient to extract the full information of the capture cross sections, the energetic position in the bandgap and the defect density, but they can be used to determine a functional dependence among some of the parameters. To construct

[1] For n-type material, see eq. (8) in [107].

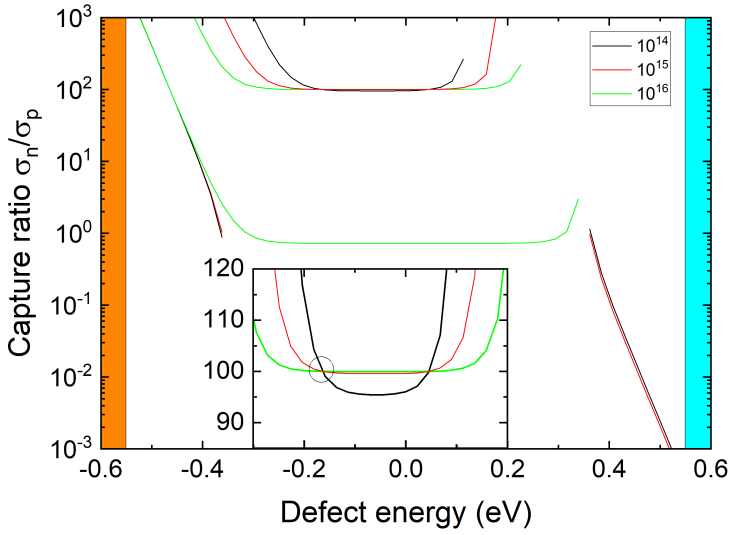


Figure F.3: Defect parameter solution surface for the curve-fits of figure F.2. The inset reproduces the region of the intersections on linear scale.

this function, we let the energetic position of the defect vary across the bandgap. The quantities n_1 and p_1 can be evaluated for each of these positions, and subsequently the fitting parameters define the ratio of the capture cross sections.

$$k = -\frac{\frac{p_1}{p_0} + \frac{b}{a+b}}{\frac{b}{a+b} - 1 + \frac{n_1}{p_0}} \tag{F.4}$$

The resulting characteristic is called defect parameter solution surface (DPSS) and shown in figure F.3 for linear fits to the defects shown in figure F.1. The rising characteristics yields the parameters of the deep defect in the range of -0.3 - $0.2\,\mathrm{eV}$ whereas the falling characteristics yield the parameters for the shallow defect close to the band edges. The DPSS characteristics for different bulk doping intersect, thus limiting the possible defect parameters from a functional dependence to two points as shown in the inset of figure F.3 for the deep defect. One of them reproduces correctly the assumed position at $0.4\,\mathrm{eV}$ above the VB and the the assumed value of 100 for the ratio of the capture cross sections. To rule out the parameters of the second intersection, one would have to carry out additional experiments. Regarding the shallow defect, its intersection at $0.1\,\mathrm{eV}$ above the VB is unfortunately not easily resolved among the steeply varying characteristics.

F.2 Boron-related defects in silicon

F.2.1 Boron-iron defect

Iron was the most notorious impurity at the time when most solar cells were processed from p-type multicrystalline silicon doped with boron. When iron is paired to boron in the form of a FeB complex, the capture cross sections for electrons and holes are similar and the state is relatively close to the edge of the conduction band. Illumination dissociates the pairs and forms interstitial $\mathrm{Fe_i}$ defects which are far more detrimental since they form defect states close to midgap. In darkness,

the FeB complex can recover when interstitial iron atoms diffuse and get captured to the boron atoms. Since the diffusion of iron in silicon is fast, recovery takes place already at room temperature, and a mild dark-annealing can be used to speed up the process.

The properties of iron-related defects have been thoroughly studied by a variety of authors [108, 109, 110]. Figure F.4 shows lifetime data for various doping concentrations and an intentional iron contamination of $3.2 \times 10^{12}\,\mathrm{cm}^{-3}$. The two contributions of shallow and deep defects are relatively well visible for low bulk doping. The shallow component is less obvious for the highly doped samples since the excess carrier concentrations in these samples hardly reach the level of the doping concentration. Thus, the ratio n/p covers only a comparatively small range.

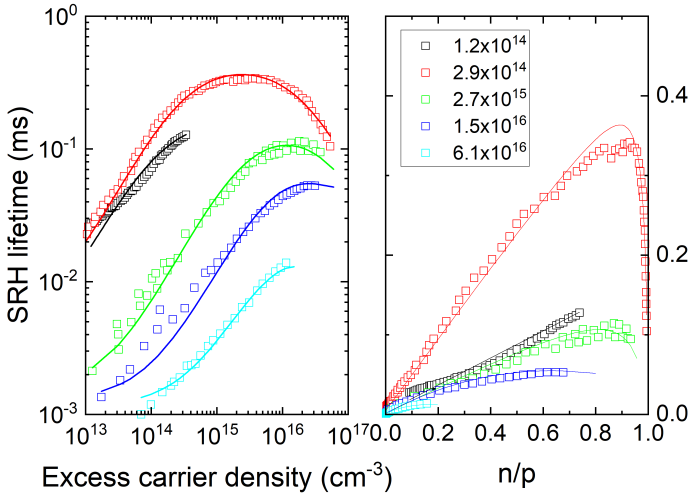


Figure F.4: Measured lifetime data of siliocn with iron contamination of $3.2 \times 10^{12}\,\mathrm{cm}^{-3}$ after light soaking for different bulk doping concentrations (left). The right panel shows the same data with respect to n/p, lines illustrate fitting functions for two defects; data digitized from [108, 109, 110].

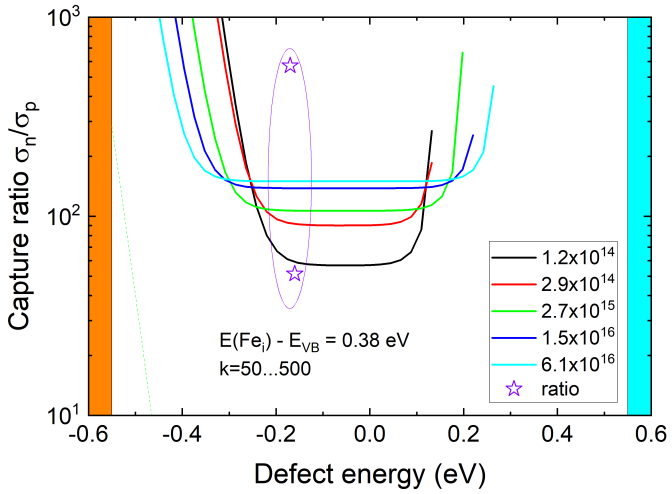


Figure F.5: Defect parameter solution surface for the curve-fits of figure F.4. Stars denote defect parameters obtained from temperature-dependent measurements [109].

The corresponding DPSS characteristics for the deep defect are shown in figure F.5. Wheres they

don't show the expected behaviour of two distinct intersections, the data still suggests energetic positions of approximately 0.3 and 0.65 eV above the VB and a capture ratio around 100. Based on temperature dependent measurements and other techniques such deep level transient spectroscopy (DLTS), literature reports that the defect state is located at 0.39 eV above the VB, but there is a large range of capture ratios between 50 and 500 as indicated by the purple stars [109].

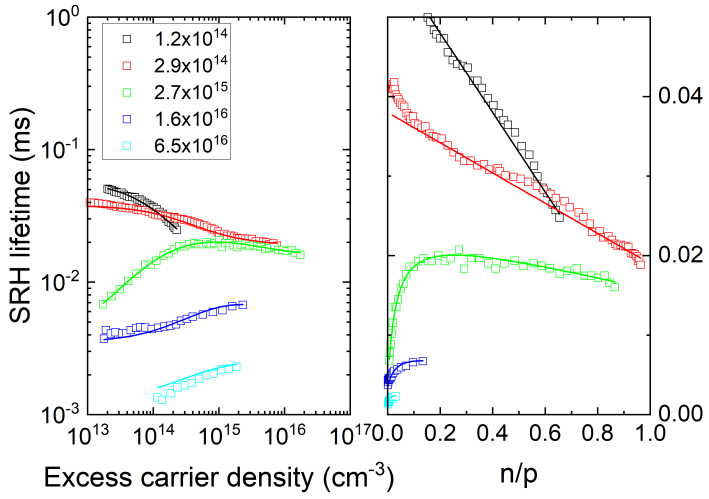


Figure F.6: Measured lifetime data of silicon with iron contamination of $3.2 \times 10^{12}\,\mathrm{cm}^{-3}$ after anneal for different bulk doping concentrations (left). The right panel shows the same data with respect to n/p, lines illustrate fitting functions for two defects ; data digitized from [108, 109, 110].

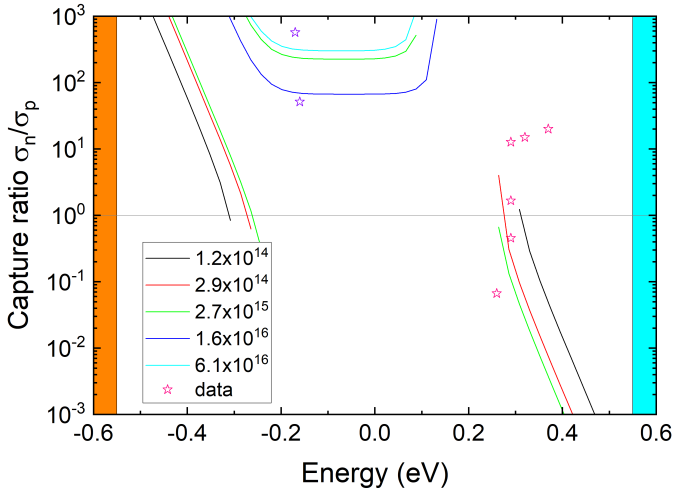


Figure F.7: Defect parameter solution surface for the curve-fits of figure F.6. Stars denote defect parameters obtained from temperature-dependent measurements [109].

The lifetime data for the FeB complex obtained after dark annealing is shown in figure F.6. For most of the shown bulk doping concentrations the characteristics are falling, suggesting that the defect is rather shallow. Once again, data for the most highly doped samples extends only over a small range of n/p ratios.

The corresponding DPSS are shown in figure F.7. The deep defect can fitted reliably only

for one of the samples, therefore no intersection can be determined. The shallow defect yields rather steeply varying characteristics and also here no intersection can be identified. Literature data scatters rather broadly at energies close to the CB as indicated by the pink stars [109], but the values with capture ratios around unity could comply with the DPSS analysis. Table F.1 summarizes the parameters for the defects associated to interstitial iron and the FeB complex.

Table F.1: Parameters for the iron-related defects in silicon [109].

	Fe_i	**FeB**
E_t	E_V+0.38 eV	E_C-0.23 eV
σ_n	3.6×10^{-15} - 4×10^{-14} cm^2	2.4×10^{-15} - 3×10^{-14} cm^2
σ_p	7×10^{-17} cm^2	1×10^{-15} - 5×10^{-14} cm^2

For p-type solar cells with Al-BSF or PERC design, it was possible to mitigate the deleterious effect of iron by careful processing. During $POCl_3$ diffusion iron was found to segregate from the bulk to the surface because the solubility of iron is enhanced in the highly doped regions by clustering or pairing to various defect species formed during the diffusion process. Since recombination in the highly doped region is ultimately dominated by Auger recombination, the segregation of iron into this region is acceptable. In the Al-BSF process a second instance of segregation was used during the formation of the rear contact since the solubility of iron is also very high in the Al-Si eutectic. Generally the segregation of impurities during the process is called "gettering"; more details can be found in a review of Liu [111].

F.2.2 Boron-oxygen defect

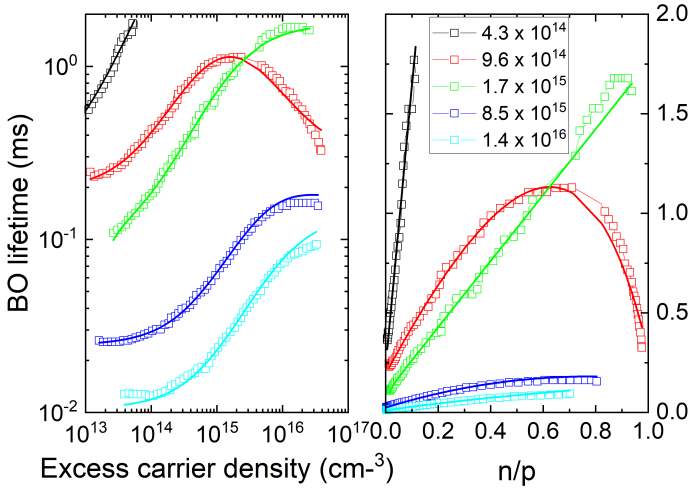


Figure F.8: Measured lifetime data of Cz silicon with oxygen contamination of 5.5×10^{17} - 7.8×10^{17} cm^{-3} after exposure to white light with intensity of one sun at 25 °C. The right panel shows the same data with respect to n/p, lines illustrate fitting functions for two defects; data digitized from [112].

