Solar Energy

Solar Energy

Fundamentals, Technology, and Systems

Klaus Jäger
Olindo Isabella
Arno H.M. Smets
René A.C.M.M. van Swaaij
Miro Zeman

Typeset in DejaVu Sans Condensed and URW Palatino with LaTeX

Preface

Hereby, we present the first version of our book *Solar Energy: Fundamentals, Technology and Systems* and hope that it will be a useful source that helps our readers to study the different topics of solar energy.

It covers the topics that are treated in the three lectures on photovoltaics (PV) that are taught at the Delft University of Technology throughout the Academic Year: PV Basics, PV Technology, and PV Systems. In addition the book also covers other forms of solar energy, in particular Solar Thermal applications and Solar Fuels. Many of the topics that are discussed in this book are also covered in the *Massive Open Online Course* (MOOC) on Solar Energy (DelftX, ET.3034TU) that is given by Arno Smets on the edX platform and starts on 1 September 2014.

The students of this MOOC are the first ones that will use this book for studying solar energy. As this is the very first version, surely several small errors will be present throughout the text. We kindly ask the students to give us feedback, so that all the errors and typos can be eliminated quickly. Please let us also know if something is not clear or should be explained more detailed. Feedback can be given on the edX website.

Preparing such a book is not possible without a lot of support from others. We want to express our special Thanks to Arianna Tozzi, who provided us with the model on estimating the effect of wind speed and irradiance on the module temperature that we present in Section 18.3. Further, we are very grateful on the information on real-life PV systems provided by Stephan van Berkel. Ravi Vasudevan and Do Yun Kim are acknowledged for their discussions on PV systems. We also want to thank Mirjam Theelen for providing feedback on the sections on CIGS and CdTe technologies. Further, we want to thank all the students that provided feedback during the 2013/14 PV Systems lecture given at Delft University of Technology.

The AUTHORS

Delft, the Netherlands
1 September 2014

Contents

About this Book

This book aims to cover all the topics that are relevant for getting a broad overview on the different aspects of *Solar Energy*, with a focus on *photovoltaics*, which is the technology that allows to convert energy transported in light directly into electrical energy.

The Organisation of this book is roughly linked to the three lectures on photovoltaics (PV), that are given at the *Faculty for Electrical Engineering, Mathematics and Computer Science* at the *Delft University of Technology* throughout the course of the academic year: PV Basics, which roughly covers the topics covered in Part II on PV Fundamentals; PV Technology which covers the topics treated in Part III; and PV Systems which is treated in Part IV.

In total, this book contains *five parts*. In the introductory Part I we provide the reader with some general facts on energy in Chapter 1, summarise the current status of PV in the world in Chapter 2 and provide a first short explanation on how solar cells work in Chapter 3.

Part II aims to cover all the physical fundamentals that are required for understanding solar cells in general and the different technologies in particular. After discussing some basics from electrodynamics in Chapter 4 and solar radiation in Chapter 5, we spend several chapters on explaining the most important concepts of semiconductor physics. After discussing the basics in Chapter 6, we elaborate on the different generation and recombination mechanisms in Chapter 7 and introduce different types of semiconductor junctions in Chapter 8. After introducing the most important parameters for characterising solar cells in Chapter 9, we conclude Part II with a discussion on the efficiency limits of photovoltaic devices in Chapter 10, from which we distill some general design rules that are very important in Part III.

The different PV technologies are discussed in Part III. After summarising the history of solar cells in Chapter 11, we discuss crystalline silicon technology, which is by far the most important PV technology, in Chapter 12. After that we take a look at the different thin-film technologies in Chapter 13. Part III is concluded with a discussion on several third generation concepts that aim to combine high efficiencies with low cost in Chapter 14.

Part IV is dedicated in the planning of real PV systems. After a short introduction on PV systems in Chapter 15, we discuss the position of the sun and its implications in great detail in Chapter 16. The different components of a PV system, starting from the modules but also including all the balance-of-system components are introduced in Chapter 17. With all this knowledge we elaborate on designing PV systems, for both off-grid and grid-connected situations in Chapter 18. This part is concluded with a discussion of the ecological and economical aspects of PV systems in Chapter 19.

Alternative methods of solar energy are discussed in Part V. In Chapter 20 we introduce different concepts related to solar thermal energy. In Chapter 21, which is the last chapter of the regular text, we discuss solar fuels, which allow to store solar energy on the long term in the form of chemical energy.

The book is concluded with an Appendix, where some derivations that are too lengthy for the book are shown.

Part I

Introduction

1 Energy

As this book is on *Solar Energy*, it is good to start the discussion with some general thoughts on *Energy*. We will begin with a quote from *The Feynman Lectures on Physics*.

> There is a fact, or if you wish, a *law*, governing all natural phenomena that are known to date. There is no known exception to this law—it is exact so far as we know. The law is called the *conservation of energy*. It states that there is a certain quantity, which we call energy, that does not change in the manifold changes which nature undergoes. That is a most abstract idea, because it is a mathematical principle; it says that there is a numerical quantity which does not change when something happens. It is not a description of a mechanism, or anything concrete; it is just a strange fact that we can calculate some number and when we finish watching nature go through her tricks and calculate the number again, it is the same.
>
> ...
>
> Energy has a large number of *different forms*, and there is a formula for each one. These are: gravitational energy, kinetic energy, heat energy, elastic energy, electrical energy, chemical energy, radiant energy, nuclear energy, mass energy. If we total up the formulas for each of these contributions, it will not change except for energy going in and out.
>
> It is important to realise that in physics today, we have no knowledge of what energy is. We do not have a picture that energy comes in little blobs of a definite amount. It is not that way. However, there are formulas for calculating some numerical quantity,

and when we add it all together it gives ... always the same number. It is an abstract thing in that it does not tell us the mechanism or the reasons for the various formulas [1].

1.1 Some definitions

We will now state some basic physical connections between the three very important physical quantities of *energy, force, and power*. These connections are taken from classical mechanics but generally valid. We start with the *force* F, which is any influence on an object that changes its motion. According to Newton's *second law*, the force is related to the acceleration a of a body via

$$\mathbf{F} = m\mathbf{a}, \tag{1.1}$$

where m is the mass of the body. The bold characters denote that **F** and **a** are vectors. The unit of force is *Newton* (N), named after Isaac Newton (1642-1727). It is defined as the force required to accelerate the mass of 1 kg at an acceleration rate of 1 $\mathrm{m/s^2}$, hence 1 N = 1 $\mathrm{kg{\cdot}m/s^2}$.

In mechanics, energy E, the central quantity of this book, is given as the product of force times distance,

$$E = \int F(s)\,\mathrm{d}s, \tag{1.2}$$

where s denotes distance. Energy is usually measured in the unit of *Joule* (J), named after the English physicist James Prescott Joule (1818-1889), which it defined as the amount of energy required applying the force of 1 Newton through the distance of 1 m, 1 J = 1 Nm.

Another important physical quantity is the *power* P, which tells us the rate of doing work, or, which is equivalent, the amount of energy consumed per time unit. It is related to energy via

$$E = \int P(t)\,\mathrm{d}t, \tag{1.3}$$

where t denotes the time. The power is usually measured in *Watt* (W), after the Scottish engineer James Watt (1736-1819). 1 W is defined as one Joule per second, 1 W = 1 J/s and 1 J = 1 Ws.

As we will see later on, 1 J is a very small amount of energy compared to the human energy consumption. Therefore, in the energy markets, such as the electricity market, often the unit *Kilowatt hour* (kWh) is used. It is given as

$$1\,\mathrm{kWh} = 1000\,\mathrm{Wh} \times 3600\,\frac{\mathrm{s}}{\mathrm{h}} = 3\,600\,000\,\mathrm{Ws}. \tag{1.4}$$

On the other hand, the amounts of energy in solid state physics, the branch of physics that we will use to explain how solar cells work, are very small. Therefore,

we will use the unit of *electron volt*, which is the energy a body with a charge of one elementary charge ($e = 1.602 \times 10^{-19}$ C) gains or looses when it is moved across a electric potential difference of 1 Volt (V),

$$1\,\mathrm{eV} = e \times 1\,\mathrm{V} = 1.602 \times 10^{-19}\,\mathrm{J}. \tag{1.5}$$

1.2 Human and world energy consumption

After this somewhat abstract definitions we will look at the *human energy consumption*. The human body is at a constant temperature of about 37°C. It hence contains *thermal energy*. As the body is continuously cooled by the surroundings, thermal energy is lost to the outside. Further, blood is pumped through the blood vessels. As it travels through the vessels, its *kinetic energy* is reduced because of internal friction and friction at the walls of the blood vessels, *i.e.* the kinetic energy is converted into heat. To keep the blood moving, the heart consumes energy. Also, if we want our body to move this consumes energy. Further, the human brain consumes a lot of energy. All this energy has to be supplied to the body from the outside, in the form of food. A grown up average body requires about 10 000 Kilojoule every day.[1] We can easily show that this consumption corresponds to an average power of the human body of 115.7 W. We will come back to this value later.

In modern society, humans do not only require energy to keep their body running, but in fact we consume energy for many different purposes. We use energy for heating the water in our houses and for heating our houses. If water is heated, its thermal energy increases, and this energy must be supplied. Further, we use a lot of energy for transportation of people and products by cars, trains, trucks and planes. We use energy to produce our goods and also to produce food. At the moment, you are consuming energy if you read this book on a computer or tablet. But also if you read this book in a printed version, you implicitly consumed the energy that was required to print it and to transport it to you place.

As we mentioned already above, energy is never produced but always converted from one form to another. The form of energy may change in time, but the total amount does not change. If we want to utilise energy to work for us, we we usually convert it from one form to another more useable form. An example is the electric motor, in which we convert electrical energy to mechanical energy.

Modern society is very much based on the capability

[1] The energy content of food usually is given in the old-fashioned unit of kilocalories (kcal). The conversion factor is 1 kcal = 4.184 kJ. An average male human requires about 2500 kcal a day.

Table 1.1: Total primary energy consumption per capita and average power used per capita of some countries in 2011 [2].

Country	Energy consumption (kWh/capita)	Average power use (W/capita)
U.S.A.	81 642	9 319
Netherlands	53 963	6 160
Germany	44 310	5 058
China	23 608	2 695
India	6 987	797

of us humans to covert energy from one form to another form. The most prosperous and technologically developed nations are also the ones which have access to and are consuming the most energy per inhabitant. Table 1.1 shows the primary energy consumption per capita and the average power consumed per capita for several countries. We see that the average U.S. citizen uses an average power of 9 319 W, which is about 80 times what his body needs. In contract, an average citizen from India only uses about 800 W, which is less then a tenths of the U.S. consumption.

Many people believe that tackling the *energy problem* is amongst the biggest challenges for human kind in the 21st century. It is a challenge because of several problems: The first challenge the human kind is facing is a supply-demand problem. The demand is continuously

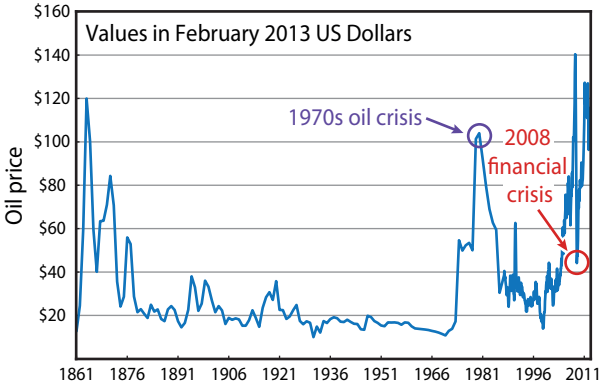

Figure 1.1: The history of the oil price normalised to the February 2013 value of the U.S. Dollar [3].

growing. The world population is still rapidly growing, and some studies predict a world population of 9 billion around 2040 in contrast to the 7 billion people living on the planet today. All these people will need energy, which increases the global energy demand. Further, in many countries the living standard is rapidly increasing like China and India, where approximately 2.5 billion people are living, which represents more than a third of the World's population. Also the increasing living standards lead to an increased energy demand.

According to the IEA World Energy Outlook 2013, the

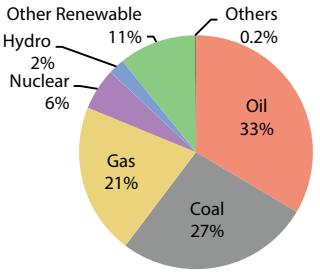

Figure 1.2: The primary energy consumption of the world by source in 2008. The total supply was 143 851 TWh [5].

global energy demand will grow about one third from 2011 to 2013 [4]. The increasing demand in energy has economic impact, as well. If there is more demand for a product, while supply does not change much, the product will get more expensive. This basic market mechanism is also true for Energy. As an example we show a plot of the annual averaged price for an oil barrel, normalised to the value of the 2008 US Dollar in Fig. 1.1. We see that prices went up during the oil crisis in the 1970s, when some countries stopped producing and trading oil for a while. The second era of higher oil prices started at the beginning of this millennium. Due to the increasing demand from new growing economies, the oil prices have been significantly increased.

A second challenge that we are facing is related to the fact that our energy infrastructure heavily depends on fossil fuels like oil, coal and gas. Fossil fuels are nothing but millions and millions of years of solar energy stored in the form of chemical energy. The problem is that humans deplete these fossil fuels much faster than they are generated through the photosynthetic process in nature. Therefore fossil fuels are not a sustainable energy source. The more fossil fuels we consume, the less easily available gas and oil resources will be available. Already now we see that more and more oil and gas is produced with *unconventional* methods, such as extracting oil from tar sands in Alberta, Canada and producing gas with fracturing such as in large parts of the United States. This new methods use much more energy to get the fossil fuels out of the ground. Further, off-shore drilling is put regions with ever larger water depths, which leads to new technological risks as we have seen in the Deepwater Horizon oil spill in the Gulf of Mexico in 2010.

A third challenge is that by burning fossil fuels we produce the so-called greenhouse gases like carbon dioxide (CO_2). The additional carbon dioxide created by human activities is stored in our oceans and atmosphere. Figure 1.3 shows the increase in carbon dioxide concentration in the Earth's atmosphere up to 2000. According to the International Panel on Climate Change (IPCC) Fifth Assessment Report (AR5),

> The atmospheric concentrations of carbon dioxide, methane, and nitrous oxide have increased to levels unprecedented in at least the last 800,000 years. Car-

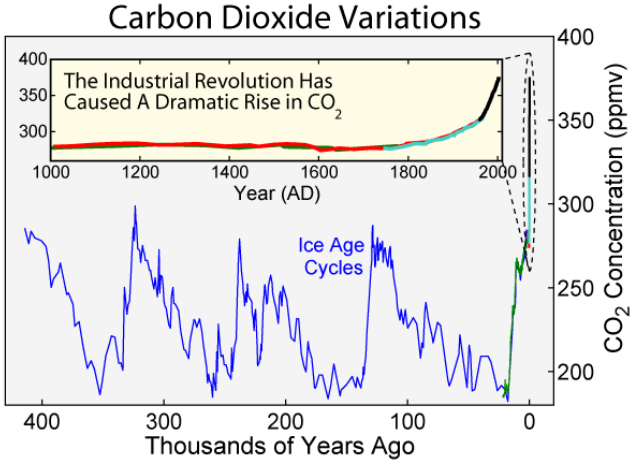

Figure 1.3: The atmospheric CO_2 content in the last 400 000 years [6].

> bon dioxide concentrations have increased by 40% since pre-industrial times, primarily from fossil fuel emissions and secondarily from net land use change emissions. The ocean has absorbed about 30% of the emitted anthropogenic carbon dioxide, causing ocean acidification [7].

Further, in the AR5 it is stated that

> Human influence on the climate system is clear. This is evident from the increasing greenhouse gas concentrations in the atmosphere, positive radiative forcing, observed warming, and understanding of the climate system [7].

and

> Human influence has been detected in warming of the atmosphere and the ocean, in changes in the global water cycle, in reductions in snow and ice, in global mean sea level rise, and in changes in some climate extremes. This evidence for human influence has grown since AR4. It is *extremely likely* that human influence has been the dominant cause of the observed warming since the mid-20th century [7].

Hence, it seems very clear that the increase in carbon dioxide is responsible for the global warming and climate change, which can have drastic consequences of the habitats of many people.

Since the beginning of the industrial revolution, mankind is heavily dependent on fossil fuels. Within a few centuries, we are using solar energy that was incident

on Earth for hundreds of millions of years, converted in to chemical energy by the photosynthetic process and stored in the form of gas, coal and oil.

Before the industrial revolution, the main source of energy was wood and biomass, which is a secondary form of solar energy. The energy source was replenished in the same characteristic time as the energy being consumed. In the pre-industrial era, mankind was basically living on a secondary form of solar energy. However, also back then the way we consumed energy was not fully sustainable. For example, deforestation due to increasing population density was already playing a role at the end of the first millennium.

1.3 Methods of Energy Conversion

Figure 1.4 shows different energy sources and the ways we utilise them. We see that usually the chemical energy stored in fossil fuels is converted to usable forms of energy via heat by burning, with an efficiency of about 90%. Using heat engines, thermal energy can be converted in to mechanical energy. Heat engines have a conversion efficiency of up to 60%. Their efficiency is ultimately limited by the Carnot efficiency limit that we will discuss in Chapter 10. The far majority of the current cars and trucks works on this principle. Mechanical energy can be converted in to electricity using electric generators with an efficiency of up to 99%. Most of the World's electricity is generated with an *turbo generator* that is connected to a steam turbine, where the coal is the major energy source. This process is explained in more detail in our discussion on solar thermal electric power in Chapter 20. Along all the process steps of making electricity out of fossil fuels, at least 50% of the initial available chemical energy is lost in the various conversion steps.

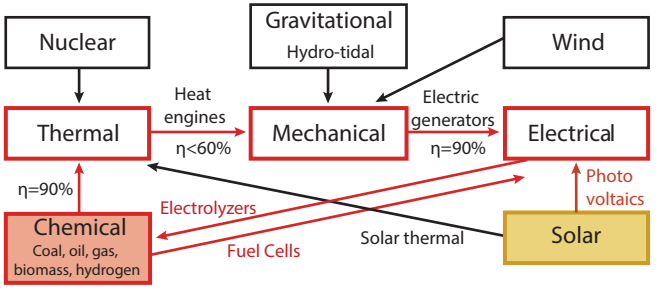

Figure 1.4: The different energy carriers and how we utilise them [8].

Chemical energy can be directly converted into electricity using a fuel cell. The most common fuel used in fuel cell technology is hydrogen. Typical conversion efficiencies of fuel cells are 60%. A regenerative fuel cell can operate in both directions and also convert electrical energy into chemical energy. Such an operation is

called *electrolysis*; typical conversion efficiencies of hydrogen electrolysers of 50-80% have been reported. We will discuss electrolysis in more detail in Chapter 21.

In *nuclear power plants*, energy is released as heat during *nuclear fission* reactions. With the heat steam is generated that drives a steam turbine and subsequently an electric generator just as in most fossil fuel power plants.

1.3.1 Renewable energy carriers

All the energy carriers discussed above are either fossil or nuclear fuels. They are not renewable because they are not "refilled" by nature, at least not in a useful amount of time. In contrast, *renewable energy carriers* are energy carriers that are replenished by natural processes at a rate comparable or faster than its rate of consumption by humans. Consequently, hydro-, wind- and solar energy are renewable energy sources.

Hydroelectricity is an example of an energy conversion technology that is not based on heat generated by fossil or nuclear fuels. The potential energy of rain falling in mountainous areas or elevated plateaus is converted into electrical energy via a *water turbine*. With *tidal pools* the potential energy stored in the tides can also be converted to mechanical energy and subsequently electricity. The kinetic energy of *wind* can be converted into mechanical energy using wind mills.

Finally, the energy contained in sunlight, called *solar energy*, can be converted into electricity as well. If this energy is converted into electricity directly using devices based on semiconductor materials, we call it *photovoltaics* (PV). The term *photovoltaic* consists of the greek word $\phi\omega\varsigma$ (phos), which means light, and -volt, which refers to electricity and is a reverence to the Italian physicist Alessandro Volta (1745-1827) who invented the battery. As we will discuss in great detail in this book, typical efficiencies of the most commercial *solar modules* are in the range of 15-20%.

Solar light can also be converted into heat. This application is called *solar thermal energy* and is discussed in detail in Chapter 20. Examples are the heating of water flowing through a black absorber material that is heated in the sunlight. This heat can be used for water heating, heating of buildings or even cooling. If concentrated solar power systems are temperatures of several hundreds of degrees are achieved, which is sufficient to generate steam and hence drive a steam turbine and a generator to produce electricity.

Next to generating heat and electricity, solar energy can be converted in to chemical energy as well. This is what we refer to as *solar fuels*. For producing solar fuels, photovoltaics and regenerative fuel cells can be combined. In addition, sunlight can also be directly converted into fuels using photoelectrochemical devices. We

will discuss solar fuels in Chapter 21.

We thus see that solar energy can be converted into electricity, heat and chemical energy. The sun has is energy source for almost all the processes that happen on the surface of our planet. Wind is a result of temperature difference in the atmosphere induced by solar irradiation, waves are generated by the wind, clouds and rain are initially formed by the evaporation of water due to sun light. As the sun is the only real energy source we have, we need to move to an era in which we start to utilise the energy provided by the sun directly to satisfy our energy needs. The aim of this book is to teach the reader how solar energy can be utilised directly.

1.3.2 Electricity

As we see in Fig. 1.5 (a), 17% of all the World's secondary energy is used as electricity, which is a form of energy that can be easily and cheaply transported with relative small losses through an electric grid. It is important to realise that without electricity modern society as we know it would not be possible. Electricity has been practically used for more than 100 years now. It provides us the energy to cook food, to wash, to do the laundry, illuminate the house and streets, and countless other applications. The access to electricity to electricity strongly determines the our living standard. Despite

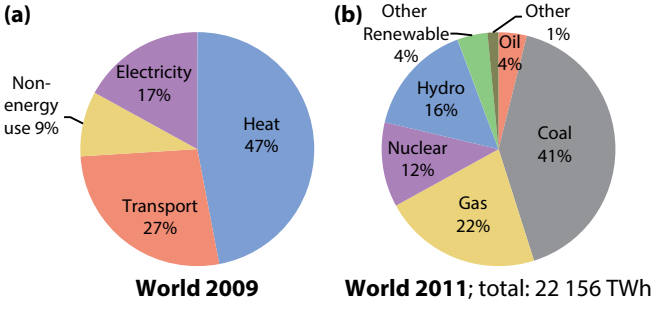

Figure 1.5: (a) The final energy consumption by energy service [9] and (b) the energy carriers used for electricity generation [10].

this importance of electricity, in 2009 still about 1.3 billion people had no access to electricity.

As we see in 1.5 (b), about 65% of the electricity is generated using fossil fuels, where coal is the dominant contributor. As coal emits about twice as much CO_2 per generated kWh as natural gas, coal power plants are a major contributor to global warming. Nuclear is responsible for 16% of the World's electricity generation. With 19%, hydroelectricity is by far the largest contributor among the renewable energy sources.

Of all the generated electricity, about 40% of the electric energy is used for residential purposes and 47% is used by industry. 13% is lost in transmission. As you can see, in 2007, transport did not play a significant role in

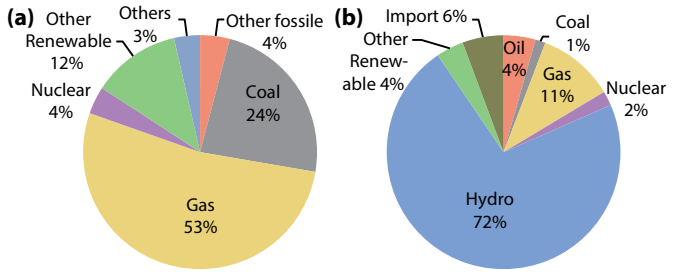

Figure 1.6: The energy mix used for electricity production in (a) the Netherlands [11] and (b) Brazil [12].

the electricity consumption. However, this is expected to change as electric cars are becoming more and more important.

Figure 1.6 shows which energy carriers are mainly used for electricity generation in the Netherlands and Brazil. We see that in the Netherlands, electricity generation heavily depends on the local gas resources, whereas in Brazil hydroelectricity is the most important resource.

2

Status and prospects of PV technology

In this Chapter we will give a brief overview on the current status of the PV technology and discuss its prospects.

In Fig. 2.1 the global production of PV modules in recent years is shown. The vertical axis represents the annual production expressed in the total produced power capacity in MW_p. The letter p denotes *peak power*, this means the maximum power a PV module can deliver if it is illuminated with the standardised AM1.5 solar spectrum, which we introduce in Section 5.5. On the horizontal axis the time is shown. We see that the solar cell production is increasing more than 40% every year, which is exponential growth.

Figure 2.2 shows the worldwide cumulative installed

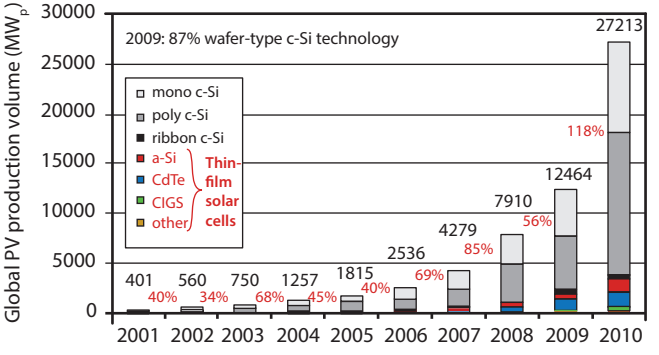

Figure 2.1: The global PV production volume in recent years REFERENCE.

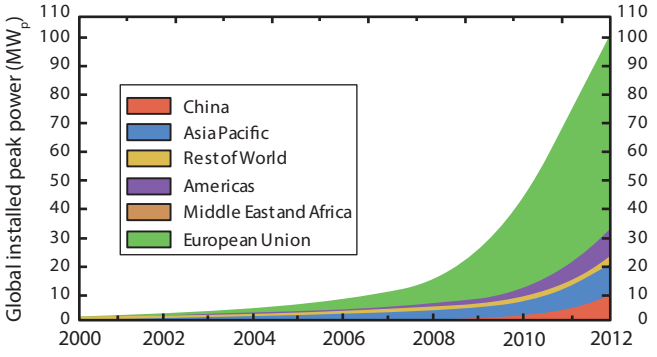

Figure 2.2: The global installed PV capacity (Data from [13]).

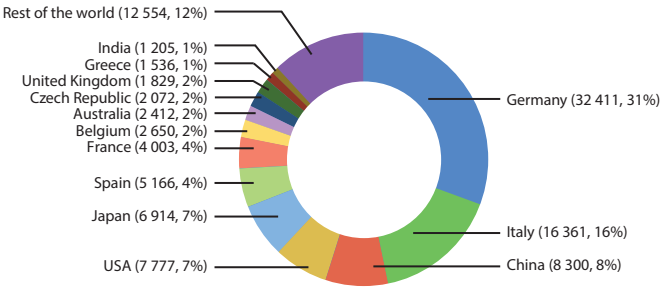

Figure 2.3: Fraction of PV installations for different countries by the end of 2012 (Data from [13]).

PV power, which is exponentially increasing in time as well. By far the largest share is installed in Europe. It is followed by the Asia Pacific Region, where most of the PV power is installed in Japan. For China we observe a very strong increase in installed PV power since 2010. By the end of 2012 the 100 GW$_p$ threshold was passed for the first time [13]. By the end of 2013, already almost 140 GW$_p$ was installed around the globe [14]. Of all the installed PV power at the end of 2013, almost one third was installed in 2013 alone!

In Fig. the installed PV power in several countries at the end of 2012 is shown. About 31% of the total PV capacity is installed in Germany. This is a result of the German government's progressive feed-in-tariff policy that was introduced in 2000 [15].[1] Considering that Germany lies within an area with a relatively low radiation level that is comparable to that of Alaska [16], the large contribution of solar electricity to Germany's electricity production indicates the promising potential of solar energy for the sunnier parts of the world.

A very strong increase also is observed in Italy, which accounts for 16% of the world wide PV capacity. China with a contribution of 8% is the fastest growing market at the moment, in 2010, China only contributed with 2% to the global PV capacity. Within the top six, we also find the United States, Japan and Spain. Their PV capacity contributes between 4% (Spain) and 7% (U.S.A.). Also Japan shows a strong growth in PV installations.

[1] We will discuss the *feed-in tariff* scheme in Chapter 19.

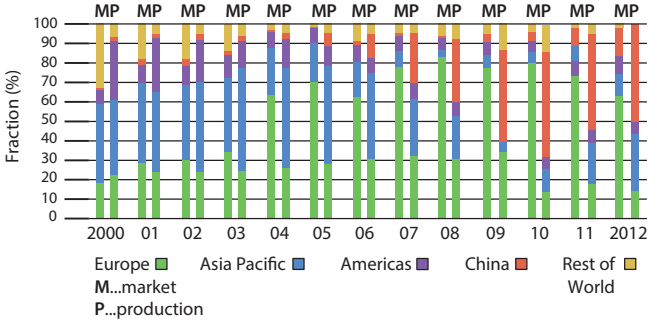

Figure 2.4: Development of the market and production shares of different PV markets since 2000 (Data from [13]).

After the *Fukushima Daiichi nuclear disaster* on 11 March 2011 the Japanese government introduced some progressive feed-in tariffs to promote and accelerate the introduction of renewable energy conversion technologies.

It is interesting to see that PV technology is not only a European affair. The local demand and supply has been changing rapidly in the last 13 years, which is illustrated in Fig. 2.4. This figure illustrates the evolution of the world-wide supply and demand of PV modules in the various regions around the world. We see that in 2000 the biggest market was Japan with a total share of 40%. In 2000, Germany introduced the *Erneuerbare Energie Gesetz* (Renewable Energy act) which induced a strong growth of the German and hence the European PV market. By 2008, Europe had a market share of more than 80%. Back then, PV was mainly a European industry. Starting from 2009, the domestic PV markets in China, the Americas (mainly U.S.) and Asia Pacific (mainly Japan) are increasing very rapidly and catching up quickly with Europe.

Figure 2.4 also shows the supply side. Up to 2005 we see that the Asia Pacific and the Europe production shares were slowly increasing, as their growth was faster than that of the other regions. Since then the picture changed drastically! Since then the Chinese production share was increasing very strongly. This can be explained by the fact that the Chinese government made huge investments in order to scale up PV module manufacturing in China. In 2012, around 60% of all PV modules were produced in China.

In 2000, the PV markets was an essentially local, meaning the European companies produced for the European market *etc*. The local demands and supplies in Asia, the Americas and Europe were in balance. In the last years, the market has become a global market. As a result, in 2012 no local balance between supply and demand existed anymore. While the majority of the demand is in Europe, the majority of the production is in China.

The demand also is strongly stimulated by the decreasing cost price of PV technology. Figure 2.5 shows the *learning curve* of PV technology. The learning curve

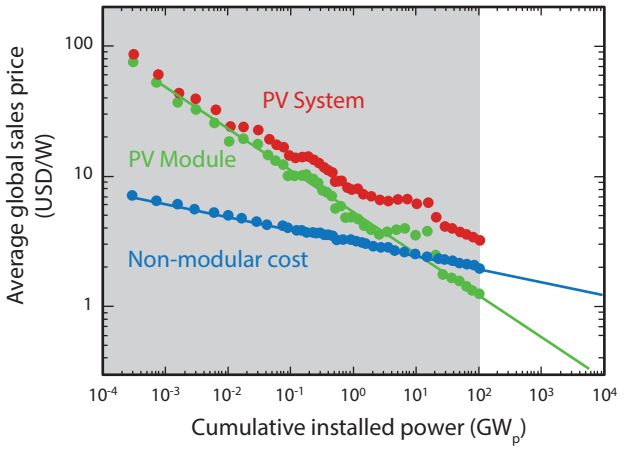

Figure 2.5: The learning curve for PV modules and PV systems (Data from [17]).

shows in a graphical way, how the cost price develops with increasing experience, where the experience is expressed by the cumulatively installed PV capacity. With more PV produced — and hence also with time — the PV industry gets more experienced. On the one hand, the industry learns to increase the *energy conversion efficiency* without increasing the cost via better and better understanding the production process and hence increasing the *production yield*. On the other hand, industry also learns to produce more efficiently, which means that the man power required per production unit can be reduced. Also, the materials and energy required for producing the PV modules becomes less and less per production unit. In addition, also up-scaling reduces the cost. Learning curves usually show an exponentially decreasing cost price, until the technology or product is fully developed.

In Fig. 2.5, the averaged global sales prices of a PV module versus the cumulative installed power up to 20 GW is shown. Note, that the points up to 20 GW (up to 2009) in the grey area are real data points, while the points in white area are extrapolation of the general trend. It is important to note that the sales prices, except for some fluctuations, follow a largely exponential decay. Currently, the average retail price of PV modules is below 1 US Dollar per Watt-peak. However, cost price of a PV system is not only determined by the module. The red dots show the decrease in the cost price of complete PV systems. While in the early

days of the PV technology, the system price was dominated by the module price, currently, the cost of the *non-modular components* of PV systems are getting more and more dominant. With non-modular components, we refer to components such as the the racking, wiring, inverter, batteries for stand-alone systems, and also the maintenance costs. All these components are discussed in detail in Chapter 17. The difference between the red and green line corresponds to the non-modular costs, which is dropping much slower than that of the PV modules.

As a consequence, PV technologies with higher energy conversion efficiencies have an advantage, as with higher efficiency less area is required to install the same PV power. As the area is directly linked to the non-modular costs, technologies with higher efficiencies require less modular costs which has an positive effect on the cost price of the complete PV system. Consequently, the c-Si PV technology, with module efficiencies ranging from 14% up to 20% has an advantage with respect to thin-film technologies, that have lower efficiencies.

In Chapter **??** we have seen that hydropower is responsible for 17% of the total worldwide electricity production while 12% of the electricity is generated in nuclear power plants. Howe do these numbers compare to solar electricity? This question is answered in Fig. 2.6 (a), where the installed capacity (in GW) of several electricity generation technologies is shown on a logarithmic scale. The figure only considers electricity generation technologies that are not dependent on fossil fuels. We see that the installed nuclear power capacity is hardly growing anymore, while the installed hydropower is still slightly growing in time. Wind is growing at a much faster rate of 20% per year. Solar has by far the largest growth rate with an annual increase of installed capacity exceeding 40% since 2008.

However, it is not fair to compare the installed power between technologies like this, because the numbers shown in the graph represent the maximum (peak) power the different technologies can generate instead of the average power they have delivered in reality. The relationship between the totally installed power and the power generated on average is called the *capacity factor* C_F. Of the technologies shown in Fig. 2.6 (a) nuclear has by far the highest capacity factor with $C_F(\text{nuclear}) = 90\%$ followed by hydropower with $C_F(\text{hydro}) = 40\%$. For wind electricity we assume $C_F(\text{wind}) = 30\%$ and for solar electricity $C_F(\text{solar}) = 15\%$. The low capacity factor for PV systems can be explained by the fact that for most geographical locations, almost half of the solar day is devoid of solar radiation at night time.

Figure 2.6 (b) shows the effective installed power corrected with the capacity factors. Currently solar energy generates about an order of magnitude less electricity than wind energy and more than two orders of mag-

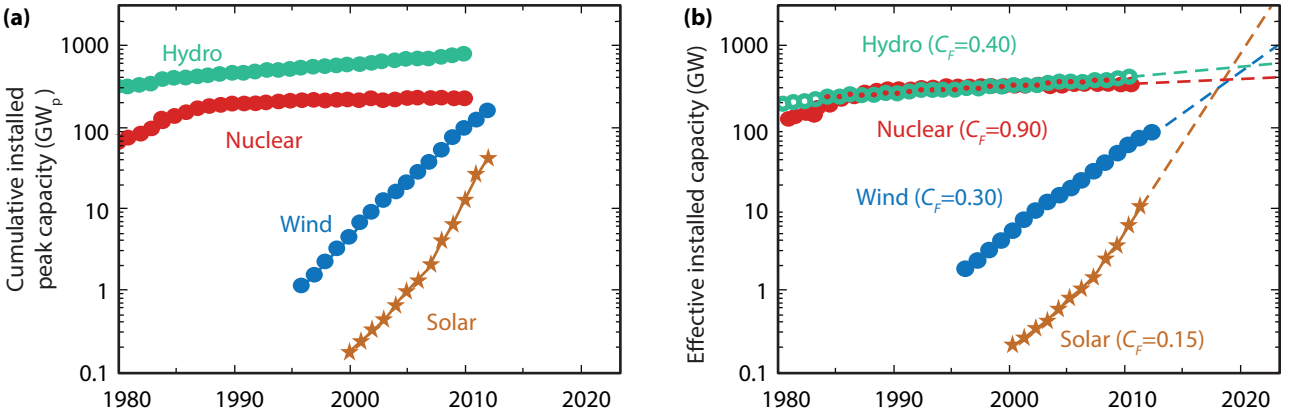

Figure 2.6: (a) Development of the installed capacity (in GW) of several electricity generation technologies since 1980. (b) The same graph corrected by the capacity factor C_F and extrapolated until 2020.

nitude less than hydro and nuclear electricity. Seen the development in recent years we however can claim that the trend in the growth of solar energy will continue the coming years. If we therefore extrapolate the trends of the last decade until 2020 we see that the installed power of solar energy will exceed nuclear, wind and hydropower by then. It is just a matter of time until solar electricity will be the most important electricity generation technology that is not based on combustion of fossil fuels.