In Cz-material the density of metallic impurities is generally lower, but in case of p-type silicon

doped with boron, a defect associated to complexes between boron and oxygen was reported. [112]. The defect state is activated by illumination, therefore effect is often called boron-oxygen light induced degradation (BO-LID). The formation of the defect is metastable, annealing at ca. 200 °C can recover the initial state. For solar cells, annealing under conditions of Fermi-level splitting, achieved either by illumination or by current injection, has been reported to permanently deactivate the majority of BO-related defects [113].

Oxygen is typically incorporated during ingot pulling, either from the quartz crucible that holds the melt, or from the ambient atmosphere that surrounds the ingot. The samples shown in figure F.8 have very similar oxygen concentrations of 5.5×10^{17} - $7.8 \times 10^{17}\,\mathrm{cm}^{-3}$ whereas the limiting lifetime τ_n at low excess carrier density scales linearly with the doping concentration N_A over almost two orders of magnitude [112]. Later work with silicon containing different amounts of oxygen contamination found that the defect density is also proportional to the square of the oxygen concentration [114], hinting at BO_2 clusters.

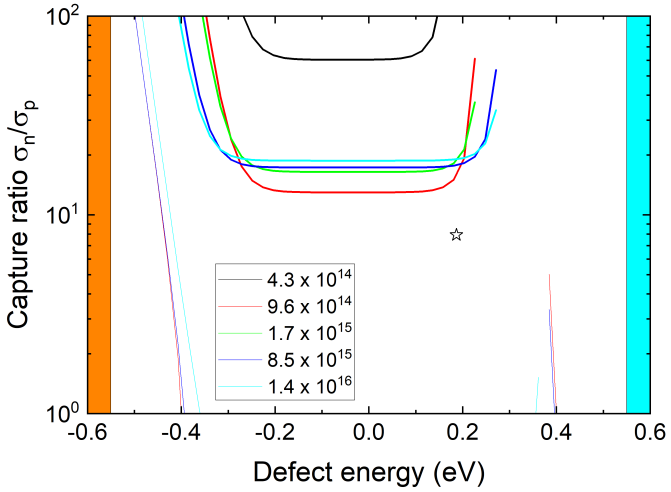


Figure F.9: Defect parameter solution surface for the curve-fits of figure F.8. The star denotes defect parameters obtained from temperature-dependent measurements [115].

Except for the one lowest bulk doping concentration, the DPSS characteristics shown in figure F.9 intersect around 0.3 eV above the VB and 0.35 eV below the CB with a capture ratio around 20. Temperature dependent measurements suggest that the defect state should be the one in the upper half of the gap, and they find slightly different parameters of $E_C - E_t = 0.41\,\mathrm{eV}$ with a capture ratio of 9 [115].

F.2.3 Mitigation strategies for boron-related defects

Owing to the technological importance of *p*-type silicon until about 2020, there have been many efforts to mitigate the boron-related defects. Below is a (non-exhaustive) list of measures that have been taken:

- Reduction of the boron concentration in the device, i.e. reduction of the doping ⇒ increase of the wafer resistivity
- Reduction of oxygen incorporation, e.g. by pulling in Ar-atmosphere and suppression of convective flow of the melt from the crucible walls to the growing ingot by magnetic confinement

- Use of alternative dopants instead of B, e.g. Ga
- Use of n-type material with phosphorus doping (was always used in heterojunction cells, and since 2020 in in TOPCon cells)
- Use of FZ wafers rather than Cz-material (mostly in research, but too costly for mass production)

F.3 Other defects in silicon

Figure F.10 collects data for common defects in silicon, showing that most of the metallic impurities have states in the lower part of the bandgap. More important for cell design, the large capture ratios observed for most of them mean that the defects preferentially capture minority carriers in p-type silicon. This explains to some extent why higher performance can be achieved with n-type material and why the market share of n-type solar cells rapidly increases.

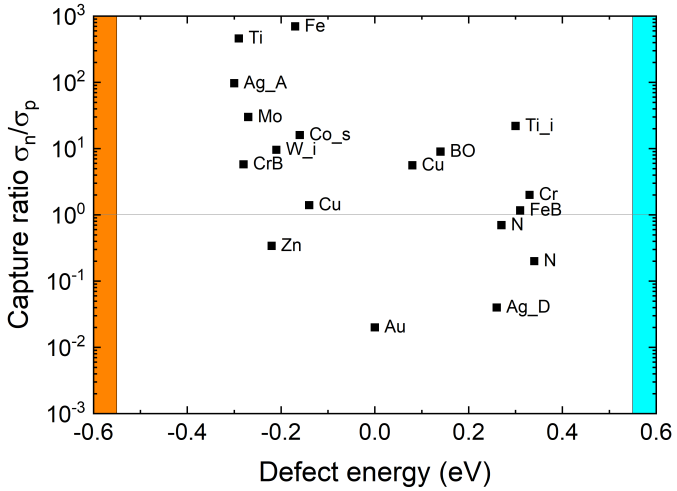


Figure F.10: Defect parameters for common impurities in silicon [116, 117].

Appendix G

The silicon/silicon oxide interface

The interface between silicon and silicon-dioxide is of great technological importance, and it is a good example to understand the properties of surface recombination. To proceed, we need to recognize two differences to bulk recombination. First, the densities of both carrier types at the interface may differ from the bulk values. Second, interface states are typically distributed continuously over the bandgap. Therefore, the recombination rate must be determined by integrating over a distribution of various defect states. The underlying procedure is called *extended SRH formalism*.

Let us consider the interface between p-type silicon and silicon oxide as illustrated in figure G.1. The interface is characterised by an interface state density D_{it} which can exchange charge with silicon, and by a fixed positive charge in the oxide which does not change.[1] If we reference to the intrinsic level E_i of the bulk far away from the interface, the hole potential is defined as negative number via $q\psi_B = E_{F,p} - E_i = -kT \cdot \ln(N_A/n_i)$.

Close to the surface, the hole density changes because holes get captured by interface states and because holes are electrostatically repelled by the fixed positive charge in the oxide. If we assume that the Fermi level remains flat, a reduction of the hole density is expressed by a downwards band bending and the potential at the interface will assume a value ψ_S which is different from the bulk. If we used aluminium oxide with a negative fixed charge instead of silicon oxide, holes would be electrostatically attracted to the interface and we would observe an upwards band bending with hole accumulation, possibly even type inversion.

To describe the interface states, we will follow an argument of Schroeder [119]. We assume that there are two types of surface state which are separated by a demarcation energy E_0. The interface is neutral when all states below E_0 contain an electron and all states above E_0 are empty. The filled states below E_0 are called "donor-like" because they can potentially provide an electron, or capture a hole. Likewise, the empty states above E_0 are called "acceptor-like" because they can capture an electron. For the remainder we will assume that E_0 is exactly at midgap position. In Figure G.1 label them D and A, respectively. Red, blue and black is used to designate positive, neutral, and negative states, the charge is additionally denoted by the superscripts.

[1] For historic reasons the density of interface states is written $D_{it}(E)$ where the subscript "*it*" denotes "interface trap" even though their position in the bandgap most often makes them recombination centres. Their density is measured per area and per energy interval in units of $\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$. By Q_{it} we count the number of charges per area in units of $[\mathrm{cm}^{-2}]$, the actual interface charge density is given by $\sigma_{it} = qQ_{it}$ in units of $\mathrm{A\,s\,cm}^{-2}$. The fixed charge Q_f is likely related to oxygen vacancies located closer than 1 nm from the interface [118]. The areal charge densities at the interface cause a discontinuity in the electric displacement $D = \epsilon E$. The potential is continuous across the interface, but there is a change in slope.

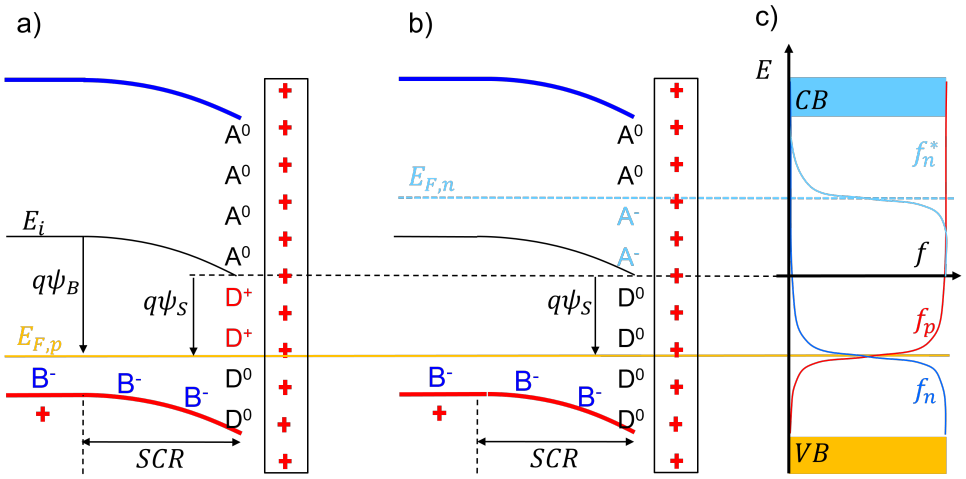


Figure G.1: Schematic band diagram of the interface between p-type c-Si and silicon oxide in darkness (panel a) and under illumination (panel b). Panel c) illustrates the occupancy functions of the interface states.

G.1 Surface potential

If we assume that the Fermi level extends flat from the p-type bulk to the surface, the donor-like interface states between the Fermi level and mid-gap can no longer retain their electrons. Instead, the electrons in these states will recombine with free holes from the VB. Since these interface states are now charged positively with respect to their natural occupancy, they are shown in red in panel a) of figure G.1. For a more precise description we should take into account the capture cross sections of the interface states and the distribution functions for electrons and holes as depicted in panel c) of figure G.1 by f_n and $f_p = 1 - f_n$, respectively. The recombined holes leave behind ionised boron atoms that build up a negative space charge region (SCR) which is associated with a downwards band-bending. Next, we consider the fixed positive charge in the silicon oxide. Its electrostatic force repels holes and forces them to move towards the bulk of the silicon crystal. We assume that the bulk is large enough to accommodate holes without a significant change in ψ_B, but the SCR gets widened in the process and thus the downwards band bending at the interface is further enhanced.

Finally, when we add illumination, the Fermi-level is no longer unique but splits into two quasi Fermi levels (QFLs) as illustrated in panel b) of figure G.1. Let us assume low injection such that $E_{F,p}$ is not changed whereas $E_{F,n} = kT \ln(\Delta n/n_i)$. Consequently, f_p will remain the same as in dark. The splitting of the QFLs in the bulk is obtained as follows:

$$\Delta E_F = E_{F,n} - E_{F,p} = kT \ln \frac{\Delta n N_A}{n_i^2} \tag{G.1}$$

Like in the case of the majority QFL, we assume a flat continuation also for the minority QFL and consequently f_n should now refer to the QFL of the electrons as illustrated by f_n^*. This gives us conflicting rules for the filling of the interface states. States below $E_{F,n}$ should be filled with electrons, states above $E_{f,p}$ should be filled with holes. To simultaneously comply with both rules, electrons and holes will continuously move from the bulk towards interface and recombine in the interface states, i.e. a recombination current will flow from the bulk towards the interface. We will

see below that there is an asymmetry in the capture cross sections such that all states below $E_{F,n}$ are rapidly filled with electrons whereas the capture of holes into the states above $E_{F,p}$ is much slower. For simplicity, panel b) in figure G.1 shows all states below $E_{F,n}$ filled with electrons. Note that those between $E_{F,p}$ and E_i are drawn in black because filling returns them to their natural occupation whereas those between E_i and $E_{F,n}$ are drawn in light blue because filling charges them negatively with respect to their natural state.

We can use the flat continuation of the QFLs to express the surface charge carrier densities in terms of the bulk values and the band bending:

$$p_s = (p_0 + \Delta n)/e^{\frac{q(\psi_s - \psi_b)}{kT}} \tag{G.2}$$

$$n_s = (n_0 + \Delta n) \cdot e^{\frac{q(\psi_s - \psi_b)}{kT}} \tag{G.3}$$

The splitting of the QFLs is still given by $\Delta E_F = kT \ln(\Delta n N_A / n_i^2)$ since the multiplying factors in the definitions of p_S and n_S cancel out.