Of course, we have to justify why solar electricity can grow much faster than the other technologies shown in Fig. 2.6. First, solar radiation is available everywhere on Earth and it is available in great abundance. The amount of solar energy incident on Earth is about 10 000 time larger than the *total* energy[2] consumption of mankind. As hydroelectricity is powered by water that is evaporated by the sun and falls on the ground as rain, it is a secondary form of solar energy. Also wind arises from temperature and pressure differences of the atmosphere and hence also is a secondary form of solar energy. As a consequence, solar energy is by far the largest available form of renewable energy.

Secondly, hydro- and nuclear electricity are *centralised* electricity generation technologies. For hydro power plants, big dams are needed. Also nuclear power plants have large power rates at about 1 GW. Building new hydro and nuclear power plants requires large public or private investments. While solar electricity can be generated in large PV parks or solarthermal power plants (see Chapter 20) as well, it has an unique advantage: PV systems can be installed decentralised on every roof. Electricity consumers can generate a least a part of their required electricity in their own homes, which makes them partially independent of the electricity market. In addition, the cost price of PV systems has dropped below grid parity in large parts of the world [18]. This means, that averaged during the lifetime of the PV system PV generated electricity is cheaper than electricity from the grid.

We believe that the installation of decentralised PV systems will be the big force behind the solar revolution in the coming years. It will change the energy landscape much faster than most people think, which is justified in Fig. 2.6 (b). As more and more people become aware of these facts, it is more likely that the growth will be further enhanced than it will be slowed down.

[2] We really mean the *total* human energy consumption and not only electricity!

3
The Working Principle of a Solar Cell

In this chapter we present a very simple model of a solar cell. Many notions presented in this chapter will be new but nonetheless the great lines of how a solar cell works should be clear. All the aspects presented in this chapter will be discussed in larger detail in the following chapters.

The working principle of solar cells is based on the *photovoltaic effect*, *i.e.* the generation of a potential difference at the junction of two different materials in response to electromagnetic radiation. The photovoltaic effect is closely related to the photoelectric effect, where electrons are emitted from a material that has absorbed light with a frequency above a material-dependent threshold frequency. In 1905, Albert Einstein understood that this effect can be explained by assuming that the light consists of well defined energy quanta, called *photons*. The energy of such a photon is given by

$$E = h\nu, \tag{3.1}$$

where h is Planck's constant and ν is the frequency of the light. For his explanation of the photoelectric effect Einstein received the Nobel Prize in Physics in 1921 [19].

The photovoltaic effect can be divided into three basic processes:

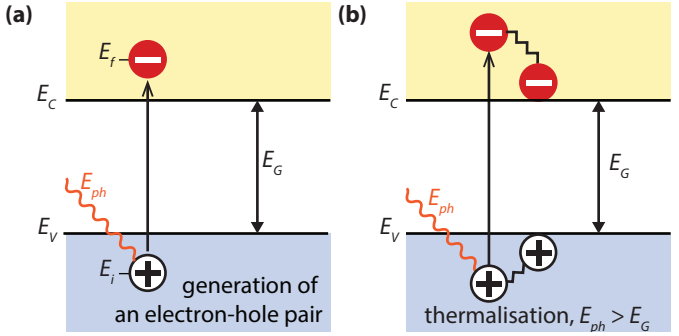

Figure 3.1: (a) Illustrating the absorption of a photon in a semiconductor with bandgap E_g. The photon with energy $E_{\text{ph}} = h\nu$ excites an electron from E_i to E_f. At E_i a hole is created. (b) If $E_{\text{ph}} > E_g$, a part of the energy is thermalised.

1. Generation of charge carriers due to the absorption of photons in the materials that form a junction.

Absorption of a photon in a material means that its energy is used to excite an electron from an initial energy level E_i to a higher energy level E_f, as shown in Fig. 3.1 (a). Photons can only be absorbed if electron energy levels E_i and E_f are present so that their difference equals to the photon energy, $h\nu = E_f - E_i$. In an ideal semiconductor electrons can populate energy levels below the so-called *valence band* edge, E_V, and above the so called *conduction band* edge, E_C. Between those two bands no allowed energy states exist, which could be populated by electrons. Hence, this energy difference is called the *bandgap*, $E_g = E_C - E_V$. If a photon with an energy smaller than E_g reaches an ideal semiconductor, it will not be absorbed but will traverse the material without interaction.

In a real semiconductor, the valence and conduction bands are not flat, but vary depending on the so-called k-vector that describes the crystal momentum of the semiconductor. If the maximum of the valence band and the minimum of the conduction band occur at the same k-vector, an electron can be excited from the valence to the conduction band without a change in the crystal momentum. Such a semiconductor is called a *direct bandgap* material. If the electron cannot be excited without changing the crystal momentum, we speak of

an *indirect bandgap* material. The absorption coefficient in an direct bandgap material is much higher than in an indirect bandgap material, thus the absorber can be much thinner [20].

If an electron is excited from E_i to E_f, a void is created at E_i. This void behaves like a particle with a positive elementary charge and is called a *hole*. The absorption of a photon therefore leads to the creation of an electron-hole pair, as illustrated in Fig. 3.2 ❶. The *radiative energy* of the photon is *converted* to the *chemical energy* of the electron-hole pair. The maximal conversion efficiency from radiative energy to chemical energy is limited by thermodynamics. This *thermodynamic limit* lies in between 67% for non-concentrated sunlight and 86% for fully concentrated sunlight [21].

The basic physics required for describing semiconductors is presented in chapter 6.

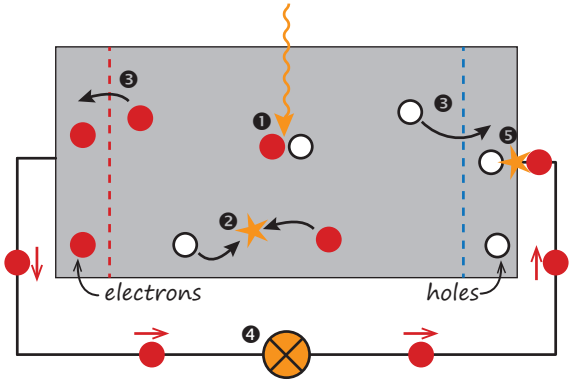

Figure 3.2: A very simple solar cell model. ❶ Absorption of a photon leads to the generation of an electron-hole pair. ❷ Usually, the electrons and holes will combine. ❸ With semi-permeable membranes the electrons and the holes can be separated. ❹ The separated electrons can be used to drive an electric circuit. ❺ After the electrons passed through the circuit, they will recombine with holes.

2. Subsequent separation of the photo-generated charge carriers in the junction.

Usually, the electron-hole pair will recombine, *i.e.* the electron will fall back to the initial energy level E_i, as illustrated in Fig. 3.2 ❷. The energy will then be released either as photon (*radiative recombination*) or transferred to other electrons or holes or lattice vibrations (*non-radiative recombination*). If one wants to use the energy stored in the electron-hole pair for performing work in an external circuit, *semipermeable membranes* must be present on both sides of the absorber, such that electrons only can flow out through one membrane and holes only can flow out through the other membrane [21], as illustrated in Fig. 3.2 ❸. In most solar cells, these membranes are formed by *n*- and *p*-type materials.

A solar cell has to be designed such that the electrons

and holes can reach the membranes before they recombine, *i.e.* the time it requires the charge carriers to reach the membranes must be shorter than their lifetime. This requirement limits the thickness of the absorber.

We will discuss generation and recombination of electrons and holes in detail in chapter 7.

3. Collection of the photo-generated charge carriers at the terminals of the junction.

Finally, the charge carriers are extracted from the solar cells with electrical contacts so that they can perform work in an external circuit (Fig. 3.2 ❹). The *chemical energy* of the electron-hole pairs is finally converted to *electric energy*. After the electrons passed through the circuit, they will recombine with holes at a metal-absorber interface, as illustrated in Fig. 3.2 ❺.

Loss mechanisms

The two most important *loss mechanisms* in single bandgap solar cells are the inability to convert photons with energies below the bandgap to electricity and thermalisation of photon energies exceeding the bandgap, as illustrated in Fig. 3.1 (b). These two mechanisms alone amount to the loss of about half the incident solar energy in the conversion process [22]. Thus, the maximal energy conversion efficiency of a single-junction solar cell is considerably below the thermodynamic limit. This *single bandgap limit* was first calculated by Shockley and Queisser in 1961 [23].

A detailed overview of loss mechanisms and the resulting efficiency limits is discussed in Chapter 10.

Part II

PV Fundamentals

4 Electrodynamic basics

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

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

4.1 The electromagnetic theory

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

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

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

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

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

The electromagnetic theory can perfectly describe how light propagates. However, it fails in explaining how light is emitted and absorbed by matter. For this purpose, quantum mechanics is required.

4.2 Electromagnetic waves

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

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

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

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

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

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

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

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

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

Thus,

$$k_z = \frac{n\omega}{c}. \tag{4.4}$$

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

$$\omega = 2\pi\nu = \frac{2\pi}{T}, \tag{4.5}$$

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

$$\lambda = \frac{2\pi}{k_z} = \frac{2\pi c}{n\omega} = \frac{c}{n\nu}. \tag{4.6}$$

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

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

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

$$\mathbf{k}\cdot\mathbf{E}_0 = \mathbf{k}\cdot\mathbf{H}_0 = \mathbf{H}_0\cdot\mathbf{E}_0 = 0. \tag{4.7}$$

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

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

where

$$Z_0 = c\mu_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 376.7\,\Omega \tag{4.9}$$

is the *impedance of free space*.

4.3 Optics of flat interfaces

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

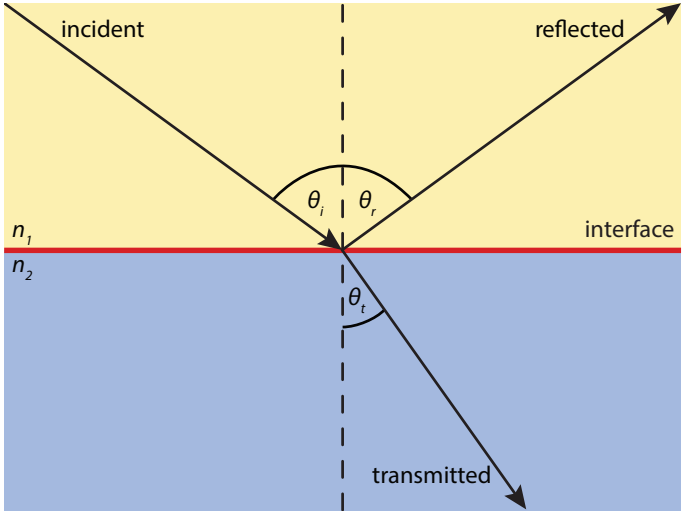

Figure 4.1: Scheme of light, reflected and refracted by a boundary.

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

$$\theta_r = \theta_i. \tag{4.10}$$

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

$$n_1 \sin\theta_i = n_2 \sin\theta_t, \tag{4.11}$$

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

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

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

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

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

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

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

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

For normal incidence this leads to

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

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

$$R + T = 1. \tag{4.16}$$

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

$$T = 1 - R = \frac{n_2 \cos\theta_t}{n_1 \cos\theta_i}\frac{1}{2}\left(t_s^2 + t_s^2\right), \tag{4.17}$$

which leads to

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

for normal incidence.

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

$$\sin\theta_{\text{crit}} = \frac{n_1}{n_2}. \tag{4.19}$$

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

4.4 Optics in absorptive media

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

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

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

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

$$\tilde{\epsilon} = \epsilon' + i\epsilon''. \tag{4.20}$$

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

$$\tilde{n} = \sqrt{\tilde{\epsilon}} = n + i\kappa, \tag{4.21}$$

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

$$\tilde{k}_z = \frac{\tilde{n}\omega}{c} = \frac{n\omega}{c} + i\frac{\kappa\omega}{c} = k_z' + ik_z''. \tag{4.22}$$

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

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

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

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

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

$$I(z) = I_0 \exp(-2k_z'' z) = I_0 \exp(-\alpha z), \tag{4.25}$$

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

$$\alpha = 2k_z'' = 2\frac{\kappa\omega}{c} = \frac{4\pi\kappa}{\lambda_0}, \tag{4.26}$$

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

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

$$\delta_p = \frac{1}{\alpha}. \tag{4.27}$$

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

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

4.5 Continuity and Poisson equations

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

4.5.1 Poisson equation

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

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

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

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

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

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

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

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

4.5.2 Continuity equation

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

$$\frac{\mathrm{d}Q_V}{\mathrm{d}t} + \oiint_S \mathbf{J}\,\mathrm{d}S = 0, \tag{4.32}$$

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

$$Q_V = \iiint_V \rho\,\mathrm{d}V. \tag{4.33}$$

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

mulation that is given by

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

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

5 Solar Radiation

5.1 The Sun

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

- a helium core (two protons and two neutrons),
- 2 positrons (the anti-particles of electrons),
- 2 neutrinos,
- electromagnetic radiation.

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

$$E = mc^2. \tag{5.1}$$

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

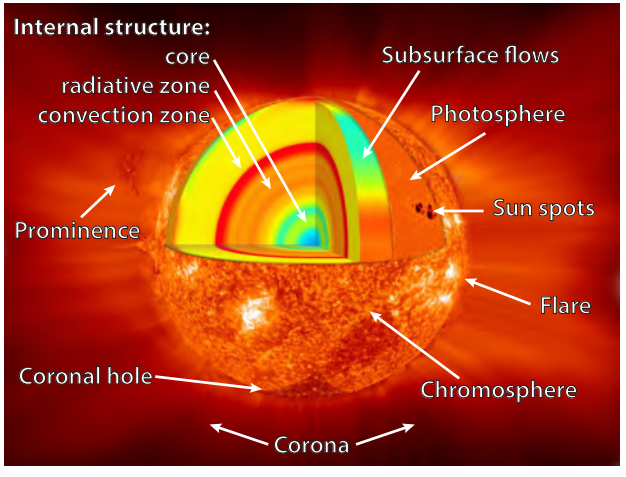

Figure 5.1: The Sun with its layer structure depicted [24].

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

The remainder of the radiation is released as electromagnetic radiation. The core of the Sun is so dense that radiation cannot travel freely but is continuously absorbed and re-emitted, such that it takes the radiation between 10 000 and 170 000 years to travel to the solar surface. The surface of the Sun is called the photosphere. It has a temperature of about 6000 K. It behaves quite perfectly as a blackbody (see section 5.3) and is the source of the solar radiation that hits the Earth. The total power density of the solar radiation at the mean Earth-sun distance on a plane perpendicular to the direction of the Sun, outside the Earth's atmosphere, is referred to as the *solar constant*. Its value is approximately 1361 W/m^2.

5.2 Radiometric properties

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

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

$$P = \frac{\mathrm{d}E}{\mathrm{d}t}. \tag{5.2}$$

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

Table 5.1: Some facts on the Sun

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

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

$$P = \int_A \int_{2\pi} L_e \cos\theta \, \mathrm{d}\Omega \, \mathrm{d}A. \tag{5.3}$$

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

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

The factor $\cos\theta$ expresses the fact that not the surface element $\mathrm{d}A$ itself is the relevant property but the projection of $\mathrm{d}A$ to the normal of the direction (θ, ϕ). This is also known as the *Lambert cosine law*.

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

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

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

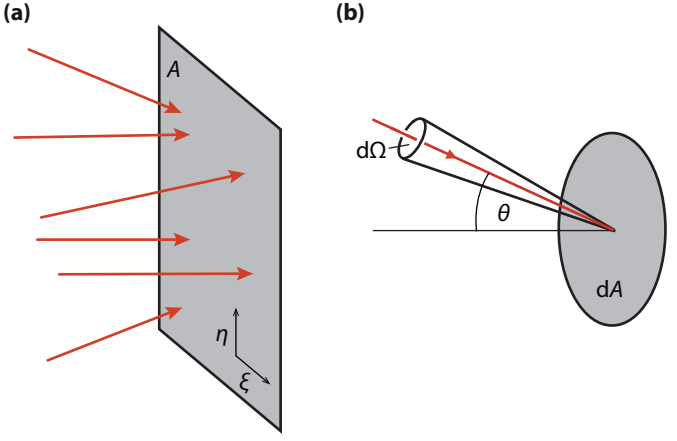

Figure 5.2: (a) Illustrating a surface A irradiated by light from various directions and (b) a surface element $\mathrm{d}A$ that receives radiation from a solid angle element $\mathrm{d}\Omega$ under an angle θ with respect to the surface normal.

properties. These are given by

$$P_\nu = \frac{\mathrm{d}P}{\mathrm{d}\nu}, \qquad P_\lambda = \frac{\mathrm{d}P}{\mathrm{d}\lambda}, \tag{5.5}$$

$$L_{e\nu} = \frac{\mathrm{d}L_e}{\mathrm{d}\nu}, \qquad L_{e\lambda} = \frac{\mathrm{d}L_e}{\mathrm{d}\lambda}, \tag{5.6}$$

etc. Their physical dimensions are

$$[P_\nu] = \mathrm{W{\cdot}Hz^{-1}} = \mathrm{W{\cdot}s}, \qquad [P_\lambda] = \mathrm{W{\cdot}m^{-1}},$$
$$[L_{e\nu}] = \mathrm{W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}s}, \qquad [L_{e\lambda}] = \mathrm{W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}m^{-1}},$$

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

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

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

The spectral power in wavelength thus can be obtained via

$$P_\lambda = \int_A \int_{2\pi} L_{e\lambda} \cos\theta \, \mathrm{d}\Omega \, \mathrm{d}A, \tag{5.8}$$

and analogously for P_ν. The radiance is given by

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

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

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

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

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

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

The irradiance also is given as

$$I_e = \frac{\partial^2 P}{\partial A}, \tag{5.11}$$

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

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

$$P_\lambda = N_{\mathrm{ph},\lambda} \frac{hc}{\lambda}, \tag{5.12}$$

and similarly for P_ν and $N_{\mathrm{ph},\nu}$. The total photon flow N is related to the spectral photon flow via

$$N_{\mathrm{ph}} = \int_0^\infty N_{\mathrm{ph},\nu} \, \mathrm{d}\nu = \int_0^\infty N_{\mathrm{ph},\lambda} \, \mathrm{d}\lambda. \tag{5.13}$$

The physical dimensions of the (spectral) photon flow are

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

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

$$\Phi_{\mathrm{ph}} = \frac{\partial^2 N_{\mathrm{ph}}}{\partial A}, \tag{5.14}$$

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

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

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

$$I_{e\lambda} = \Phi_{\mathrm{ph},\lambda} \frac{hc}{\lambda}, \tag{5.15}$$

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

5.3 Blackbody radiation

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

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

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

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

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

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

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

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

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

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

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

where $c \approx 2.998 \cdot 10^8$ m/s is the speed of light *in vacuo* and $h \approx 6.626 \cdot 10^{-34}$ m^{2}kg/s is the nowadays called *Planck constant*. Via Eq. (5.7) we find the *Planck law* expressed as a function of the frequency ν,

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

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

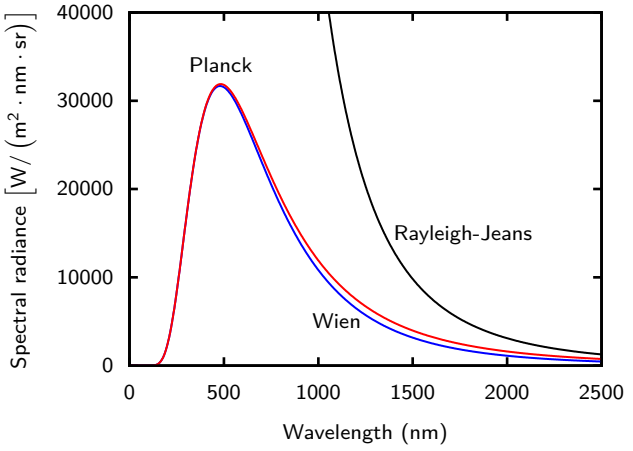

Figure 5.3: The blackbody spectrum at 6000 K as calculated with the Wien approximation, the Rayleigh-Jeans law and the Planck law.

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

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

For short wavelength,

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

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

For long wavelength we can use the approximation

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

which directly results in the Rayleigh-Jeans law.

The total radiant emittance of a black body is given by

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

where

$$\sigma = \frac{2\pi^5 k_B^4}{15 c^2 h^3} \approx 5.670 \cdot 10^{-8} \, \frac{\mathrm{J}}{\mathrm{sm^2K^4}} \tag{5.20}$$

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

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

$$\lambda_{\max} T = b \approx 2.898 \cdot 10^{-3}\,\mathrm{m \cdot K}. \tag{5.21}$$

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

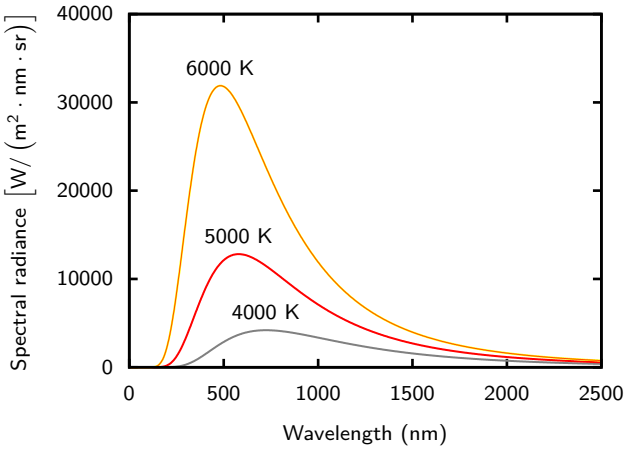

Figure 5.4: The blackbody spectrum at three different temperatures

5.4 Wave-particle duality

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

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

$$E_{\mathrm{ph}} = h\nu. \tag{5.22}$$

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

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

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

5.5 Solar spectra

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

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

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

We can calculate Ω_{Sun} with

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

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

$$\Omega_{\mathrm{Sun}} \approx 68\,\mu\mathrm{sr}. \tag{5.25}$$

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

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

$$\text{AM} := \frac{1}{\cos\theta}. \tag{5.26}$$

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

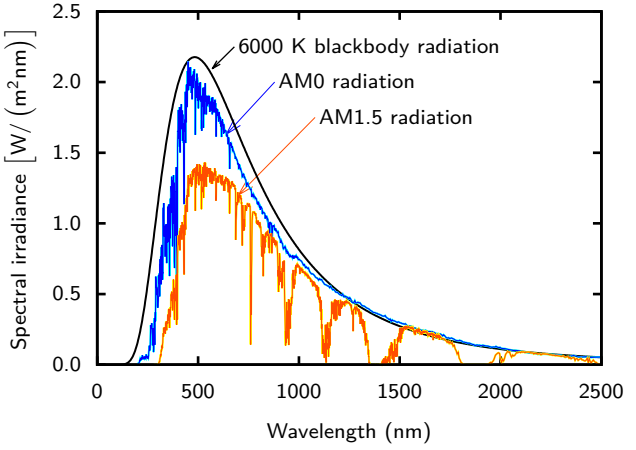

Figure 5.5: Different solar spectra: the blackbody spectrum of a blackbody at 6000 K, the extraterrestrial AM0 spectrum and the AM1.5 spectrum.

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

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

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

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

6 Basic Semiconductor Physics

6.1 Introduction

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

The first successful solar cell was made from crystalline silicon (c-Si), which still is by far the most widely used PV material. Therefore we shall use c-Si as an example to explain the concepts of semiconductor physics that are relevant to solar cell operation. This discussion will give us a basic understanding of how solar cells based on other semiconductor materials work.

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

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

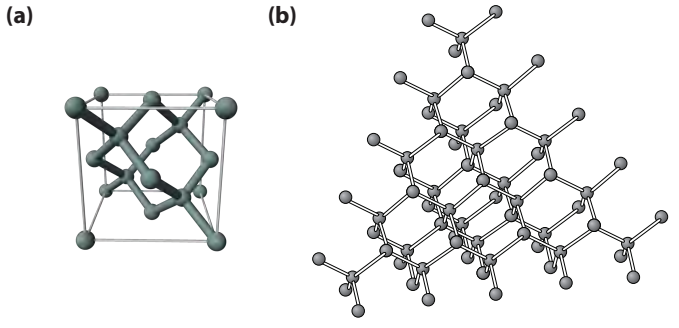

Figure 6.1: (a) A diamond lattice unit cell representing a unit cell of single crystal Si [25], (b) the atomic structure of a part of single crystal Si.

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

3. The lifetime τ and the diffusion length L of the *excess carriers* characterise the recombination-generation processes, discussed in Chapter 7.

4. The band gap energy E_g, and the complex refractive index $n - ik$, where k is linked to the absorption coefficient α, characterise the ability of a semiconductor to absorb electromagnetic radiation.

6.2 Atomic structure

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

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

In the ground state, two valence electrons live in the 3s orbital and the other two are present in the three 3p orbitals (p_x, p_y and p_z). In that state only the two electrons in the 3p orbitals can form bond as the 3s orbital is full. In a silicon crystal, where every atom is *symmetrically* connected to four others, the Si atoms are present as so-called *sp_3 hybrids*. The 3p and 3s orbitals are mixed forming 4 sp_3 orbitals. Every of these four orbitals is occupied by one electron that can form a covalent bond with a valence electron from a neighbour-

ing atom.

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

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

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

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

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

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

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

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

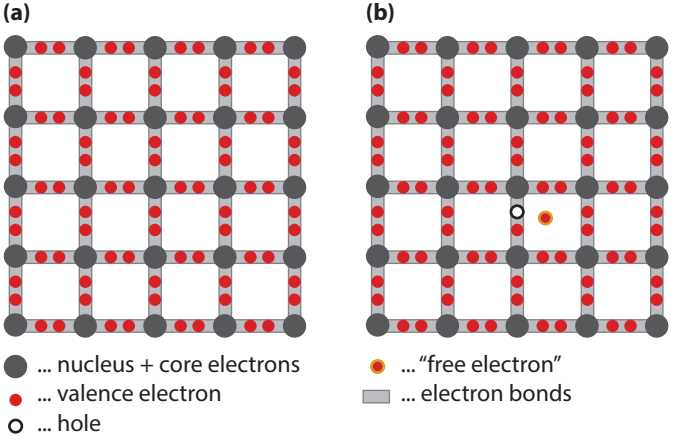

Figure 6.2: The bonding model for c-Si. (a) No bonds are broken. (b) A bond between two Si atoms is broken resulting in a free electron and hole.

6.3 Doping

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

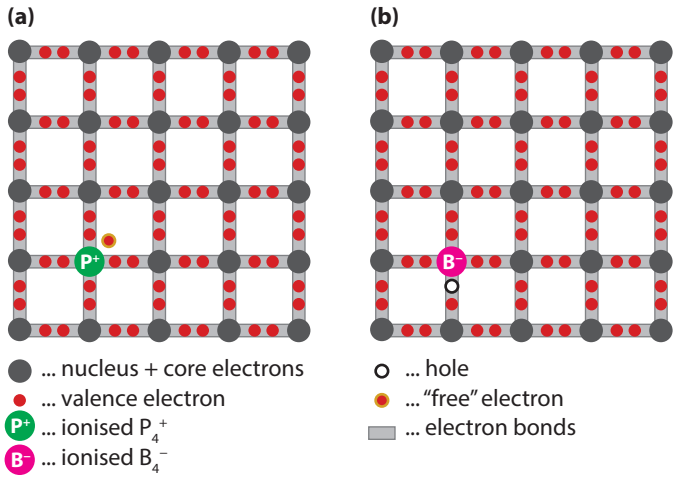

Figure 6.3: The doping process illustrated using the bonding model. (a) A phosphorus (P) atom substitutes a Si atom in the lattice resulting in the positively-ionised P atom and a free electron, (b) A boron (B) atom substitutes a Si atom resulting in the negatively ionised B atom and a hole.

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

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

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

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

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

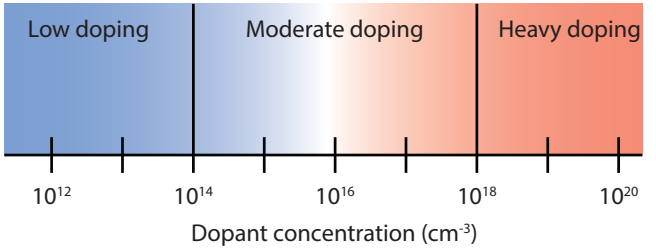

Figure 6.4: The range of doping levels used in c-Si.

6.4 Carrier concentrations

6.4.1 Intrinsic semiconductors

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

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

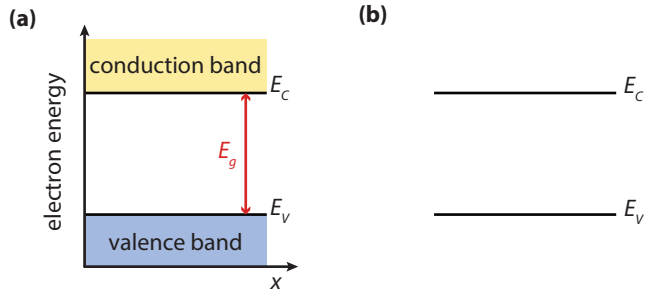

Figure 6.5: (a) The basic energy band diagram and (b) widely used simplified version of the energy diagram.

In order to determine the carrier concentration one has to know the function of density of allowed energy states of electrons and the occupation function of the allowed energy states. The density of energy states function, $g(E)$, describes the number of allowed states per unit volume and energy. The occupation function is the well known Fermi-Dirac distribution function, $f(E)$, which describes the ratio of states filled with an electron to total allowed states at given energy E. In an isolated Si atom, electrons are allowed to have only discrete energy values. The periodic atomic structure of single crystal silicon results in the ranges of allowed energy states for electrons that are called *energy bands* and the excluded energy ranges, forbidden gaps or *band gaps*. Electrons that are liberated from the bonds determine the charge transport in a semiconductor. Therefore, we further discuss only those bands of energy levels, which concern the valence electrons. Valence electrons, which are involved in the covalent bonds, have their allowed energies in the *valence band* (VB) and the allowed energies of electrons liberated from the covalent bonds form the *conduction band* (CB). The valence band is separated from the conduction band by a band of forbidden energy levels. The maximum attainable valence-band energy is denoted E_V, and the minimum attainable conduction-band energy is denoted E_C. The energy difference between the edges of these two bands is called the band gap energy or band gap, E_g, and it is an important material parameter.

$$E_G = E_C - E_V. \tag{6.1}$$

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

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

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

to E_V are given by

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

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

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

The Fermi-Dirac distribution function is given by

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

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

The carriers that contribute to charge transport are electrons in the conduction band and holes in the valence band. The total concentration of electrons in the conduction band and the total concentration of holes in the valence band is obtained by multiplying density of states function with the distribution function and integrating across the whole energy band,

$$n = \int_{E_C}^{E_{top}} g_c(E) f(E) \mathrm{d}E, \tag{6.4a}$$

$$p = \int_{E_{bottom}}^{E_V} g_v(E) \left[1 - f(E)\right] \mathrm{d}E, \tag{6.4b}$$

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

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

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

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

$$N_C = 3.22 \times 10^{19}\,\mathrm{cm}^{-3}, \tag{6.6a}$$

$$N_V = 1.83 \times 10^{19}\,\mathrm{cm}^{-3}. \tag{6.6b}$$

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

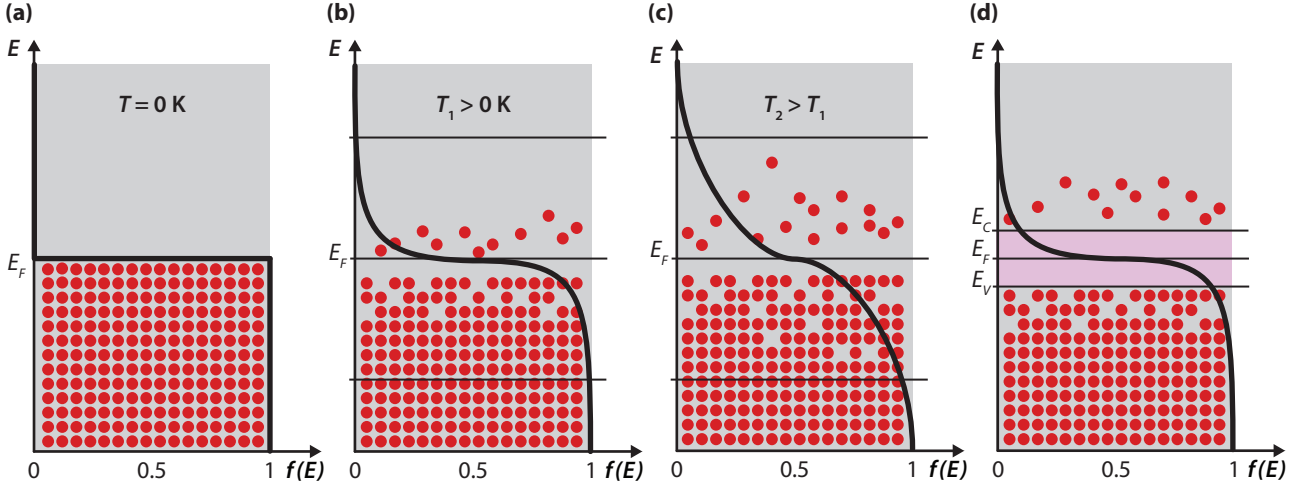

Figure 6.6: The Fermi-Dirac distribution function. (a) For $T = 0$ K, all allowed states bellow the Fermi level are occupied by two electrons. (b, c) At $T > 0$ K not all states below the Fermi level are occupied and there are some states above the Fermi level that are occupied. (d) In an energy gap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

isfied the semiconductor is referred to as a*nondegenerate* semiconductor.

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

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

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

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

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

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

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

6.4.2 Doped semiconductors

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

$$\rho = e\left(p + N_D^+ - n - N_A^-\right), \tag{6.10}$$

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

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

$$p + N_D^+ - n - N_A^- = 0. \tag{6.11}$$

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

dopant atoms. We therefore may assume

$$N_D^+ \approx N_D \qquad \text{and} \qquad N_A^- \approx N_A, \tag{6.12}$$

and hence

$$p + N_D - n - N_A = 0, \tag{6.13}$$

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

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

$$p + N_D - n = 0, \tag{6.14}$$

Under the assumption that

$$N_D \approx N_D^+ \approx n \tag{6.15}$$

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

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

In case of a p-type material almost all acceptor atoms N_A are ionised at room temperature and accept an electron and leaving a hole in the valence band. Under the assumption that $N_A = 0$, Eq. (6.13) becomes

$$p - n - N_A = 0. \tag{6.17}$$

Under the assumption that

$$N_A \approx N_A^- \approx p \tag{6.18}$$

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

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

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

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

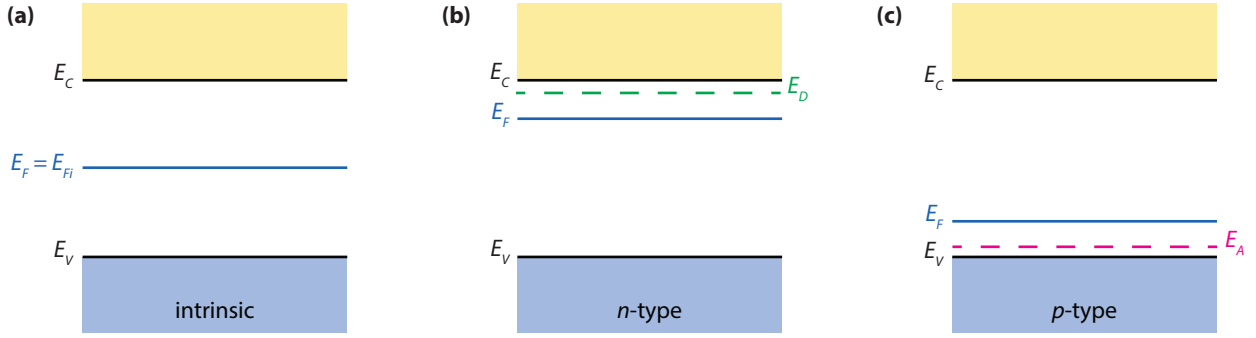

Figure 6.7: A shift of the position of the Fermi energy in the band diagram and the introduction of the allowed energy level into the bandgap due to the doping.

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

The position of the Fermi level in an *n*-type semiconductor can be calculated using Eqs. (6.5a) and (6.15) and in a *p*-type semiconductor using Eqs. (6.5b) and (6.18), respectively,

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

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

6.5 Transport properties

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

Example

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

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

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

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

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

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

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

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

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

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

6.5.1 Drift

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

$$\mathbf{v}_{dn} = -\mu_n \xi, \tag{6.21a}$$
$$\mathbf{v}_{dp} = \mu_p \xi. \tag{6.21b}$$

The proportionality factor is called mobility μ. It is a central parameter that characterises electron and hole transport due to drift. Although the electrons move in the opposite direction to the electric field, because the charge of an electron is negative the resulting electron drift current is in the same direction as the electric field. This is illustrated in Fig. 6.8.