G.2 Charge neutrality at the interface

We can proceed to find the value of the potential at the surface, ψ_S. To this end, we determine the total interface charge density as function of ψ_S, and in a second step we will request charge neutrality and solve for value of ψ_S that ensures it. The interface charge density is composed as follows:

G.2.1 Charge in the semiconductor

The charge density in the semiconductor is normally expressed with respect to a volume. To obtain an areal density, we have to integrate it between the surface and a position where the semiconductor is neutral. This is similar to the derivation of the hole accumulation region in appendix E, except that here the region close to the surface is depleted of holes because of the positive fixed charge in the oxide. In case of a large fixed charge we can even attract electrons from the bulk and form an inverted surface with electron accumulation. The cases of accumulation, depletion and inversion are described in a wealth of literature on MOS transistors. However, they are usually operated in dark whereas lifetime measurements use optical injection of charge carriers. Consequently, the literature on surface voltage measurements may be more helpful, e.g. the book by Kronik and Shapira [120]. For an excess carrier density Δn by photogenerated carriers, the charge in the semiconductor as function of the band bending V_b is then given by their eq. (2.52):[2]

$$Q_s = -\operatorname{sgn}(\psi_S - \psi_B) \cdot \sqrt{\frac{2\epsilon\epsilon_0 kT}{q}} \times \sqrt{p_0\left(e^{-\frac{qV_b}{kT}} + \frac{qV_b}{kT} - 1\right) + n_0\left(e^{\frac{qV_b}{kT}} - \frac{qV_b}{kT} - 1\right) + \Delta n\left(e^{-\frac{qV_b}{kT}} + e^{\frac{qV_b}{kT}} - 2\right)} \tag{G.4}$$

Compared to eq. E.4 that was derived for a p-type semiconductor, we find that the equation above is simply a generalisation that adds in the root the terms for n-type doping and for injection.

[2] In the literature, the prefactor is often extended to contain the Debye-length. We refrain from doing so because there are inconsistencies with its definition, and whether to use intrinsic carrier density or the doping density.

The band bending V_b and the surface potential ψ_S are related via $\psi_S = V_b + \psi_B$, and the expression is multiplied by $-\operatorname{sgn}(\psi_S - \psi_B)$ to account correctly for the type of charge.

Let us briefly discuss an approximation for a p-type semiconductor for surface potentials between ψ_B and zero. In the first round bracket, the exponential and the "-1" can be neglected for positive qV_b, i.e. for all ψ_S larger than ψ_B. The second of the round brackets in eq. (G.4) can be neglected for p-type material. In the third round bracket, we can simplify by expanding the exponential functions. For fourth order we obtain:

$$Q_s \approx -\sqrt{\frac{2\epsilon\epsilon_0 kT}{q}\left(p_0\frac{qV_b}{kT} + \Delta n\left(\frac{qV_b}{kT}\right)^2 + \Delta n\frac{1}{12}\left(\frac{qV_b}{kT}\right)^4\right)} \tag{G.5}$$

For $\Delta n = 0$, we obtain the depletion approximation which we could have obtained more easily from $Q_{SCR} = -N_A w_{SCR}$, using the width of a one-sided SCR according to eq. (6.4) with the assumption that $N_D >> N_A$.

The impact of the various approximations is shown by the dashed lines in figure G.2. In dark, the depletion approximation holds over a surprisingly large region around 0 eV. For the illuminated case, our approximations appear to hold in the relevant region where the surface charge changes sign. Note that this may be a fortuitous coincidence since even a fourth-order expansion is not yet a good approximation for an exponential function.

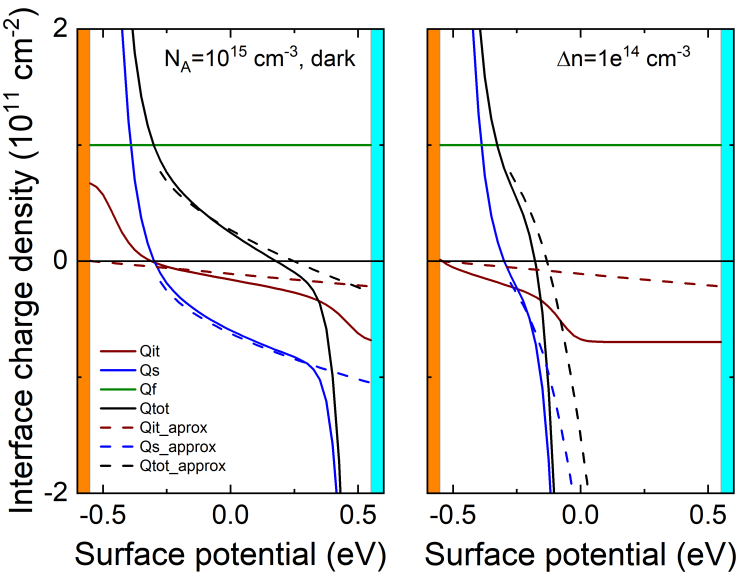


Figure G.2: Interface charge as function of the surface potential ψ_S for $N_A = 1 \times 10^{15}\,\mathrm{cm}^{-3}$, $D_{it} = 2 \times 10^{10}\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$, and a fixed charge of $1 \times 10^{11}\,\mathrm{cm}^{-2}$ for the cases of darkness (left) and injection of $\Delta n = 1 \times 10^{14}\,\mathrm{cm}^{-3}$ (right).

G.2.2 Interface charge

To calculate the charge density of the interface states we need to know their distribution over the bandgap as well as their capture cross sections for electrons and holes. Figure G.3 shows data of the Si/SiO_2 interface [102]. The density D_{it} is characterised by two exponential tails extending from the VB and from the CB whose asymmetry in the slopes resembles the band tails of amorphous

silicon [121]. The region around midgap is almost constant at $D_{it} = 2 \times 10^{10}\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$, but for calculations later in this chapter we will allow that it varies with doping concentration as shown in the right panel. The capture cross sections for electrons and holes can be approximated by broad Gaussians and we note that σ_p is about two orders of magnitude less than σ_n. The overlaid bars denote discrete values that we use later in this chapter.

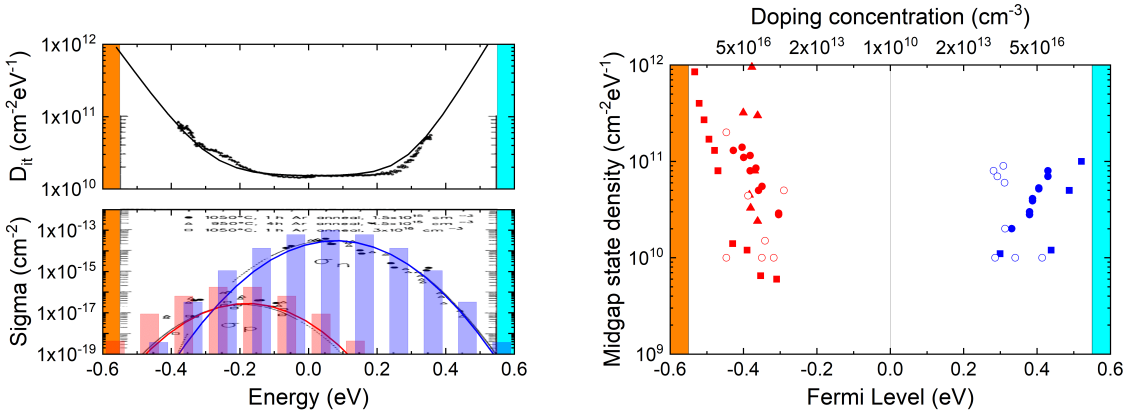


Figure G.3: Density of interface states (upper left panel) and their capture cross sections (lower left panel) with respect to the position within the bandgap (after [102]). The red and blue bars denote discrete values for σ_p and σ_n used for the treatment described later in this chapter. The right panel shows the dependence of D_{it} at midgap on doping concentration; data from [122] (full squares), [123] (circles), [124] (full triangles).

The charge density of the interface states is determined by integrating over the bandgap the occupancies of the acceptor-like states and of the donor-like states:

$$Q_{it} = -\int_{E_V}^{E_C} D_A(E) f_A(E) dE + \int_{E_V}^{E_C} D_D(E) f_D(E) dE \tag{G.6}$$

Here, $f_A(E)$ is the electron occupancy of acceptor states and $f_D(E)$ is the hole occupancy of donor states. These distribution functions are defined with the capture cross sections σ_p and σ_n shown in figure G.3, and p_1 and n_1 are the same as eqns. (5.18) and (5.19) for SRH recombination, except that we have to use ψ_S to express the Fermi-level of the holes at the interface.

$$f_A(E) = \frac{\sigma_n(E) n_S + \sigma_p(E) p_1}{\sigma_n(E)(n_S + n_1) + \sigma_p(E)(p_S + p_1)} \tag{G.7}$$

$$f_D(E) = \frac{\sigma_n(E) n_1 + \sigma_p(E) p_S}{\sigma_n(E)(n_S + n_1) + \sigma_p(E)(p_S + p_1)} \tag{G.8}$$

Figure G.1 illustrates that f_A and f_D resemble Fermi-Dirac functions. Let us first discuss the case without illumination. For our case of a p-type material, f_A is illustrated by the red curve. It tells us that the electron occupancy of the acceptor states above E_i is essentially zero, i.e. we can ignore the first integral. From the blue curve denoting f_D we see that the hole occupancy of the donor states switches from zero to unity at the Fermi-level, in other words, these donor-like states lost all their electrons.

When we add illumination, the occupation of the states between the QFLs is given by a steady state between the capture of electrons and holes. Keeping in mind that the large values of σ_n fill

all states up to $E_{F,n}$ as illustrated in panel b) of figure G.1, we can approximate Q_{it} by integrating over the negatively charged states between E_i and $E_{F,n}$.

$$Q_{it} \approx -\int_{E=E_i}^{E_{F,n}} D_{it} dE \approx -D_{it}(E_{F,n} - E_i) = -D_{it}(\psi_S + \Delta E_F) \tag{G.9}$$

Here, we further simplified by assuming a constant D_{it} that can be taken out of the integral. If we set $\Delta E_F = 0$ in the final expression, it yields correctly the positive interface charge illustrated in panel a) of figure G.1.

The impact of the various approximations is shown by the red curves in figure G.2. For the dark case, the slope around 0 eV appears right, but the full line is offset downwards. This is a consequence of the band tails which have a significant impact as the band edges are approached. Upon illumination, the dashed curve is correctly offset towards the right, but its slope is no longer correct.

G.2.3 Fixed charge

As the fixed charge in the SiO_2 layer is located very close to the interface, we assume that it amounts to an areal charge density equal to Q_f at a position $x = d_f$. As the name suggests, it is fixed and does not vary with the surface potential. The oxide in the positive half-space is assumed to be charge-free.

G.2.4 Gate charge

Even though gate electrodes are normally not used in solar cells, we add this section because metal oxide semiconductor (MOS) junctions are widely used to investigate the properties of the Si/SiO_2 interface. To this end, the oxide is normally grown to a thickness that renders it insulating, a metallic gate electrode is applied, and the wafer is contacted by an Ohmic contact. Thus, the surface potential can be changed by applying an external bias voltage V_g to the gate electrode. Similar to the preceding sections, we have to find an expression for the charge Q_g on the gate electrode in relation to the gate voltage and the surface potential. We can do that by expressing the charge densities of the oxide and the gate electrode by $\rho = qQ_f \cdot \delta(x - d_f) + qQ_g \cdot \delta(x - d_{ox})$. Putting this into the Poisson equation, we can easily integrate once to obtain an expression for the electric field $E(x)$. If we want to impose the condition that $E(d_{ox}) = 0$, we should start the integration from a variable position x, and fix the upper limit at the end of the oxide layer at d_{ox}.

$$\underbrace{E(d_{ox})}_{=0} - E(x) = \frac{q}{\epsilon_{ox}\epsilon_0}(Q_f \Theta(x - d_f) + Q_g \Theta(x - d_{ox})) \tag{G.10}$$

Here, $\Theta(x)$ is the Heaviside function. We can now integrate a second time from 0 to d_{ox} to obtain the potential difference between these positions:

$$V_g - \psi_s = \frac{q}{\epsilon_{ox}\epsilon_0}(Q_f d_f + Q_g d_{ox}) \tag{G.11}$$

Alternatively, if the fixed charge is assumed as volume density that extends uniformly between $x = 0$ and $x = d_f$, the first term in the parenthesis to the right of eq. (G.11) should be replaced by $(Q_f d_f^2)/2$ [102]. In both cases, if the charge is very close to the interface at $x = 0^+$, it will not enter explicitly into the expression for the gate charge and we can rearrange eq. (G.11) as follows:[3]

[3]Even though the fixed charge drops out of this expression, it does still add to the total charge, resulting in a discontinuity of the electric field and a kink in the potential.

$$Q_g = -\frac{\epsilon_{ox}\epsilon_0(\psi_s - V_g)}{qd_{ox}} \tag{G.12}$$

G.3 Surface recombination velocity

The total interface charge is obtained by summing up all contributions, i.e. $Q_{tot} = Q_s + Q_{it} + Q_f$ (the contribution of Q_g should be added in case of a configuration with gate electrode). The result is illustrated in figure G.2. The black lines represent the total surface charge Q_{tot}. Since the interface must be neutral, we can find the desired value of the surface potential ψ_S graphically at the position where the black curve goes through zero.

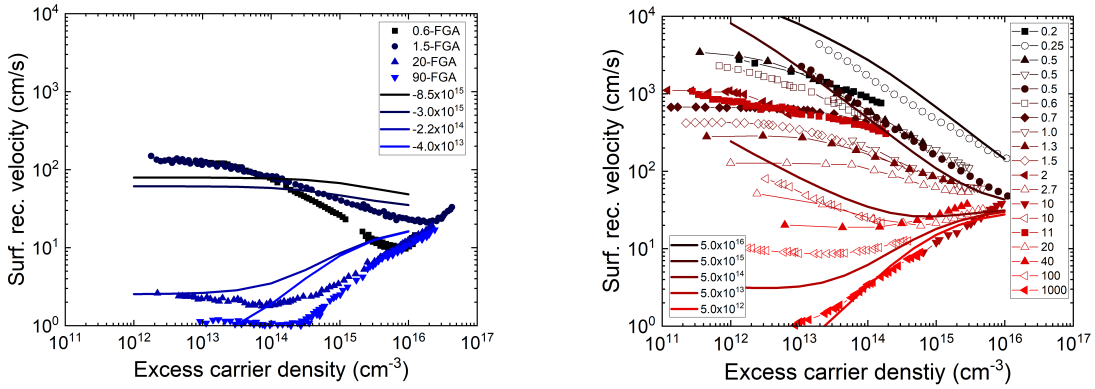


Figure G.4: Surface recombination velocities of the Si/SiO_2 interface for different doping densities (symbols), using n-type (left) and p-type material (right). Overlaid lines illustrate modelled data. Data for n-type material digitized from [125]. For p-type material, resistivities between 1000 - 0.2 Ω cm correspond to doping densities of 3.4×10^{14} - 8.5×10^{16} cm^{-3}. Data digitized from [102] (squares), [126] (circles), [127] (up triangles), [128] (down triangles), [129] (diamonds), and [130] (left triangles).

Having determined the surface potential ψ_S and thus the amount of band bending, we are finally in a position to determine the surface recombination velocity. Similar to the charge of the interface states, we have to use an integral to account for the continuous distribution of interface states and the energy dependence of their capture cross sections.

$$S = R/\Delta n \tag{G.13}$$

$$= \frac{1}{\Delta n}\int_{E_V}^{E_C} \frac{D_{it}(E)v_{th}\cdot(p_s n_s - n_i^2)}{(n_s + n_1)/\sigma_p(E) + (p_s + p_1)/\sigma_n(E)}dE \tag{G.14}$$

Figure G.4 collects experimental data of the SRV for the Si/SiO_2 interface of both polarities. As a first observation we note that the SRV is globally higher for p-type material. This is primarily related to the the larger capture cross section for electrons as illustrated in figure G.3, similar to chapter F where it appears that also most bulk defects have a higher capture rate for electrons. For the interface defects the SRV is likely further increased by the broader tail of D_{it} density on the side of the VB.

The thick solid lines in figure G.4 are obtained by using eq. (G.4) for the charge in the semiconductor eq. (G.6) for the interface charge density Q_{it}. The expression of D_{it} is based on figure G.3, but the plateau value at midgap was kept variable and the integral was replaced by

a sum with discrete values for σ_p and σ_n as illustrated by the bars in figure G.3. The model yields a tolerable agreement with measured data of both polarities for a D_{it} plateau values of 2.0×10^{10} - $3.9 \times 10^{10}\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$ for n-type material where higher values correspond to higher doping concentrations, and $3.0 \times 10^{10}\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$ for p-type material. For the fixed charge a value of $Q_f = 5 \times 10^{10}\,\mathrm{cm}^{-2}$ was used which is well within the range of values reported for the Si/SiO_2 interface.

G.4 Surface recombination at highly doped surfaces

Surface passivation is often applied to the highly doped surfaces such as those of n-type emitters, the Al-BSF, or of the boron-doped emitters of modern TOPCon cells. Figure G.5 reproduces SRV data from figure G.4 at $1 \times 10^{15}\,\mathrm{cm}^{-3}$ and plots it with respect to the wafer doping concentration. The figure includes data for higher doping concentrations that were obtained on wafers with phosphorus and boron diffusion profiles. In this case, the SRV data is shown with respect to the surface concentration; the values were determined by measuring the depth profile of the dopant density and the saturation current density j_0. Subsequently, the EDNA tool of PV-Lighthouse was used to generate the function j_0 as a function of S, and finally S was read from the characteristic [95].

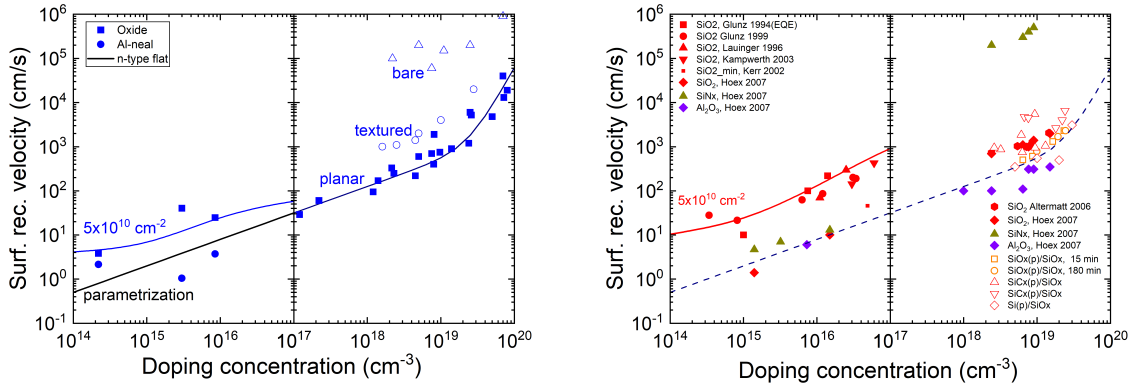


Figure G.5: Variation of S_{eff} with doping concentration for n-Si and (left) and for p-Si (right). For low doping, values were taken from figure G.4 at $\Delta n = 1 \times 10^{15}\,\mathrm{cm}^{-3}$. Data for high n-doping digitized from [95], the black line illustrate a parametrization given therein. Data for high p-doping represent digitized from [105, 131], the dashed line reproduces the parametrization and open symbols represent passivating tunnel oxide contacts [132, 133].

The modelled data for n-type and p-type material appears to agree generally with the experimental SRV data. If we would extend the modelled data, it appears that n-type surfaces are underestimated whereas p-type surfaces are overestimated, but we have keep in mind that data for high doping corresponds to higher excess carrier density and that we should not extend the model beyond $1 \times 10^{17}\,\mathrm{cm}^{-3}$ as it is based on non-degenerate statistics.

For highly doped n-type surfaces, the figure also that textured and bare surfaces have much higher SRVs than planar ones. For highly doped p-type surfaces it shows the impact of the fixed charge Q_f. Whereas the higher positive fixed charge in SiN_x w.r. to SiO_x is rather detrimental for p-type silicon, the negative fixed charge in AlO_x provides excellent passivation. Finally, the figure includes data for tunnel contacts that are briefly discussed in the next section. It appears that their SRVs behave very similarly to thick oxides, suggesting that the behaviour of the oxides

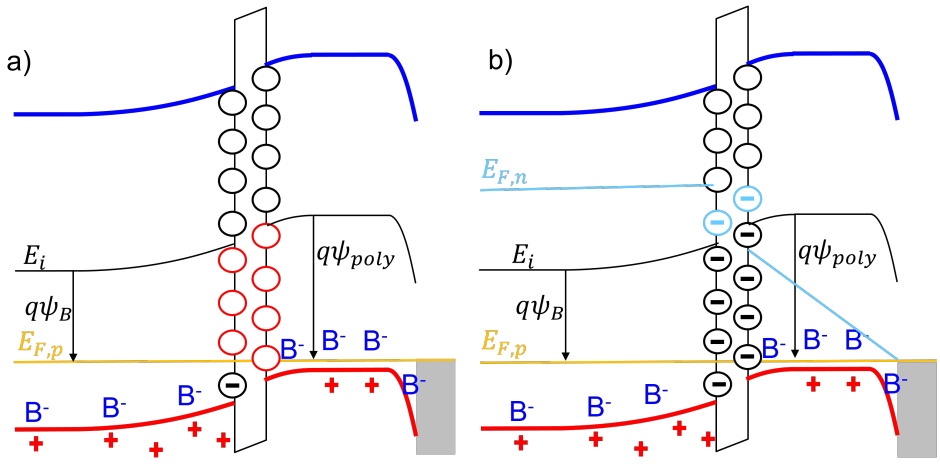


Figure G.6: Schematic band diagram of a tunnel contact between p-type c-Si, silicon oxide, and a p-type poly-layer in darkness (panel a). Illumination splits the QFLs as illustrated by the pale blue $E_{F,n}$ (panel b); as the electron capture cross section of the Si/SiO_2 interface states is larger than the one for holes, all states up to $E_{F,n}$ are shown as filled. The colours of the interface states depict again their charge with respect to a charge-neutrality condition at midgap.

is dominated by the first few atomic layers.

G.5 Tunnel oxide contact

In figure G.5 there is rather surprising finding if we look at the regio of high p-type doping densities; different from the full symbols that represent B-diffused surfaces covered with a thick layer of a dielectric, the open symbols represent tunnel oxide contacts that contain only a thin oxide layer of ca. 1.5 nm in combination with a highly p-doped layer of polycrystalline silicon. Apparently, also in these structures the properties of the interface are governed by the oxide layer, whereas the covering layer of polycrystalline silicon has only a minor impact.

The band diagram of a p-type tunnel oxide contact to a p-type wafer is illustrated in figure G.6. If the oxide layer is thin enough for tunnelling, charge carriers will exchange across the oxide until an equilibrium is reached and the Fermi levels are aligned. This yields an upward band bending with accumulation of holes in the wafer and a downwards bending with depletion in the highly doped layer. Choosing once again a geometry where the semiconductor fills the negative half-space with the interface at $x = 0$, we can copy from above the expressions for Q_S and Q_{it} in terms of the potential $\psi_s = \psi(0)$, and we can add Q_f.

Having added up all charges up to a position $x = 0^+$, we can use Gauss' law to calculate the field in the oxide layer:

$$\epsilon_{ox}\epsilon_0 E_{ox} = q(Q_s + Q_{it} + Q_f) \tag{G.15}$$

The value E_{ox} applies to a location just beyond the position of the fixed charge. As the bulk of the oxide does not contain charges, the field remains constant to $x = d_{ox}^-$, a position just before the highly doped region. Using the constant value of E_{ox}, we can project the potential drop associated to this field and the thickness of the oxide. This yields a relation between the potentials at $x = 0$ and $x = d_{ox}$.

$$E_{ox} = -\Delta\psi/d_{ox} = -(\psi(d_{ox}) - \psi(0))/d_{ox}. \tag{G.16}$$

With the help of $\psi(d_{ox})$ we can then calculate the charge in the interface states at $x = d_{ox}$. We have to use Gauss' law once more to determine the field at $x = d^+$ just outside the oxide in the poly-layer. This allows us to determine the charge in the depletion region of the highly doped layer. Subsequently, we request once again the condition of charge neutrality for the full interface to find the value of $\psi(0)$ that fulfils this condition, and in a final step we plug this into the integral for the SRV.