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

$$\mathbf{J}_{n,\,\text{drift}} = -en\mathbf{v}_{dn} = en\mu_n \xi, \tag{6.22a}$$
$$\mathbf{J}_{p,\,\text{drift}} = ep\mathbf{v}_{dp} = ep\mu_p \xi. \tag{6.22b}$$

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

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

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

$$\mu_n \approx 1\,360\,\text{cm}^2\text{V}^{-1}\text{s}^{-1},$$
$$\mu_p \approx 450\,\text{cm}^2\text{V}^{-1}\text{s}^{-1}.$$

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

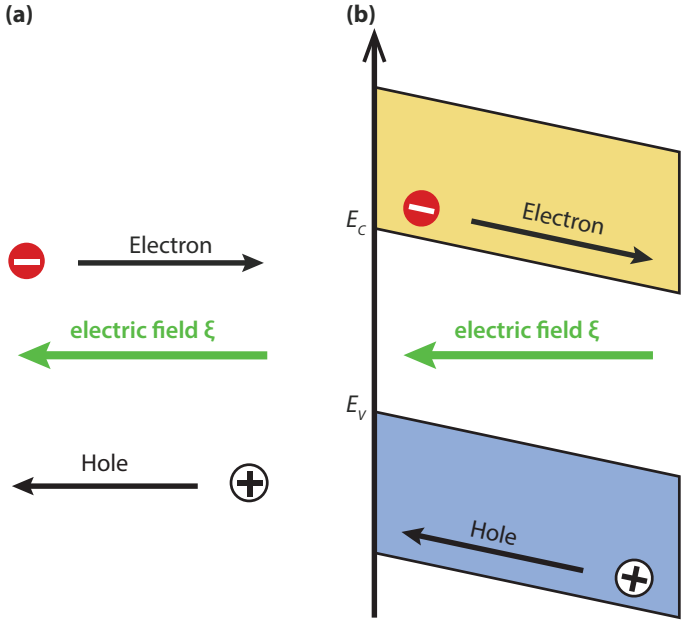

Figure 6.8: Visualisation of (a) the direction of carrier fluxes due to an electric field and (b) the corresponding band diagram.

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

6.5.2 Diffusion

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

$$\mathbf{J}_{n,\,\mathrm{diff}} = \quad eD_n\nabla_n, \tag{6.24a}$$
$$\mathbf{J}_{p,\,\mathrm{diff}} = -eD_p\nabla_p, \tag{6.24b}$$

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

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

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

Example

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

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

by the *Einstein relationship* that is given by

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{e}. \tag{6.26}$$

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

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

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

6.5.3 Continuity equations

As we have discussed in Section 4.5, charge is a conserved quantity. Now we will have a look at how the continuity equations for the electrons and holes can be formulated. We have to take drift, diffusion, and recombination as well as generation processes into account; the latter are discussed in Chapter 7.

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

$$\frac{\partial n}{\partial t} = \left.\frac{\partial n}{\partial t}\right|_{\mathrm{drift}} + \left.\frac{\partial n}{\partial t}\right|_{\mathrm{diff}} + \left.\frac{\partial n}{\partial t}\right|_{R} + \left.\frac{\partial n}{\partial t}\right|_{G}, \tag{6.28a}$$

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

$$\frac{\partial p}{\partial t} = \left.\frac{\partial p}{\partial t}\right|_{\mathrm{drift}} + \left.\frac{\partial p}{\partial t}\right|_{\mathrm{diff}} + \left.\frac{\partial p}{\partial t}\right|_{R} + \left.\frac{\partial p}{\partial t}\right|_{G}. \tag{6.28b}$$

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

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

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

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

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

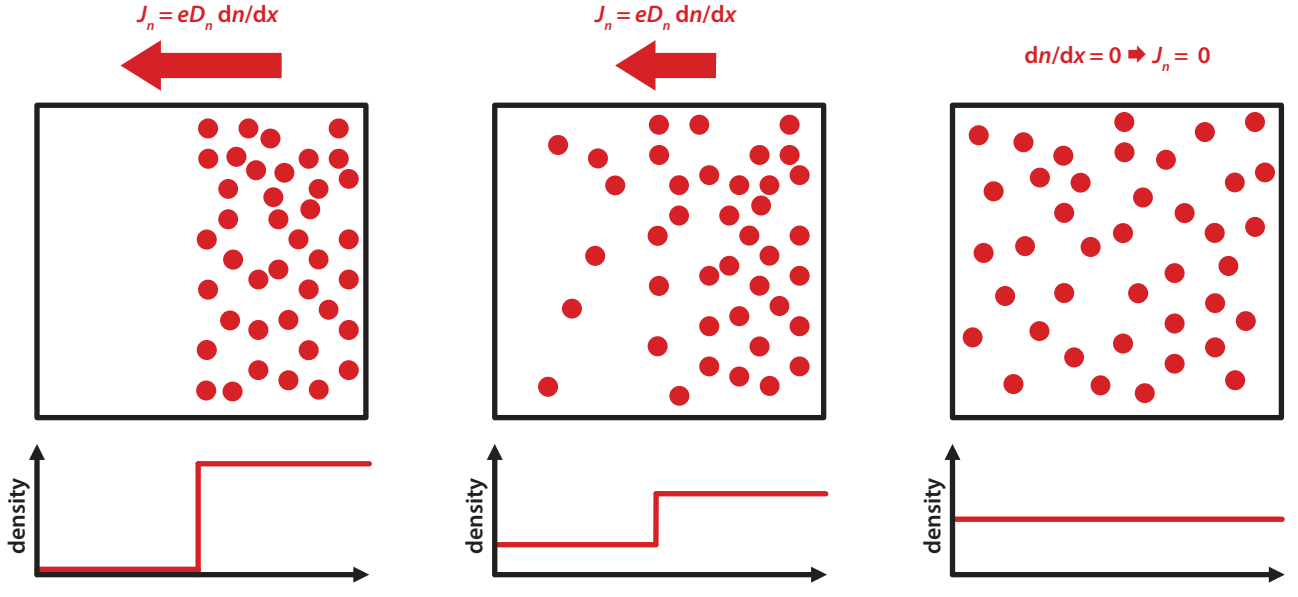

Figure 6.9: Visualisation of electron diffusion.

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

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

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

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

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

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

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

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

7

Generation and Recombination of Electron-Hole Pairs

7.1 Introduction

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

$$np > n_i^2. \tag{7.1}$$

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

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

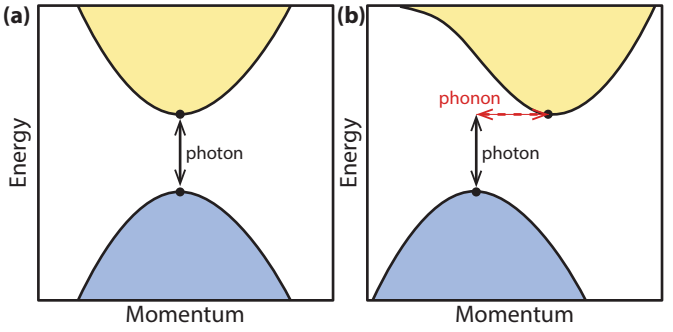

Figure 7.1: Illustrating the dispersion diagram of (a) an direct bandgap semiconductor and (b) a indirect bandgap semiconductor.

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

Before we can actually start with the treatment on the different generation and recombination mechanisms, we mention that we have to distinguish between *direct* and *indirect* semiconductors. Figure 7.1 shows the energy-momentum space of the electrons, which also is called the *electronic dispersion diagram*. On the vertical axis the energy state in the electronic bands is plotted. On the horizontal axis the momentum of the charge carrier is shown. This momentum is also called the *crystal momentum*, it is also related to the *wave vector* $\mathbf{k}$ of the electron. It is important to realise that the position of the valence and conduction band may differ in different directions of the lattice coordination. We can understand this by realising that the crystal can look very different if we look at it from different directions. Hence, also the energy levels in that electrons and holes can propagate across the crystal depend on the direction. The dispersion diagram of silicon is discussed in more detail in Chapter 12.

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

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

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

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

7.2 Bandgap-to-bandgap processes

Generation and recombination processes that happen from bandgap to bandgap are also called *direct* generation and recombination. They are much more likely to happen in direct bandgap materials, as there no change in momentum is required for an electron that is excited into the conduction band. These processes most usually are *radiative*, which means that a photon is absorbed when an electron-hole pair is created, and a photon is emitted if electron-hole pairs recombine directly.

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

7.2.1 Radiative Generation

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

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

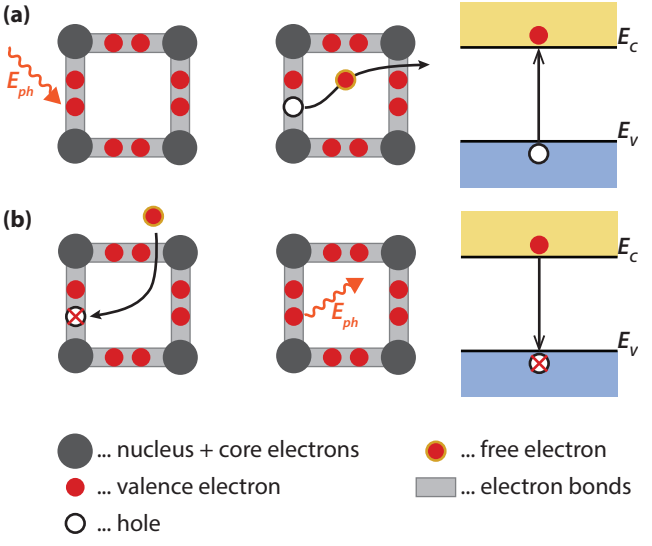

Figure 7.2: Visualisation of bandgap-to-bandgap (a) generation and (b) recombination processes using the bonding model and the energy band diagram.

the absorber,

$$\Phi(x,\lambda) = \Phi_0(\lambda)\exp\left[-\alpha(\lambda)x\right], \tag{7.2}$$

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

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

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

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

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

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

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

via

$$G_L(x) = \eta_g A(x). \tag{7.6}$$

Hence,

$$A(x) = \int_{\lambda_1}^{\lambda_2} \Phi_0(\lambda)\alpha(\lambda)\exp\left[-\alpha(\lambda)x\right]\,\mathrm{d}\lambda. \tag{7.7}$$

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

$$\left.\frac{\partial n}{\partial t}\right|_{\text{light}} = \left.\frac{\partial p}{\partial t}\right|_{\text{light}} = G_L. \tag{7.8}$$

7.2.2 Direct recombination

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

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

$$np = n_i^2 \tag{7.9}$$

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

$$R_{th} = G_{th}. \tag{7.10}$$

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

$$R^* = \beta np, \tag{7.11}$$

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

$$R_{th} = \beta n_0 p_0. \tag{7.12}$$

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

Example

Let us calculate the total absorption in a $d = 300\ \mu$m *thick c-Si wafer for light with a wavelength of 500* nm*. The optical constants of c-Si this wavelength are: refractive index is* $n = 4.293$*, extinction coefficient* $k = 0.045$ *and absorption coefficient* $\alpha = 1.11 \times 10^4\ \text{cm}^{-1}$*. The incident irradiance is 1000* Wm^{-2}*.*
First we calculate the photon flux density at 500 nm *corresponding to the irradiance of* 1000 Wm^{-2}*. Using Eq. (7.3) we obtain*

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

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

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

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

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

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

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

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

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

situation. The total recombination and generation rates are given by

$$R^* = \beta np = \beta(n_0 + \Delta n)(p_0 + \Delta p), \tag{7.13}$$
$$G = G_{th} + G_L, \tag{7.14}$$

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

$$\Delta n = n - n_0, \tag{7.15a}$$
$$\Delta p = p - p_0. \tag{7.15b}$$

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

$$G_L = R^* - G = R_d, \tag{7.16}$$

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

$$G_L = R_d = \beta(np - n_0 p_0). \tag{7.17}$$

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

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

where

$$\tau_{pd} = \frac{1}{\beta n_0} \tag{7.19}$$

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

$$p - p_0 = G_L \tau_{pd}. \tag{7.20}$$

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

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{p(t) - p_0}{\tau_{pd}}. \tag{7.21}$$

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

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

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

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

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

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

where the lifetime of the electrons is given by

$$\tau_{nd} = \frac{1}{\beta p_0}. \tag{7.24}$$

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

$$R_{\text{tot}} = R_1 + R_2 + \ldots \tag{7.25}$$

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

$$R_{\text{tot}} = \frac{p - p_0}{\tau_{p1}} + \frac{p - p_0}{\tau_{p2}} + \ldots = \frac{p - p_0}{\tau_{p,\text{tot}}}. \tag{7.26}$$

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

$$\frac{1}{\tau_{p,\text{tot}}} = \frac{1}{\tau_{p1}} + \frac{1}{\tau_{p2}} + \ldots \tag{7.27}$$

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

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

$$L_n = \sqrt{D_n \tau_n} \quad \text{for electrons in a } p\text{-type material,} \tag{7.28a}$$

$$L_p = \sqrt{D_p \tau_p} \quad \text{for holes in an } n\text{-type material,} \tag{7.28b}$$

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

7.3 Shockley-Read-Hall Recombination

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

Example

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

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

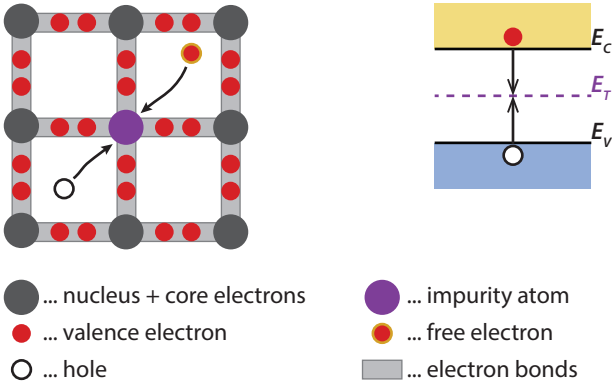

Figure 7.3: Visualisation of Shockley-Read-Hall recombination using the bonding model and the energy band diagram.

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

We distinguish between two kinds of traps: first, *donor-type* traps that are neutral when they contain an electron and positively charged when they do not contain an electron. Secondly, *acceptor-type* traps that are negatively charged when they contain and electron and neutral when they do not contain an electron.

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

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

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

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

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

$$\tfrac{1}{2}m_n^* v_{th,n}^2 = \tfrac{3}{2}kT \qquad \tfrac{1}{2}m_p^* v_{th,p}^2 = \tfrac{3}{2}kT, \tag{7.29}$$

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

The electron capture cross section σ_n describes the effectiveness of the trap state to capture an electron. It is a measure of how close an electron has to come to the trap to be captured. It has the unit of area, cm^2. Similarly, σ_p describes the effectiveness of a trap state to capture a hole.

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

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

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

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

According to the Fermi-Dirac statistics the carrier dis-

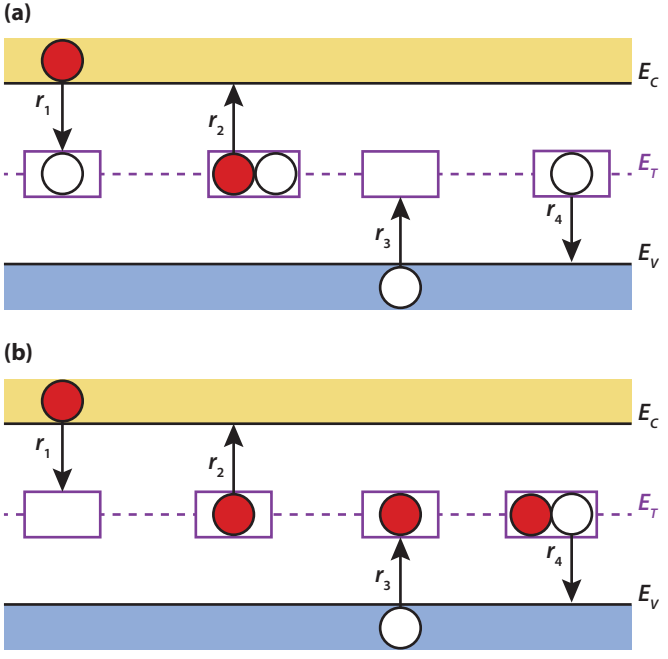

Figure 7.4: Schematic illustration of the processes involved with SRH recombination in a single-electron trap state for (a) a donor-type and (b) an acceptor-type trap.

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

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

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

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

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

where E_T is the trap energy.

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

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

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

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

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

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

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

$$R_{\text{SRH}} = \frac{\mathrm{d}n}{\mathrm{d}t} = \frac{\mathrm{d}p}{\mathrm{d}t} = r_1 - r_2 = r_3 - r_4. \tag{7.34}$$

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

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

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

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

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

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

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

and hence

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

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

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

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

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

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

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

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

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

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

The lifetime of the minority carriers due to Shockley-Read-Hall recombination therefore is indirectly proportional to the trap density N_T. Hence, for a good semiconductor device it is crucial to keep N_T low.

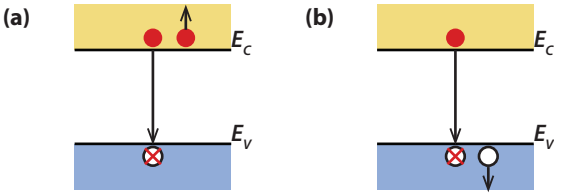

Figure 7.5: Schematic illustration of Auger recombination with (a) two electrons and (b) two holes involved.

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

7.4 Auger Recombination

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

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

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

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

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

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

respectively, where C_n and C_p are the proportionality constants that are strongly dependent on the temperature [26]. R_{eeh} is dominant when the electrons are the majority charge carriers, while R_{ehh} is dominant when the holes are the majority charge carries. Adding them leads to the total Auger recombination rate,

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

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

$$R_{\text{eeh}} = C_n N_D^2 p. \tag{7.44}$$

Hence, the lifetime can be approximated with

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

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

$$R_{\text{ehh}} = C_p N_A^2 n. \tag{7.46}$$

The lifetime then is

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

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

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

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

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

7.5 Carrier concentration in non-equilibrium

When a semiconductor is illuminated additional electrons and holes are generated in the material by the absorption of photons. The *photo-generated carriers* interact with the semiconductor lattice. The extra energy that the electron-hole pairs receive from the photons with energies larger than the band gap of the semiconductor is released into the lattice in form of heat. After this so called *thermalisation* process, which is very fast and takes approximately 10^{-12} s, the carrier concentrations achieve a steady-state. In this non-equilibrium state the electron and hole concentrations are different than those in the equilibrium state. In non-equilibrium states two Fermi distributions are used to describe the electron and hole concentrations. One Fermi distribution with the *quasi-Fermi energy for electrons*, E_{FC}, describes the occupation of states in the conduction band with electrons. Another Fermi distribution with the *quasi-Fermi energy for holes*, E_{FV}, describes the occupation of states in the valence band with electrons, and therefore determines also the concentration of holes. Using the band diagram with the quasi-Fermi levels the process of creation of electron-hole pairs and their subsequent thermalisation that describe the carrier concentration under illumination is illustrated in Fig. 7.6. The difference between the quasi-Fermi levels is the electrochemical energy, μ_{eh}, of the generated electron-hole pairs which represents the measure for the conversion efficiency of solar radiation.

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

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

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

It then follows that under non-equilibrium conditions

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

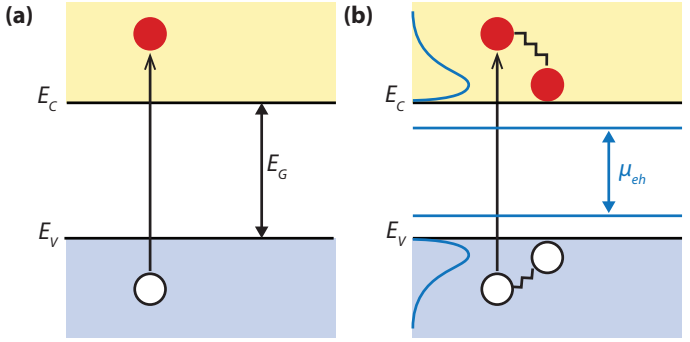

Figure 7.6: Thermalisation of photo-generated electron-hole pairs resulting in non-equilibrium charge-carrier concentrations described by the quasi-Fermi levels.

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

$$\mathbf{J}_N = n\mu_n \nabla E_{FC}, \tag{7.51a}$$
$$\mathbf{J}_P = p\mu_p \nabla E_{FV}. \tag{7.51b}$$

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

$$\frac{\mathrm{d}E_{FC}}{\mathrm{d}x} \neq 0 \qquad \text{or} \qquad \frac{\mathrm{d}E_{FV}}{\mathrm{d}x} \neq 0,$$

the current is flowing inside the semiconductor. By checking the position dependence of the quasi-Fermi levels in an energy band diagram, one can easily determine whether current flows inside the semiconductor.

8 Semiconductor Junctions

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

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

8.1 *p-n* homojunctions

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

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

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

$$n = n_{n0} \approx N_D, \tag{8.1a}$$

$$p = p_{n0} \approx n_i^2 \Big/ N_D \,. \tag{8.1b}$$

For the isolated p-type semiconductor we have

$$p = p_{p0} \approx N_A, \tag{8.2a}$$

$$n = n_{p0} \approx n_i^2 \Big/ N_A \,. \tag{8.2b}$$

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

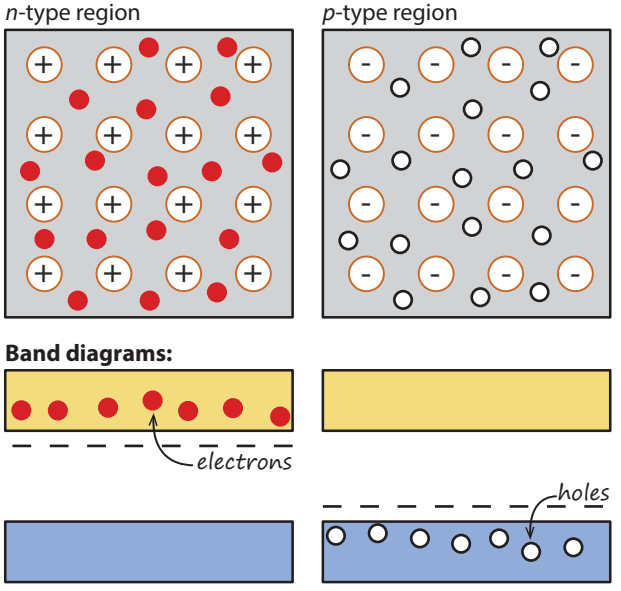

Figure 8.1: Schematic representation of an isolated n-type and p-type semiconductor and corresponding band diagrams.

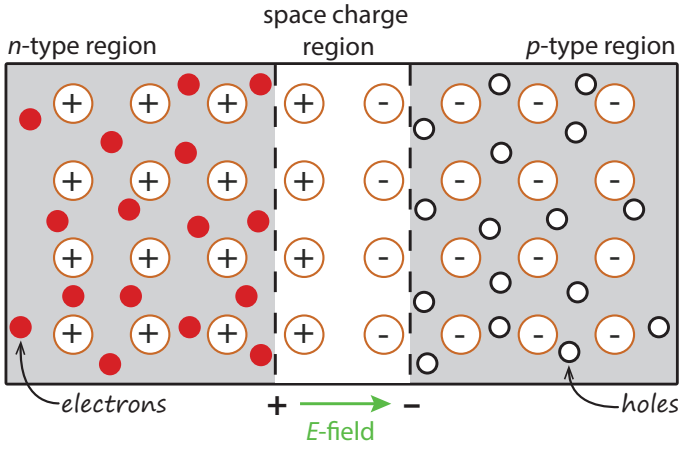

Figure 8.2: Formation of a space-charge region, when *n*-type and *p*-type semiconductors are brought together to form a junction. The coloured part represents the space-charge region.

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

The space charge around the metallurgical junction results in the formation of an internal electric field which forces the charge carriers to move in the opposite direction than the concentration gradient. The diffusion currents continue to flow until the forces acting on the charge carriers, namely the concentration gradient and the internal electrical field, compensate each other. The driving force for the charge transport does not exist any more and no net current flows through the *p-n* junction.

8.1.2 The *p-n* junction under equilibrium

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

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

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

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

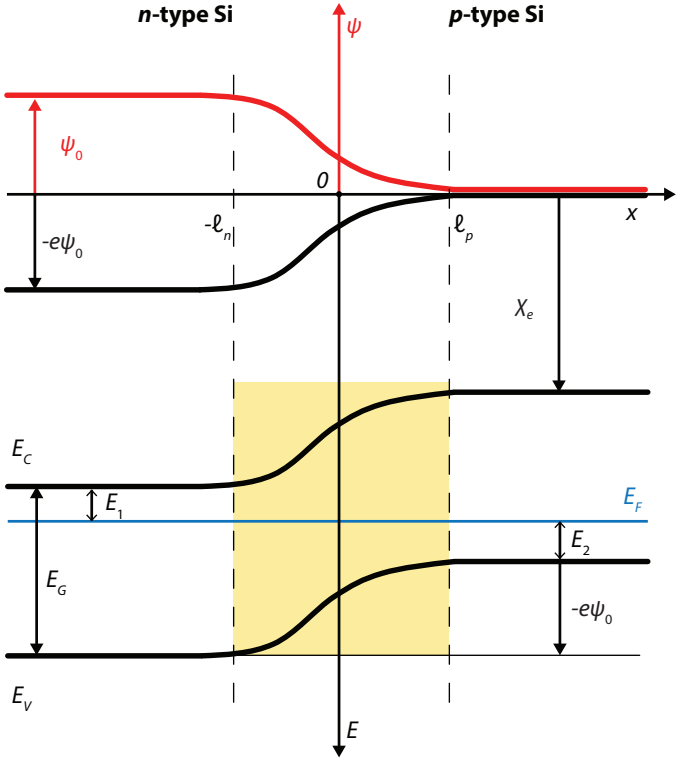

Figure 8.3: Energy-band diagram of the *p-n* junction. The electrostatic potential profile (red curve) is also presented in the figure.

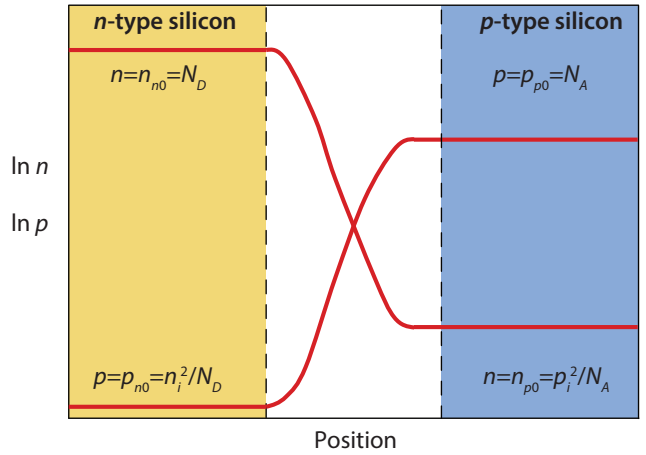

Figure 8.4: Concentrations profile of mobile charge carriers in a *p*-*n* junction under equilibrium.

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

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

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

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

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

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} = -\frac{\mathrm{d}\,\xi}{\mathrm{d}x} = -\frac{\rho}{\varepsilon_r\varepsilon_0}, \tag{8.4}$$

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

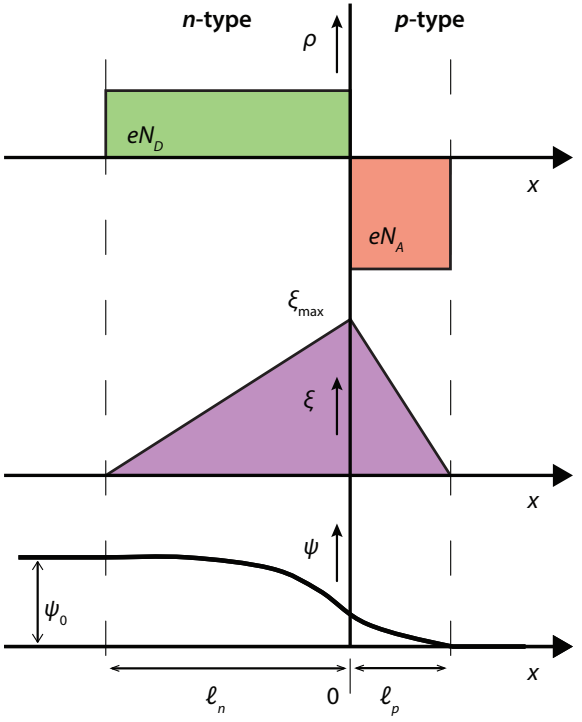

Figure 8.5: The space-charge density $\rho(x)$, the electric field $\xi(x)$, and the electrostatic potential $\psi(x)$ across the depletion region of a *p*-*n* junction under equilibrium.

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

$$\xi = \frac{1}{\varepsilon_r \varepsilon_0} \int \rho \, \mathrm{d}x. \tag{8.5}$$

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

$$\xi\left(-\ell_n\right) = \xi\left(\ell_p\right) = 0, \tag{8.6}$$

we obtain as solution for the electric field

$$\xi\ (x) = \frac{e}{\varepsilon_r \varepsilon_0} N_D\ \left(\ell_n + x\right) \quad \text{for} \quad -\ell_n \leq x \leq 0, \tag{8.7a}$$

$$\xi\ (x) = \frac{e}{\varepsilon_r \varepsilon_0} N_A\ \left(\ell_p - x\right) \quad \text{for} \quad 0 \leq x \leq \ell_p. \tag{8.7b}$$

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

$$N_A \ell_p = N_D \ell_n. \tag{8.8}$$

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

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

$$\psi = -\int \xi \mathrm{d}x. \tag{8.9}$$

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

$$\psi(\ell_p) = 0. \tag{8.10}$$

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

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

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

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

$$\psi_0 = \psi(-\ell_n) - \psi(\ell_p) = \psi(-\ell_n). \tag{8.12}$$

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

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

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

$$e\psi_0 = E_G - E_1 - E_2. \tag{8.14}$$

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

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

We can write

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

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

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

we can rewrite Eq. (8.15) and obtain

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

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

$$\ell_n = \sqrt{\frac{2\varepsilon_r\,\varepsilon_0}{q}\,\psi_0\,\frac{N_A}{N_D}\left(\frac{1}{N_A+N_D}\right)}, \tag{8.17a}$$

$$\ell_p = \sqrt{\frac{2\varepsilon_r\,\varepsilon_0}{q}\,\psi_0\,\frac{N_D}{N_A}\left(\frac{1}{N_A+N_D}\right)}. \tag{8.17b}$$

The total space-charge width, W, is the sum of the partial space-charge widths in the n- and p-type semiconductors and can be calculated with

$$W = \ell_n + \ell_p = \sqrt{\frac{2\varepsilon_r\,\varepsilon_0}{e}\,\psi_0\,\left(\frac{1}{N_A}+\frac{1}{N_D}\right)}. \tag{8.18}$$

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

$$\xi_{\max} = \sqrt{\frac{2e}{\varepsilon_r\,\varepsilon_0}\,\psi_0\,\left(\frac{N_A\, N_D}{N_A+N_D}\right)}. \tag{8.19}$$

Example

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

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

P-type region:

$$p = p_{p0} \approx N_A = 10^{16}\,\mathrm{cm}^{-3}.$$

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

N-type region:

$$n = n_{n0} \approx N_A = 10^{18}\,\mathrm{cm}^{-3}.$$

$$p = p_{n0} = n_i^2 \Big/ n_{n0} = \left(1.5 \cdot 10^{10}\right)^2 \Big/ 10^{18} = 2.25 \cdot 10^2\,\mathrm{cm}^{-3}$$

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

N-type region:

$$E_F - E_C = -kT \ln\left(N_C/n\right) = -0.0258 \ln\left(3.32 \cdot 10^{19} \Big/ 10^{18}\right) = -0.09\,\mathrm{eV}.$$

P-type region:

$$E_F - E_C = -kT \ln\left(N_C/n\right) = -0.0258 \ln\left(3.32 \cdot 10^{19} \Big/ 2.24 \cdot 10^{4}\right) = -0.90\,\mathrm{eV}.$$

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

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

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

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

$$W = \sqrt{\frac{2\varepsilon_r \varepsilon_0}{e} \psi_0 \left(\frac{1}{N_A} + \frac{1}{N_D} \right)} = \sqrt{\frac{2 \cdot 11.7 \cdot 8.854 \cdot 10^{-14}}{1.602 \cdot 10^{-19}} \cdot 0.81 \left(\frac{1}{10^{16}} + \frac{1}{10^{18}} \right)} = 3.25 \cdot 10^{-5}\,\mathrm{cm} = 0.325\,\mu\mathrm{m}.$$

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

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

$$\xi_{\max} = \sqrt{\frac{2e}{\varepsilon_r \varepsilon_0} \psi_0 \left(\frac{N_A N_D}{N_A + N_D} \right)} = \frac{2 \cdot 1.602 \cdot 10^{-19}}{11.7 \cdot 8.854 \cdot 10^{-14}} \cdot 0.81 \left(\frac{10^{16}\, 10^{18}}{10^{16} + 10^{18}} \right) = 5 \cdot 10^{4}\,\mathrm{V\,cm^{-1}}.$$

8.1.3 The *p-n* junction under applied voltage

When an external voltage, V_a, is applied to a *p-n* junction the potential difference between the *n*- and *p*-type regions will change and the electrostatic potential across the space-charge region will become $(\psi_0 - V_a)$. Remember that under equilibrium the built-in potential is negative in the *p*-type region with respect to the *n*-type region. When the applied external voltage is negative with respect to the potential of the *p*-type region, the applied voltage will increase the potential difference across the *p-n* junction. We refer to this situation as *p-n* junction under *reverse-bias voltage*. The potential barrier across the junction is increased under reverse-bias voltage, which results in a wider space-charge region. The band diagram of the *p-n* junction under reverse-biased voltage is presented in Fig. 8.6 (a). Under external voltage the *p-n* junction is not under equilibrium any more and the concentrations of electrons and holes are described by the quasi-Fermi energy for electrons, E_{FC}, and the quasi-Fermi energy for holes, E_{FV}, respectively. When the applied external voltage is positive with respect to the potential of the *p*-type region, the applied voltage will decrease the potential difference across the *p-n* junction. We refer to this situation as *p-n* junction under *forward-bias voltage*. The band diagram of the *p-n* junction under forward-biased voltage is presented in Fig. 8.6 (b). The potential barrier across the junction is decreased under forward-bias voltage and the space charge region becomes narrower. The balance between the forces responsible for diffusion (concentration gradient) and drift (electric field) is disturbed. Lowering the electrostatic potential barrier leads to a higher concentration of minority carriers at the edges of the space-charge region compared to the situation in equilibrium. This process is referred to as minority-carrier *injection*. This gradient in concentration causes the diffusion of the minority carriers from the edge into the bulk of the quasi-neutral region.

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

$$J = J_{\text{rec}} - J_{\text{gen}} = 0 \qquad \text{for } V_a = 0 \text{ V}. \tag{8.20}$$

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

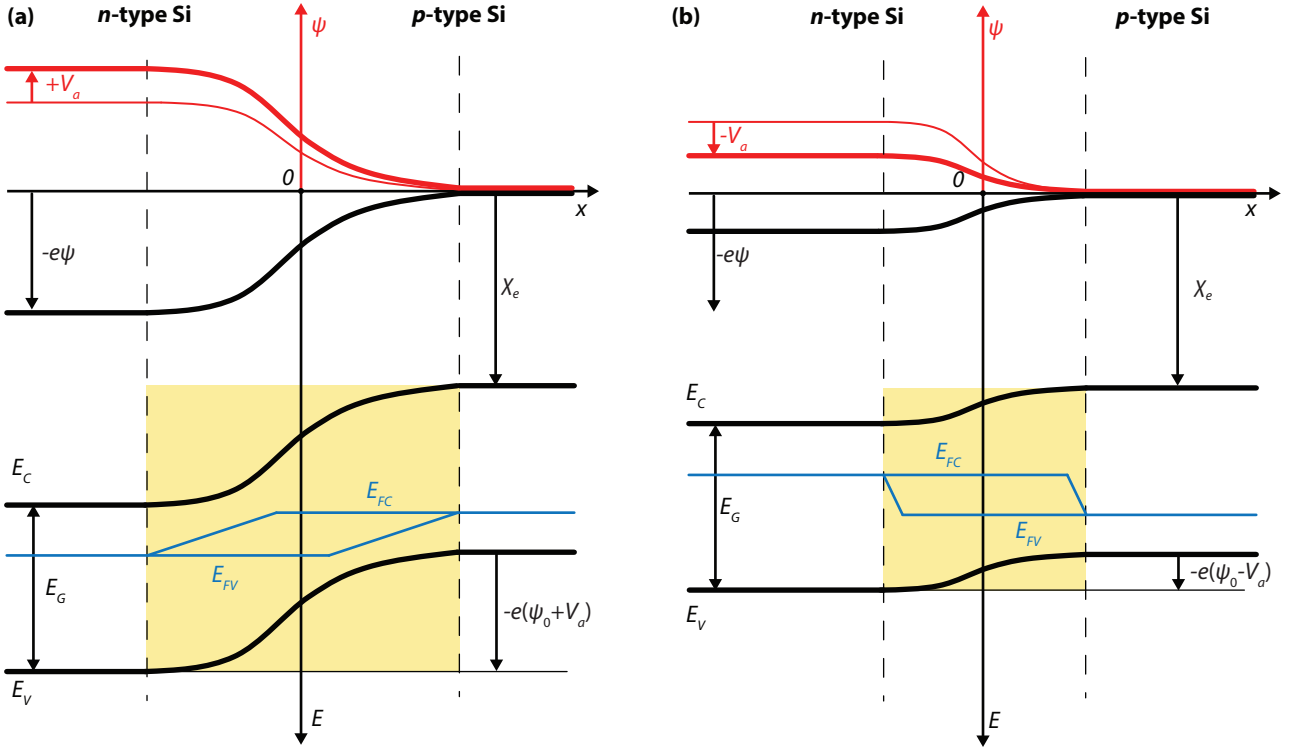

Figure 8.6: Energy band diagram and potential profile (in red colour) of a *p*-*n* junction under (a) reverse bias, and (b) forward bias.