Appendix H

Metal-semiconductor contacts

The junction between a metal and a semiconductor was intensively investigated by Schottky in the late 1920ies. He concluded that mobile carriers are removed from the region close to the interface, leaving behind ionized dopants that form a space charge region. Eventually, he formulated the depletion approximation which is a central concept in semiconductor science [134]. To construct the band diagram of figure H.1 we assume that once the materials are in contact, the Fermi-levels of the metal and the semiconductor must align. At the same time, the offset between the Fermi-level of the metal and the CB of the semiconductor remains. We call this offset the *Schottky barrier* which is given by $q\phi_b = W_m - \chi$.[1] Consequently, the bands in the semiconductor must bend upwards by a an amount that we call *built-in voltage* V_{bi}. It is given by the band-bending in equilibrium and we can calculate it by subtracting from $q\phi_b$ the difference $E_C - E_F$.

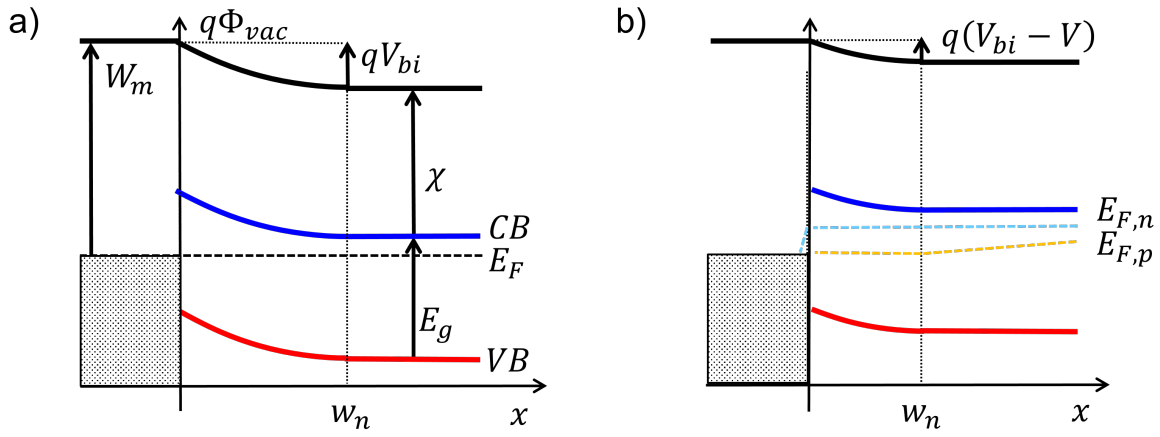


Figure H.1: Band diagram of a Schottky junction in equilibrium a) and under forward bias b).

H.1 Current-voltage characteristic

Different from the p-n junction discussed in chapter 6, the metal-semiconductor junction works purely through he transport of majority carriers. If we look at the forward biased junction in

[1] We denote by W_m the work function which is the energy needed to remove an electron from the metal. $W_m = q\phi_m$ relates it to the potential difference ϕ_m between the a state at the Fermi-level of the metal to the vacuum state.

figure H.1 b), electrons are injected into the semiconductor somewhere at the right and move towards the edge of the SCR.

Note that the depletion approximation assumes that all free carriers are gone from the SCR whereas in reality a small fraction is still left and we can use eq. (3.14) to calculate their concentration from the variation of $E_C(x)$.[2] In theory the barrier height at the interface is given by $E_C|_{x=0} - E_F = q\phi_b = W_m - \chi$. In reality, we should use the effective barrier heights of figure H.2.

Once the electrons cross the interface, they do not become minority carriers, but they join directly into the pool of conduction electrons of the metal. Under reverse bias this works very similarly; electrons from the metal have to overcome the barrier, after that they are injected as majority carriers into the CB of the n-type material. Both current directions are thus dominated by majority carriers. In case of a junction with a p-type semiconductor, a forward bias injects holes into the metal. Just as electrons can get injected into empty states above the Fermi level of the metal, there is no problem to inject holes into filled states below the Fermi level. In this case, we use eq. (3.15) to define the barrier with respect to the valence band edge.

To understand the electronic transport, it is best to start with the situation of equilibrium depicted in figure H.1 a). Since we are dealing with majority carriers, the time between two scattering events is the relaxation time τ_r. Thus, the mean free path of the carriers is very short and we can assume that the transport across the junction is dominated by the carriers that are located exactly at the interface. Let us now discuss the three main possibilities, starting from low energies:

1. Even though we are considering an n-type semiconductor here, there are still a few occupied hole states in the VB. On their energy level, they can easily recombine with electrons in the filled bands of the metal. We cannot transport large currents through this channel, but it is an almost ideal scenario for recombination of minority carriers. This explains why we usually assume infinite SRV for Schottky contacts, and when we discuss solar cells based on Schottky junctions below, this is the explanation for their low V_{oc} values.

2. A little higher up at the position of the Fermi-level, there are many electron states in the metal, but due to the position in the bandgap, there are no allowed states in the semiconductor that could contribute to the transport.

3. At yet higher energy levels, we have electrons in the mobile states of the CB which can easily interact with the empty states of the metal. Note that not all of these states are empty even though metals are generally depicted with a sharp Fermi energy. This is an idealized case which corresponds to the occupation statistics $T = 0\,\mathrm{K}$. In reality, the probability distribution tails upwards and there are some electrons excited into states with higher energy. In equilibrium, the rate of electrons that are emitted from the CB into the metal is exactly the same as the rate of electrons emitted from from highly excited states in the metal into the semiconductor.

Thus, the third of these transport paths will dominate and in equilibrium the current from the semiconductor into the metal will the same as the current from the metal into the semiconductor [135, 136]. Let us start with the component that flows from the semiconductor towards the metal. We can express a current density as product of the charge q, the density of charge carries n

[2] The expression for the electron density n at the interface as function of ϕ_b was also used by Schottky in his derivation of the current-voltage characteristic, but then his treatment differed from later interpretations as he assumed drift transport [134]. To this end, he used Gauss' law to determine the magnitude of the drift field at the interface from the charge density of the SCR, i.e. $\epsilon_0 \epsilon E|_{x=0} = qN_D w_n$.

and their velocity v. Looking at figure H.1, this is carried by the electrons that have a positive v_x-component. We can related v_x to the wave-vector component through $v_x = \hbar k_x/m^*$. After multiplying with the statistical distribution function, we integrate in $\vec{k}$-space over all states with a positive k_x component:[3]

$$j_x = \frac{1}{4\pi^3}\int_{k_x>0} q\frac{\hbar k_x}{m} f(E(\vec{k}))d^3k \tag{H.1}$$

We can approximate with Maxwell-Boltzmann statistics and we express the total energy of the electrons by a kinetic term and their electrostatic energy ϕ_b:

$$E = \frac{\hbar^2}{2m}\left(k_x^2 + k_y^2 + k_z^2\right) + q\phi_b \tag{H.2}$$

Using cylindrical coordinates along the x-direction, we define a transversal wave-vector in the yz-plane by $k_\perp = \sqrt{k_y^2 + k_z^2}$. This allows us to evaluate the integral as follows:[4]

$$\begin{aligned} j_x &= \frac{q\hbar}{4\pi^3}\int_0^\infty k_x e^{\frac{-\hbar^2}{2mkT}k_x^2}\cdot dk_x \int_0^\infty k_\perp e^{\frac{-\hbar^2}{2mkT}k_p erp^2}\cdot 2\pi dk_\perp e^{-\frac{q\phi_b}{kT}} && \text{(H.3)}\\ &= \frac{4\pi m^* k^2 T^2}{h^3} e^{-\frac{q\phi_b}{kT}} && \text{(H.4)} \end{aligned}$$

If we apply a forward bias, the QFLs will split such that $E_{F,n}$ remains flat up to the interface and then drops very rapidly within the metal. Thus, the forward bias V reduces the barrier on the side of the semiconductor to a value of $\phi_b - qV$. Next, we'd have to derive a relation for the electrons in the metal. We can get away more easily by simply stating that in equilibrium the magnitudes of the forward- and of the reverse-flowing components must be the same. As there is virtually no band bending in the metal, a bias voltage will not change this component. Finally, we can add up the resulting currents of which only one depends on the bias voltage:

$$\begin{aligned} j(V) &= \frac{4\pi m^* k^2 T^2}{h^3} e^{-q(\phi_b - V)/kT} - \frac{4\pi m^* k^2 T^2}{h^3} e^{-q\phi_b/kT} && \text{(H.5)}\\ &= \underbrace{A^* T^2 e^{(-q\phi_b/kT)}}_{j_0}\left(e^{qV/kT} - 1\right) && \text{(H.6)} \end{aligned}$$

An expression for a saturation current density preceding the round brackets was first noted by Richardson in 1914, but his arguments were based on kinetic gas theory [137]. A more adequate treatment with quantum-mechanics yields eq. (H.6) with the a Richardson constant $A^* = 4\pi m^* q k^2/h^3$. The underlying charge transport is called thermionic emission (TE). It relies purely on the barrier-height ϕ_b and holds only for weakly doped semiconductors. We can make it more accurate by taking into account that the energy surface is actually an ellipsoid with different effective masses [136].

For the case of high doping concentrations, the barrier height remains unchanged, but the depletion region becomes very narrow. Thus, there is a chance that charge-carriers tunnel through the

[3]The integral resembles eq. 3.9, but here the velocity component v_x introduces an additional dependence on direction.

[4]We can capture the physics of the process by assuming $j = qnv$, using the thermal velocity $v_{th} = \sqrt{3kT/m^*}$ and $n = N_C \exp(-\phi_b/kT)$. Assuming further that only half of the carriers move towards the interface, we obtain a pre-factor with the same physical quantities, but numerically off by a factor of approximately two.

barrier. For historic reasons this tunnelling process is called field emission (FE) as its mathematical treatment is very similar to the one of field emission of electrons from a heated metal electrode into free space. For intermediate doping concentrations there is a mixed process where charge carriers are excited over a part of the barrier and proceed by tunnelling through the remainder. This transport process is called thermionic field emission (TFE) [138].

H.2 Junction capacitance and Schottky barrier

Figure H.1 illustrates that the width of the SCR is reduced when a forward bias is applied. In addition to the forward current flowing through the junction, we note that the ionised donor states close to the edge of the SCR change their charge state from positive to neutral. In other words, the charge of the SCR is reduced from $qN_D w_n(0)$ to $qN_D w_n(V_{fwd})$, and we can retrieve the charge by reverting to zero bias. Likewise, under reverse bias we increase the width of the SCR, resulting in more positively charged dopant states. Effectively, the SCR acts like a capacitance. For an SCR with constant doping concentration, we can express this as follows:

$$\begin{aligned} C &= \frac{dQ}{dV} = \frac{d(-qN_D w_n(V))}{dV} \\ &= -qN_D \frac{d}{dV}\sqrt{\frac{2\epsilon\epsilon_0(V_{bi}-V)}{qN_D}} \\ &= \sqrt{\frac{\epsilon\epsilon_0 qN_D}{2(V_{bi}-V)}} \end{aligned} \tag{H.7}$$

The eq. (H.7) gives a powerful tool to access the doping concentration N_D and the built-in voltage V_{bi}. If $1/C^2$ is plotted versus V (a so-called Mott-Schottky plot), the doping concentration N_D can be determined from the slope and the built-in voltage V_{bi} is given by the extrapolated intersect on the abscissa. With the found doping concentration we can calculate $E_C - E_F$, it is then easy to determine the Schottky barrier.

For the example of c-Si with its value of $\chi = 4.05\,\text{eV}$, there are many metals with low work-function for which the the barrier height $q\phi_b$ would become negative. This means that there would not be a depletion region but an accumulating junction that would be perfectly Ohmic. In reality, all known metals form depleting junctions with silicon due to the formation of interface states or a thin interfacial layer [139]. Thus, the barrier-heights as defined above remain a theoretical concept, in reality we must use effective barrier-heights obtained by experiment. Figure H.2 shows that they vary between 0.4 eV for Mg and 0.85 eV for Ir and Pt, even though the work-functions of these metals span a range of more than 2 eV [139, 119].

Schottky-junctions with p-type semiconductors are treated similarly to the n-type Schottky junctions discussed so far, but their barriers are defined with respect to E_V. For example, gold with its electron barrier of 0.8 eV has thus a hole barrier of only 0.35 eV.