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

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

This assumption is called the *Boltzmann approximation*.

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

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

The external net-current density can be expressed as

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

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

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

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

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

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

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

$$J(V_a) = -J_0, \tag{8.26}$$

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

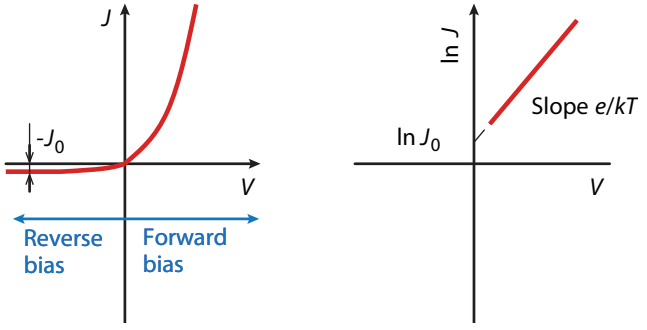

Figure 8.7: *J-V* characteristic of a *p-n* junction; (a) linear plot and (b) semi-logarithmic plot.

8.1.4 The *p-n* junction under illumination

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

When a load is connected between the electrodes of the

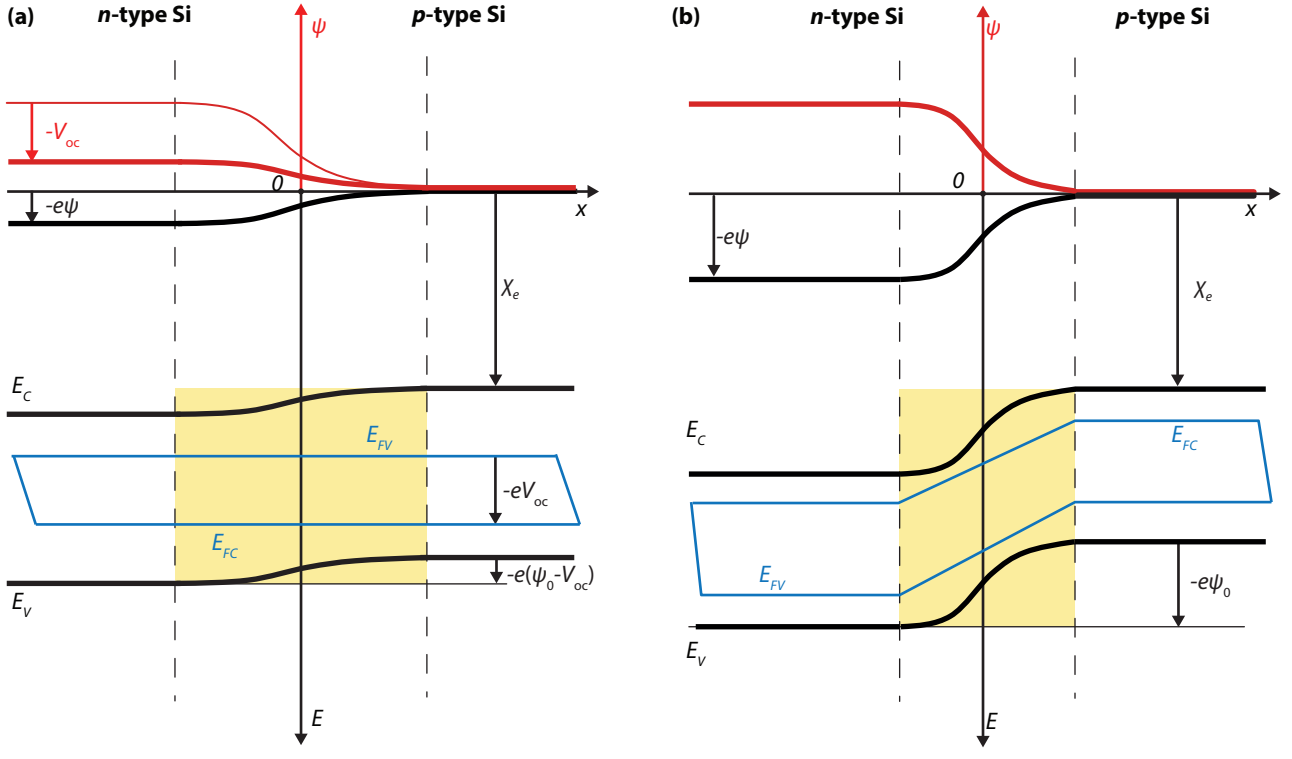

Figure 8.8: Energy band diagram and electrostatic-potential (in red colour) of an illuminated *p-n* junction under the (a) open-circuit and (b) short-circuit conditions.

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

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

Both the dark and illuminated *J-V* characteristics of the *p-n* junction are represented in Fig. 8.9. Note, that in the figure the superposition principle is reflected. The illuminated *J-V* characteristic of the *p-n* junction is the same as the dark *J-V* characteristic, but it is shifted down by the photo-generated current density J_{ph}. The detailed derivation of the photo-generated current density of the *p-n* junction is carried out in Appendix B.2. and its value under uniform generation rate, G, is

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

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

8.2 *p-n* heterojunctions

8.3 Metal-semiconductor junctions

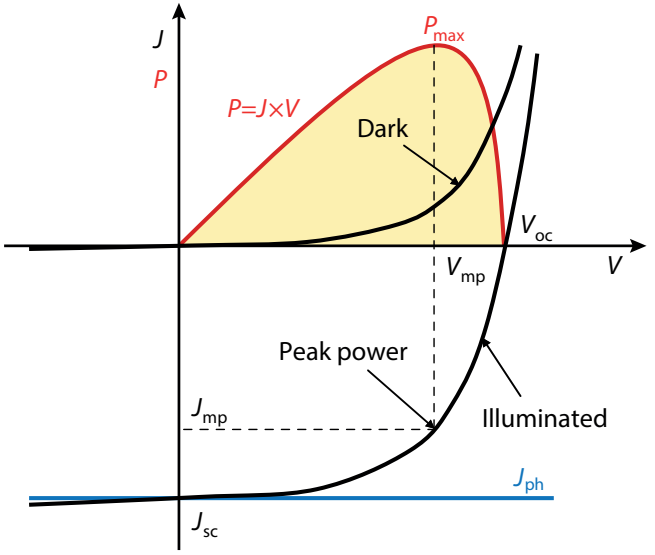

Figure 8.9: *J-V* characteristics of a *p-n* junction in the dark and under illumination.

9 Solar Cell Parameters and Equivalent Circuit

9.1 External solar cell parameters

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

9.1.1 Short-circuit current density

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

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

(8.28), which shows that in case of an ideal diode (for example no surface recombination) and uniform generation, the critical material parameters that determine J_{ph} are the diffusion lengths of minority carriers. Crystalline silicon solar cells can deliver under an AM1.5 spectrum a maximum possible current density of 46 mA/cm^2. In laboratory c-Si solar cells the measured J_{sc} is above 42 mA/cm^2, while commercial solar cell have an J_{sc} exceeding 35 mA/cm^2.

9.1.2 Open-circuit voltage

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

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

This equation shows that V_{oc} depends on the saturation current of the solar cell and the photo-generated current. While J_{ph} typically has a small variation, the key effect is the saturation current, since this may vary by orders of magnitude. The saturation current density, J_0, depends on the recombination in the solar cell. Therefore, V_{oc} is a measure of the amount of recombination in the device. Laboratory crystalline silicon solar cells have a V_{oc} of up to 720 mV under the standard AM1.5 conditions, while commercial solar cells typically have V_{oc} exceeding 600 mV.

9.1.3 Fill factor

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

$$FF = \frac{J_{mp}V_{mp}}{J_{sc}V_{oc}}. \tag{9.2}$$

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

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

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

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

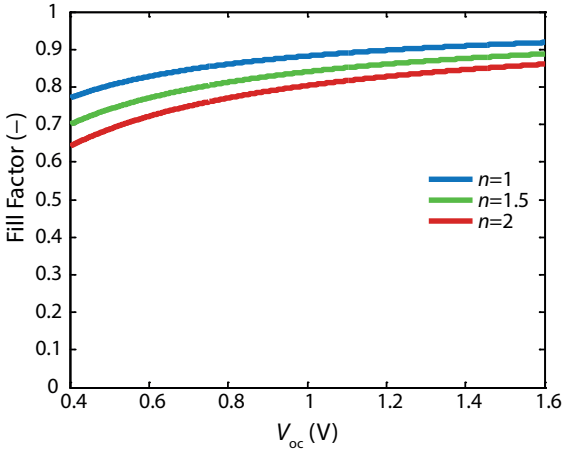

Figure 9.1: The FF as a function of V_{oc} for a solar cell with ideal diode behaviour.

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

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

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

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

9.1.4 Conversion efficiency

The *conversion efficiency* is calculated as the ratio between the maximal generated power and the incident power. The irradiance value P_{in} of 1000 W/m^2 for the AM1.5 spectrum has become a standard for measuring

the conversion efficiency of solar cells,

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

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

Example

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

$$
\begin{aligned}
J_{\text{ph}} &= 350\,\text{Am}^{-2}.\\
N_A &= 10^{17}\,\text{cm}^{-3} = 10^{23}\,\text{m}^{-3}.\\
N_D &= 10^{19}\,\text{cm}^{-3} = 10^{25}\,\text{m}^{-3}.\\
L_N &= 500 \cdot 10^{-6}\,\text{m}.\\
L_P &= \ 10 \cdot 10^{-6}\,\text{m}.\\
D_N &= (kT/e)\mu_n = 0.0258\,\text{V} \cdot 1000 \cdot 10^{-4}\,\text{cm}^2\text{V}^{-1}\text{s}^{-1} = 2.58 \cdot 10^{-3}\,\text{m}^2\text{s}^{-1}.\\
D_P &= (kT/e)\mu_p = 0.0258\,\text{V} \cdot \ 100 \cdot 10^{-4}\,\text{cm}^2\text{V}^{-1}\text{s}^{-1} = 2.58 \cdot 10^{-4}\,\text{m}^2\text{s}^{-1}.
\end{aligned}
$$

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

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

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

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

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

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

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

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

Hence,

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

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

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

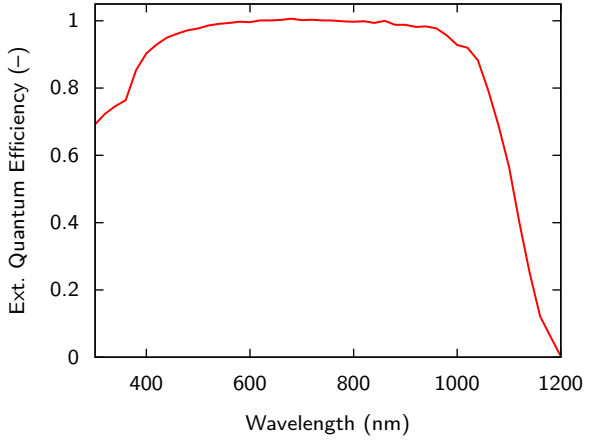

Figure 9.2: The external quantum efficiency of a high-quality crystalline silicon based solar cell.

9.2 The external quantum efficiency

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

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

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

Figure 9.2 illustrates a typical EQE for a high quality crystalline silicon based solar cell. We can identify the major optical loss mechanisms for such a solar cell: For short wavelengths only a small fraction of the light is converted into electron-hole pairs. Most photons are already absorbed in the layers that the light traverses prior to the absorber layer. For long wavelengths, the penetration depth[2] of the light exceeds the optical thickness of the absorber. Then the absorber itself becomes transparent so that most of the light leaves the solar cell before it can be absorbed. We can see that for this type of solar cells the EQE is close to 1 for a broad wavelength band. Hence, in this band almost all absorbed photons are converted into electron-hole pairs that can leave the solar cell.

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

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

Measuring the EQE

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

As already seen in Eq. (9.5), EQE(λ) is proportional to the the current divided by the photon flux. While the current can be easily determined using an Ampere meter, the photon flux must be determined indirectly. This is done by performing a measurement with a calibrated photodetector (or solar cell), whose EQE is known. Via this measurement we find

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

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

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

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

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

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

$$J_{sc} = -e \int_{\lambda_1}^{\lambda_2} \mathrm{EQE}(\lambda) \Phi_{\mathrm{AM1.5}}(\lambda)\, \mathrm{d}\lambda. \tag{9.8}$$

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

9.3 The equivalent circuit

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

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

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

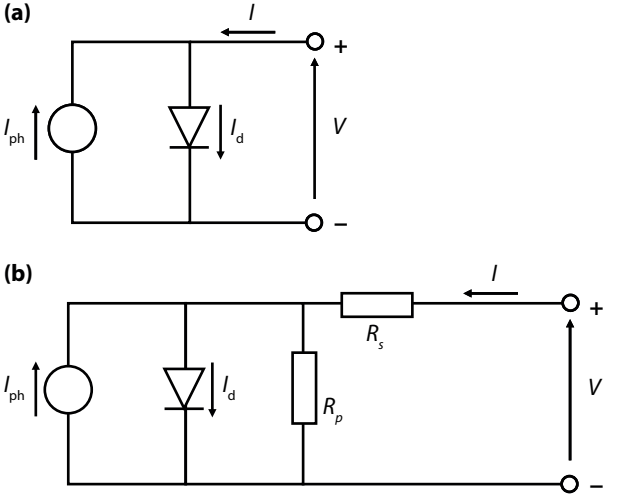

Figure 9.3: The equivalent circuit of an (a) ideal solar cell and a (b) solar cell with a series resistance R_s and a shunt resistance R_p.

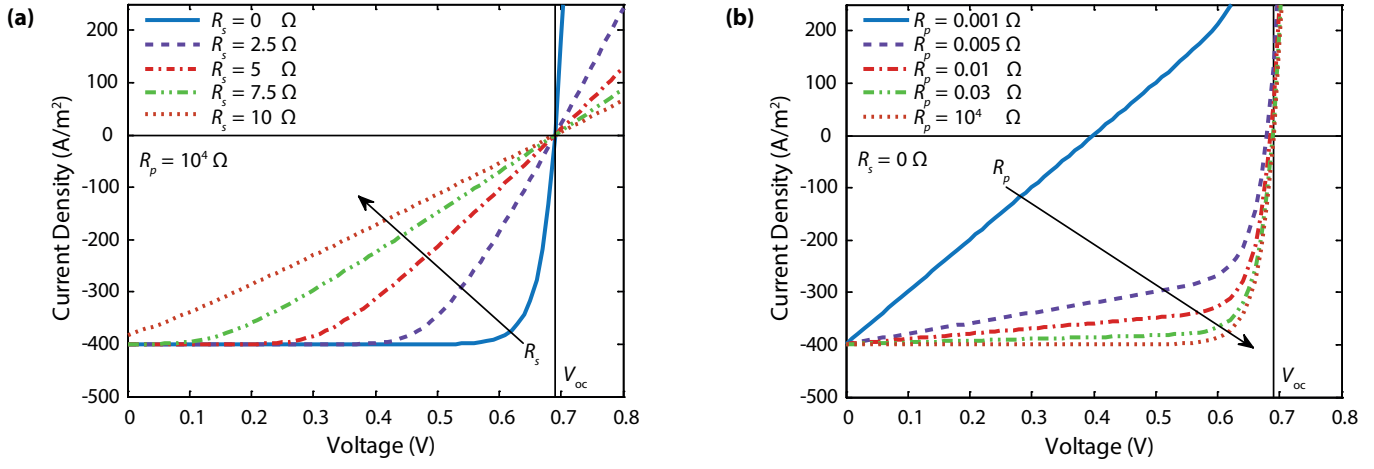

Figure 9.4: Effect of the (a) series resistance and (b) parallel resistance on the *J-V* characteristic of a solar cell.

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

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

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

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

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

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

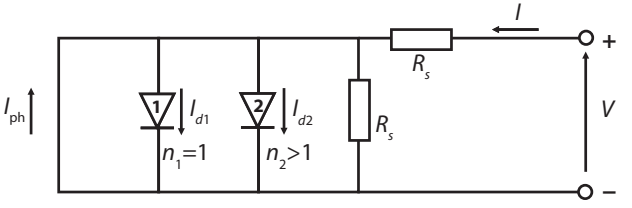

Figure 9.5: The equivalent circuit of a solar cell based on the two-diode model.

10 Losses and Efficiency Limits

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

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

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

10.1 The thermodynamic limit

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

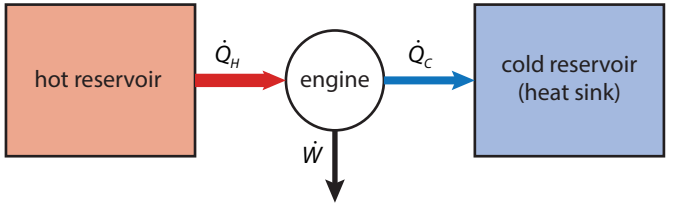

Figure 10.1: Illustrating the major heat flows in a generic heat engine.

by

$$\eta = \frac{\dot{W}}{\dot{Q}_h}. \tag{10.1}$$

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

$$\eta_{\text{Carnot}} = 1 - \frac{T_C}{T_H}. \tag{10.2}$$

For a Carnot engine, the entropy does not increase. Note that all the temperatures must be given in a temperature scale where the absolute zero takes the value 0. For example, the Kelvin scale is such a scale. From Eq. (10.2) we can already see two important trends that are basically true for every heat engine, *i.e.* also steam engines of combustion engines. The efficiency increases, if the higher temperature T_H is increased and/or the lower temperature T_C is decreased.

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

$$\eta_{\text{TD}} = 1 - \frac{T_C}{T_A}. \tag{10.3}$$

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

by

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

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

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

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

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

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

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

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

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

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

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

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

10.2 The Shockley-Queisser Limit

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

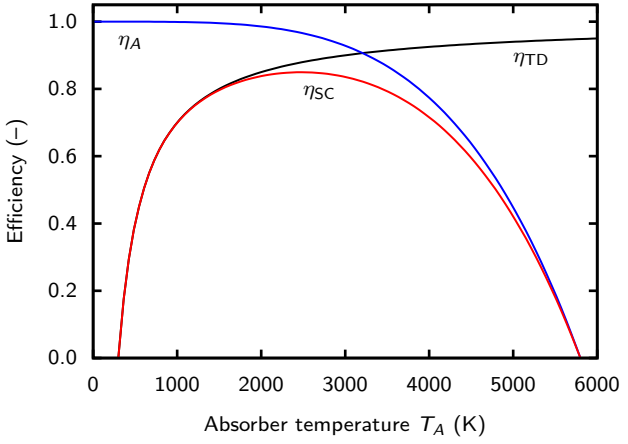

Figure 10.2: The absorber efficiency η_A, the thermodynamic efficiency η_{TD} and the combined solar cell efficiency η_{SC} under full concentration for a solar temperature of 5800 K and an ambient temperature of 300 K.

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

10.2.1 Spectral mismatch

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

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

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

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

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

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

$$p_{\text{use}} = \frac{E_g \int_0^{\lambda_g} \Phi_{\text{ph},\lambda}\, \mathrm{d}\lambda}{\int_0^{\lambda_g} \frac{hc}{\lambda} \Phi_{\text{ph},\lambda}\, \mathrm{d}\lambda}. \tag{10.10}$$

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

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

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

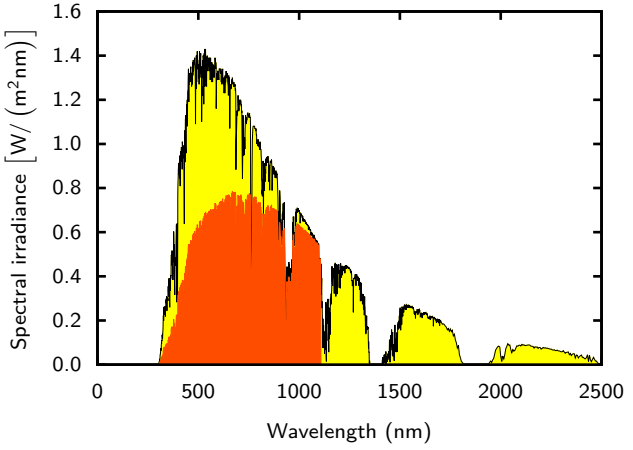

Figure 10.3: The fraction of the AM1.5 spectrum that can be converted into a usable energy by a crystalline silicon solar cell with $E_g = 1.12$ eV.

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

10.2.2 Detail balance limit of the efficiency

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

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

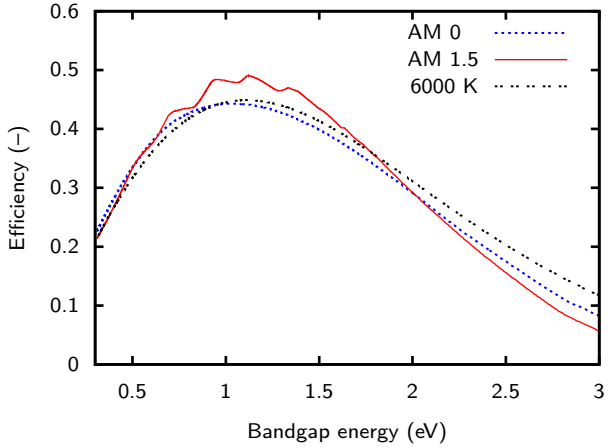

Figure 10.4: The ultimate conversion efficiency for the black body spectrum at 6000 K, the AM0 and AM1.5 solar radiation spectra, limited only by the spectral mismatch as a function of the band gap of a semiconductor absorber in single junction solar cells.

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

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

$$\eta = \frac{J_{\mathrm{ph}} V_{\mathrm{oc}}\, FF}{P_{\mathrm{in}}}. \tag{10.12}$$

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

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

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

$$J_{\mathrm{ph}} = -\frac{e}{E_g} P_{\mathrm{in}} \eta_{\mathrm{ult}} = -\frac{P_{\mathrm{in}} \eta_{\mathrm{ult}}}{V_g}. \tag{10.14}$$

Let us now define the *bandgap utilisation efficiency* η_V

that is given by

$$\eta_V = \frac{V_{oc}}{V_g} \tag{10.15}$$

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

$$\eta = \eta_{ult}\eta_V FF. \tag{10.16}$$

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

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

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

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

Using the equation for the *blackbody radiance* $L_{e\lambda}^{BB}$ as given in Eq. (5.18a) we find for the radiative recombination current

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

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

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

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

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

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

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

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

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

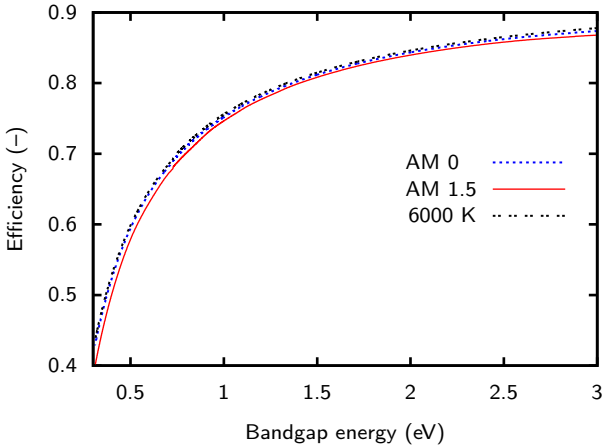

Figure 10.5: The bandgap utilisation efficiency η_V for the black body spectrum at 6000 K, and the AM0 and AM1.5 solar spectra.

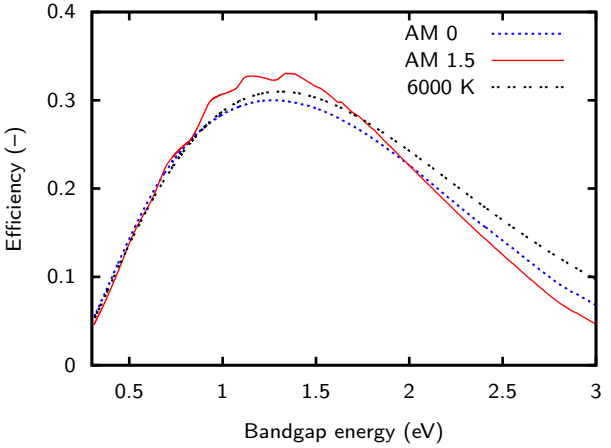

Figure 10.6: The Shockley-Queisser efficiency limit for the black body spectrum at 6000 K, and the AM0 and AM1.5 solar spectra.

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

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

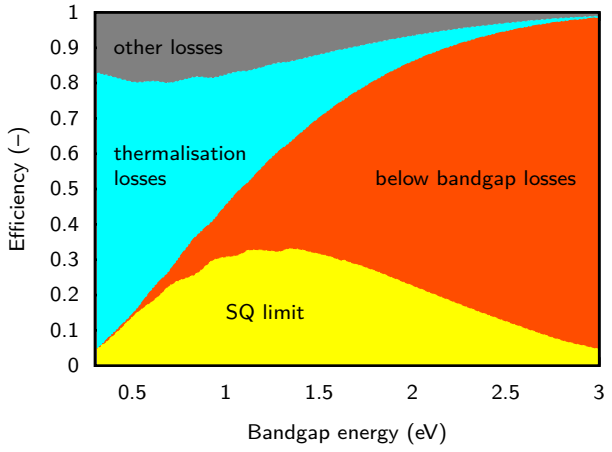

Figure 10.7: The major loss mechanisms in the Shockley Queisser limit. For this calculation the AM1.5 spectrum was used as incident light.

10.2.3 Efficiency limit for silicon solar cells

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

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

10.3 Other Losses

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

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

10.3.1 Optical losses

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

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

As we will discuss in more detail in Chapter 12, in most c-Si solar cells thin metal strips are placed on the front side of the solar cell that serve as front electrode. The metal-covered area does not allow the light to enter the solar cell because it reflects or slightly absorbs the incident light. The area that is covered by the electrode effectively decreases the active area of the solar cell. When we denote the total area of the cell as A_{tot} and the cell area that is not covered by the electrode as A_f, the fraction of the active area of the cell is determined by the ratio

$$C_f = \frac{A_f}{A_{\text{tot}}}, \tag{10.21}$$

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

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

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

10.3.2 Solar cell collection losses

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

The contributions of both the electronic and optical properties of the solar cell materials to the photovoltaic performance are taken into account in the absolute external quantum efficiency that we already defined in Chapter 9. The EQE can be approximated by

$$\mathrm{EQE}(\lambda) = (1 - R^*)\mathrm{IQE_{op}}(\lambda)\eta_g(\lambda)\mathrm{IQE_{el}}(\lambda), \qquad (10.22)$$

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

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

$$J_{\mathrm{sc}} = J_{\mathrm{ph}}(1 - R^*)\mathrm{IQE}^*_{\mathrm{op}}\eta^*_g\mathrm{IQE}^*_{\mathrm{el}}C_f, \qquad (10.23)$$

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

10.3.3 Additional limiting factors

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

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

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

10.3.4 Conversion efficiency

Again, we start with the expression for the efficiency

$$\eta = \frac{J_{sc} V_{oc}\, FF}{P_{in}}. \tag{10.24}$$

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

$$\begin{aligned}\eta &= \frac{\eta_{ult} V_{oc}\, FF}{V_g}(1-R^*)\mathrm{IQE}^*_{op}\eta^*_g \mathrm{IQE}^*_{el} C_f \\ &= p_{abs} p_{use}(1-R^*)\mathrm{IQE}^*_{op}\eta^*_g \mathrm{IQE}^*_{el} C_f \eta_V FF,\end{aligned} \tag{10.25}$$

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

$$\eta = \underbrace{\frac{\int_0^{\lambda_g} \frac{hc}{\lambda}\Phi_{ph,\lambda}\,d\lambda}{\int_0^{\infty} \frac{hc}{\lambda}\Phi_{ph,\lambda}\,d\lambda}}_{1} \underbrace{\frac{E_g \int_0^{\lambda_g} \Phi_{ph,\lambda}\,d\lambda}{\int_0^{\lambda_g} \frac{hc}{\lambda}\Phi_{ph,\lambda}\,d\lambda}}_{2} \underbrace{(1-R^*)}_{3} \underbrace{\mathrm{IQE}^*_{op}\eta^*_g}_{4} \underbrace{\mathrm{IQE}^*_{el}}_{5} \underbrace{\frac{A_f}{A_{tot}}}_{6} \underbrace{\frac{eV_{oc}}{E_g}}_{7} \underbrace{FF}_{8}. \tag{10.26}$$

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

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

photons,

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

10.4 Design Rules for Solar Cells

Now, as we extensively have discussed all the factors that limit the efficiency of a solar cell we are able to distill three *design rules*. We will use this design rules in Part III when we discuss different PV technologies. The three design rules are

1. Utilisation of the band gap energy,
2. Spectral utilisation,
3. Light trapping.

We now will take a closer look at each of these design rules.

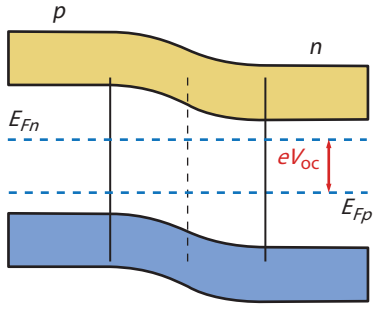

Figure 10.8: The open circuit voltage of a solar cell is determined by the splitting of the quasi-Fermi levels.

10.4.1 Bandgap utilisation

As we have seen earlier in this chapter, the open circuit voltage V_{oc} is always below the voltage V_g corresponding to the bandgap, which we characterised with the bandgap utilisation efficiency η_V. The open circuit voltage is determined by the extent to which the quasi-Fermi levels are able to split, which is limited by the charge carrier recombination mechanisms. We will discuss that various PV materials have different recombination mechanism that limit the utilisation of the band gap energy.

Let us now discuss the first design rule, *bandgap utilisation*. Figure 10.8 shows a *p*-*n*-junction, with the *p*-doped region on the left and the *n*-doped region right. Fur-

ther the quasi-Fermi levels are depicted. The extent of splitting between the quasi-Fermi levels determines the open circuit voltage.

The open circuit voltage, for example expressed in Eq. (10.17) also can be expressed in terms of the generation rate G_L, the life time τ_0 of the minority charge carriers and the intrinsic density of the charge carriers n_i in the semiconductor material,

$$V_{oc} = \frac{2kT}{e} \ln\left(\frac{G_L \tau_0}{n_i}\right). \tag{10.27}$$

The derivation of this equation is out of the scope of this book.

Let us now take a closer look on Eq. (10.27). If we increase the irradiance, or in other words, the generation rate of charge carriers, the open circuit voltage is increased. This is a welcome effect which is utilised in concentrator photovoltaics, which we will briefly discuss in Section 13.2. Secondly, we see that the lifetime τ_0 plays an important role. The larger the lifetime of the minority charge carrier, the larger the open circuit voltage can be. Or in other words, the longer the lifetime, the larger the possible splitting between the quasi-Fermi levels and the larger the fraction of the bandgap energy that can be utilised.

The lifetime of the minority charge carrier is determined by the recombination rate. As discussed in Chapter 7, we have to consider three different recombination mechanisms: radiative, Shockley-Read-Hall, and Auger recombination. While radiative and Auger recombination depend on the semiconductor itself, SRH recombination is proportional to the density of traps or impurities in the semiconductor. In the three recombination mechanisms energy and momentum are transferred from charge carriers to phonons or photons.

The efficiency of the different recombination processes depends on the nature of the band gap of the used semiconductor material used. We distinguish between *direct* and *indirect* bandgap semiconductors. Crystalline silicon is an indirect band gap material. The radiative recombination in an indirect band gap material is inefficient and recombination will be dominated by the Auger mechanism. For direct band gap materials such as GaAs under moderate illumination conditions, radiative recombination will be the dominant loss mechanism of charge carriers. For very high illumination conditions, Auger recombination starts to play a role as well.

To summarise, we find that in the defect rich solar cells, the open circuit voltage is limited by the SRH recombination. In low-defect solar cells based on indirect band gap materials, the open circuit voltage is limited by Auger recombination. In low-defect solar cells based on direct band gap materials, the open circuit voltage is limited by radiative recombination.

Besides band gap utilisation, it also is important to discuss the relationship between the maximum thickness for the absorber layer of a solar cell and the dominant recombination mechanism. As we have seen in Chapter 7, the recombination mechanism also affects the diffusion length of the minority charge carrier. The diffusion length L_n of minority electrons is given by

$$L_n = \sqrt{D_n \tau_n}, \tag{10.28}$$

where D_n is the diffusion coefficient and τ is the lifetime of the minority charge carrier. Similarly, we can formulate the diffusion length for minority holes, L_p.

It is important to realise that the thickness of the absorber layer should not exceed the diffusion length. To understand this requirement, we consider photons that would penetrate far into the absorber layer being absorbed. We want these charge carriers to be separated at the *p-n* junction or at the back contact. If the distance of these charge carriers from the *p-n* junction or the back contact is exceeding the diffusion length, the excited charge carriers will recombine within the typical diffusion length before arriving at the *p-n* junction or back contact. In other words, their lifetime is too short. This means that all charge carriers generated at a distance larger than the diffusion length from the *p-n* junction or the back contact cannot be collected and hence are lost. If the charge carriers are generated with diffusion length, they can be collected. This means that the diffusion length of the minority charge carrier, limits the maximum thickness of the solar cell.

To summarise this section, the open circuit voltage is limited by the dominant recombination mechanism. The dominance of radiative, Auger or Shockley-Reed-Hall recombination depends on the type of semiconductor materials used in the solar cell and the illumination conditions. We will discuss several different cases on Part III on PV technology.

10.4.2 Spectral utilisation

The spectral utilisation is mainly determined by the choice of materials from which the solar cell is made of. As we have seen in Section 10.2, and mainly Eq. (10.13), the photocurrent is determined by the bandgap of the material. For a bandgap of 0.62 eV corresponding to a wavelength of 2000 nm, we could theoretically generate a short circuit current density of 62 mA/cm^2. If we consider c-Si, having a band gap of 1.12 eV (1107 nm), we arrive at a theoretical current density of 44 mA/cm^2.

The optimal bandgap for single-junction solar cells is determined by the Shockley-Queisser limit, as illustrated in Fig. 10.6. For single junction solar cells, semiconductor material as such silicon, gallium arsenide and cadmium telluride have a band gap close to the optimum.

In Part III we will discuss various concepts that allow to surpass the Shockley-Queisser limit. Here, we will briefly discuss the concept of *multijunction solar cells*. In this devices, solar cells with different bandgaps are stacked on top of each other. As illustrated in 10.9, the excess energy can be reduced significantly, and the spectral utilisation will improve.

10.4.3 Light trapping

The third and last design rule that we discuss is *light trapping*. In an ideal solar cell, all light that is incident on the solar cell should be absorbed in the absorber layer. As we have discussed in Section 4.4, the intensity of light decreases exponentially as it travels through an absorptive medium. This is described by the Lambert-Beer law that we formulated in Eq. (4.25),

$$I(d) = I_0 \exp(-\alpha d). \tag{10.29}$$

From the Lambert-Beer law it follows that at the side, at which the light is entering the film, more light is absorbed in reference to the back side. The total fraction of the incident light absorbed in the material is equal to the light intensity entering the absorber layer minus the intensity transmitted through the absorber layer,

$$I^{\text{abs}}(d) = I_0[1 - \exp(-\alpha d)]. \tag{10.30}$$

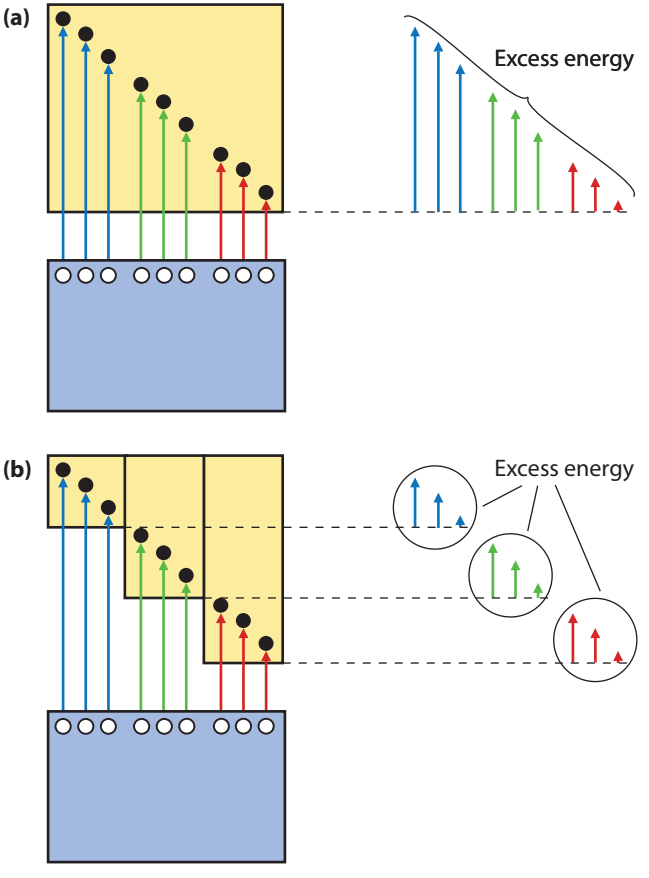

Figure 10.9: Illustrating the lost excess energy in (a) a single-junction and (b) a multi-junction solar cells.

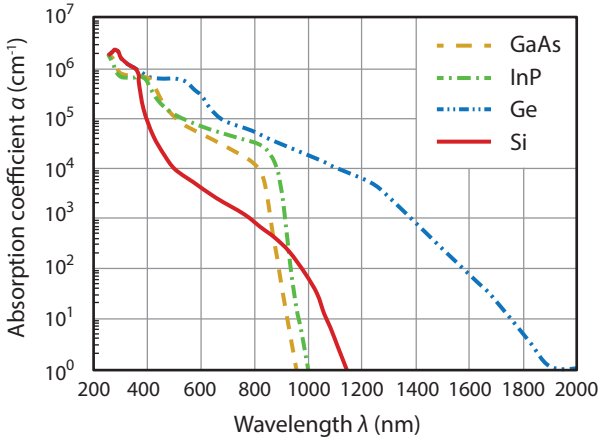

Figure 10.10: Absorption coefficients of different semiconductors.

Ideally, we would like to absorb a solar cell 100% of the incident light. Such an absorber is called *optically thick* and has a transmissivity very close to 0. As we can see from the Lambert-Beer law, this can be achieved by either absorbers with a large thickness d or with very large absorption coefficients α.

Figure 10.10 shows the absorption coefficients for four different semiconductor materials: germanium (Ge), silicon (Si), gallium arsenide (GaAs) and indium phosphide (InP). We notice that germanium has the lowest band gap. It starts to absorb at long wavelengths, which corresponds to a low photon energy. GaAs has the highest band gap, as it starts to absorb light at the smallest wavelength, or highest photon energy. Secondly, if we focus on the visible spectral part from 300 nm up 700 nm, we see that the absorption coefficients of InP and GaAs are significantly higher than for silicon. This is related to the fact that InP and GaAs are direct band gap materials as discussed earlier. Materials with an indirect band gap have smaller absorption coefficients. Only in the very blue and ultraviolet part below 400 nm, Si has a direct band gap transition. Silicon is a relatively poor absorber. Therefore for the same fraction of light thicker absorber layers are required in comparison to GaAs.

In general, for all semiconductor materials the absorption coefficient in the blue is orders of magnitude higher than in the red. Therefore the penetration depth of blue light into the absorber layer is rather small. In crystalline silicon, the blue light is already fully absorbed within a few nanometers. The red light requires an absorption path length of 60 µm to be fully absorbed. The infrared light is hardly absorbed, and after an optical path length of 100 µm only about 10% of the light intensity is absorbed.

As the absorption of photons generates excited charge carriers, the wavelength dependence of the absorption coefficient determines the local generation profile of the charge carriers. At the front side where the light enters

the absorbing film, the generation of charge carriers is significantly higher than at the back side. It follows that the EQE values measured in the blue correspond to charge carriers generated close to the front of the solar cell, whereas the EQE in the red part represents charge carriers generated throughout the entire absorber layer.

Further, it is important to reduce the optical loss mechanisms such as *shading losses*, *reflection*, and *parasitic absorption* that we already discussed in Section 10.3.1.

For reducing the reflection, anti-reflective coatings (ARC) can be used. Light that is impinging onto a surface between two media with different refractive indices will always be partly reflected and partly transmitted. In order to reduce losses, it is important to minimise these reflective losses.

The first method is based on a clever utilisation of the Fresnel equations that we discussed in Section 4.3. For understanding how this can work, we first will take a look at interfaces with silicon, the most common used material for solar cells. Let us consider light of 500 nm wavelengths falling onto an air-silicon interface perpendicularly. At 500 nm, the refractive index of air is $n_0 = 1$ and that of silicon is $n_s = 4.3$. With the Fresnel equations we hence find that the optical losses due to reflection are significant with 38.8%.

The reflection can be significantly reduced by introducing an interlayer with a refractive index n_1 with a

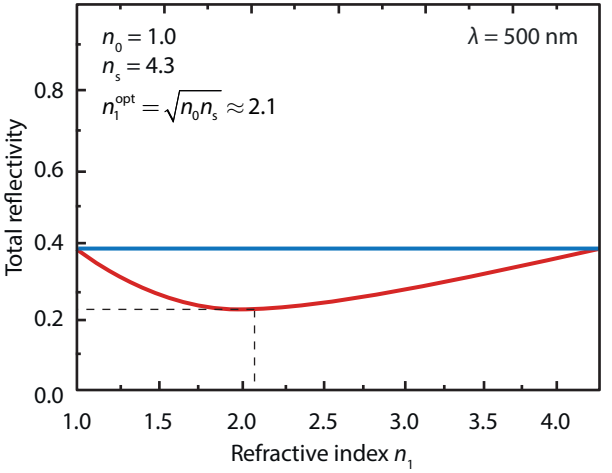

Figure 10.11: Illustrating the effect of an interlayer with refractive index n_1 in between n_0 and n_s on the reflectivity.

value in between that of n_0 and n_s. If no multiple reflection or interference is taken into account, it can easily be shown that the reflectivity becomes minimal if

$$n_1 = \sqrt{n_0 n_s}. \tag{10.31}$$

This is also seen in Fig. (10.11), where n_1 takes all the values in between n_0 and n_s. In this example, including a single interlayer can reduce the reflection at the interface from 38.8% down to 22.9%. If more than one interlayers are used, the reflection can be reduced even further. This technique is called *refractive index grading*.

In another approach constructive and destructive interference of light is utilised. In Chapter 4 we already discussed that light can be considered as an electromagnetic wave. Waves have the interesting properties that they can interfere with each-other, they can be *superimposed*. For understanding this we look at two waves A and B that have the same wavelength and cover the same portion of space,

$$A(x,\ t) = A_0 e^{ikx - i\omega t}, \tag{10.32a}$$
$$B(x,\ t) = B_0 e^{ikx - i\omega t + i\phi}. \tag{10.32b}$$

The letter ϕ denotes the *phase shift* between A and B and is very important. We can superimpose the two waves by simply adding them

$$\begin{aligned} C(x,\ t) &= A(x,\ t) + B(x,\ t) \\ &= A_0 e^{ikx - i\omega t} + B_0 e^{ikx - i\omega t + i\phi} \\ &= A_0 e^{ikx - i\omega t} + B_0 e^{ikx - i\omega t} e^{i\phi} \\ &= \left(A_0 + B_0 e^{i\phi}\right) e^{ikx - i\omega t}. \end{aligned} \tag{10.33}$$

The amplitude of the superimposed wave is thus

$$C_0 = A_0 + B_0 e^{i\phi}. \tag{10.34}$$

Depending of the phase shift, the superimposed wave will be stronger or weaker than A and B. If ϕ is *in phase*, *i.e.* a multiple of 2π, *i.e.* $0, 2\pi, 4\pi, \ldots$, we will have maximal amplification of waves,

$$C_0 = A_0 + B_0 \cdot 1 = A_0 + B_0. \tag{10.35}$$

This situation is called *constructive interference*.

But if ϕ is *in antiphase*, *i.e.* from the set $\pi, 3\pi, 5\pi, \ldots$ we have maximal attenuation of the waves, or *destructive interference*,

$$C_0 = A_0 + B_0 \cdot (-1) = A_0 - B_0. \tag{10.36}$$

Based on this principle, we can design an anti-reflection coating, as illustrated in Fig. 10.12. The green wave shows the reflection back from the first interface and the red wave shows the wave which is reflected back

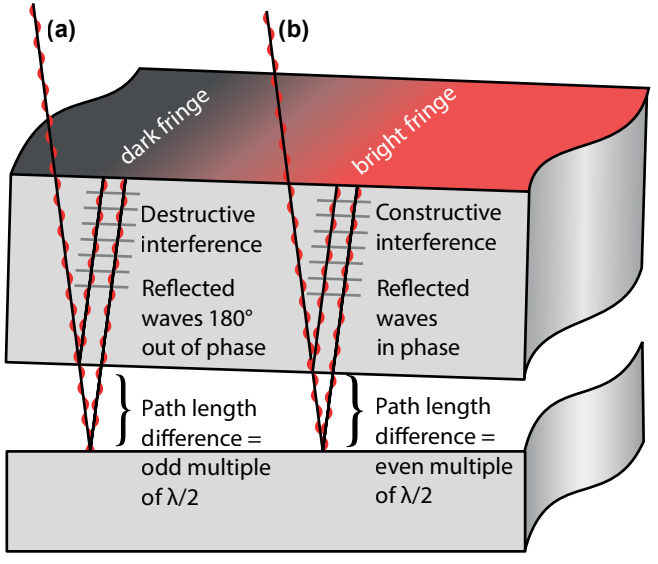

Figure 10.12: Illustrating the working principle of an anti-reflective coating based on interference [34].

from the second interface. If we look at the two waves coupled out of this system, they appear to be in anti-phase. As a result the total amplitude of the electric field of the outgoing wave is smaller and hence the total irradiance coupled out of the system is smaller as well.

It can be shown easily that the have antiphase, when the product of the refractive index and thickness of the interlayer is equal to the wavelength divided by four,

$$n_1 d = \frac{\lambda_0}{4}. \tag{10.37}$$

Using an antireflection coating based on interference demands that the typical length scale of the interlayer thickness must be in the order of the wavelength.

The last approach that we discuss for realising anti-reflective coating is using *textured interfaces*. Here, we consider the case where the typical length scales of the surface features are larger than the typical wavelength of light. In this case, which is also called the *geometrical limit*, the reflection and transmission of the light rays are fully determined by the Fresnel equations and Snell's law.

The texturing helps to enhance the coupling of light into the layer. For example, for light that is perpendicular incident, light that is reflected at one part of the textured surface can be reflected into angles in which the trajectory of the light ray is incident a second time

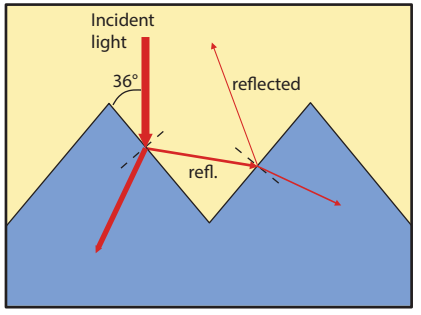

Figure 10.13: Illustrating the effect of texturing.

somewhere else on the interfaces, as illustrated in Fig. 10.13. Here, another fraction of the light will be transmitted into the layer and effectively less light will be reflected, when compared to a flat interface in-between the same materials.

In summary, we have discussed three types of anti-reflective coatings: Rayleigh films with intermediate refractive indexes, anti-reflection coatings based on destructive interference and enhanced in-coupling of light due to scattering at textured interfaces.

At the end of this chapter, we want to discuss something that is very important for thin-film solar cells. Because of reducing production cost but also because of reducing bulk recombination it is desirable to have the absorber layer as thin as possible. On the other hand, it should be *optically thick* in order to absorb as much light as possible. In principle, if the light can be reflected back and force inside the absorber until everything is absorbed. However, at every internal reflection part of the light is transmitted out of the film.

But if the light would travel through the layer at an angle larger than the critical angles of the front and back interfaces of the absorber, it could stay there until everything is absorbed without any loss. Unfortunately, it is very difficult if not impossible to design such a solar cell.

Note that textured interfaces do not help in this case. For the same reason more light can be coupled into the absorber using such a texturisation, also more light is coupled out of the absorber!

Part III

PV Technology

11

A Little History of Solar Cells

We will start our discussion on PV technology with a brief summary of the *history of solar energy* in general and of photovoltaics in particular. Already in the seventh century BCE, humans used magnifying glasses to concentrate sunlight and hence to make fire. Later, the ancient Greeks and Romans used concentrating mirrors for the same purpose.

In the 18th century the Swiss physicist Horace-Bénédict de Saussure build heat traps, which are a kind of miniature green houses. He constructed hot boxes, consisting of a glass box, within another bigger glass box, with a total number of up to five boxes. When exposed to direct solar irradiation, the temperature in the innermost box could rise up to values of 108°C; warm enough to boil water and cook food. These boxes can be considered as the World's first solar collectors.

In 1839, the French phycisist Alexandre-Edmond Becquerel, discovered the photovoltaic effect at an age of only 19 years. He observed this effect in an electrolytic cell, which was made out of two platinum electrodes, placed in an *electrolyte*. An electrolyte is an electrically conducting solution; Becquerel used silver chloride dissolved in an acidic solution. Becquerel observed that the current of the cell was enhanced when his setup was irradiated with sunlight.

In the 1860s and 1870s, the French inventor Augustin Mouchot developed solar powered steam engines using the World's first *parabolic trough* solar collector, that we will discuss in Chapter 20. Mouchot's motivation was

his believe that the coal resources were limited. At that time, coal was *the* energy source for driving steam enginges. However, as coal became cheaper, the French government decided that solar energy was too expensive and stopped funding Mouchet's research.

In 1876, the British natural philosopher William Grylls Adams together with his student Richard Evans Day demonstrated the photovoltaic effect in a junction based on platinum and the semiconductor selenium, however with a very poor performance. Seven years later, the American inventor Charles Fritts managed to make a PV-device based on a gold-selenium junction. The energy conversion efficiency of that device was 1%.

In 1887, the German physicist Heinrich Hertz discovered the photo-electric effect, already briefly mentioned in Chapter 3. In this effect, electrons are emitted from a material that has absorbed light with a wavelength shorter than a material-dependent threshold frequency. In 1905 Albert Einstein published a paper in which he explained the photoelectric effect with assuming that light energy is being carried with quantised packages of energy [19], which we nowadays call *photons*.

In 1918 the Polish chemist Jan Czochralski invented a method to grow high-quality crystalline materials. This technique nowadays is very important for growing monocrystalline silicon used for high-quality silicon solar cells that we will study in detail in Chapter 12.

The development of the c-Si technology started in the second half of the 20th century.

In 1953, the American chemist Dan Trivich was the first one to perform theoretical calculations on the solar cell performance for materials with different bandgaps.

The real development of solar cells as we know them today, started at the *Bell Laboratories* in the United states. In 1954, their scientists Daryl M. Chapin, Calvin S. Fuller, and Gerald L. Pearson, made a silicon-based solar cell with an efficiency of about 6% [35]. In the same year, Reynolds *et al.* reported on the photovoltaic effect for cadmium sulfide (CdS), a II-VI semiconductor [36].

In the mid and late 1950s several companies and laboratories started to develop silicon-based solar cells in order to power satellites orbiting the Earth. Among these were RCA Corporation, Hoffman Electronics Corporation but also the Unites States Army Signal Corps. In these days, research on PV technology was mainly driven by supplying space applications with energy. For example, the American satellite *Vanguard 1*, which was launched by the U.S. Navy in 1958, was powered by solar cells from Hoffman Electronics. It was the fourth artificial Earth satellite and the first one to be powered with solar cells. It was operating until 1964 and still is orbiting Earth. In 1962 Bell Telephone Laboratories launched the first solar powered telecommunications satellite and in 1966 NASA launched the first Or-

biting Astronomical Observatory, which was powered by a 1 kW photovoltaic solar array.

In 1968, the Italien scientist Giovanni Francia built the first concentrated-solar power plant near Genoa, Italy. The plant was able to produce 1 MW with superheated steam at 100 bar and 500°C.

In 1970, the Soviet physicist Zhores Alferov developed solar cells based on a gallium arsenide heterojunction. This was the first demonstrator of a solar cell based on III-V semiconductor materials that we will discuss in Section 13.2. In 1976, Dave E. Carlson and Chris R. Wronski developed the first thin-film photovoltaic devices based on amorphous silicon at RCA Laboratories. We will discuss this technology in Section 13.3. In 1978, the Japanese companies SHARP and Tokyo Electronic Application Laboratory bring the first solar powered calculators on the market.

Because of the oil crisis, induced by the OPEC oil embargo in 1973 and as a consequence a sharply rising oil price, the public interest in photovoltaic technology for terrestrial application was increasing in the 1970s. In that time, PV technology moved from a niche technology for space application to a technology applicable for terrestrial applications. In the late 1970s and 1980s many companies started to develop PV modules and system for terrestrial applications.

In 1980 the first thin film solar cells based on a copper-sulfide/cadmium-sulfide junction was demonstrated with a conversion efficiency above 10% at the University of Delaware. In 1985, crystalline silicon solar cells with efficiencies above 20% were demonstrated at the University of New South Wales in Australia.

From 1984 through 1991 the World's largest solar thermal energy generating facility in the world was built in the Mojave Desert in California. It consits of 9 plants with a combined capacity of 354 Megawatts.

In 1991 the first high efficiency Dye-sensitized solar cell was published by the École polytechnique fédérale de Lausanne in Switzerland by Michael Grätzel and coworkers. The Dye-sensitized solar cell is a kind of photo-electrochemical system, in which a semiconductor material based on molecular sensitizers, is placed between a photoanode and an electrolyte. We will introduce this technology in Section 13.6.

In 1994, the U.S. *National Renewable Energy Laboratory* placed in Golden, Colorado, demonstrated a concentrator solar cell based on III-V semiconductor materials. Their cell based on a indium-gallium-phosphide/gallium-arsenide tandem junction exceeded the 30% conversion limit.

In 1999, the total global installed photovoltaic power passed 1 GW_p. Starting from about 2000, environmental issues and economic issues started to become more and more important in the public discussion, which re-

newed the public interest in solar energy. Since 2000, the PV market therefore transformed from a regional market to a global market, as discussed in Chapter 2. Germany took the lead with a progressive feed-in tariff policy, leading to a large national solar market and industry [15].

Since about 2008, the Chinese government has been heavily investing in their PV industry. As a result, China has been the dominant PV module manufacturer for several years now. In 2012 the world-wide solar energy capacity surpassed the magic barrier of 100 GW_p [13]. Between 1999 and 2012, the installed PV capacity hence has grown with a factor 100. In other words, in the last 13 years, the average annual growth of the installed PV capacity was about 40%.

12 Crystalline Silicon Solar Cells

As we already discussed in Chapter 6, most semiconductor materials have a *crystalline lattice* structure. As a starting point for our discussion on crystalline silicon PV technology, we will take a closer look at some properties of the crystal lattice.

12.1 Crystalline silicon

In such a lattice, the atoms are arranged in a certain pattern that repeats itself. The lattice thus has *long-range order* and *symmetry*. However, the pattern is not the same in every direction. In other words, if we make a large cut through the lattice, the various planes you can make do not look the same.

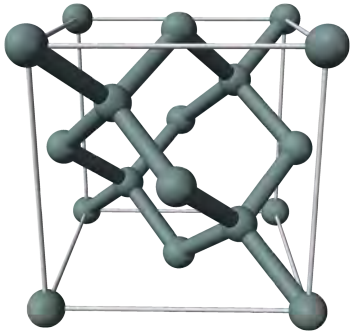

Figure 12.1: A unit cell of a diamond cubic crystal [25].

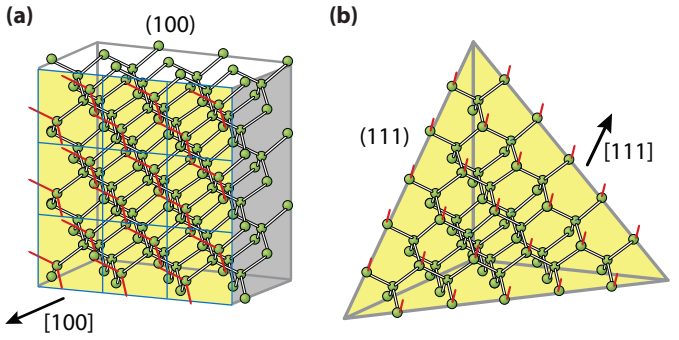

Figure 12.2: The (a) 100 and the (b) 111 surfaces of a silicon crystal.

Crystalline silicon has a density of 2.3290 g/cm^3 and a *diamond cubic crystal structure* with a lattice constant of 543.07 pm, as illustrated in Fig. 12.1. Figure 12.2 shows two different sections through a crystalline silicon lattice, which originally consisted out of three by three by three unit cells. The first surface shown in Fig. 12.2 (a) is the *100 surface*, whose surface normal is the 100 direction. At a 100 surface, each Si atom has two back bonds and two valence electrons pointing to the front. The second surface, shown in Fig. 12.2 (b), is the *111 surface*. Here, every Si atom has three back bonds and one valence electron pointing towards the plane normal.

To understand the importance of these directions, we look at the *electronic band dispersion diagram* for silicon,

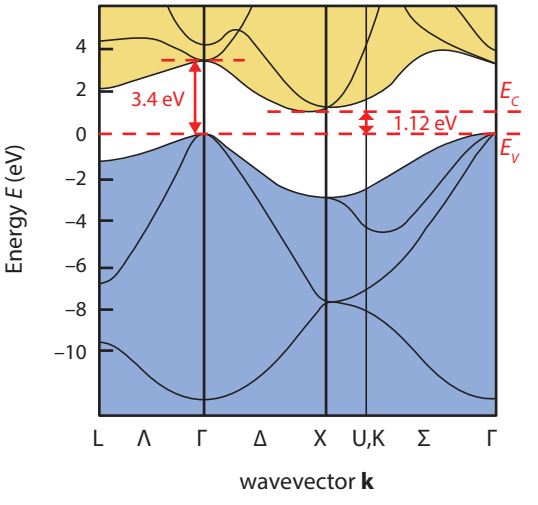

Figure 12.3: The band diagram of crystalline silicon.

shown in Fig. 12.3. On the vertical axis, the energy position of the valence and conduction bands is shown. The horizontal axis shows the crystal momentum, *i.e.* the momentum of the charge carriers. The white area represents the energy levels in the forbidden band gap. The band gap of silicon is determined by the lowest energy point of the conduction band at X, which corresponds to the 100 direction, and the highest energy value of the valence band, at Γ. The *band gap energy* is the difference between those two levels and is equal 1.12 eV, or 1107 nm, when expressed in wavelengths. 1107 nm is in the infrared part of the spectrum of light. This bandgap is an *indirect bandgap*, because the charge carriers must change in energy *and* momentum to be excited from the valence to the conduction band. As we can see, crystalline silicon has a direct transition as well. This transition has an energy of 3.4 eV, which is equivalent to a wavelength of 364 nm, which is in the blue spectral part.

Because of the required change in momentum, for for an indirect band gap material it is less likely that a photon with an energy exceeding the bandgap can excite the electron, with respect to a direct bandgap material like gallium arsenide (GaAs) or indium phosphite (InP). Consequently, the absorption coefficient of crystalline silicon is significantly lower than that of direct band gap materials, as we can see in Fig. 12.4. While in the visible part of the spectrum c-Si absorbs less than the GaAs and InP, below 364 nm, it absorbs

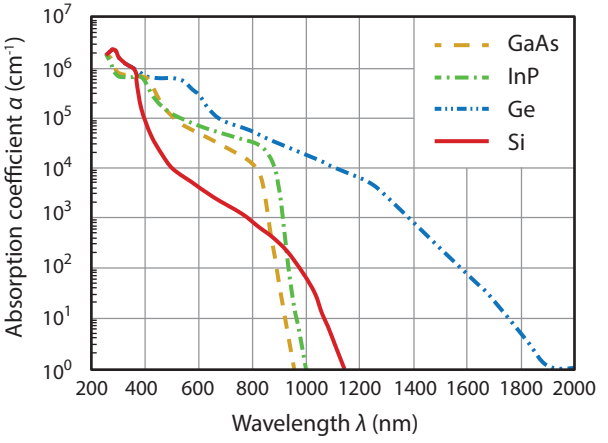

Figure 12.4: Absorption coefficients of different semiconductors.

just as much as GaAs and InP, because there silicon has a direct band to band transition as well. Germanium (Ge), also indicated in the figure, is an indirect bandgap material, just like silicon. The bandgap of Ge is 0.67 eV, which means it already starts to absorb light at wavelengths shorter than 1850 nm. In the visible part of the spectrum, germanium has some direct transitions as well.

Let us now take another look at the design rules for solar cells that we introduced in Section 10.4. First, we look at *spectral utilisation.* The c-Si band gap of 1.12 eV means that in theory we can generate a maximum short circuit current density of 45 mA per square centimeter. Let us now consider the second design rule, *i.e. light trapping*. First, we look at a wavelength around 800 nm, where c-Si has an absorption coefficient of 1000 cm^{-1}. Using the Lambert-Beer law [Eq. (4.25)], we easily can calculate that for realising an an absorption of 90% of the incident light at 800 nm, the required absorption path length is 23 µm. Secondly, we look at 970 nm wavelength, where c-Si has an absorption coefficient of 100 cm^{-1}. Hence, an absorption path length of 230 µm is required to absorb 90% of the light. 230 µm is a typical thickness for silicon wafers. This calculation demonstrates that the light trapping techniques become important for crystalline silicon absorber layers above a wavelength of about 900 nm.

Let us now consider the design rule of *bandgap utilisation*, which is determined by the recombination losses. As silicon is an indirect band gap material, only Auger recombination and Shockley-Read-Hall (SRH) recombination will determine the open circuit voltage, while radiative recombination can be neglected. Considering SRH recombination, the recombination rate of the charge carriers is related to the electrons trapped at defect states. When looking at the defect density in the bulk of silicon, we can differentiate between two major types of silicon wafers: *monocrystalline silicon* and *multicrystalline silicon*, which is also called *polycrystalline silicon*.

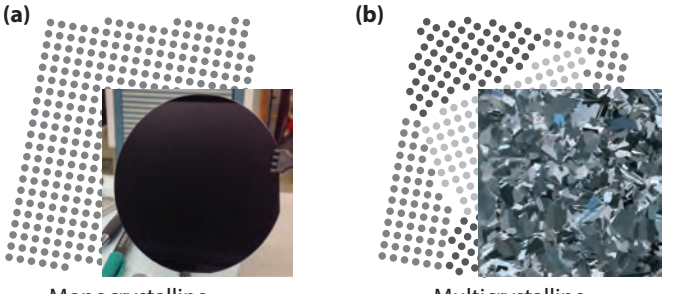

Figure 12.5: Illustrating a (a) monocrystalline and a (b) multicrystalline silicon wafer.

Monocrystalline silicon, also known as called single-crystalline silicon, is a crystalline solid, in which the crystal lattice is continuous and unbroken without any

grain boundaries over the entire bulk, up to the edges. In contrast, polycrystalline silicon, often simply abbreviated with polysilicon, is a material that consists of many small crystalline grains, with random orientations. Between these grains grain boundaries are present. Figure 12.5 shows two pictures of monocrystalline and multicrystalline wafers. While a monocrystalline silicon wafer has one uniform color, in multicrystalline silicon, the various grains are clearly visible for the human eye. At the grain boundaries we find lattice mismatches, resulting in many defects at these boundaries. As a consequence, the charge carrier lifetime for polycrystalline silicon is shorter than for monocrystalline silicon, because of the SRH recombination. The more grain boundaries in the material, the shorter the lifetime of the charge carriers. Hence, the grain size plays an important role in the recombination rate.

Figure 12.6 shows the relationship between the open circuit voltage and the average grain size for various solar cells developed around the world, based on the multicrystalline wafers [37]. The larger the grain size, the longer the charge carrier lifetimes and the larger the band gap utilisation and hence the open circuit voltage will be. On the right hand side of the graph the open circuit voltages of various solar cells, based on monocrystalline wafers, is shown. As monocrystalline silicon has no grain boundary, much larger open circuit voltages can be obtained.

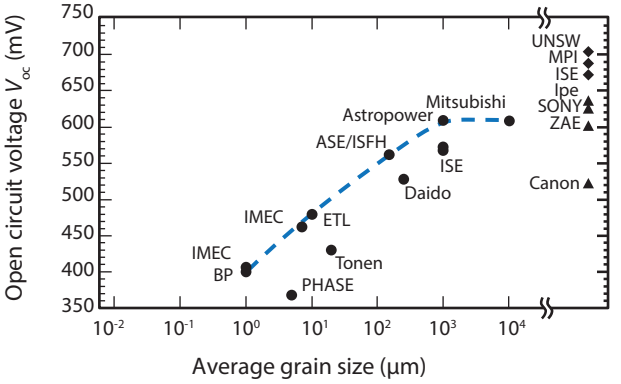

Figure 12.6: The relationship between the open circuit voltage and the average grain size [37]

12.2 Production of silicon wafers

After the initial considerations on designing c-Si solar cells, we now will discuss how monocrystalline and multicrystalline silicon wafers can be produced. In Fig. 12.7 we illustrate the production process of monocrystalline silicon wafers.

The lowest quality of silicon is the so-called metallurgical silicon, which is made from *quartzite*. Quartzite is a rock consisting of almost pure silicon dioxide (SiO_2). For producing silicon the quartzite is molten in a submerged-electrode arc furnace by heating it up

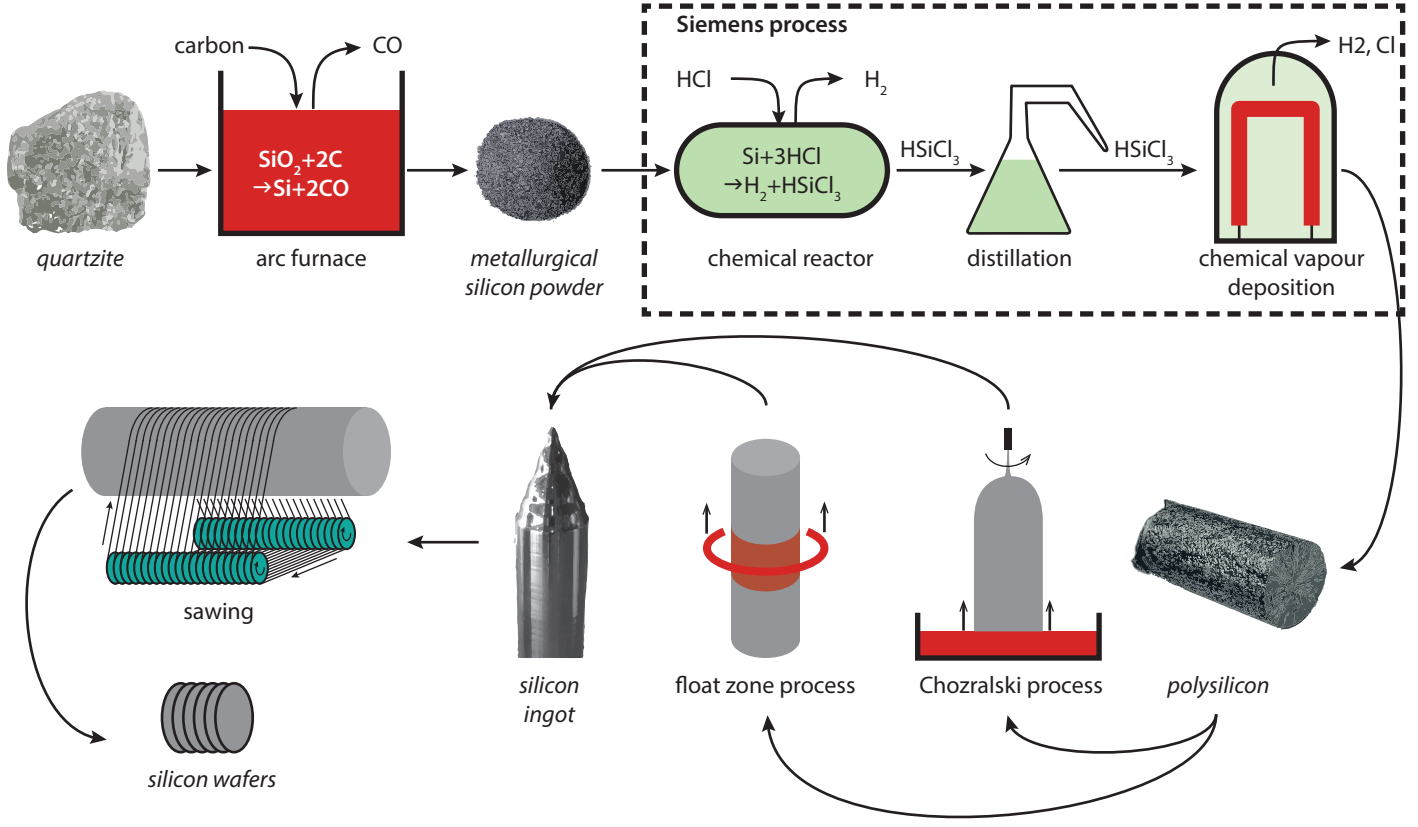

Figure 12.7: Illustrating the production process of monocrystalline silicon wafers.

to around 1900°C, as illustrated in Fig. 12.7. Then, the molten quartzite is mixed with carbon. As a carbon source, a mixture of coal, coke and wood chips is used. The carbon then starts reacting with the SiO_2. Since the reactions are rather complex, we will not discuss them in detail here. The overall reaction how ever can be written as

$$SiO_2 + 2\,C \rightarrow Si + 2\,CO \tag{12.1}$$

As a result, carbon monoxide (CO) is formed, which will leave the furnace in the gas phase. In this way, the quartzite is purified from the silicon. After the reactions are finished, the molten silicon that was created during the process is drawn off the furnace and solidified. The purity of metallurgic silicon, shown as a powder in Fig. 12.7, is around 98% to 99%.

About 70% of the worldwide produced metallurgical silicon is used in the aluminium casting industry for make aluminium silicon alloys, which are used in automotive engine blocks. Around 30% are being used for make a variety of chemical products like silicones. Only around 1% of metallurgical silicon is used as a raw product for making electronic grade silicon.

The silicon material with the next higher level of purity is called polysilicon. It is made from a powder of metallurgical silicon in the *Siemens process*. In the process, the metallurgical silicon is brought into a reactor and exposed hydrogen chloride (HCl) at elevated temperatures in presence of a catalyst. The silicon reacts with the hydrogen chloride,

$$Si + 3\,HCl \rightarrow H_2 + HSiCl_3, \tag{12.2}$$

leading to the creation of trichlorosilane ($HSiCl_3$). This is a molecule that contains one silicon atom, three chlorine atoms and one hydrogen atoms. Then, the trichlorosilane gas is cooled and liquified. Using *distillation*, impurities with boiling points higher or lower than $HSiCl_3$ are removed. The purified trichlorosilane is evaporated again in another reactor and mixed with hydrogen gas. There, the trichlorosilane is decomposed at hot rods of highly purified Si, which are at a high temperature in between around 850°C and 1050°C. The Si atoms are deposited on the rod whereas the chlorine and hydrogen atoms are desorbed from the rod surface back in to the gas phase. As a result a pure silicon material is grown. This method of depositing silicon on the rod is one example of *chemical vapour deposition* (CVD). As the exhaust gas still contains chlorosilanes and hydrogen, these gasses are recycled and used again: Chlorosilane is liquified, distilled and reused. The hydrogen is cleaned and thereafter recycled back in to the reactor. The Siemens process consumes a lot of energy.

Another method for producing poly silicon granules is using *Fluid Bed Reactors* (not shown in Fig. 12.7). This process is operated at lower temperatures and con-

sumes much less energy. Polycrystalline silicon can have an purity as high as 99.9999%, or in other words only one out of million atoms is an atom different from Si.

The last approach we briefly mention is that of *upgraded metallurgical silicon* (not shown in Fig. 12.7). In this process metallurgical silicon is chemically refined by blowing gasses through the silicon melt removing the impurities. Although processing is cheap, the silicon purity is lower than that achieved with the Siemens or the Fluid Bed reactor approaches.

Now we introduce two methods that are industrially used for making monocrystalline silicon *ingots*, *i.e.* large cylinders of silicon that consist of one crystal only. This means that inside the ingot no grain boundaries are present. Such a monocrystalline ingot and both the methods are sketched in Fig. 12.7.

The first method we discuss is based on the *Czochralski process* that was discovered by the Polish scientist Jan Czochralski in 1916. In this method, highly purified silicon is melted in a crucible at typical temperature of 1500°c. Intentionally boron or phosphorus can be added for making p-doped or n-doped silicon, respectively. A seed crystal that is mounted on rotating shaft is dipped in to the molten silicon. The orientation of this seed crystal is well defined; it has either is either 100 or 111 oriented. The melt solidifies at the seed crystal and adopts the orientation of the crystal. The crystal is rotating and pulled upwards slowly, allowing the formation of a large, single-crystal cylindrical column from the melt — the ingot. For successfully conducting the process, temperature gradients, the rate of pulling the shaft upwards and the rotational speed must be well controlled. Due to improved process control throughout the years, nowadays ingots of diameters of 200 mm or even 300 mm with lengths of up to two meters can be fabricated. To prevent the incorporation of impurities, the Czochralski this process takes place in an inert atmosphere, like argon gas. The crucible is made from quartz, which partly dissolves in the melt as well. Consequently, monocrystalline silicon made with the Czochralski method has a relatively high oxygen level.

The second method to make monocrystalline silicon is the *float zone* process, which allows fabricating ingots with extremely low densities of impurities like oxygen and carbon. As a source material, a polycrystalline rod made with the Siemens process is used. The end of the rod is heated up and melted using an induction coil operating at radio frequency (RF). The molten part is then brought in contact with the seed crystals, where it solidifies again and adopts the orientation of the seed crystal. Again 100 or 111 orientations are being used. As the molten zone is moved along the polysilicon rod, the single-crystal ingot is growing as well. Many impurities remain in and move along with the molten zone. Nowadays, during the process nitrogen is intentionally added in order to improve the control over mi-

crodefects and the mechanical strength of the wafers. One advantage of the float—zone technique is that the molten silicon is not in contact with other materials like quartz, as this is the case when using the Czochralski method. In the float-zone process the molten silicon is only in contact with the inert gas like argon. The silicon can be doped by adding doping gasses like diborane (B_2H_6) and phosphine (PH_3) to the inert gas to get *p*-doped and *n*-doped silicon, respectively. The diameter of float-zone processed ingots generally is not larger than 150 mm, as the size is limited by surface tensions during the growth.