H.3 Schottky solar cells

In the early 1970ies, Schottky junctions were considered for solar cells because evaporation of a metal promised to be a simple process for the formation of a junction. Eventually it turned out that their performance was very low. To understand why, let us start with the electron current of eq. (H.6). As usual, we can interpret the term before the round brackets as saturation current

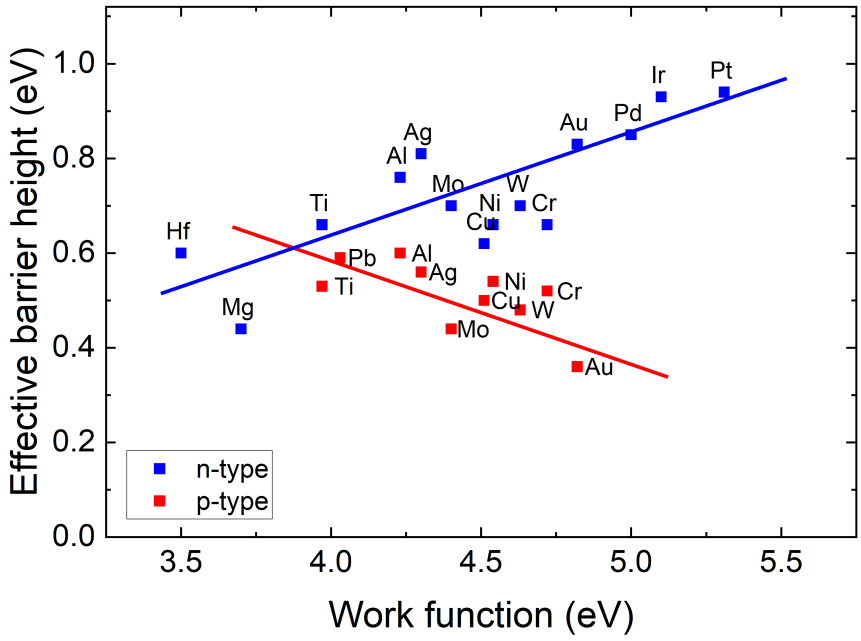


Figure H.2: Effective barrier heights between common metals and n- as well as p-type crystalline silicon (after [119]).

density j_0. Figure H.3 illustrates it for two typical values of the Schottky barrier. Regarding the saturation current of the holes, figure H.1 suggests that it is due to injection of minority carriers into the semiconductor, and thus we can use the result that we determined for the p-n junction. The comparison in figure H.3 shows that the total saturation current will be completely dominated by the electron current, and if we assume the superposition principle with a photocurrent of $-j_L$, we can use eq. (6.21) to estimate the V_{oc} values shown on the right axis.

Clearly, Schottky junctions are limited by their large saturation current denstites that yield low V_{oc}. The most obvious mitigation is an increased barrier, but figure H.2 suggests that this is not possible with metals. A different route was mentioned in section 10.3 where we discussed that a thin oxide layer offers a possibility to manipulate the saturation current density. For an oxide thickness of δ, the $j(V)$ characteristic of a metal-insulator-semiconductor (MIS) junction reads [141]:

$$j(V) = A^* T^2 e^{(-q\phi_b/kT)} e^{-\frac{4\pi}{h}\sqrt{2m_t^*\chi_t}\cdot\delta} \left(e^{qV/kT} - 1\right) \tag{H.8}$$

Here, χ_t is height of the tunnelling barrier and m_t^* is the effective tunnelling mass. These are based on the somewhat questionable assumption that the bands known from the infinite crystal also form in a layer that is only a few nm thick, such that one could define an energy surface and extract from its curvature an effective mass, the so-called tunnelling mass m_t.

The insertion of a tunnelling oxide was tried and the resulting devices did have better V_{oc} up to 400 mV, but they suffered from low photocurrents because of the poor transmission of the full-area metal contact at the front [140]. Improved photocurrents were eventually obtained by applying a metallisation grid with the MIS contact underneath the grid whereas the rest of the surface was covered with a passivating oxide. This design fixed the transparency issue, but it had limitations of its own because it suffered from poor lateral transport. A workaround to that issue was the

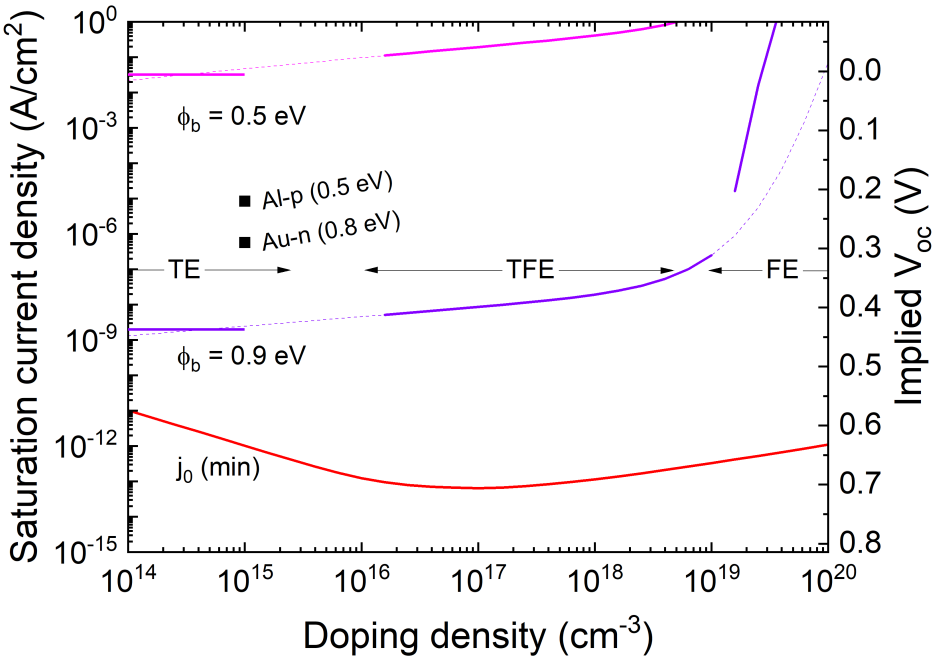


Figure H.3: Saturation current densities of Schottky junctions for thermionic emission (TE) and increasing amounts of tunnelling (TFE and FE). Pink and purple curves illustrate barriers of 0.4 and 0.9 eV, respectively. The red line reproduces $j_{0,p}$ related to diffusion of minority carriers (holes) from figure D.2. The right scale illustrates the implied V_{oc}, symbols show measured V_{oc} values of Schottky solar cells (references of MIS cells without oxide, [140]).

use of a large fixed charge in the insulating layer. For example, compared to SiO_2, the density of positive fixed charge in Si_3N_4 is much higher. If applied to the surface of a p-type wafer, the charge is strong enough to repel majority carriers from the interface, and to attract electrons. For strongly charged SiN_x, electrons can even become majority carriers in a thin inversion layer close to the surface which is then used for lateral transport. Even though the fixed charge seemed to be stable, MIS cells were substantially improved with a diffused junction in a configuration called MINP (metal insulator n-p junction, [47]). Even so, they they were soon abandoned in favour of p-n junction cells. Currently we are seeing a return of tunnel oxides where they are used as n-type contact in the TOPCon design.

H.4 Contact resistivity

Schottky junctions may not be very suitable for carrier collection, but they are still present in the metallic contacts of solar cells. Thus, we need to know their contact resistivity. This is obtained by evaluating the derivative of eq. (H.6) [142]:

$$\rho_c = \left(\left. \frac{dj}{dV} \right|_{V=0} \right)^{-1}$$

This yields the contact resistivity for lowly doped semiconductors, for higher doping we should use the corresponding expressions for TFE and FE. Figure H.4 collects some data for the contact

materials frequently used in silicon solar cells. Symbols present measured data, the overlaid lines are the theoretical characteristics for TFE and FE at low and high doping-concentrations, respectively. Especially for weakly doped silicon this would result in rather high contact resistivity.

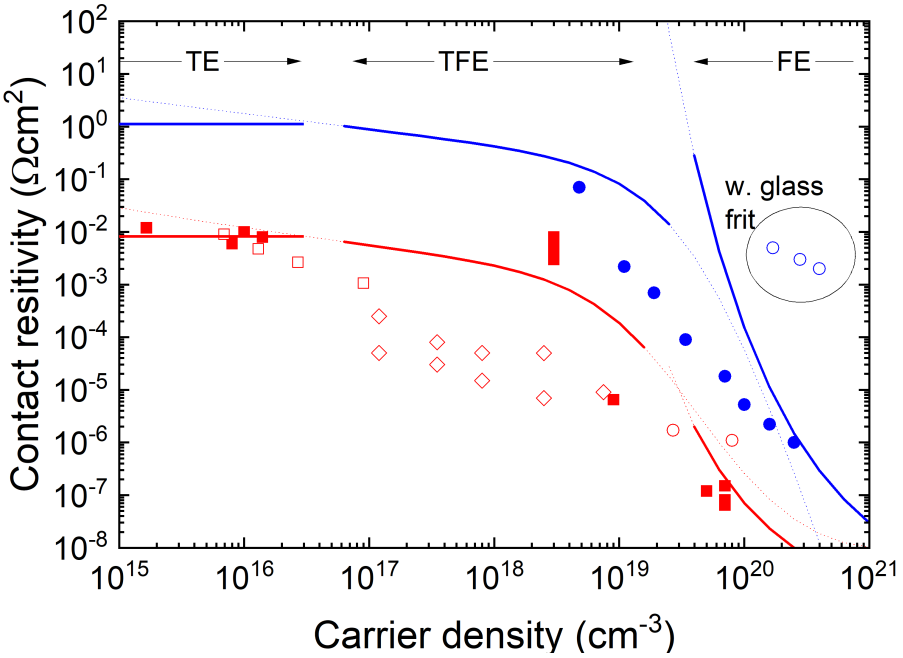


Figure H.4: Contact resitivites of Ag with n-doped silicon (blue) and of Al with p-doped silicon (red). The lines illustrate the limiting values of the eq. (H.6) for thermionic emission (TE) as well as the characteristics of thermionic field emission (TFE) and tunnelling (FE).

The blue curve illustrates the contact resistivities between n-type silicon and silver which has an effective barrier of 0.5 eV. In solar cells, the contact of silver with the n-layer is usually made with a mixture of silver and glass frit. After processing, a thin layer of insulating glass frit remains at the interface, and only occasionally small silver crystals make contact with silicon. Thus, the actual contact area is very small and the macroscopically measured contact resistivity is increased by several orders of magnitude as depicted by the open blue symbols.

The red curve in figure H.4 shows the properties of the contact between p-type silicon and aluminium. The data is reasonably well fitted with an effective barrier of 0.4 eV which is less than the value shown in figure H.2. One possible explanation is that evaporated metal contacts are often annealed after their fabrication. In case of Al, this can be sufficient to incorporate Al into silicon where it forms a p-type dopant. On small scale, the metal might thus be in contact with material of higher doping than the nominal doping of the wafer.

Bibliography and references

[1] W. Shockley. "The theory of p-n junctions in semiconductors and p-n junction transistors". In: *Bell System Technical Journal* 28.3 (1949), pp. 435–489.

[2] M. A. Green. "Accuracy of analytical expressions for solar cell fill factors". In: *Solar Cells* 7.3 (1982), pp. 337–340.

[3] W. Shockley and H. J. Queisser. "Detailed balance limit of efficiency of p-n junction solar cells". In: *Journal of applied physics* 32.3 (1961), pp. 510–519.

[4] L. C. Hirst and N. J. Ekins-Daukes. "Fundamental losses in solar cells". In: *Progress in Photovoltaics: Research and Applications* 19.3 (2011), pp. 286–293.

[5] A. Richter, M. Hermle, and S. W. Glunz. "Reassessment of the limiting efficiency for crystalline silicon solar cells". In: *IEEE journal of photovoltaics* 3.4 (2013), pp. 1184–1191.

[6] S. Schäfer and R. Brendel. "Accurate calculation of the absorptance enhances efficiency limit of crystalline silicon solar cells with lambertian light trapping". In: *IEEE Journal of Photovoltaics* 8.4 (2018), pp. 1156–1158.

[7] A. De Vos. "Detailed balance limit of the efficiency of tandem solar cells". In: *Journal of Physics D: Applied Physics* 13.5 (1980), p. 839.

[8] O. Dupré, R. Vaillon, and M. A. Green. "Physics of the temperature coefficients of solar cells". In: *Solar energy materials and solar cells* 140 (2015), pp. 92–100.

[9] M. Vellini, M. Gambini, and V. Prattella. "Environmental impacts of PV technology throughout the life cycle: Importance of the end-of-life management for Si-panels and CdTe-panels". In: *Energy* 138 (2017), pp. 1099–1111.

[10] V. Fthenakis and E. Leccisi. "Updated sustainability status of crystalline silicon-based photovoltaic systems: Life-cycle energy and environmental impact reduction trends". In: *Progress in Photovoltaics: Research and Applications* 29.10 (2021), pp. 1068–1077.

[11] A. Müller et al. "A comparative life cycle assessment of silicon PV modules: Impact of module design, manufacturing location and inventory". In: *Solar Energy Materials and Solar Cells* 230 (2021), p. 111277.