Not only monocrystalline silicon ingots, but also *multicrystalline silicon ingots*, which consist of many small crystalline grains, can be fabricated (not shown in Fig. 12.7). This can be made by melting highly purified silicon in a dedicated crucible and pouring the molten silicon in a cubic shaped *growth crucible*. There, the molten silicon solidifies in to multi-crystalline ingot in a process called *silicon casting*. If both melting and solidification is done in the same crucible it is referred to as *directional solidification*. Such multicrystalline ingots can have a front surface area of up to 70×70 cm^2 and a height of up tp 25 cm.

Now, as we know how to produce monocrystalline and multicrystalline ingots we will discuss how to make *wafers* out of them. The process that is used to make the wafers is *sawing*, as illustrated in Fig. 12.7. Logically, sawing will damage the surface of the wafers. Therefore, this processing step is followed by an polishing step. The biggest disadvantage of sawing is that a significant fraction of the silicon is lost as *kerf loss*, which usually is determined by the thickness of the wire or saw used for sawing. Usually, it is in the order of 100 µm. As typical wafers used in modern solar cells have thicknesses in the order of 150 µm up to 200 µm, the kerf loss is very significant.

A completely different process for making silicon wafers is the *silicon ribbon* technique (not shown in Fig. Fig. 12.7). As this technique does not include any sawing step, no kerf loss occurs. In the silicon ribbon technique a string is used that is resistant agains high temperatures. This string is pulled up from a silicon melt. The silicon solidifies on the string and hence a sheet of crystalline silicon is pulled out of the melt. Then, the ribbon is cut into wafers. Before the wafers can be processed further in order to make solar cells, some surface treatments are required. The electronic quality of ribbon silicon is not as high as that of monocrystalline.

To summarise, we discussed how to make metallurgical silicon out of quartzite and how to fabricate polysilicon. We have seen that monocrystalline ingots are made using either the Czochralski or the float-zone process, while multicrystalline ingots are made using a casting method. Wafers are fabricated by sawing these ingots. A method that does not have any kerf losses in the rib-

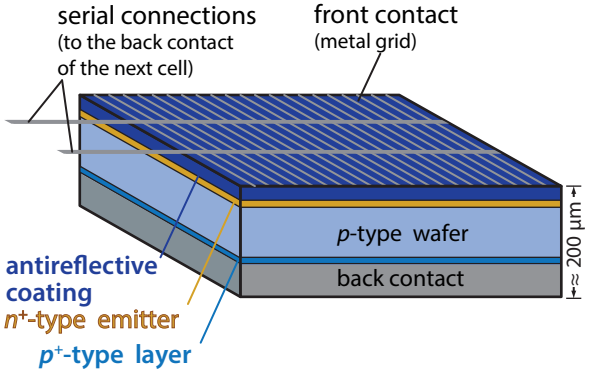

Figure 12.8: Scheme of a modern crystalline silicon cell.

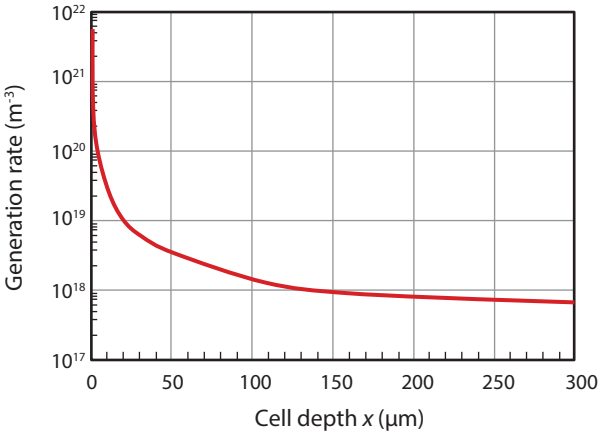

Figure 12.9: The generation profile in crystalline silicon.

bon silicon approach.

12.3 Fabricating solar cells

Before we start with the actual discussion on how c-Si solar cells are fabricated, we briefly discuss the operating principles of c-Si solar cells. Especially, we discuss several technical aspects that play an important role in the collection of the light, excitation of charge carriers and the reduction of optical losses.

In Chapter 8 we discussed how an illuminated *p*-*n* junction can operate as a solar cell. In the illustrations used there, both the *p*-doped and *n*-doped regions have the same thickness. This is not the case in real c-Si devices. For example, the most conventional type of c-Si solar cells is built from a *p*-type silicon wafer, as sketched in Fig. 12.8. However, the *n*-type layer on the top of the *p*-wafer is much thinner than the wafer; it typically has a thickness of around 1 µm. Often, this layer is called the *emitter* layer. As mentioned before, the whole wafer has typically thicknesses in between 100 and 300 µm.

Because of the Lambert-Beer law [Eq. (4.25)], the intensity if light inside the silicon bulk decays exponentially

with the depth. Hence also the generation profile will show an exponential decay as discussed in Section 7.2.1 and shown in Fig. 12.9. Therefore, the largest fraction of the light is absorbed close to the front surface of the solar cell. In the first 10 µm by far the most charge carriers are generated. By making the front emitter layer very thin, a large fraction of the light excited charge carriers generated by the incoming light are created within the diffusion length of the *p*-*n* junction.

The emitter layer can be created with a *solid state diffusion* process. In this process, the wafers are placed in a furnace at around 850°C, where the dopant atoms are present in atmosphere in the form of phosphine (PH_3). These atoms react at these high temperatures with the surface. Further, at these temperatures the phosphine atoms are mobile in the silicon crystal. Based on Fick's law, the therefore can diffuse into the wafer,

$$J(x) = -D\frac{\mathrm{d}n}{\mathrm{d}x}. \tag{12.3}$$

This law states that the particle flux J is proportional to the negative gradient of the particle density n. D is the diffusion constant. The diffusion process has to be controlled such that the dopants penetrate into the solid to establish the desired emitter thickness.

Now we are discuss the collection of charger carriers in a crystalline silicon solar cell. The crucial components that play a role in charge collection are the emitter layer, the metal front contacts and the back contact.

First, the emitter layer: At the *p*-*n* junction the minority charge carriers, which are excited by the light, are separated at the *p*-*n* junction: the minority electrons in the *p*-layer drift to the *n*-layer, where they have to be collected. Since the silicon *n*-emitter is not sufficiently conductive we have to use the much more conductive metal contacts, which are placed on top of the emitter layer. Very often, the metal contacts are made of the cheap metal aluminium.

This means that the electrons have to diffuse laterally through the emitter layer to the electric front contact to be collected. Which factors are important for a good transport of the electrons to the contact? First, the lifetime of the charge carriers needs to be high. A high lifetime guarantees large open circuit voltages, or in other words the optimal utilisation of the band gap energy. For increasing the lifetime, recombination losses must be reduced as much as possible.

Recombination not only reduces the V_{oc}, it also limits the collected current. As mentioned earlier, in silicon two recombination mechanisms are present: Shockley-Read-Hall recombination and Auger recombination. First, let us take a look at Shockley-Read-Hall recombination. A bare c-Si surface contains many defects, because the surface silicon atoms have some valence electrons that cannot make molecular orbitals due to the absence of neighbouring atoms. These valence orbitals containing only one electron at the surface act

like defects. They are also called *dangling bonds*. At the dangling bonds, the charge carriers can recombine through the SRH process. The probability and speed at which charge carriers can recombine is usually expressed in terms of the *surface recombination velocity.* Since a large fraction of the charge carriers are generated close to the front surface, a high surface recombination velocity at the emitter front surface will lead to significant charge carriers losses and consequently lower short-circuit current densities. In high-quality monocrystalline silicon wafers, for example, no defect-rich boundaries are present in the bulk. Thus, the lifetime of charge carriers is limited by the recombination processes at the wafer surface.

In order to reduce the surface recombination two approaches are used. First, the defect concentration on the surface is reduced by depositing a thin layer of a different material on top of the surface. This material partially restores the bonding environment of the silicon atoms. In addition, the material must be an insulator, it must force the electrons to remain in and move through the emitter layer. Typical materials used for this *chemical passivation* layers are silicon oxides (SiO_x) and silicon nitride (Si_xN_y).[1] A silicon oxide layer can be formed by heating up the silicon surface in an oxygen-rich atmosphere, leading to the oxidation of the surface silicon atoms. Si_3N_4 can be deposited using plasma-enhanced chemical vapour deposition (PE-CVD) that we will discuss in more detail in Chapter 13.

A second approach for reducing the surface recombination velocity is to reduce the minority charge carrier density near the surface. As the surface recombination velocity is limited by the minority charge carriers density, it is beneficial to have the minority charge carrier density at the surface as low as possible. By increasing the doping of the emitter layer, the density of the minority charge carriers can be reduced. This results in lower surface recombination velocities. However, this is in competition with the diffusion length of the minority charge carriers. The blue part of the solar spectrum leads to generation of many charge carrier very close to the surface, *i.e.* in the emitter layer. For utilising these light excited minority charge carriers, the diffusion length of the holes has to be large enough to reach the depletion zone at the *p-n* junction as indicated by the blue arrow. However, increasing the doping levels leads to a decreasing diffusion length of the minority holes in the emitter. Therefore, too high doping levels or too thick emitter layers would result in a poor blue response or – in other words – low External Quantum Efficiency (see Chapter 9) values in the blue part of the spectrum. Such an emitter layer could be called "dead layer" as the light excited minority charge carriers can not be collected.

[1] Stoichiometric silicon oxide and silicon nitride have the chemical formula SiO_2 and Si_3N_4, respectively. As the layers used for passivation are not stoichiometric, we indicate the elementary fractions with x and y.

As a next step, we take a closer look at the metal-emitter interface. Because electrons must be easily conducted from the emitter to the metal, insulating passivation layers such as SiO_x or Si_xN_y cannot be used. Therefore, the metal-semiconductor interface has more defects and hence an undesirably high interface recombination velocity. Additionally, a metal-semiconductor junction induces a barrier for the majority charge carriers. It is, however, out of the scope of this book to discuss the physics of this barrier in more detail. We just must keep in mind that this high barrier will give rise to a higher contact resistance. Again high doping levels can reduce the recombination velocity at the metal-semiconductor interface and reduce the contact resistance. In order to minimise the recombination at the interface defects as much as possible, the area of the metal-semiconductor interface must be minimised and the emitter directly below the interface should be heavily doped, which is indicated with n^{++}. The sides of the metal contacts are buried in the insulting passivation layer. The area below the contact has been heavily doped. These two approaches reduce the recombination and collection losses at the metal contact.

The solar cell shown in Fig. 12.8 has a classic metal grid pattern on top. We see two high ways for the electrons in the middle of top surface of the solar cell. They are called *busbars*. The small stripes going from the busbars to the edges of the solar cell are called the *fingers*. Let now R be the resistance of such a finger. If L, W, and H are the length, width and height of the fingers, respectively, and ρ is the resistivity of the metal, R is given by

$$R = \rho \frac{L}{WH}. \tag{12.4}$$

This equation shows that the longer the path length for an electron, the larger the resistance the electrons experience. Further, he smaller the cross-section $(W \cdot H)$ of the finger, the larger the resistance will be. Note, that the resistance of the contacts will act as a series resistance in the equivalent electric circuit. Larger series resistance will result in lower fill factors of the solar cell as discussed in Chapter 9. Hence, electrically large cross-sections of the fingers are desirable.

To arrive at a metallic contact, electrons in the emitter have to travel laterally through the emitter. Because of the resistivity of the n-type silicon, the charge carriers in the emitter layer also experience a resistance. It can be shown that the power loss due to the resistivity of the emitter layer scales with the spacing between two fingers to the power 3. EXERCISE???

As the metal contacts are at the front surface, they act as unwelcome shading objects, or in other words light incident on the metallic front contact area cannot be absorbed in the PV-active layers. Therefore, the contact area should be kept as small as possible, which is in competition with the fact that the finger cross section should be maximised. So basically, the finger

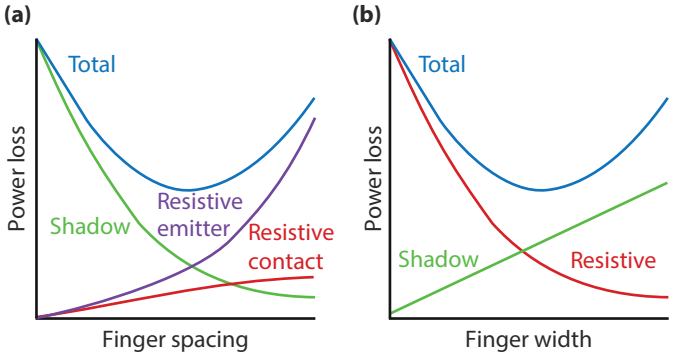

Figure 12.10: Power loss due to (a) spacing and (b) widths of the metallic fingers on top of a c-Si solar cell.

height should be as high as possible while the finger width should you would like to have a small as possible width and a high as possible height to comply with these requirements.

We can see that several effects compete with other. Fig. 12.10 (a) shows the relationship between power losses and finger spacing. With increasing finger spacing the power losses of the solar cell decreases because of less shading. On the other hand, the losses due to the increased resistivity in the emitter layer increase. Hence, there is an optimal spacing distance at that the power loss is minimal. A similar plot can be made for the power loss in dependence of the finger width; it is shown in Fig. 12.10 (b). The larger W the larger shading losses will be. But with increasing W the resistance decreases. Again, here an optimum exists at which the power losses are minimal. We see that optimising the front contact pattern is a complex interplay between the finger width and spacing.

For designing the back contact we find similar issues. Just as the electrons are to be collected in the front n-type layer, the holes are to be collected at the back contact. Electrons are the only charge carriers that exist in metal. Therefore the holes have to recombine with the electrons at the back semiconductor-metal interface. If the distance between the p-n interface and the back contact is smaller than the typical diffusion length of the minority electrons, the minority electrons can be lost at the defects of the back contact interface because of SRH recombination

Several methods can be used to reduce this loss. First, the area between the metal contact and the semiconductor can be reduced, just as for the front contact. To do this, point contacts can be used, while the rest of the rear surface is passivated by an insulating passivation layer, similar as we already discussed for the emitter front surface. The recombination loss of electrons at the back contact can be further reduced by a introducing a *back surface field*. Above the point contacts a highly doped p-doped region is placed, which is indicated by p^+.

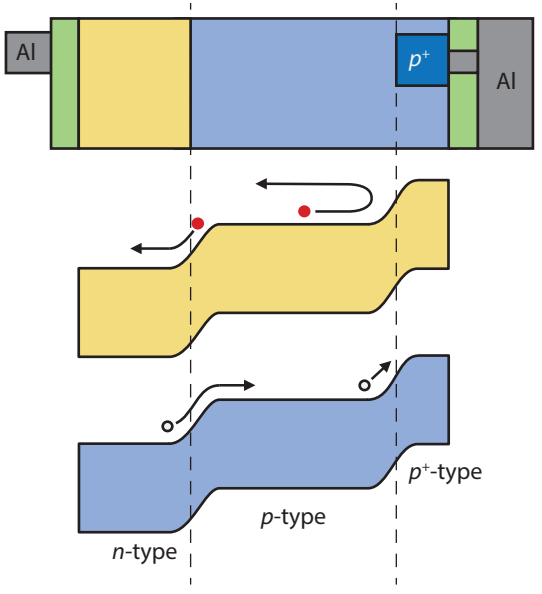

Figure 12.11: Effect of the back surface field illustrated in a band diagram.

To understand how the back surface field works, we take a look at the band diagram shown in Fig. 12.11. The interface between the normally doped p-region and the highly doped p^+-region acts like an n-p junction. Here, this junctions acts as a barrier that prevents minority electrons in the p-region from diffusing to the back surface. Further, the space charged field behaves like a passivation layer for the defects at the back contact interface and allows to have higher minority carrier densities in the p-doped bulk.

After this thorough discussion on managing the charge carriers, we now take a closer look on managing the photons in a crystalline silicon solar cell. Several optical loss mechanisms must be addressed. These are shading, reflection losses, parasitic absorption losses in the non-PV active layers, and transmission through the back of the solar cell. As already mentioned, *shading losses* are caused by the metallic front contact grid.

Secondly, *reflection* from the front surface is an important loss mechanisms. We briefly mention two approaches to design *anti-reflective coatings* (ARC) for reducing these losses. First, reflection can be minimised based on the *Rayleigh film* principle: by putting a film with a refractive index smaller than that of silicon wafer between the cell and the wafer losses can be reduced. The optimal value for the refractive index of the intermediate layer equals the square root of the product

of refractive indexes of the two other media,

$$n_{\text{opt}} = \sqrt{n_{\text{air}} n_{\text{Si}}}. \tag{12.5}$$

At a wavelength of 500 nm, the optimal refractive index for a layer in-between air and silicon is 2.1. Note that in practice a solar cell in encapsulated under a glass or polymer plate, which will have a beneficial effect on the refractive index grading as well, reducing the reflection losses further. Secondly, using the concept of destructive interference, the thickness and refractive index of an antireflection coating can be chosen such that in a certain wavelength range the reflection is minimised. This happens when the light reflected from the air-ARC interface is in anti-phase with the light reflected from the ARC-Si interface, as discussed in Chapter 4. We find that the thickness of such a layer should be *a quarter of the* wavelength in the layer,

$$d_{\text{ARC}} = \frac{\lambda_0}{4n}, \tag{12.6}$$

where λ_0 denotes the wavelength *in vacuo* and n is the refractive index of the antireflective layer. For a refractive index of 2.1, a layer thickness of 60 nm would lead to destructive interference at 500 nm. As mentioned earlier, a typical material for passivation is silicon nitride. we have discussed earlier, one of the typical passivation layers of standard crystalline silicon is silicon nitride. Fig. 12.12 (a) shows a multicrystalline wafer without any ARC. It appears silverish, which means

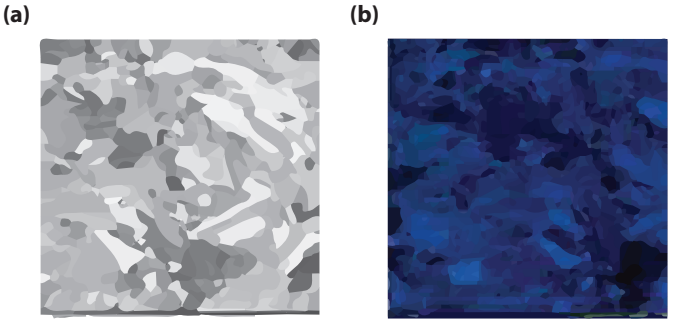

Figure 12.12: A multicrystalline silicon wafer (a) without and (b) with an antireflective coating from silicon nitride.

that it is highly reflective. In Fig. 12.12 (b) a similar wafer is shown after it was passivated with Si_xN_y. We see that it has a dark-blue appearance, hence its reflection is much lower. Interestingly enough, the refractive index of Si_xN_y at 500 nm is in the range of 2 to 2.2, close to the optimum mentioned earlier. The blue appearance indicates that reflection in the blue is stronger than at other wavelengths.

The reflection losses also can be minimised by texturing the wafer surface. Light that is reflected at the textured surface can be reflected at angles in which the trajectory of the light ray is incident somewhere else on the surface, where it still can be coupled into the silicon. Additionally, the scattering at textured surfaces will couple the light under angles different from the

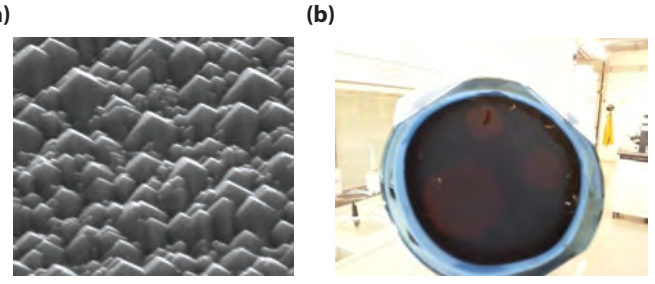

Figure 12.13: (a) A *textured* multicrystalline Si wafer and (b) a wafer with a *black silicon* surface.

interface normal into to the wafer. Therefore, the average path length of the light in the absorber will be increased, which leads to stronger absorption of the light. For realising the textures, usually wet etching approaches are used. Interestingly enough, 100 surfaces are etched much faster than the 111 surfaces. Such a behaviour is called *unisotropic etching*. Thus, when a 100 wafer is etched, the 100 surface will be etched away, while the (*slanted*) 111 surface facets will remain. Therefore, we obtain a wafer with a pyramid structure consisting of 111 facets only, as illustrated in Fig. 12.13 (a) that shows such a wafer that was processed in the DIMES Research Center in Delft. With a different etching approach, it is even possible to make silicon to appear completely black, as shown in Fig. 12.13 (b). This is called *black silicon*.

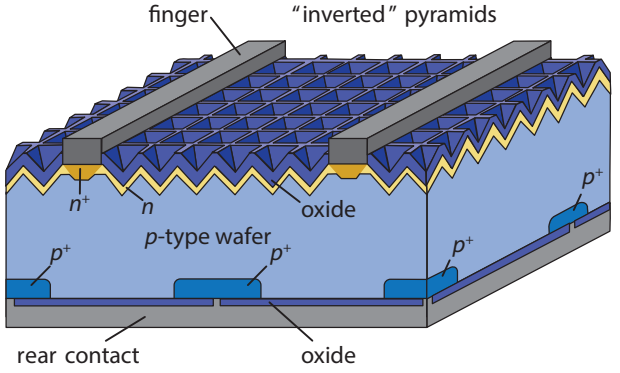

Figure 12.14: Illustrating the structure of a PERL solar cell.

12.4 High-efficiency concepts

In the last section of this chapter we discuss three examples of high-efficiency concepts based on crystalline silicon technology. As already discussed earlier, different types of silicon wafers with different qualities can be used. Naturally, for achieving the highest efficiencies, the bulk recombination must be as small as possible. Therefore the high efficiency concepts are based on monocrystalline wafers.

12.4.1 The PERL concept

The first high efficiency concept was developed in the late 1980s and the early 1990s at the *University of New South Wales* in the group of Martin Green. in the late eighties and early nineties. Figure 12.14 shows an illustration of the PERL concept, which uses a *p*-type float zone silicon wafer. With this concept, conversion efficiencies of 25% were achieved [38]. The abbreviation PERL stands for *Passivated Emitter Rear Locally*. This abbreviation indicates two important concepts that have been integrated in this technology: First, the optical losses of the PERL solar cell at the front side are minimised using three techniques:

1. The top surface of the solar cell is textured with inverted-pyramid structures. This microscopic texture allows a fraction of the reflected light to be incident on the front surface for a second time, which enhances the total amount of light coupled in to the solar cell.
2. The inverted pyramid structure is covered with a double-layer anti-reflection coating (ARC), which results in an extremely low top surface reflection. Often a double layer coating of magnesium fluoride (MgF_2) and zinc-sulfide (ZnS) is used as an antireflection coating.
3. The contact area at the front side has to be as small as possible, to reduce the shading losses. In the PERL concept the very thin and fine metal fingers are processed using photolithography technology.

Secondly, the emitter layer is smartly designed. As discussed in the previous sections, the emitter should be highly doped underneath the contacts, which in the PERL concept is achieved by heavily phosphorus diffused regions. The rest of the emitter is moderately doped, or in other words lightly diffused, to preserve an excellent blue response. The emitter is passivated with a silicon oxide layer on top of the emitter to suppress surface recombination as much as possible. With the PERL concept, the surface recombination velocity could be reduced so far that open circuit voltages with values of above 700 mV could be obtained.

At the rear surface of the solar cell, point contacts have been used in combination with thermal oxide passivation layer, which passivates the non-contacted area, and hence reduces the undesirable surface recombination. A highly doped boron region, created by local Boron diffusion, operates as a local back surface field, to limit the recombination of the minority electrons at the metal back contact.

The PERL concept includes some expensive processing steps. Therefore, the Chinese company *Suntech* developed in collaboration with University of New South Wales a more commercially viable crystalline silicon wafer technology, which is inspired on the PERL cell configuration.

12.4.2 The interdigitated back contact (IBC) solar cell

A second successful cell concept is the *interdigitated back contact* (IBC) solar cell. The main idea of the IBC concept is to have no shading losses at the front metal contact grid at all. All the contacts responsible for collecting of charge carriers at the *n*- and *p*-side are positioned at the back of the crystalline wafer solar cell. A sketch of such a solar cell is shown in Fig. 12.15 (a).

An advantage of the IBC concept is that monocrystalline float-zone *n*-type wafers can be used. This is interesting because *n*-type wafers have some interesting advantages with respect to *p*-type wafers. First, the *n*-type wafers do not suffer from light induced degradation. In *p*-type wafers both boron and oxygen are present, which under light exposure start to make complexes that act like defects. This light induced degradation causes a reduction of the power output with 2-3% after the first weeks of installation. No such effect is present in *n*-type wafers. The second advantage is that *n*-type silicon is not that sensitive for impurities like for example iron impurities. As a result, less efforts have to be made to fabricate high quality *n*-type silicon, and thus high quality *n*-type silicon can be processed cheaper than *p*-type silicon. On the other hand, *p*-doped wafers have the advantage that the boron doping is more homogeneously distributed across the wafer as this is possible for *n*-type wafers. This means that

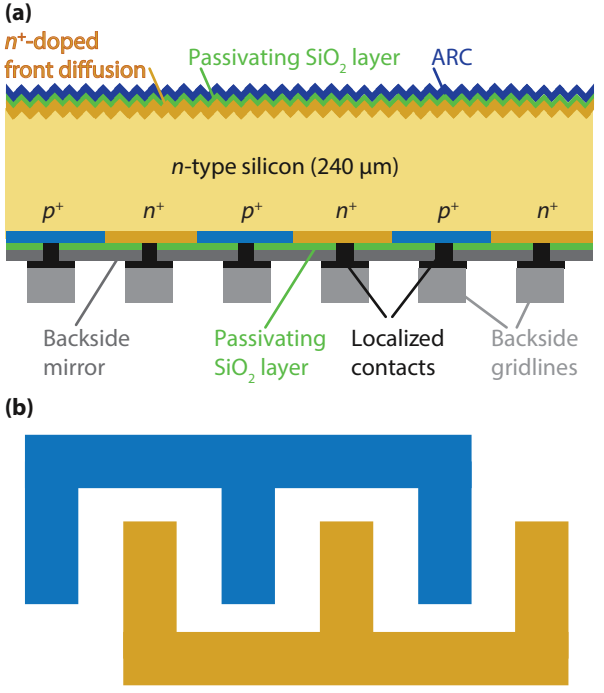

Figure 12.15: (a) Illustrating the structure of an IBC solar cell and (b) the contacts on its back.

within one n-type wafer the electronic properties can vary, which lowers the yield of solar cell production based on n-type monocrystalline wafers.

Besides that IBC cells are made from n-type wafers, they lack one large p-n junction. Instead, IBC cells have many localised junctions. The holes are separated at a junction between the p^+ silicon and the n-type silicon, whereas the electrons are collected using n^+-type silicon. The semiconductor-metal interfaces are kept as small as possible in order to reduce the undesired recombination at this defect-rich interface. Another advantage is that the cross section of the metal fingers can be made much larger, because they are at the back and therefore do not cause any shading losses. Thus, resistive losses at the metallic contacts can be reduced. Since both electric contacts are on the back side, it contains two metal grids, as illustrated in Fig. 12.15 (b). The passivation layer should made from a low-refractive-index material such that it operates like a backside mirror. It will reflect the light above 900 nm, which is not absorbed during the first pass back in to the absorber layer. Thus, this layer enhances the absorption path length.

At the front side of the IBC cell losses of light-excited charge carriers due to surface recombination is suppressed by a *front surface field* similar to the back surface field discussed earlier. This field is created with a highly doped n^+ region at the front of the surface. Thus, a n^+-n junction is created that acts like a n-p junction. It will act as a barrier that prevents the light-excited minority holes in the n-region from diffusing towards the front surface. The front surface field behaves like a passivation for the defects at the front interface and allows to have higher levels for the hole minority density in the n-doped bulk.

Reflective losses at the front side are reduced in a similar way as for PERL solar cells: Deposition of double layered antireflection coatings and texturing of the front surface.

The IBC concept is commercialised by the U.S. company *SunPower Corp.* They have achieved high solar cell efficiencies of 24.2%.

12.4.3 Hetero junction (HIT) solar cells

The third high-efficiency concept are *heterojunction with intrinsic thin layer* (HIT) solar cells. Before we discuss the technological details, we briefly recall the principles of heterojunctions, that already were discussed in Section 8.2.

Homojunctions, which are present in all the c-Si solar cell types discussed so far in this chapter, are fabricated by different doping types within the same semiconductor material. Hence, the band gap in the p- and n-regions is the same. A junction consisting of a p-doped

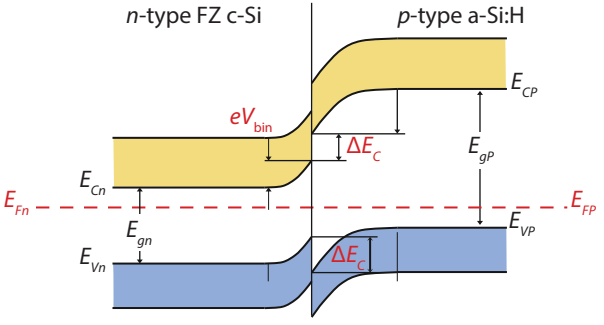

Figure 12.16: Illustrating the band diagram of a heterojunction.

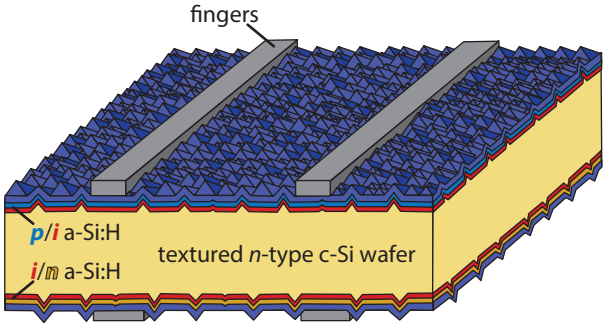

Figure 12.17: Illustrating the structure of a HIT solar cell.

semiconductor material and an *n*-doped semiconductor made from another material is called a *heterojunction*. In HIT cells, the heterojunction is formed in between two different silicon-based semiconductor materials: On the one hand, we use a *n*-type float zone monocrystalline silicon wafer. The other material is hydrogenated amorphous silicon (a-Si:H), that we will discussed in more detail in Section 13.3. a-Si:H is a silicon material in which the atoms are not ordered in a crystalline lattice but in a disordered lattice. For the moment we only must keep in mind that a-Si:H has a band gap of around 1.7 eV which is considerably higher than that of c-Si (1.12 eV).

In Fig. 12.16 a band diagram of a heterojunction in between *n*-doped crystalline silicon and *p*-doped amorphous silicon in the dark and thermal equilibrium is sketched. We see that next to the induced field because of the space charge region, some local energy steps are introduced. These steps are caused by the two different bandgaps for the *p* and *n* regions. The valence band is higher positioned in the *p*-type amorphous silicon than in the *n*-type crystalline silicon. This will allow the minority charge carriers in the *n*-type c-Si, the holes, to drift to the *p*-type silicon. However, the holes experience a small barrier. While they could not travel across such a barrier in classical mechanics, quantum mechanics allows them to *tunnel* across this barrier.

Let us now take a closer look to HIT solar cells. The HIT concept was developed by the Japanese company

Sanyo, that is currently a part of the Japanese *Panasonic Corp.* As we can see in Fig. 12.17, the HIT cell configuration has two junctions: The junction at the front side is formed using a thin layer of only 5 nanometers of intrinsic amorphous silicon which is indicated by the color red. A thin layer of *p*-doped amorphous silicon is deposited on top and here indicated with the color blue. The heterojunction forces the holes to drift to the *p*-layer. At the rear surface a similar junction is made: First, a thin layer of intrinsic amorphous silicon is deposited on the wafer surface, indicated by red. On top of the intrinsic layer an *n*-doped amorphous silicon is deposited indicated, which is indicated by the yellow color.

As discussed earlier, for high quality wafers, like this *n*-type float zone monocrystalline silicon wafer, the recombination of charge carriers at the surface determines the charge carrier lifetime. The advantage of the HIT concept is that the amorphous silicon acts as a very good passivation layer. With this approach the highest possible charge carrier lifetimes are accomplished. Thus, c-Si wafer based heterojunction solar cells have the highest achieved open circuit voltage among the different crystalline silicon technologies. The current record cell[2] has an open circuit voltage of 0.74 V and en efficiency of 25.6%.

How do the charge carriers travel to the contacts? The conductive properties of the *p*-doped amorphous silicon are relatively poor. While in the homojunction solar cells the lateral diffusion to the contacts takes place in the emitter layer, in a HIT solar cell this occurs through a transparent conducting oxide (TCO) material, like indium tinoxide (ITO), which is deposited on top of the *p*-doped layer. TCOs are discussed in more detail in Section 13.1.

The same contacting scheme is applied at the *n*-type back side. This means that this solar cell can be used in a bifacial configuration: it can collect light from the front, and scattered and diffuse light falling on the back-side of the solar cell. Another important advantage of the HIT technology is that the amorphous silicon layers are deposited using cheapplasma enhanced chemical vapour deposition (PE-CVD) technology at low temperatures, not higher than 200°C. Therefore, making the front surface and back surface fields in HIT solar cells is very cheap.

In this chapter we discussed many aspects of crystalline silicon solar cell technology. Before the solar cells can be installed, they must be packed as a *PV module*. How such modules are made is discussed in Part IV on PV systems in Section 17.1.

[2] As of May 2014.

13
Thin-film solar cells

In Chapter 12 we have discussed the PV technology based on c-Si wafers, which currently is by far the dominant PV technology. It is very likely that it will stay dominant for a long time to go.

In this chapter we will look at an alternative family of technologies, namely *thin film technologies*, also referred to as the *second generation* PV technology. These solar cells are made from films that are much thinner than the wafers that form the base for first generation PV. According to Chopra *et al.* [40], 'a *thin film* is a film that is created *ab initio* by the random nucleation process of individually condensing/reacting atomic/ionic/ molecular species on a substrate. The structural, chemical, metallurgical and physical properties of such a material are strongly dependent on a large number of deposition parameters and may also be thickness dependent.'

Thin-film solar cells were expected to become cheaper than first generation solar cells. However, due to the current price decline in wafer based solar cells thin-film solar cells have not yet become interesting from an economic point of view yet.[1] In general thin-film cells have a lower efficiency than c-Si solar cells. GaAs is an exception to this rule of thumb [38]. In contrast to wafer based silicon solar cells that are self-supporting, thin-film solar cells require a carrier that gives them mechanical stability. Usual carrier materials are glass, stainless steel or polymer foils. It is thus possible to produce flexible thin-film solar cells.

[1] According to the *PHOTON module price index*, the price for wafer-based modules has decreased around 40% within one year as of 25 May 2012 [41].

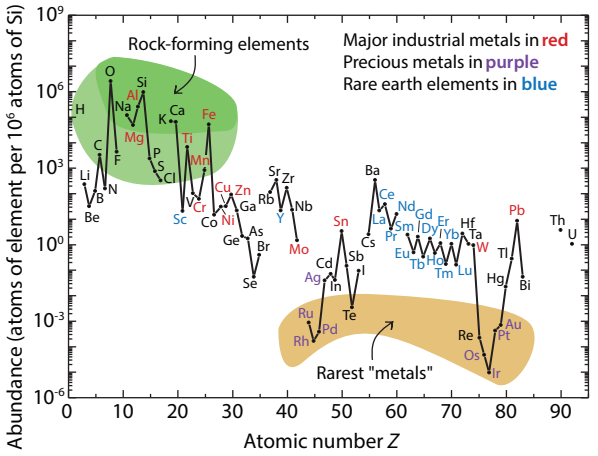

Figure 13.1: The abundance of elements (as atomic fraction) in the Earth's upper crust. Data provided by the USGS [39].

In thin-film solar cells the active semiconductor layers are sandwiched between a transparent conductive oxide (TCO) layer and the electric back contact. Often a back reflector is introduced at the back of the cell in order to minimise transmissive solar cell losses. As we will see in this chapter, many different semiconductors are used for thin-film solar cells. Figure 13.1 shows the abundance of metals in the Earth's crust. Some semiconductors require very rare elements such as *indium* (In), *selenium* (Se), or *tellurium* (Te). For terawatt scale photovoltaics, solar cells should be based on abundant elements only.

Thin-film PV technologies easily can fill a book on its own, see for example the book edited by Poortmans and Arkhipov [42]. In this chapter we therefore only can give a general introduction into the different thin-film technologies. We will focus on the working principles of the various devices, the current status and the future challenges of the various technologies. But before we start with this discussion, we will begin with a short introduction on *transparent conducting oxides* (TCOs).

13.1 Transparent conducting oxides

Due to the paramount importance of the TCO layer for the solar cell performance we briefly discuss its main

properties. The TCO layer acts as electric front contact of the solar cell. Furthermore, it guides the incident light to the active layers. It therefore should be both highly conductive and highly transparent in the active wavelength range. The first resistance measurements on thin-films of what we nowadays call TCOs were published by Bädeker in 1907 [43].

Typical TCO layers are made from fluor-doped tin oxide (SnO_2:F), aluminium-doped zinc oxide (ZnO:Al), boron-doped zinc oxide (ZnO:B), hydrogen-doped (hydrogenated) indium oxide (In_2O_3:H) [44], and indium tin oxide, which is a mixture of about 90% indium oxide (In_2O_3) and 10% tin oxide (SnO_2). These films are processed using either sputtering (ZnO:Al), low-pressure-chemical vapour deposition (LP-CVD, used *e.g.* for ZnO:B), metal-organic chemical vapour deposition (MO-CVD), or atmospheric-pressure chemical vapour deposition (AP-CVD, used *e.g.* for SnO_2:F).

Figure 13.2 shows the transmission, reflection and absorption spectra of a flat ZnO:Al layer. Following Kluth, we divide this spectrum into three parts [45]: For short wavelength, the transmission is very low due to the high absorption of light with energies higher than the bandgap. For longer wavelength, with photon-energies below the bandgap, the transmission is very high. We here see interference fringes that can be used to determine the film thickness. After a broad highly transmissive wavelength band, the absorption increases again. This

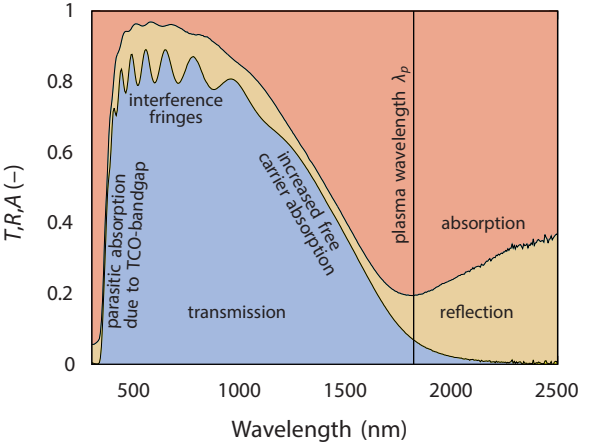

Figure 13.2: Transmission, reflection and absorption of a ZnO:Al layer (d = 880 nm).

absorption is called free carrier absorption an can be explained with the Drude theory of metals that was developed in 1900 [46, 47]. In this model, the frequency-dependent electric permittivity is given by

$$\begin{aligned}\epsilon(\omega) &= \left[n(\omega) - i\tilde{k}(\omega)\right]^2 \\ &= 1 + \chi(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\frac{\omega}{t}},\end{aligned} \tag{13.1}$$

where $\chi(\omega)$ is the dielectric susceptibility, t is the relaxation time,[2] $n - i\tilde{k}$ is the complex refractive index and ω_p denotes the *plasma frequency* that is given by

$$\omega_p = \frac{\mathfrak{N}e^2}{\epsilon_0 m_e^{*2}}. \tag{13.2}$$

Here, $\mathfrak{N}$ is the density of free charge carriers, e is the elementary charge, ϵ_0 is the permittivity of vacuum and m_e^* is the effective electron mass in the TCO layer.

The real and imaginary part of the susceptibility are given by

$$(\Re\chi)(\omega) = -\omega_p^2 \frac{t^2}{\omega^2 t^2 + 1}, \tag{13.3a}$$

$$(\Im\chi)(\omega) = \omega_p^2 \frac{t/\omega}{\omega^2 t^2 + 1}. \tag{13.3b}$$

[2] The relaxation time denotes the average time between two collisions, *i.e.* two abrupt changes of velocity, of the electrons.

If $\omega t \gg 1$, ϵ can be simplified to

$$\epsilon(\omega) \approx 1 - \frac{\omega_p^2}{\omega^2}, \tag{13.4}$$

while the imaginary part is negligible.

If this approximation is valid around ω_p, the material is transparent for $\omega > \omega_p$ ($\epsilon > 0$). For $\omega < \omega_p$, ϵ becomes negative, *i.e.* the refractive index is purely imaginary and the material therefore has a reflectivity of 1. In this case, the material changes dramatically from transparent to reflective, as ω_p is crossed.

If the approximation $\omega t \gg 1$ is not valid, $\Im\chi$ cannot be neglected. The imaginary part will increase with decreasing frequency, *i.e.* with increasing wavelengths the absorption increases. For wavelengths longer than the plasma wavelength the material becomes more reflective, what we also see in Fig. 13.2. For application in solar cells, the TCO should be highly transparent in the active region of the absorber. Therefore the plasma-wavelength should at least be longer than the bandgap wavelength of the absorber. On the other hand the plasma frequency is proportional to the free carrier density $\mathfrak{N}$. A longer plasma wavelength therefore corresponds to a lower $\mathfrak{N}$. Finding an optimum between high transparency and high carrier densities is an important issue in designing TCOs for solar cell applications.

Even though the Drude model gives a good approxima-

tion of the free carrier related phenomena in TCOs, this model often is too simple. Therefore several authors used extended Drude models with more parameters [48–50].

Of all TCO materials currently available, the trade off between transparency and conductivity is best for indium tin oxide [51]. However, indium is a rare earth element with a very low abundance of 0.05 ppm in the Earth's crust, similar to the abundance of silver (0.07 ppm) and mercury (0.04 ppm) [52], which makes it less preferable for cheap large-scale PV applications. Therefore other TCO materials are thoroughly investigated and used in industry. Amongst them are aluminium doped zinc oxide, boron doped zinc oxide and fluorine doped tin oxide. The abundances of the used elements are: aluminium: 7.96%, zinc: 65 ppm, boron: 11 ppm, fluorine: 525 ppm, and tin: 2.3 ppm [52].

The morphology of selected TCO samples

As we have briefly mentioned in Chapter 10 and will discuss in more detail in Chapter **??**, *light management* is crucial for designing solar cells. Very often, the TCO layers in thin-film solar cells have nano-textured surfaces. The incident light is scattered at these interfaces such that the average photon path length in the absorber layer is enhanced. Thus, more light can be absorbed and the photocurrent density can be increased. However, the texture may also influence the growth of layers that are deposited onto the TCO and thus alter the electrical properties, which may lead to a reduction in V_{oc} and the fill factor. Further, the increased interface area due to the texture may lead to more surface recombination and also reduce the voltage.

Figure 13.3 shows the surface morphology, rms roughness σ_r and correlation length ℓ_c of some nano-textured TCO samples made from three different materials. Fluorine-doped AP-CVD tin oxide (SnO_2:F) of Asahi U-type [54] and boron-doped LP-CVD zinc-oxide (ZnO:B) of so-called 'B-type' from PV-LAB of the École polytechnique fédérale de Lausanne (EPFL), Switzerland [55], obtain their nano-structure already during the deposition process. The nano-structure is due to the crystal growth of the TCO layers and has a *pyramid-like* shape.

In contrast, RF-sputtered, aluminium-doped zinc oxide (ZnO:Al) is flat after deposition ($\sigma_r \approx 3$ nm, depending on the deposition condition als higher). To obtain a nano-texture, ZnO:Al is etched in a 0.5% HCl solution [56, 57]. This etching process leads to *crater-like* features. The lateral size of these craters is influenced by the grain size of the zinc oxide crystals. Fig. 13.3 (c) and (d) shows etched ZnO:Al with broad craters. Fig. 13.3 (e) and (f) shows etched ZnO:Al with narrow craters.

The surface morphologies shown in Fig. 13.3 were obtained with *atomic force microscopy* (AFM). Also the two statistical parameters rms roughness σ_r and correlation

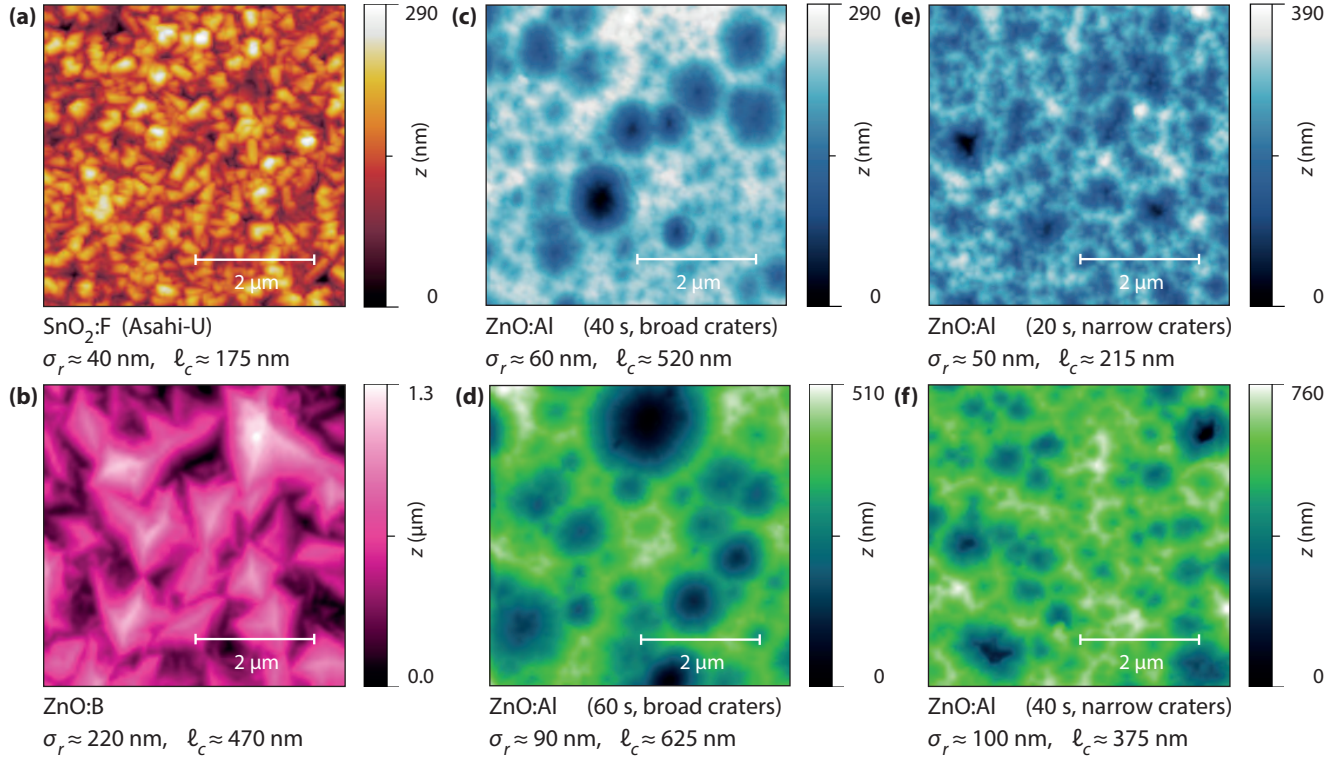

Figure 13.3: The morphology, rms roughness, and correlation length for some selected TCO samples [53].

length ℓ_c were obtained with AFM scans of 256×256 points over an area of 20×20 µm^2. AFM and the determination of the statistical parameters are explained in Appendix C.

13.2 The III-V PV technology

Of all the thin-film technologies, the III-V (speak three-five) PV technology results in the highest conversion efficiencies under both one sun standard test conditions and concentrated sun conditions. Therefore, it is mainly used in space technology and in solar concentrator technology.

As some concepts use a crystalline germanium or an GaAs wafer as substrate, it might not be considered as a real thin film PV technology, in contrast to thin-film silicon, CdTe, CIGS or organic PV, which we will discuss later in this chapter. However, the III-V based absorber layers themselves can be considered as thin compared to the thickness of crystalline silicon wafers.

The III-V materials are based on the elements with *three valence electrons* like aluminium (Al), gallium (Ga) or indium (In) and elements with *five valence electrons* like phosporus (P) or arsenic (As). Various different semiconductor materials such as *gallium arsenide* (GaAs), *gallium phosphide* (GaP), *indium phosphide* (InP), *indium arsenide* (InAs), and more complex alloys like GaInAs, GaInP, AlGaInAs and AlGaInP have been explored.

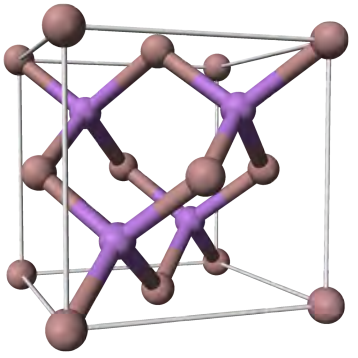

Figure 13.4: A unit cell of a gallium arsenide crystal in the zinc blend structure [58].

We now will take a closer look on GaAs, which is the III-V semiconductor most used for solar cells. The GaAs has a *zinc blend crystal structure*, illustrated in Fig. 13.4. It is very similar to the diamond cubic crystal structure with the difference that it has atoms of alternating elements at its lattice sites. In contrast to c-Si, where every Si atom has four neighbours of the same kind, in GaAs, every Ga has four As neighbours, while every As atom has four Ga neighbours. When compared to silicon, GaAs has a slightly larger lattice constant of 565.35 pm,

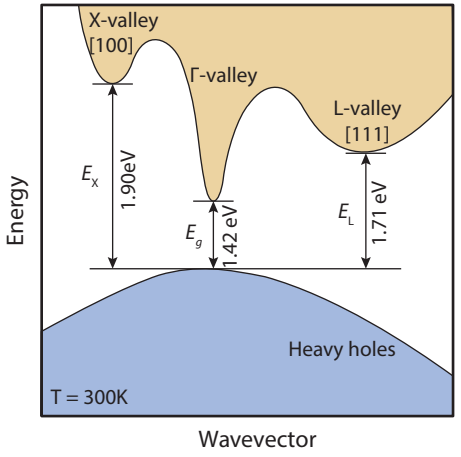

Figure 13.5: The electronic band dispersion diagram of gallium arsenide.

and significant denser than silicon with a density of 5.3176 g/cm^3 (Si: 543.07 pm, 2.3290 g/cm^3). Note that both Ga and As are are roughly twice as dense as silicon. In contrast to silicon, which is highly abundant, the abundance of gallium in the Earth's crust is only about 14 ppm [59]. GaAs therefore is a very expensive material. Arsenic is highly toxic; it is strongly suggested that GaAs is carcinogenic for humans [60].

Figure 13.5 shows the electronic band dispersion diagram of gallium arsenide. We can easily see that GaAs is a *direct band gap* material, *i.e.* , the highest energy level in the valence band is vertically aligned with the lowest energy level in the conduction band. Hence, only transfer of energy is required to excite an electron from the valence to the conduction band, but no transfer of momentum is required. The band gap of GaAs is 1.424 eV [20]. Here we look again at the absorption coefficient versus the wavelength. Consequently, as you can see in Fig. 12.4, the absorption coefficient of GaAs is significantly larger than that of silicon. The same is true for InP, another III-V material that also is shown in Fig. 12.4. Because of the high absorption coefficient. Because of the high absorption coefficient, the same amount of light can be absorbed in a film more than one order of magnitude as thin when compared to silicon. Another advantage of the direct III-V semiconductor materials their sharp band gap. Above E_g, the absorption coefficient increases quickly.

Let us now take a look at the utilisation of the band gap energy. Since GaAs is a direct band gap material, *radiative recombination processes* become important. On the other hand, Shockley-Read-Hall recombination can be kept at a low level because III-V films can be deposited tih *epitaxy processes* that result in high purity films.

13.2.1 Multi-junction cells

III-V PV devices can reach very high efficiencies because they are often based on the multi-junction concept, which means that more than one band gap is used. As discussed in Chapter 10, the maximum theoretical efficiency of single-junction cells is described by the Shockley-Queisser limit. A large fraction of the energy of the energetic photons are lost as heat, while photons with energies below the band gap are lost as they are not absorbed.

For example, if we use a low bandgap material, a large fraction of the energy carried by the photons is not used. However, if we use more band gaps, the same amount of photons can be used but less energy is wasted as heat. Thus, large parts of the solar spectrum and larges part of the energy in the solar spectrum can be utilised at the same time, if more than one *p-n* junctions are used.

In Fig. 13.6 a typical III-V triple junction cell is illustrated. As substrate, a germanium (Ge) wafer ie used. From this wafer, the *bottom cell* is created. Germanium has a band gap of 0.67 eV. The *middle cell* is based on GaAs and has a band gap of about 1.4 eV. The *top cell* is based on GaInP with a band gap in the order of 1.86 eV.

Let now take a closer look on how a multi-junction solar cell works. Light will enter the device from

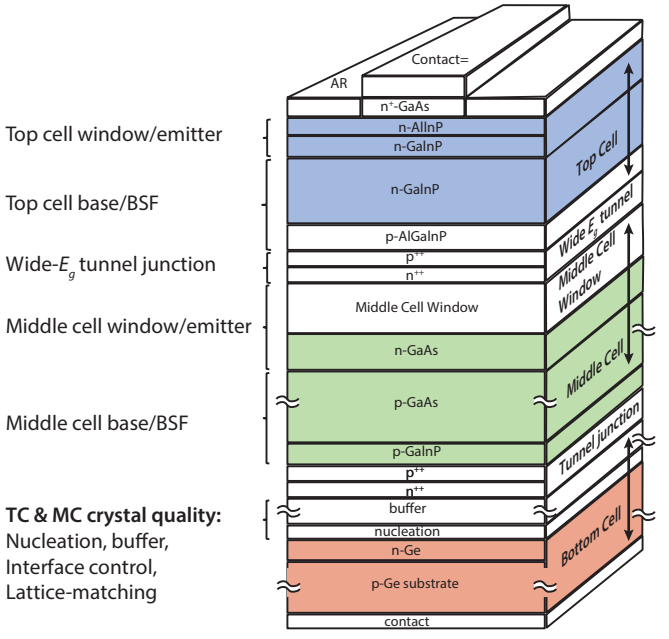

Figure 13.6: Illustrating a typical III-V triple junction solar cell.

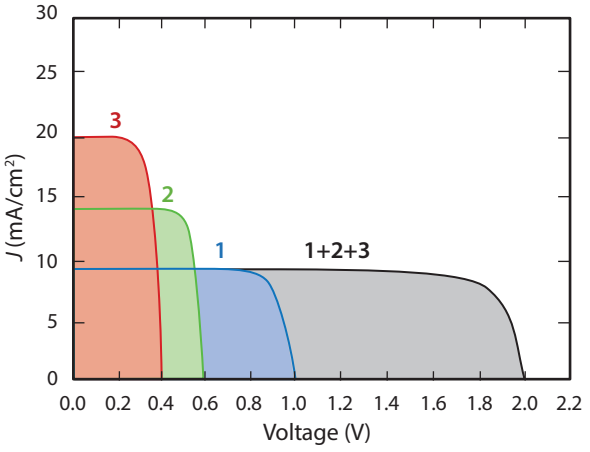

Figure 13.7: The *J-V* curves of the three junctions used in the III-V triple junction cell and *J-V* curves of the III-V triple junction cell (in black).

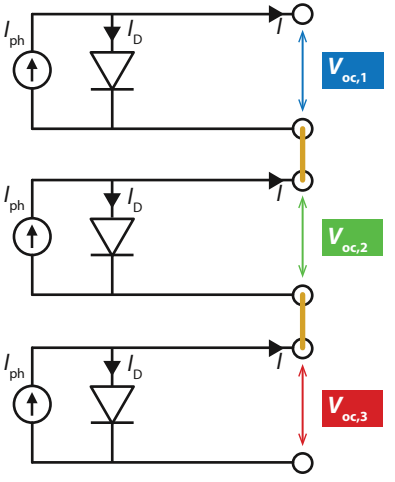

Figure 13.8: The equivalent circuit of the three junctions connected in series.

the top. As the spectral part with the most energetic photons like blue light has the smallest penetration depth in materials, the junction with the highest band gap always acts as the top cell. On the other hand, as the near infrared light outside the visible spectrum has the longest penetration depth, the bottom cell is the cell with the lowest band gap.

Figure 13.7 shows the *J-V* curve of the three single *p-n* junctions. We observe *p-n* junction 1 has the highest open circuit voltage and the lowest short circuit current density, which means that this *p-n* junction has the highest band gap. In contrast, *p-n* junction 3 has a low open circuit voltage and a high current density, consequently it has the lowest band gap. *p-n* junction 2 has a band gap in between. Hence, if we are designing a triple-junction cell out of these three junctions, junction 1 will act as the top cell, junction 2 will act as the middle cell and junction 3 will act as the bottom cell.

For understanding how the *J-V* curve of the triple junction looks like, we take a look at the equivalent circuit. Every *p-n* junction in the multijunction cell can be represented by the circuit of a single-junction cell, as discussed in chapter 9. As the three junctions are stacked onto each other, they are connected to each other *in series*, as illustrated in Fig. 13.8. In a series connection, the voltages of the individual cell add up in the triple junction cell. Further, the current density in a series connection is equal over the entire solar cell, hence the current density is determined by the *p-n* junction generating the lowest current. The resulting *J-V* curve is also illustrated in Fig. 13.7. We see that the voltages add up and the current is determined by the cell delivering the lowest current.

Figure 13.9 (a) shows a typical band diagram of such triple junction. The top cell with high band gap is shown at the left hand side and the bottom cell is at the right hand side. However, this band diagram does not represent reality. If we would place three *p-n* junctions in series, for example the *p*-layer of the top cell and the *n*-layer of the middle cell would form a *p-n* junction in the reverse direction than the *p-n* junctions of the three single junction solar cells. These reverse junctions would significantly lower the voltage of the total triple junction.

For preventing the creation of such reverse junctions, so-called *tunnel junctions* are included. These tunnel

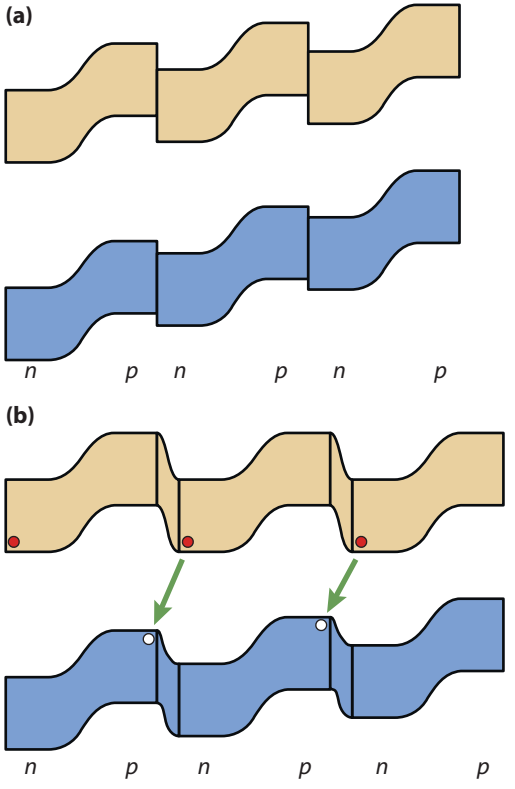

Figure 13.9: The band diagram of the III-V triple junction cell (a) without and (b) with tunnel junctions.

junctions align the valence band at one side with the conduction band at the other side of the tunnel junction, as illustrated in Fig. 13.9 (b). They have a high band gap to prevent any parasitic absorption losses. Further, tunnel junctions are relatively thin and have an extremely narrow depletion zone. As a result, the slopes of the valence and conduction band are so steep that the electrons from n-layer can tunnel through the small barrier to the p-layer, where they recombine with the holes. It is important to have the tunnel junctions with a low resistance such that the voltage loss across them is low.

In the triple junction cell, two tunnel junctions are present. First, a tunnel junction via that the holes in the p-layer of the top cell have to recombine with the electrons of the n-layer of the middle cell. Secondly, a tunnel junction where the holes in the p-layer of the middle cell recombine with the electrons of the n-layer of the bottom cell. The electrons in the top cell n-layer are collected at the front contact and the hole in the p-layer of the bottom cell are collected at the back contact. It is important to realise that the recombination current at the tunnel junctions represents the current density of the complete triple junction.

Let us take a look at the external parameters of a typical lattice matched triple junction solar cell from SpectroLab, which is a subsidiary of The Boeing Company. Because this cell was developed for space applications, it is not tested under AM1.5 conditions, but under AM0 conditions with an irradiance of 135 $\mathrm{mW/cm^2}$. The total open circuit voltage is 2.6 Volts. The short circuit current density is 17.8 $\mathrm{mA/cm^2}$, which corresponds to a spectral utilisation up to the lowest band gap of 0.67 eV of

$$J_{\mathrm{tot}} = 3 \cdot 17.8\ \mathrm{mA/cm^2} = 53.4\ \mathrm{mA/cm^2}. \tag{13.5}$$

The maximal theoretical current that could be generated when absorbing AM0 for all wavelength shorter than 1850 nm, which corresponds to 0.67 eV, is 62 $\mathrm{mA/cm^2}$. The average EQE of the solar cell thus is 86%, which is an very impressive value. Further, this cell has a very high fill factor of 0.85, which means that it has a conversion efficiency of 29.5% under AM0 illumination.

Figure 13.10 shows the EQE of the subcells of a typical III-V triple junction cell. We note that the shape of the spectral utilisation of all the curves approaches the shape of block functions, which would be the most ideal shape. These block shapes are possible because the III-V materials have very sharp band gap edges and high absorption coefficients. We see that the bottom cell generates much more current than the middle and top cells. Hence, a lot of current is lost in the bottom cell.

This ineffective use of the near infrared part can be reduced using quadruple junctions instead of triple junctions. In these cell, an additional cell is placed in-between the middle and bottom cells of the triple junc-

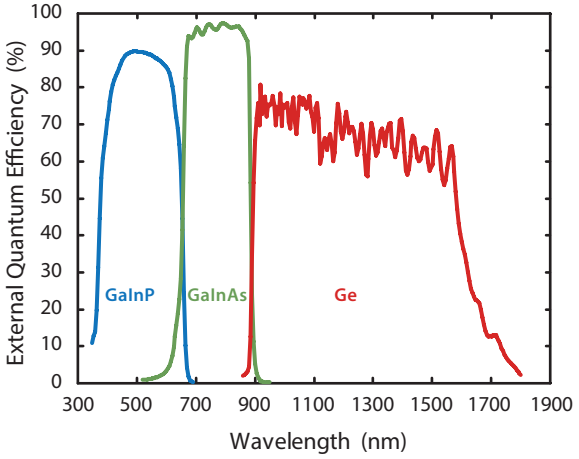

Figure 13.10: The external quantum efficiency of a three-junction III-V solar cell (Data from SPECTROLAB).

tion. The spectral utilisation can be even increased further by moving to multi-junction solar cells consisting of 5 or even 6 junctions. The major challenge for these cells is that lattice matching can not be guaranteed anymore. Lattice-mismatched multi-junctions are called *metamorphic multijunctions*. They require buffer layers that have a profiling in the lattice constant, going from the lattice constant of one *p-n* junction to the lattice constant of the next *p-n* junction. Using this technology, Spectrolab could demonstrated an impressive 38.8% conversion efficiency of a 5-junction solar cell under 1-sun illumination [38].

The III-V PV technology is very expensive. Hence, such cells are mainly used for space-applications and in concentrator technology, where high performance is more important that the cost. For reducing the cost per Watt peak, a *solar concentrator* can be used. In such a concentrator, the irradiance of a large area is concentrated onto the small solar device. One advantage of concentrated sunlight is that the open circuit voltage is increasing with the increasing irradiance as well, as long as Auger recombination of light excited charge carriers does not become dominant. It is, however, very important that the small solar cell devices are actively cooled, as their performance decreases with increasing temperatures. This temperature dependence is caused by stronger recombination and hence higher dark currents and is discussed in more detail in Section 18.3. Additionally, concentrator systems require sun-tracking systems, as only

direct sunlight can be concentrated. The tracking systems ensure that the optical concentrator system tracks the sun and guarantees the optimal light concentration on the small PV device during the entire day. Hence, the non-modular costs of concentrator systems are larger than that of conventional PV system based on c-Si or other thin-film technologies.

The current world record conversion efficiency for all solar cell technologies on lab scale is 44.4% for a three-junction III-V solar cell under 302-fold concentrated sun light conditions. This result was achieved by Sharp [38].

13.2.2 Processing of III-V semiconductor materials

As already mentioned above, high quality III-V semiconductor materials can be deposited using *epitaxy* deposition methods. In this method, crystalline *o*verlayers are deposited on a crystalline substrate, such that they adopt the crystal lattice structure of the substrate. The *precursor* atoms from that the layers are grown are provided by various elemental sources. For example, if GaAs is deposited with epitaxy, Ga and As atoms are directed to a growth surface under ultra-high vacuum conditions. For growing III-V semiconductors, germanium substrates usually are used. On this substrate the GaAs crystalline lattice is grown layer-by-layer and adopts the structure of the crystalline substrate.

As the layer-by-layer growth process is very slow, it allows the deposition of compact materials without any vacancy defects. Furthermore, processing at high vacuum conditions prevents the incorporations of impurities. Hence, III-V semiconductors can be deposited up to a very high degree of purity. Dopants can be added to make it *n*- or *p*-type.

Typically, III-V semiconductor layers are deposited by using *metal-organic chemical vapour deposition* (MOCVD). Typical precursor gasses are trimethylgallium [$Ga(CH_3)_3$], trimethylindium [$In(CH_3)_3$], trimehtylaluminium [$Al_2(CH_3)_6$] , arsine gas (AsH_3) and phosphine gas (PH_3). Surface reactions of the metal-organic compounds and hydrides, which contain the required metallic chemical elements, create the right conditions for the epitaxial crystalline growth. Epitaxial growth is a very expensive process. Similar techniques are also used in the microchip production process.

A big challenge of depositing III-V semiconductor materials is that the lattice constants of the various materials are different, as seen in Fig. 13.11. We see that every III-V semiconductor has a unique bandgap-lattice constant combination. Hence, interfaces between different III-V materials show a lattice mismatch, as illustrated in Fig. 13.12. Because of this mismatch, not every valence electron is able to make a bond with a neighbour. This problem can be solved by *lattice matching*, as this also is done in the triple junction cell discussed in Section

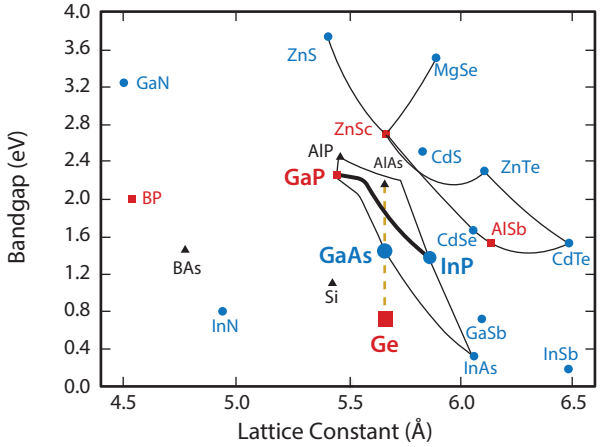

Figure 13.11: The bandgap and lattice constant of various III-V semiconductor materials.

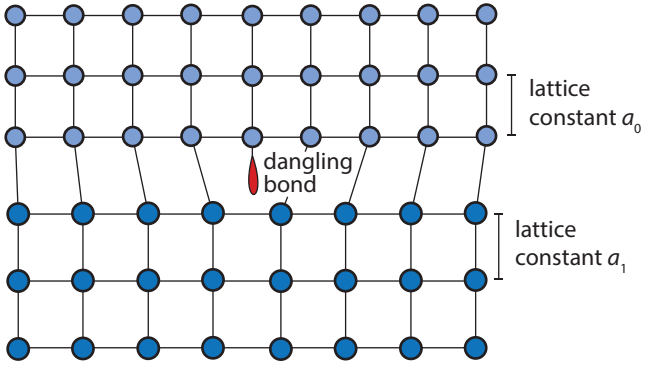

Figure 13.12: Illustrating the lattice mismatch at an interface between two different III-V materials.

13.2.1. For understanding what this means, we take another look at the phase diagram shown in Fig. 13.12. The triple junction is processed on a *p*-type germanium substrate, the bottom cell is a Ge cell. One top it is best to place a junction that has a higher band gap but the same lattice constant. As we can see in the plot, GaAs has exactly the same lattice constant as Germanium. Therefore, between GaAs and Germanium interfaces can be made without any coordination defects related to mismatched lattices. For reasonable current matching the desired band gap of the top cell should be around 1.8 eV. However, we see that no alloys, based on solely two elements, exist. But if we take a mixture of gallium, indium and phosphorus, we can make a III-V alloys

with a band gap of 1.8 eV material and with matching lattice constant. Consequently, triple junction cells based on GaInP, GaAs and Germanium can be fully lattice matched.

13.3 Thin-film silicon technology

In this section we are going to discuss the thin-film silicon PV. An advantage of thin-film silicon solar cells is that they can be deposited on glass substrates and even on flexible substrates.

13.3.1 Thin-film silicon alloys

Thin-film silicon materials usually are deposited with *chemical vapour deposition* (CVD) processes that we will discuss in more detail in Section 13.3.3. In chemical vapour deposition different *precursor gasses* are brought into the reaction chamber. Due to chemical reactions, layer is formed on the substrate. Depending on the used precursors and other deposition parameters such as the gas flow rate, pressure, and temperature, various different alloys with different electrical and optical parameters can be deposited. We will discuss the most important alloys in the following paragraphs.

We start with two alloys consisting of silicon and hydrogen: *hydrogenated amorphous silicon* (a-Si:H) and *hydrogenated nanocrystalline silicon* (nc-Si:H), which is also known as microcrystalline silicon. Usually in discussions the term 'hydrogenated' is left out for simplicity. *Hydrogenated* means that some of the valence electrons in the silicon lattice are *passivated* by hydrogen, which is indicated by the ':H' in the abbreviation. The typical atomic hydrogen content of these alloys is from 5% up to about 15%. The hydrogen passivates most defects in the material, resulting in a defect density around 10^{16} cm^{-3} [61], which is suitable for PV applications. Pure amorphous silicon (a-Si) would have an extremely high defect density ($> 10^{19}$ cm^{-3}) [42], which would result in fast recombination of photo-excited excess carriers. Similarly, we can make alloys from *germanium* and hydrogen: *hydrogenated amorphous germanium* (a-Ge:H) and *hydrogenated nanocrystalline germanium* (nc-Ge:H).

Let us now take a look at alloys of silicon with four valence electrons with other elements with four valence electrons, *carbon and germanium*. In thin-film silicon solar cells, both hydrogenated amorphous and nanocrystalline *silicon-germanium alloys* (a-SiGe:H and nc-SiGe:H) are being used. Silicon is also mixed using the four valence electron material, carbon, leading to hydrogenated amorphous *silicon carbide* (a-SiC:H).

Another interesting alloy is obtained when oxygen with six valence electrons is incorporated into the lattice:

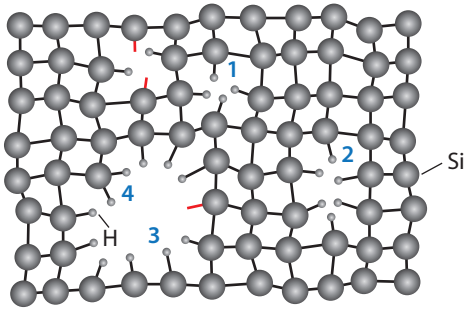

Figure 13.13: Illustrating the atomic structure of amorphous silicon with four typical defects: (1) monovacancies, (2) divacancies, (3) nanosized voids with mono- and (4) dihydrides. With kind permission of M. A. Wank [62].

hydrogenated nanocrystalline silicon oxide is often used in thin-film silicon solar cells.

All these alloys can be doped, usually boron is used as a *p* dopant while phosphorus is the most common *n* dopant.

Many of the alloys mentioned above are present as *amorphous* materials. It is hence important to discuss the structure of amorphous lattices. In this discussion we will limit ourselves to amorphous silicon, since it is the best studied amorphous semiconductor and the general properties of the other amorphpous alloys are similar. In Chapter 12 we thoroughly discussed crystalline silicon, which has an ordered lattice in which the orientation and structure is repeated in all directions. For amorphous materials this is not the case, but the lattice is disordered as it is illustrated in Fig. 13.13. This figure shows a so-called *continuous random network* (CRN). On atomic length scales, or also called *short-range order*, the atoms still have a tetrahedral coordination structure, just like crystalline silicon. But the silicon bond angles and silicon-silicon bond lengths are slightly distorted with respect to a crystalline silicon network. However, at larger length scales, or also referred to as long range, the lattice does not look crystalline anymore. Not all silicon atoms have four silicon neighbours, but some valence electrons form dangling bonds, similar to the unpassivated surface electrons already discussed in Chapter 12. Recent studies show that these dangling bonds are not distributed homogeneously throughout the amorphous lattice, but that they group in divacancies, multivacancies or even nanosized voids *smets:2007*. The surfaces of these volume deficiencies are passivated with hydrogen.

Another phase of hydrogenated silicon alloys is the *nanocrystalline* lattice. It is a heterogeneous material with a structure that even is more complex than that of amorphous materials. Nanocrystalline silicon consists of small grains that have a crystalline lattice and are a few tens of nanometres big. These grains are embedded in a tissue of hydrogenated amorphous silicon. Figure 13.14 shows the various phases of thin-film sil-

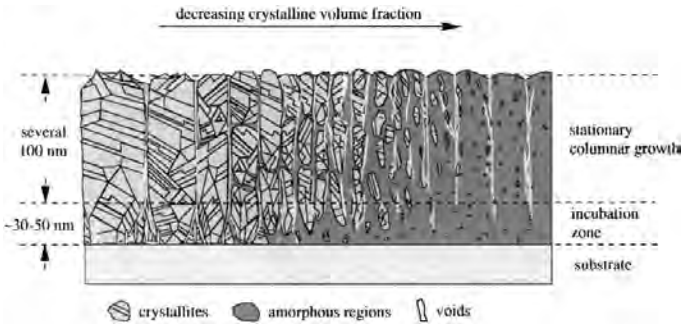

Figure 13.14: The various phases of thin-film silicon with a highly crystalline phase on the left and amorphous silicon on the right[63].

icon [63]. On the left-hand side, a fully crystalline phase is shown, which is close to that of poly-crystalline silicon, except that is has much more cracks and pores in it. On the right hand side, the phase represents the amorphous lattice. Please note again, that the terms microcrystalline and nanocrystalline refer to the same material. Going from right to left, the amorphous phase is changing in to a mixed phase with a few small crystalline grains to a phase which is dominated by large crystalline grains and a small fraction of amorphous tissue. Research has shown that the best nanocrystalline bulk materials used in solar cells has a network close to the *amorphous-nanocrystalline silicon transition region* and its crystalline volume fraction is in the order of 60%.