[12] M. M. de Wild-Scholten. "Energy payback time and carbon footprint of commercial photovoltaic systems". In: *Solar Energy Materials and Solar Cells* 119 (2013), pp. 296–305.

[13] M. A. Green et al. "Solar cell efficiency tables (Version 65)". In: *Progress in Photovoltaics: Research and Applications* 33 (2025), pp. 3–15.

[14] B. O'regan and M. Grätzel. "A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO2 films". In: *nature* 353.6346 (1991), pp. 737–740.

[15] G. Dresselhaus, A. Kip, and C. Kittel. "Cyclotron resonance of electrons and holes in silicon and germanium crystals". In: *Physical Review* 98.2 (1955), p. 368.

[16] E. O. Kane. "Energy band structure in p-type germanium and silicon". In: *Journal of Physics and Chemistry of Solids* 1.1-2 (1956), pp. 82–99.

[17] J. R. Chelikowsky and M. L. Cohen. "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors". In: *Physical Review B* 14.2 (1976), p. 556.

[18] P. Altermatt et al. "Reassessment of the intrinsic carrier density in crystalline silicon in view of band-gap narrowing". In: *Journal of Applied Physics* 93.3 (2003), pp. 1598–1604.

[19] P. Drude. "Zur Elektronentheorie der Metalle". In: *Annalen der physik* 306.3 (1900), pp. 566–613.

[20] P. Drude. "Zur elektronentheorie der Metalle; II. Teil. Galvanomagnetische und thermomagnetische Effecte". In: *Annalen der physik* 308.11 (1900), pp. 369–402.

[21] J. Y. Seto. "The electrical properties of polycrystalline silicon films". In: *Journal of Applied Physics* 46.12 (1975), pp. 5247–5254.

[22] K. Smith et al. "The solar cells and their mounting". In: *Bell System Technical Journal* 42.4 (1963), pp. 1765–1816.

[23] H. Deckman et al. "Optically enhanced amorphous silicon solar cells". In: *Applied Physics Letters* 42.11 (1983), pp. 968–970.

[24] A. Wang, J. Zhao, and M. Green. "24% efficient silicon solar cells". In: *Applied Physics Letters* 57.6 (1990), pp. 602–604.

[25] T. Trupke et al. "Temperature dependence of the radiative recombination coefficient of intrinsic crystalline silicon". In: *Journal of Applied Physics* 94.8 (2003), pp. 4930–4937.

[26] D. Matsakis et al. "A renaming proposal: The Auger–Meitner effect". In: *Physics Today* 72.9 (2019), pp. 10–11.

[27] A. Hangleiter and R. Häcker. "Enhancement of band-to-band Auger recombination by electron-hole correlations". In: *Physical Review Letters* 65.2 (1990), p. 215.

[28] M. J. Kerr, A. Cuevas, and P. Campbell. "Limiting efficiency of crystalline silicon solar cells due to Coulomb-enhanced Auger recombination". In: *Progress in Photovoltaics: Research and Applications* 11.2 (2003), pp. 97–104.

[29] A. Richter et al. "Improved quantitative description of Auger recombination in crystalline silicon". In: *Physical Review B* 86.16 (2012), p. 165202.

[30] T. Niewelt et al. "Reassessment of the intrinsic bulk recombination in crystalline silicon". In: *Solar Energy Materials and Solar Cells* 235 (2022), p. 111467.

[31] W. Shockley and W. Read Jr. "Statistics of the recombinations of holes and electrons". In: *Physical Review* 87.5 (1952), p. 835.

[32] R. N. Hall. "Electron-hole recombination in germanium". In: *Physical Review* 87.2 (1952), p. 387.

[33] G. Kousik, Z. Ling, and P. Ajmera. "Nondestructive technique to measure bulk lifetime and surface recombination velocities at the two surfaces by infrared absorption due to pulsed optical excitation". In: *Journal of applied Physics* 72.1 (1992), pp. 141–146.

[34] A. Sproul. "Dimensionless solution of the equation describing the effect of surface recombination on carrier decay in semiconductors". In: *Journal of Applied Physics* 76.5 (1994), pp. 2851–2854.

[35] J. Brody, A. Rohatgi, and A. Ristow. "Review and comparison of equations relating bulk lifetime and surface recombination velocity to effective lifetime measured under flash lamp illumination". In: *Solar Energy materials and Solar Cells* 77.3 (2003), pp. 293–301.

[36] I. Anfimov et al. "On the Problem of Determining the Bulk Lifetime by Photoconductivity Decay on the Unpassivated Samples of Monocrystalline Silicon". In: *Russian Microelectronics* 46.8 (2017), pp. 585–590.

[37] E. Yablonovitch and T. Gmitter. "Auger recombination in silicon at low carrier densities". In: *Applied Physics Letters* 49.10 (1986), pp. 587–589.

[38] R. M. Swanson et al. "Point-contact silicon solar cells". In: *IEEE Transactions on Electron Devices* 31.5 (1984), pp. 661–664.

[39] A. Blakers et al. "18-percent efficient terrestrial silicon solar cells". In: *IEEE electron device letters* 5.1 (1984), pp. 12–13.

[40] A. Blakers and M. Green. "20% efficiency silicon solar cells". In: *Applied Physics Letters* 48.3 (1986), pp. 215–217.

[41] T. Tiedje et al. "Limiting efficiency of silicon solar cells". In: *IEEE Transactions on electron devices* 31.5 (1984), pp. 711–716.

[42] M. Green et al. "Improvements in silicon solar cell efficiency". In: *Solar cells* 17.1 (1986), pp. 75–83.

[43] O. Breitenstein et al. "Interpretation of the commonly observed IV characteristics of c-Si cells having ideality factor larger than two". In: *2006 IEEE 4th World Conference on Photovoltaic Energy Conference*. Vol. 1. IEEE. 2006, pp. 879–884.

[44] Y. Sakaguchi et al. "Production of high purity silicon by carbothermic reduction of silica using AC-arc furnace with heated shaft". In: *ISIJ international* 32.5 (1992), pp. 643–649.

[45] G. Sævarsdottir, H. Kvande, and T. Magnusson. "Greenhouse gas emissions from silicon production-development of carbon footprint with changing energy systems". In: *Proc. Infacon XVI*. 2021.

[46] P. D. Miller, J. Lee, and I. B. Cutler. "The reduction of silica with carbon and silicon carbide". In: *Journal of the American Ceramic Society* 62.3-4 (1979), pp. 147–149.

[47] M. Green et al. "19.1% efficient silicon solar cell". In: *Applied Physics Letters* 44.12 (1984), pp. 1163–1164.

[48] M. D. Lammert and R. J. Schwartz. "The interdigitated back contact solar cell: A silicon solar cell for use in concentrated sunlight". In: *IEEE Transactions on Electron Devices* 24.4 (1977), pp. 337–342.

[49] A. W. Blakers et al. "22.8% efficient silicon solar cell". In: *Applied Physics Letters* 55.13 (1989), pp. 1363–1365.

[50] V. Borisenko and S. Yudin. "Steady-state solubility of substitutional impurities in silicon". In: *physica status solidi (a)* 101.1 (1987), pp. 123–127.

[51] R. B. Fair and J. Tsai. "A quantitative model for the diffusion of phosphorus in silicon and the emitter dip effect". In: *Journal of The Electrochemical Society* 124.7 (1977), p. 1107.

[52] C. Ballif et al. "Silver thick-film contacts on highly doped n-type silicon emitters: structural and electronic properties of the interface". In: *Applied Physics Letters* 82.12 (2003), pp. 1878–1880.

[53] T. Röder et al. “Add-on laser tailored selective emitter solar cells”. In: *Progress in Photovoltaics: Research and Applications* 18.7 (2010), pp. 505–510.

[54] G. Agostinelli et al. “Local contact structures for industrial PERC-type solar cells”. In: *Proceedings of the 20th European Photovoltaic Solar Energy Conference*. 2005, pp. 942–945.

[55] S. Gatz et al. “19.4%-efficient large-area fully screen-printed silicon solar cells”. In: *physica status solidi (RRL)–Rapid Research Letters* 5.4 (2011), pp. 147–149.

[56] T. Dullweber and J. Schmidt. “Industrial silicon solar cells applying the passivated emitter and rear cell (PERC) concept - A review”. In: *IEEE Journal of Photovoltaics* 6.5 (2016), pp. 1366–1381.

[57] K. Park, D. Seo, and J. Lee. “Conductivity of silver paste prepared from nanoparticles”. In: *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 313 (2008), pp. 351–354.

[58] S. Rane et al. “Firing and processing effects on microstructure of fritted silver thick film electrode materials for solar cells”. In: *Materials Chemistry and Physics* 82.1 (2003), pp. 237–245.

[59] K. Yoshikawa et al. “Silicon heterojunction solar cell with interdigitated back contacts for a photoconversion efficiency over 26%”. In: *Nature energy* 2.5 (2017), p. 17032.

[60] H. De Graaff and J. G. De Groot. “The SIS tunnel emitter: A theory for emitters with thin interface layers”. In: *IEEE Transactions on Electron Devices* 26.11 (1979), pp. 1771–1776.

[61] A. Blakers and M. Green. “678-mV open-circuit voltage silicon solar cells”. In: *Applied Physics Letters* 39.6 (1981), pp. 483–485.

[62] J. Gan and R. Swanson. “Polysilicon emitters for silicon concentrator solar cells”. In: *IEEE Conference on Photovoltaic Specialists*. IEEE. 1990, pp. 245–250.

[63] F. Feldmann et al. “Tunnel oxide passivated contacts as an alternative to partial rear contacts”. In: *Solar Energy Materials and Solar Cells* 131 (2014), pp. 46–50.

[64] F. Haase et al. “Laser contact openings for local poly-Si-metal contacts enabling 26.1%-efficient POLO-IBC solar cells”. In: *Solar Energy Materials and Solar Cells* 186 (2018), pp. 184–193.

[65] D. Yan et al. “Phosphorus-diffused polysilicon contacts for solar cells”. In: *Solar Energy Materials and Solar Cells* 142 (2015), pp. 75–82.

[66] D. Yan et al. “23% efficient p-type crystalline silicon solar cells with hole-selective passivating contacts based on physical vapor deposition of doped silicon films”. In: *Applied Physics Letters* 113.6 (2018).

[67] 2. MacFarlane et al. “Fine structure in the absorption-edge spectrum of Si”. In: *Physical Review* 111.5 (1958), p. 1245.

[68] Y. P. Varshni. “Temperature dependence of the energy gap in semiconductors”. In: *physica* 34.1 (1967), pp. 149–154.

[69] C. Thurmond. “The standard thermodynamic functions for the formation of electrons and holes in Ge, Si, GaAs, and GaP”. In: *Journal of the Electrochemical Society* 122.8 (1975), p. 1133.

[70] V. Alex, S. Finkbeiner, and J. Weber. “Temperature dependence of the indirect energy gap in crystalline silicon”. In: *Journal of Applied Physics* 79.9 (1996), pp. 6943–6946.

[71] R. Pässler. "Dispersion-related description of temperature dependencies of band gaps in semiconductors". In: *Physical Review B* 66.8 (2002), p. 085201.

[72] G. Nilsson and G. Nelin. "Study of the homology between silicon and germanium by thermal-neutron spectrometry". In: *Physical Review B* 6.10 (1972), p. 3777.

[73] C. Hamaguchi and C. Hamaguchi. *Basic semiconductor physics*. Vol. 9. Springer, 2010.

[74] S. L. Adler. "Quantum theory of the dielectric constant in real solids". In: *Physical Review* 126.2 (1962), p. 413.

[75] H. D. Barber. "Effective mass and intrinsic concentration in silicon". In: *Solid-State Electronics* 10.11 (1967), pp. 1039–1051.

[76] M. A. Green. "Intrinsic concentration, effective densities of states, and effective mass in silicon". In: *Journal of Applied Physics* 67.6 (1990), pp. 2944–2954.

[77] R. Humphreys. "Valence band averages in silicon: Anisotropy and non-parabolicity". In: *Journal of Physics C: Solid State Physics* 14.21 (1981), p. 2935.

[78] J. Hensel, H. Hasegawa, and M. Nakayama. "Cyclotron resonance in uniaxially stressed silicon. II. Nature of the covalent bond". In: *Physical Review* 138.1A (1965), A225.

[79] D. Klaassen, J. Slotboom, and H. De Graaff. "Unified apparent bandgap narrowing in n-and p-type silicon". In: *Solid-State Electronics* 35.2 (1992), pp. 125–129.

[80] X. Aymerich-Humet, F. Serra-Mestres, and J. Millan. "A generalized approximation of the Fermi–Dirac integrals". In: *Journal of applied physics* 54.5 (1983), pp. 2850–2851.

[81] W. Spitzer and H. Fan. "Infrared absorption in n-type silicon". In: *Physical Review* 108.2 (1957), p. 268.

[82] P. Schumann et al. "Silicon optical constants in the infrared". In: *Journal of the Electrochemical Society* 118.1 (1971), p. 145.

[83] P. Schmid. "Optical absorption in heavily doped silicon". In: *Physical Review B* 23.10 (1981), p. 5531.

[84] P. P. Altermatt et al. "Injection dependence of spontaneous radiative recombination in c-Si: experiment, theoretical analysis, and simulation". In: *NUSOD'05. Proceedings of the 5th International Conference on Numerical Simulation of Optoelectronic Devices, 2005.* IEEE. 2005, pp. 47–48.

[85] M. Rüdiger et al. "Parameterization of free carrier absorption in highly doped silicon for solar cells". In: *IEEE transactions on electron devices* 60.7 (2013), pp. 2156–2163.

[86] P. Altermatt et al. "Injection dependence of spontaneous radiative recombination in crystalline silicon: Experimental verification and theoretical analysis". In: *Applied physics letters* 88.26 (2006), p. 261901.

[87] H. Sai et al. "Very thin crystalline silicon cells: A way to improve the photovoltaic performance at elevated temperatures". In: *Progress in Photovoltaics: Research and Applications* 29.10 (2021), pp. 1093–1104.

[88] H. Lin et al. "Silicon heterojunction solar cells with up to 26.81% efficiency achieved by electrically optimized nanocrystalline-silicon hole contact layers". In: *Nature Energy* (2023), pp. 1–11.

[89] H. Sai, H. Umishio, and T. Matsui. "Very Thin (56 μm) Silicon Heterojunction Solar Cells with an Efficiency of 23.3% and an Open-Circuit Voltage of 754 mV". In: *Solar RRL* 5.11 (2021), p. 2100634.

[90] J. Fossum and E. Burgess. "Silicon solar cell development for concentrated-sunlight, high-temperature applications". In: *12th Photovoltaic Specialists Conference*. 1976, pp. 737–743.

[91] H. Elgamel et al. "640 mV open-circuit voltage multicrystalline silicon solar cells: role of base doping on device parameters". In: *Solar energy materials and solar cells* 36.1 (1995), pp. 99–105.

[92] J. Zhao, A. Wang, and M. A. Green. "24· 5% Efficiency silicon PERT cells on MCZ substrates and 24· 7% efficiency PERL cells on FZ substrates". In: *Progress in photovoltaics: research and applications* 7.6 (1999), pp. 471–474.

[93] S. Glunz, J. Dicker, and P. Altermatt. "Band gap narrowing in p-type base regions of solar cells". In: *17th EU Photovoltaic Energy Conference*. 2001, pp. 1391–1395.

[94] M. Kerr et al. "Surface recombination velocity of phosphorus-diffused silicon solar cell emitters passivated with plasma enhanced chemical vapor deposited silicon nitride and thermal silicon oxide". In: *Journal of applied physics* 89.7 (2001), pp. 3821–3826.

[95] P. P. Altermatt et al. "Numerical modeling of highly doped Si: P emitters based on Fermi–Dirac statistics and self-consistent material parameters". In: *Journal of Applied Physics* 92.6 (2002), pp. 3187–3197.

[96] H. De Graaff, J. Slotboom, and A. Schmitz. "The emitter efficiency of bipolar transistors: Theory and experiments". In: *Solid-State Electronics* 20.6 (1977), pp. 515–521.

[97] C. Selvakumar. "Simple general analytical solution to the minority carrier transport in heavily doped semiconductors". In: *Journal of applied physics* 56.12 (1984), pp. 3476–3478.

[98] P. Altermatt, A. Schenk, and G. Heiser. "A simulation model for the density of states and for incomplete ionization in crystalline silicon. I. Establishing the model in Si: P". In: *Journal of Applied Physics* 100.11 (2006), p. 113715.

[99] B. Min et al. "Limitation of industrial phosphorus-diffused emitters by SRH recombination". In: *Energy Procedia* 55 (2014), pp. 115–120.

[100] J. Reynolds and A. Meulenberg Jr. "Measurement of diffusion length in solar cells". In: *Journal of Applied Physics* 45.6 (1974), pp. 2582–2592.

[101] A. Goetzberger, B. Voß, and J. Knobloch. *Sonnenenergie: Photovoltaik, Teubner Studienbücher*. 1997.

[102] A. G. Aberle, S. Glunz, and W. Warta. "Impact of illumination level and oxide parameters on Shockley–Read–Hall recombination at the Si-SiO2 interface". In: *Journal of Applied Physics* 71.9 (1992), pp. 4422–4431.

[103] J. Hauser and P. Dunbar. "Minority carrier reflecting properties of semiconductor high-low junctions". In: *Solid State Electronics* 18.7 (1975), pp. 715–716.

[104] M. Rüdiger et al. "Effect of incomplete ionization for the description of highly aluminum-doped silicon". In: *Journal of applied physics* 110.2 (2011), p. 024508.

[105] B. Hoex et al. "Excellent passivation of highly doped p-type Si surfaces by the negative-charge-dielectric Al 2 O 3". In: *Applied Physics Letters* 91.11 (2007), p. 112107.

[106] B. Rose and H. Weaver. "Determination of effective surface recombination velocity and minority-carrier lifetime in high-efficiency Si solar cells". In: *Journal of applied physics* 54.1 (1983), pp. 238–247.

[107] J. D. Murphy et al. "Parameterisation of injection-dependent lifetime measurements in semiconductors in terms of Shockley-Read-Hall statistics: An application to oxide precipitates in silicon". In: *Journal of Applied Physics* 111.11 (2012).

[108] J. E. Birkholz et al. "Electronic properties of iron-boron pairs in crystalline silicon by temperature-and injection-level-dependent lifetime measurements". In: *Journal of Applied Physics* 97.10 (2005).

[109] D. MacDonald et al. "Doping dependence of the carrier lifetime crossover point upon dissociation of iron-boron pairs in crystalline silicon". In: *Applied Physics Letters* 89.14 (2006).

[110] S. Rein. *Lifetime spectroscopy: a method of defect characterization in silicon for photovoltaic applications*. Vol. 85. Springer Science & Business Media, 2005.

[111] A. Liu, S. P. Phang, and D. Macdonald. "Gettering in silicon photovoltaics: A review". In: *Solar Energy Materials and Solar Cells* 234 (2022), p. 111447.

[112] J. Schmidt and A. Cuevas. "Electronic properties of light-induced recombination centers in boron-doped Czochralski silicon". In: *Journal of Applied Physics* 86.6 (1999), pp. 3175–3180.

[113] B. J. Hallam et al. "Hydrogen passivation of BO defects in Czochralski silicon". In: *Energy Procedia* 38 (2013), pp. 561–570.

[114] S. Rein et al. "Quantitative correlation of the metastable defect in Cz-silicon with different impurities". In: *3rd World Conference onPhotovoltaic Energy Conversion, 2003. Proceedings of.* Vol. 2. IEEE. 2003, pp. 1048–1052.

[115] S. Rein and S. Glunz. "Electronic properties of the metastable defect in boron-doped Czochralski silicon: Unambiguous determination by advanced lifetime spectroscopy". In: *Applied Physics Letters* 82.7 (2003), pp. 1054–1056.

[116] D. Macdonald and L. Geerligs. "Recombination activity of interstitial iron and other transition metal point defects in p-and n-type crystalline silicon". In: *Applied Physics Letters* 85.18 (2004), pp. 4061–4063.

[117] L. E. Mundt et al. "Spatially resolved impurity identification via temperature-and injection-dependent photoluminescence imaging". In: *IEEE Journal of Photovoltaics* 5.5 (2015), pp. 1503–1509.

[118] H. Fukuda et al. "Process dependence of the SiO2/Si (100) interface trap density of ultrathin SiO2 films". In: *Journal of applied physics* 72.5 (1992), pp. 1906–1911.

[119] D. K. Schroder and D. L. Meier. "Solar cell contact resistance—A review". In: *IEEE Transactions on electron devices* 31.5 (1984), pp. 637–647.

[120] L. Kronik and Y. Shapira. "Surface photovoltage phenomena: theory, experiment, and applications". In: *Surface science reports* 37.1-5 (1999), pp. 1–206.

[121] T. Sakurai and T. Sugano. "Theory of continuously distributed trap states at Si-SiO2 interfaces". In: *Journal of applied physics* 52.4 (1981), pp. 2889–2896.

[122] J. Snel. "The doped Si/SiO2 interface". In: *Solid-State Electronics* 24.2 (1981), pp. 135–139.

[123] S. Peterström. "Si-SiO2 interface trap density in boron-and phosphorus-implanted silicon". In: *Applied physics letters* 63.5 (1993), pp. 672–674.

[124] E. Kameda et al. "Interface state density in n-MOSFETs with Si-implanted gate oxide measured by subthreshold slope analysis". In: *Solid-State Electronics* 43.3 (1999), pp. 565–573.

[125] M. J. Kerr. "Surface, emitter and bulk recombination in silicon and development of silicon nitride passivated solar cells". PhD thesis. The Australian National University, 2002.

[126] H. Kampwerth, S. Rein, and S. W. Glunz. "Pure experimental determination of surface recombination properties with high reliability". In: *3rd World Conference onPhotovoltaic Energy Conversion, 2003. Proceedings of.* Vol. 2. IEEE. 2003, pp. 1073–1076.

[127] S. W. Glunz et al. "Injection-level-dependent recombination velocities at the Si-SiO2 interface for various dopant concentrations". In: *Journal of Applied Physics* 75.3 (1994), pp. 1611–1615.

[128] S. W. Glunz et al. "Field-effect passivation of the SiO 2 Si interface". In: *Journal of Applied Physics* 86.1 (1999), pp. 683–691.

[129] T. Lauinger et al. "Record low surface recombination velocities on 1 Ω cm p-silicon using remote plasma silicon nitride passivation". In: *Applied physics letters* 68.9 (1996), pp. 1232–1234.

[130] A. Stephens, A. Aberle, and M. Green. "Surface recombination velocity measurements at the silicon–silicon dioxide interface by microwave-detected photoconductance decay". In: *Journal of Applied Physics* 76.1 (1994), pp. 363–370.

[131] P. P. Altermatt et al. "The surface recombination velocity at boron-doped emitters: comparison between various passivation techniques". In: *Proceedings of the 21st European Photovoltaic Solar Energy Conference*. WIP Renewable Energies Dresden Germany. 2006, pp. 647–650.

[132] G. Nogay et al. "Interplay of annealing temperature and doping in hole selective rear contacts based on silicon-rich silicon-carbide thin films". In: *Solar Energy Materials and Solar Cells* 173 (2017), pp. 18–24.

[133] P. Wyss et al. "A mixed-phase SiO x hole selective junction compatible with high temperatures used in industrial solar cell manufacturing". In: *IEEE Journal of Photovoltaics* 10.5 (2020), pp. 1262–1269.

[134] W. Schottky. "Vereinfachte und erweiterte Theorie der Randschicht-Gleichrichter". In: *Zeitschrift für Physik* 118.9 (1942), pp. 539–592.

[135] H. A. Bethe. "Theory of the boundary layer of crystal rectifiers". In: *MIT Radiation Laboratory Report 43-12* (1942), pp. 1–26.

[136] C. Crowell. "The Richardson constant for thermionic emission in Schottky barrier diodes". In: *Solid-State Electronics* 8.4 (1965), pp. 395–399.

[137] O. Richardson. "LXVII. The distribution of the molecules of gas in a field of force, with applications to the theory of electrons". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 28.167 (1914), pp. 633–647.

[138] F. A. Padovani and R. Stratton. "Field and thermionic-field emission in Schottky barriers". In: *Solid-State Electronics* 9.7 (1966), pp. 695–707.

[139] A. Cowley and S. Sze. "Surface states and barrier height of metal-semiconductor systems". In: *Journal of Applied Physics* 36.10 (1965), pp. 3212–3220.

[140] K. K. Ng and H. Card. "A comparison of majority-and minority-carrier silicon MIS solar cells". In: *IEEE Transactions on Electron Devices* 27.4 (1980), pp. 716–724.

[141] H. Card and E. Rhoderick. "Studies of tunnel MOS diodes I. Interface effects in silicon Schottky diodes". In: *Journal of Physics D: Applied Physics* 4.10 (1971), p. 1589.

[142] A. Y. C. Yu. “Electron tunneling and contact resistance of metal-silicon contact barriers”. In: *Solid-state Electronics* 13.2 (1970), pp. 239–247.