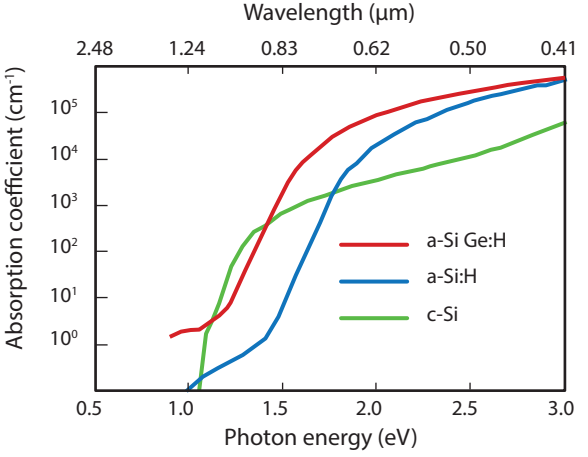

Figure 13.15: The absorption coefficients of different thin-film silicon materials.

The band gap of nanocrystalline silicon is close to that of crystalline silicon ($\approx$ 1.12 eV) due to the crystalline network in the grains. The band gap of amorphous silicon is in the order of 1.6 up to 1.8 eV, which can be tuned by the amount of hydrogen incorporated in to the silicon network. It is larger than that of crystalline silicon because of the distortions in bond angles and bond lengths. It is out of the scope of this book to discuss the reasons for this increase in band gap in more detail. An important consequence of a disordered amorphous lattice is that the electron momentum is poorly defined in contrast to crystalline silicon. As we discussed in Chapter 12, both energy and momentum transfer are needed to excite and electron from the valence band to the conduction band. Hence, crystalline silicon is an indirect band gap material. This is not true for amorphous silicon, which is a direct band gap material. Therefore the absorptivity of a-Si:H is much higher than that of c-Si, as we can see in Fig. 13.15. We see that the absorption coefficient for amorphous silicon in the visible spectrum is much larger that that of crystalline silicon. In some wavelength regions it is about two orders of magnitude larger, which means that much thinner silicon films can be used in reference to the typical wafers in crystalline silicon solar cells. In the figure, also data for amorphous silicon-germanium are shown. a-SiGe:H has lower band gap and even higher absorption coefficient in the visible. Its band gaps are in the range of 1.4 up to 1.6 eV. Amorphous silicon carbide alloys have band gaps of 1.9 eV and larger. Finally, nanocrystalline silicon oxides have band gaps exceeding 2 eV.

13.3.2 The design of thin-film silicon solar cells

The first successful a-Si:H solar cell with an efficiency of 2.4% was reported by Carlson and Wronski in 1976 [64]. We now will discuss how such solar cells can be designed and what are the issues in designing them.

When compared to c-Si, hydrogenated amorphous and nanocrystalline silicon films have a relatively high defect density of around 10^{-16} cm^{-3}. As mentioned before, because of the disordered structure not all valence electrons are able to make bonds with the neighbouring atoms. The dangling bonds can act as defects that limit the lifetime of the light excited charge carriers. This is described with Shockley-Read-Hall recombination, which controls the diffusion length. Because of the high defect density the diffusion length of charge carriers in hydrogenated amorphous silicon is only 100 nm up to 300 nm. Hence, the transport of charge carriers in a thick absorber can not rely on diffusion.

Therefore, amorphous silicon solar cells are *not* based on a *p-n* junction like wafer based c-Si solar cells. Instead, they are based on a *p-i-n* junction, which

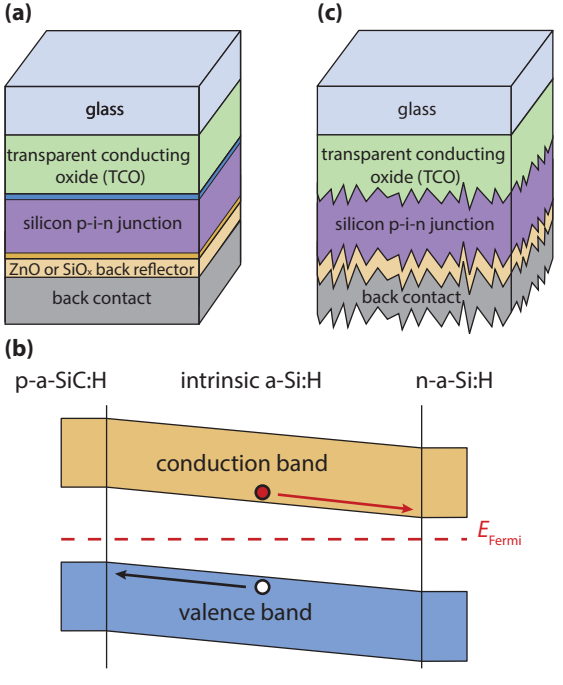

Figure 13.16: Illustrating the (a) layer structure and the (b) band diagram of an amorphous silicon solar cell. (c) Thin-film silicon solar cells hava nanotextured interfaces.

means that an intrinsic (undoped) layer is sandwiched between thin p-doped and n-doped layers, as illustrated in Fig. 13.16 (a). While the i-layer is several hundreds of nanometres thick, the doped layers are only about 10 nm thick. Between the p- and n-doped layers a built-in electric field across the intrinsic absorber layer is created. This is illustrated in the electronic band diagram shown in Fig. 13.16 (b). If the layers are not connected to each other, the Fermi level in the p- n layers is closer to the valence and the conduction bands, respectively. For the intrinsic layer, it is in the middle of the band gap. When the p, i, and n layers are connected to each other, the Fermi level has to be the same throughout the junction, if it is in the dark and under thermal equilibrium. This creates a slope over the electronic band in the intrinsic film as we can see in the illustration. This slope reflects the built-in electric field.

Another way to look at the field is to consider the i-layer as an dielectric in between to charged plates (the doped layers). Neglecting any edge effects, the electric field in the i layer is given by

$$E = \frac{\sigma}{\epsilon\epsilon_0}, \tag{13.6}$$

where σ is the charge density on the doped layers and ϵ is the electric permittivity of the i-layer material. Note that the electric field in the i layer is constant. Follow-

ing Eq. (4.29), we find the voltage in the i layer to be

$$U(x) = -\int E(x)\,\mathrm{d}x = U_0 - \frac{\sigma}{\epsilon\epsilon_0}x. \qquad (13.7)$$

We see that the voltage and hence the energy in the i layer is linear, just as shown in the band diagram of Fig. 13.16 (b).

Because of the electric field, the light excited charge carrier will move through the intrinsic layer. As discussed in Section 6.5.1, the holes move up the slope in the valence band towards the p layer and the electrons move down the slope in the conduction band towards the n layer. Such a device, where electronic drift because of an electric field is the dominant transport mechanism is called a *drift device*. In contrast, a wafer based crystalline silicon solar cell, as discussed in Chapter 12, can be considered as a *diffusion device*.

Note, that because of the intrinsic nature of the absorber layer the hole and electron density is in the same order of magnitude. On the other hand, in the p layer the holes are the majority charge carriers and the dominant transport mechanism is diffusion. Similarly, in the n layer the electrons are the majority charge carriers and again diffusion in the dominant transport mechanism. Because of the low diffusion length, both p and n layers must be very thin thin.

Now we take another look at Fig. 13.16 (a). The cell sketched there is deposited in *superstrate* configuration. This means that the layer that is passed first by the incident light in the solar cell is also deposited first in the production process. Also the term *p-i-n* layer refers to this superstrate configuration, as it indicates the order of the depositions: because in thin-film silicon holes have a significantly lower mobility than electrons, the p layer is in front of the n layer and hence the p layer is deposited first. As then the generation rate is higher close to the p layer, more holes can reach it.

In superstrate thin-film silicon solar cells, usually glass is used as a superstrate because it is highly transparent and can easily handle all the chemical and physical conditions in that all the depositions are carried out. Before the *p-i-n* layers can be deposited, a *transparent front contact* has to be deposited. Usually *transparent conducting oxides* are used, that are discussed in more detail in Section 13.1.

Another possible configuration is the *substrate configuration*. There, either the substrate acts like a back contact or the back contact is deposited on the substrate. Consequently, no light will pass through the substrate. Thin-film silicon solar cells deposited in the substrate configuration are also called *n-i-p* cells as the n layer is deposited before the i layer and the p layer.

Usually, thin-film silicon solar cells have no flat interfaces, as shown in Fig. 13.16 (a), but nano-textured interfaces, as illustrated in Fig. 13.16 (c). These textured interfaces scatter the incident light and hence prolong

the average path length of the light through the absorber layer. Therefore, more light can be absorbed and the photocurrent can be increased. This is an example of *light management* that is discussed in more detail in Chapter **??**.

For the *p* layer, often not amorphous silicon but higher band gap materials such as silicon carbide or silicon oxides are used in order to minimise *parasitic absorption* mainly in the blue part of the spectrum. Usually, boron is used as a dopant. Also for the *n* layer nowadays often silicon oxides are used, but sometimes still a-Si:H is used. Phosphorus is the main dopant used here. The *n*-SiO_x:H is very transparent. Therefore it also can be used as a back reflector structure when its thickness is chosen such that destructive interference occurs at the *i*-*n* interface minimising the electric field strength and hence parasitic absorption in the *n* layer. Additionally, also several tenth of nm thick TCO can be used with the same purpose. Further, a metallic back reflector that also acts as the electric back contact is used. Because of its attractive cost, mainly Al is used for this purpose. However, also silver that has a higher reflectivity can be used, yet is it more expensive.

The band gap of hydrogenated amorphous silicon is in the order of 1.75 eV, hence it only is absorptive for wavelength shorter than 700 nm. The highest current densities achieved in single junction amorphous silicon solar cell are 17 up to 18 $mA{\cdot}cm^{-2}$, whereas the maximum theoretic current that could be achieved up to 700 nm is in the order of 23 $mA{\cdot}cm^{-2}$. Thus, the EQE averaged over the spectrum is in the order of 74 to 77%.

The highest open circuit voltages that have been experimentally achieved are in the order of 1.0 eV. With respect to a band gap of 1.75 eV the band gap utilisation is quite low because of the high levels of SRH recombination and the relative broad valence and conduction-band tails. The highest stabilised efficiencies of a single junction solar cells is 10.1% [38]. It was obtained by the research Oerlikon Solar Lab in Switzerland, which currently is a subsidiary of the Japanese Tokyo Electron Ltd.

Besides a-Si:H, also nanocrystalline silicon films are used for the intrinsic absorber layers in *p*-*i*-*n* solar cell as well. The spectral utilisation of the nc-Si:H is better than that of amorphous silicon because of the lower band gap of nc-Si:H. However, to utilise the spectral part from 700 up to 950 nm, thicker films are required, because of the indirect band gap of the silicn crystallites. Typical intrinsic film thicknesses are in between 1 µm and 3 µm. The current nc-Si:H record cell has a short circuit current density of 28.8 $mA{\cdot}cm^{-2}$, an open circuit voltage of 523 mV and an efficiency of 10.8% [38]. This result was achieved by the *Japanese Institute of Advanced Industrial Science and Technology* (AIST).

Neither a-Si:H nor nc-Si:H has an optimal spectral util-

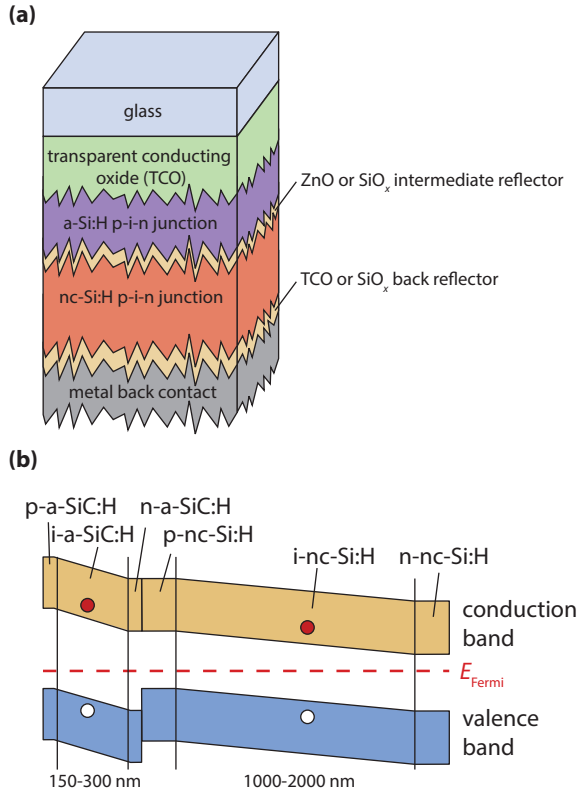

Figure 13.17: Illustrating the (a) layer structure and the (b) band diagram of an micromorph silicon solar cell.

isation. Therefore also in thin-film technology the *multijunction* approach is used, just like for III-V solar cells. Probably the most studied concept is the *micromorph* concept illustrated in Fig. 13.17 (a), which is a double junction concept consisting of an a-Si:H and a nc-Si:H junction. Like before, the solar cell with the highest bandgap is used as a top cell that converts the most energetic photons into electricity, while the lower bandgap material is used for the bottom cell and converts the lower energetic photons.

Figure 13.17 (b) shows the typical band diagram of such a micromorph solar cell, that is also called a *tandem solar cell*. On the left hand side the electronic band diagram of the amorphous silicon top cell is shown; on the right hand side the electronic band diagram of nanocrystalline silicon bottom cell is shown. The blue and green short-wavelength light is absorbed in the top cell, where electron-hole pairs are generated. Similarly, the red and infrared long-wavelength light is absorbed in the bottom cell, where also electron-hole pairs are generated. Let us take a closer look on the two electron-hole pairs excited in the top and bottom cells, respectively. The hole generated in the amorphous top cell moves to the top p layer and the electron excited in the bottom cell drifts to the bottom n layer. Both can be collected at the front and back contacts. However, the electron excited in the top cell drifts to the top n layer and the hole generated in the bottom cell drifts to the p layer. Just as for the III-V multijunction devices, the

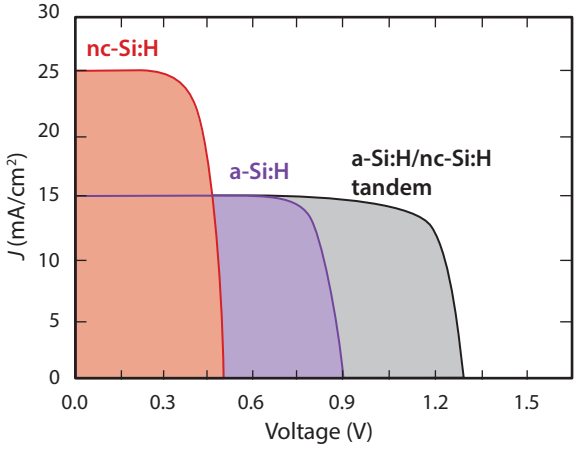

Figure 13.18: The *J-V* curve of a micromorph solar cell and its isolated subcells.

electrons and holes have to recombine at a *tunnel recombination junction* between the *n* layer of the top cell and the *p* layer of the bottom cell. Often a very thin and defect rich layer is used for this purpose. Just as for the III-V multijunction solar cells, the total current density is equal to that of the junction with the lowest current density. Therefore, for an optimised multijunction cell all current densities in the various sub cells have to be matched in order to achieve the best spectral utilisation.

Figure 13.18 shows the *J-V* curves of a single junction a-Si:H solar cell and of a single junction nc-Si:H solar cell. Let us assume that the V_{oc} of the high band gap a-Si:H top cell has an open circuit voltage of 0.9 V and a relatively low short circuit density of 15 mA·cm^{-2}, whereas the low band gap material of nc-Si:H has a lower open circuit voltage of 0.5 Volts and a higher short circuit current density of 25 mA·cm^{-2}. If we make a tandem of both junctions, the resulting current density of the double junction is lower than the currents in both bottom cells. Because the open circuit voltage is approximately proportional to $\ln(J_{ph}/J_0)$, the open circuit voltages of the junctions in a tandem cell will be slightly lower than for similar single junction cells. The total current utilisation of the tandem cell is determined by the bottom cell, *i.e.* 25 mA·cm^{-2}. Given the examples of the single junctions here, the best current matching of both cells would deliver 12.5 mA·cm^{-2}. The current tandem record cell has an efficiency of 12.3% and was manufactured by the Japanese company *Kaneka* [38].

Just as for the III-V technology, also with thin-film technology multijunction cells with more than two junctions can be made. For example the former US company *United Solar Ovonic LLC* made a thin-film silicon based triple junction device with an a-Si:H top cell, an a-SiGe:H middle cell and a nc-Si:H bottom cell, illustrated in Fig. 13.19 (a). As illustrated in this figure, also various other combinations for triple junctions can be made, for example a-Si:H/nc-Si:H/nc-Si:H. Figure 13.19

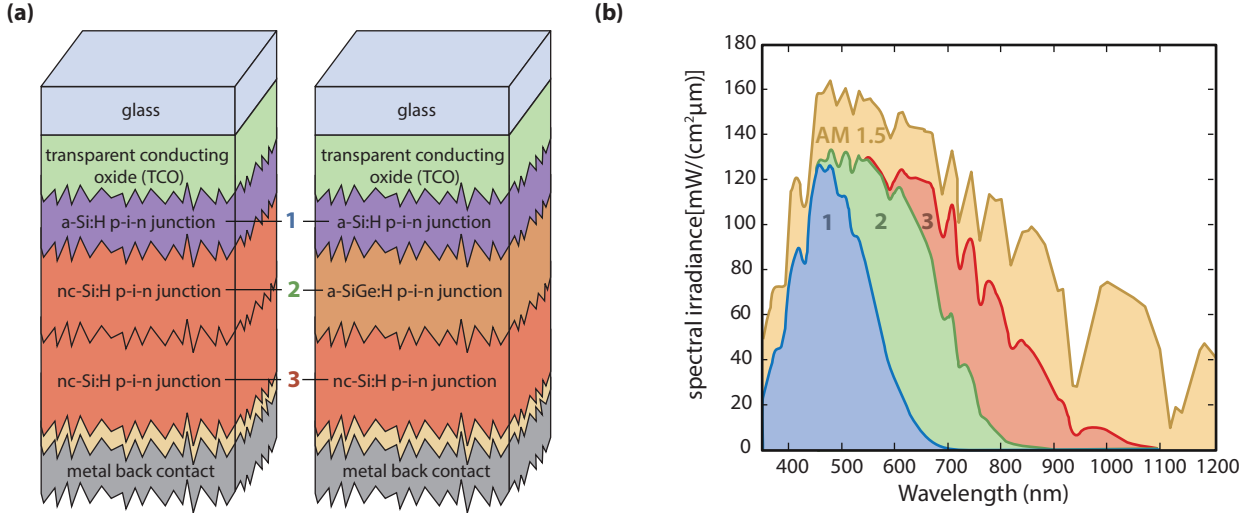

Figure 13.19: (a) Two possible structures and (b) the typical spectral utilisation (EQE·AM1.5) of a triple-junction thin-film silicon cell.

(b) shows the spectral utilisation[3] of the three junctions and the total cell of the record device by United Solar. In difference to the EQE of the multijunction III-V cell (Fig. 13.10), here the individual cells have various overlaps instead of being block functions as for the III-V device. Here, light with wavelengths below 450 nm is utilised by the top cell only, around 550 nm light is utilised by the top cell and middle cell, wavelengths at 650 nm are utilised by all three junctions, and wavelengths above 900 nm are utilised by the bottom cell only. Consequently, optimising thin film silicon multijunction solar cells is a complex interplay between the various thicknesses and light trapping concepts used in the solar cell. The *initial efficiency* of the current record triple junction cell by Uni Solar Ovonic is 16.3%.

The term *initial* refers to a big issue for amorphous thin-film silicon alloys: they suffer from *light induced degradation* which leads to a reduction of the efficiency. For example, for the initial 16.3% record cell the efficiency drops below the 13.4% of the current *stabilised* record cell. This record cell consists of an a-Si:H/nc-Si:H/nc-Si:H stack and was produced by the South Korean *LG Corp.* [38].

Light induced degradation, which is also referred to as the *Staebler-Wronski effect* (SWE), is one of the biggest challenges for the thin film solar cells. It was already discovered one year after the first a-Si:H solar cells were made in 1977 [65]. Because of the recombination of light excited charge carriers, *meta-stable* defects are created in the absorber layers. The increased defect density leads to increased bulk charge carrier recombination, which mainly affects the performance of the amorphous solar cells. After about 1000 hours of illumination, the efficiency of amorphous thin-film solar cells stabilises at around 85% - 90% of the initial efficiency and stays stable for the rest of its lifetime. If the SWE could be tackled, thin-film silicon devices easily could achieve stable efficiencies well above 16%.

Just as for III-V cells, current matching is very important for thin-film silicon multijunction cells. First, nanotextured surfaces scatter the incident light in order to enhance the average photon path length and hence to increase the absorption in the various absorber layers of the multijunction cell. Scattering becomes more important for the bottom cell because its nc-Si:H film is the thickest layer in the device has to absorb most of the red and infrared, where nc-Si:H is not that absorptive. Secondly, intermediate reflector layers are used as a tool to redistribute the light between the junctions above and below them. More specifically, the top and bottom junctions are separated with a low reflective index material, such as nc-SiO_x:H. Because of the refractive index difference of the top absorber and the nc-SiO_x:H, more light is reflected back into the top cell. Thus, the a-Si:H top cell can be made thinner, which makes it less sensitive to light-induced degradation.

[3] The spectral utilisation is given as EQE·AM1.5.

13.3.3 Making thin-film silicon solar cells

Let us now discuss how lab-scale thin-film silicon solar cells are made in the DIMES Laboratory in Delft, the Netherlands. Before the deposition can start, the sample has to be cleaned in a *ultrasonic cleaning bath.* During cleaning, dirt and dust particles are removed, which could lead to a *shunt* between the front and back contact in the final solar cell. Now, the TCO layer can be deposited. In Delft, we can deposit ZnO:Al or ITO with sputtering processes. After sputtering, ZnO:Al can be etched with acids in order to achieve a crater-like structure for light scattering. Alternatively, the sample can already be covered with a TCO layer, for example SnO_2:F from the Japanese *Asahi* company, which already has a pyramid-like structure for light scattering because of its deposition process. Also on SnO_2:F a 5-10 nm thick ZnO:Al layer is deposited that protects the SnO_2:F from being reduced by hydrogen-rich plasma present during the deposition of the thin-film silicon layers. During sputtering, the zinc oxide target is bombarded using an ionised noble gas like Argon. The generated aluminium zinc oxide species are sputtered into the chamber and deposited on to the substrate.

Before the thin-film silicon layers are deposited, a thin Al strip is deposited on the side of the sample, that will act as electric front contact when the cells are measured.

Processing of the different thin-film silicon layers often happens in multi-chamber setups that allow to process each layer in a separate chamber and therefore can prevent cross-contamination, *e.g.* from *p* and *n* dopants, which would reduce the quality of the layers. After the sample is mounted on a suitable *substrate holder*, it enters the setup via a *load lock*, in which the substrate is brought under low pressure before it's moved into the processing chambers. This avoids the processing chambers to be contaminated with various unwelcome atoms and molecules present in ambient air. Then the sample is brought into a central chamber, from where all process chambers can be accessed.

A typical precursor gas for depositing thin-film Si layers is silane (SiH_4), which often is diluted with hydrogen (H_2). The precursor gasses are brought into the process chamber at low pressure. A radio-frequency (RF) or very-high-frequency (VHF) bias voltage between two electrodes is used to generate a plasma. This plasma leads to dissociation of the SiH_4 atoms into radicals such as SiH_3, SiH_2 or SiH. This radicals react with the substrate, where a layer is deposited. With increasing the H_2 content, the material becomes more nanocrystalline, as sketched in Fig. 13.14.

As dopants, mainly diborane (B_2H_6) is used for *p* layers and phosphine (PH_3) is used for the *n* layers. For silicon carbide layers, additionally to the silane, methane (CH_4) is used. Silicon oxide layers can be made by combining silane with carbon dioxide (CO_2).

After the *p-i-n* junction is deposited, the sample is covered with a *mask*, which defines the areas on that the metallic back contacts are deposited. In Delft, we deposit silver with *evaporation*. Little pieces of silver are put in a boat that is heated by a very high current that is flowing through it. The silver melts and evaporated silver particles can move freely through a vacuum until they hit the sample, where a silver layer is deposited.

Now the solar cells are ready, every metallic square defines a little solar cell. When they are measured, the small metal strip is used as front contact and the back of the metallic square is the electric back contact.

Of course, such a configuration is not suitable for large-scale thin-film PV modules. How such a module is made, is discussed in Section 17.1 during our general discussion on PV modules.

13.3.4 Crystalline silicon thin-film solar cells

At the end of our discussion on thin-film silicon technology we briefly discuss *crystalline silicon thin-film solar cells*, where the advantages of crystalline silicon technology and thin-film silicon technology should be combined [66]. This means, that high-quality crystalline films of only several tenths of micrometers that are positioned on a substrate or a superstrate are used as absorbers.

Different approaches are investigated for reaching this goal: Kerf-less wafering techniques are investigated where thin wafers can be prepared without any kerf loss. The wafers then are transferred on a glass substrate. With this technique, efficiencies of 10.0% were obtained with a 50 µm thick absorber [66].

Alternatively, films of large-grain nanocrystalline silicon or amorphous silicon can be deposited. The amorphous films can be crystallised after deposition, for example in a thermal annealing step. Because of the low electrical quality of the defect-rich silicon films, the maximally achieved efficiency is 10.5% and was achieved by the Chinese company CSG Solar [66].

The highest efficiencies are reached when the thin crystalline layers are prepared in an epitaxy process, just as for the high-performance III-V solar cells. The epitaxy films then are transferred on glass. The current record cell made with this method has an efficiency of 20.1% on a 43 µm thick substrate and was made by the American company SOLEXEL [38].

13.4 Chalcogenide solar cells

The third class of thin-film solar cells that we discuss are the large class of *chalcogenide solar cells*, where our

focus mainly will be on CIGS and cadmium telluride (CdTe) solar cells. The term *chalcogenides* refers to all chemical compounds consisting of at least one *chalcogen* anion from the group 16 (also known as *group VI*) with at least one or more electropositive elements. Five elements belong to group 16: oxygen (O), sulphur (S), selenium (Se), tellurium (Te), and the radioactive polonium (Po). Typically, oxides are not included in discussions of chalcogenides. Because of its radioactivity, compounds with Po are of only very limited relevance for semiconductor physics.

13.4.1 Chalcopyrite solar cells

The first group of chalcogenide solar cells that we discuss are *chalcopyrite solar cells*. The name of this class of materials is based on *chalcopyrite* (copper iron disulfide, $CuFeS_2$). Like all the chalcopyrites, its forms *tetragonal* crystals, as illustrated in Fig. 13.20.

Many chalcopyrites are semiconductors. As they consist of elements from groups I, III, and VI, they are also called I-III-VI semiconductors or *ternary* semiconductors. In principle, all these combinations can be used:

$$\begin{pmatrix} \text{Cu} \\ \text{Ag} \\ \text{Au} \end{pmatrix} \begin{pmatrix} \text{Al} \\ \text{Ga} \\ \text{In} \end{pmatrix} \begin{pmatrix} \text{S} \\ \text{Se} \\ \text{Te} \end{pmatrix}_2$$

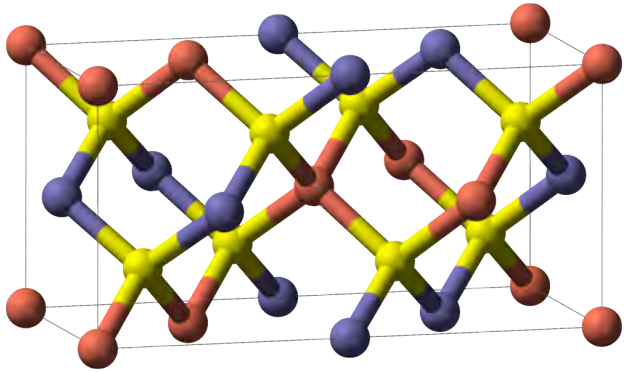

Figure 13.20: The crystal structure of *copper indium diselenide*, a typical chalcopyrite. The colors indicate copper (red), selenium (yellow) and indium (blue). For *copper gallium diselenide*, the In atoms are replaced by Ga atoms [67].

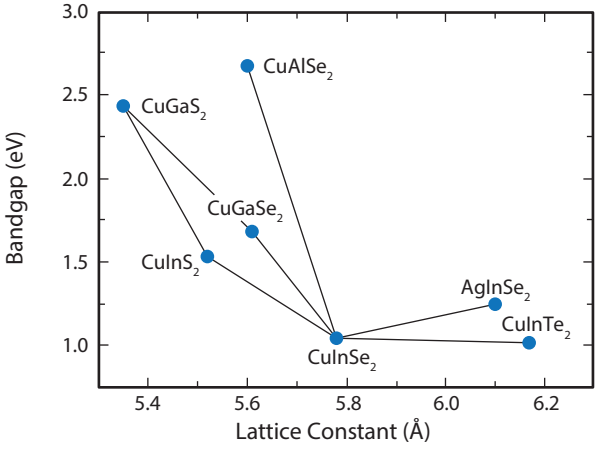

Figure 13.21: The bandgap vs. the lattice constant for several chalcopyrite materials. Data taken from [42].

Figure 13.21 shows the bandgap vs. the lattice structure for several chalcopyrite materials.

The most common chalcopyrite used for solar cells is a mixture *copper indium diselenide* ($CuInSe_2$, CIS) and copper gallium diselenide ($CuGaSe_2$, CGS). This mixture is called *copper indium gallium diselenide* [$Cu(In_xGa_{1-x})Se_2$, CIGS], where the x can vary between 0 and 1. Several research groups and companies use a compound that also contains sulfur; it is called *copper indium gallium diselenide/disulfide* [$Cu(In_xGa_{1-x})(Se_yS_{1-y})_2$, CIGSS], where y is a number in between 0 and 1.

The physical properties of CIGS(S) are rather complex and many different views exist on these properties among scientists. $CuInSe_2$ has a band gap of 1.0 eV, the bandgap of $CuInS_2$ is 1.5 eV and $CuGaSe_2$ has a band gap of 1.7 eV. By tuning the In:Ga ratio x and the Se:S ratio y, the band gap of CIGS can be tuned from 1.0 eV to 1.7 eV. As CIGS(S) is a a direct band gap semiconductor material, it has a large absorption coefficient, hence an absorber thickness of 1-2 µm is sufficient for absorbing a large fraction of the fraction of the light with energies above the band gap energy. Also the typical electron diffusion length is the order of a few micrometers. CIGS(S) is a p-type semiconductor, the p-type character resulting from intrinsic defects in the material that amongst others are related to Cu deficiencies. The many different types of defects in CIGS(S) and their properties are a topic of ongoing research.

Figure 13.22 (a) illustrates a typical CIGS solar cell structure deposited on glass, which acts as a substrate. On top of the glass a *molybdenum layer* (Mo) of typically 500 nm thick is deposited, which acts as the electric back contact. Then, the p-type CIGS absorber layer is deposited with a thickness up to 2 µm. Onto the p-CIGS layer, a thin n-CIGS layer is deposited, for example an indium/gallium rich $Cu(In_xGa_{1-x})_3Se_5$ alloy. The pn-junction is formed by stacking a thin *cadmium sulfide* (CdS) *buffer layer* of around 50 nm thickness onto

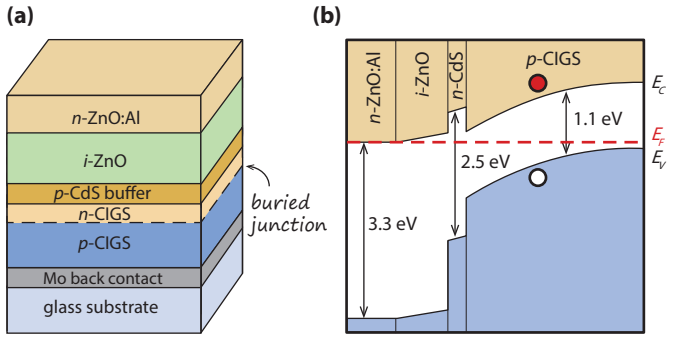

Figure 13.22: The (a) layer structure and (b) band diagram of a typical CIGS solar cell.

the CIGS layers. The *n*-type region is extended with the TCO layer, that also is of *n* type: First an intrinsic *zinc oxide* (ZnO) layer deposited followed by a layer of Al doped ZnO. The Al is used as an *n* dopant for the ZnO. Similar to thin film silicon technology, the *n*-type TCO acts as the transparent front contact for the solar cell.

Figure 13.22 (b) shows the electronic band diagram of a CIGS solar cell. The light enters the cell from the left, via the ZnO. The *p*-type CIGS absorber layers used in industrial modules typically has a band gap of 1.1-1.2 eV, which is achieved using $Cu(In_xGa_{1-x})Se_2$ with $x \approx 0.3$ [68]. The *n*-type CdS buffer layer has a band gap of 2.5 eV. The band gaps of the *n*- and *p*-type materials are different, which means that such CIGS cells can be can be considered as heterojunctions. The band gap of ZnO is very large with values of 3.2 eV or even higher, which minimises the parasitic absorption losses in this device [69].

The defect density at the surface is higher than in the bulk, which could be a loss mechanism for the minority charge carriers. This recombination can be reduced by placing an *n*-type CIGS layer between the *p*-type CIGS and the *n*-type CdS layers, as already mentioned above. The *pn*-junction within the CIGS is called a *buried junction*. There the electron-hole pairs are separated. In *p*-CIGS, which is Cu-deficient, the dominant recombination mechanism is Shockley-Read-Hall recombination in the bulk. In contrast, in Cu rich CIGS films the SRH recombination at the CIGS/CdS interface becomes dominant.

A very important issue in the development of CIGS solar cells is the role of *sodium* (Na). A low contamination with sodium appears to reduce the recombination in *p*-type CIGS materials because of better recombination of the grain boundaries. The reduced recombination rate results in a higher band gap utilisation and thus a higher open circuit voltages. Typically, the optimal concentration of sodium in the CIGS layers is about 0.1%. Often, as a Na source soda-lime glass is used, that is also the substrate for the solar cell. If no soda-lime glass is used, the Na has to be intentionally added during the deposition process. The scientific

question why Na significantly improves several properties of the CIGS films is still not completely solved. Currently, also the influence of *potassium* (K) is heavily investigated.

Fabrication of CIGS solar cells

CIGS films can be deposited with various different technologies. Because many of these activities are developed within companies, not much detailed information is available on many of these processing techniques. One example of a processing technique is *co-evaporation* under vacuum conditions. Using various crucibles of copper, gallium, selenium and indium, the precursors in are co-evaporated onto a substrate in various steps. Two approaches can be used: First, co-evaporation on a heated substrate. During the process an additional selenium source is used, such that a CIGS film can be formed. The second approach is sputtering onto a substrate at room temperature. After that, the substrate is *thermally annealed* under presence of selenium vapour, such that the CIGS structure can be formed. Alternatively, a selenium rich layer can be deposited on top of the initially deposited alloy, followed by an annealing step. Because of the variety and complexity of the reactions taking place during the *'selenisation'* process, the properties of CIGS are difficult to control. Companies that use or have used co-evaporation process are Würth Solar, Global Solar and Ascent Solar Technologies. Among CIGS companies using sputter approaches are Solar Frontier, Avancis, MiaSolé and Honda Soltec.

An alternative approach to produce CIGS layers is based on a *wafer bonding* technique. Two different films are deposited onto a substrate and a superstrate, repspectively. Then, the films are pressed together on high pressure. During annealing, the film is released from the superstrate and a CIGS film remains on a substrate.

Non-vacuum techniques can be based on depositing nanoparticles of the precursor materials on a substrate after which the film is sintered. During the *sintering process* films are made out of powder. For achieving this, the powder is heated up to a temperature below the melting point. At such high temperatures, atoms in the particles can diffuse across the particle boundaries. The particles thus fuse together forming one solid. This process has to be followed by a *selenisation step*.

An important advantage of the CIGS PV technology is that on it has achieved the highest conversion efficiencies among the different thin-film techologies except the III-V technology. The current record for lab-scale CIGS solar cells processed on glass is 20.8% and was achieved by the German research institute ZSW [38]. This record cell has a V_{oc} of 0.757 V, a J_{sc} of 34.77 mA/cm^2, and a FF of 79.2%. The world record on flexible substrates has been obtained at the Swiss Fed-

eral Laboratories for Materials Science and Technology (EMPA). On flexible polymer foil they achieved a conversion efficiency of 20.4% [70].

For making CIGS modules, the interconnection is done similar as for other thin-film technologies and is discussed in detail in Section 17.1. As generally true for the different PV technologies, the record efficiencies of modules are significantly lower than that of lab-scale cells. The record efficiencies of 1 m^2 modules are in the order of 13%, whereas the aperture-area efficiencies are just above 14% as confirmed by NREL. The German manufacturer Manz AG has presented an 15.9% aperture area efficiency and a total area efficiency of 14.6%. The Japanese company Solar Frontier claims a 17.8% aperture area efficiency on a small module of 900 cm^2 size.

Despite the very high conversion efficiencies, the CIGS technology faces several challenges. As CIGS is a deposited in a complex deposition processes, it is challenging to perform large area deposition with a high *production yield*. The production yield is the percentage of all the modules coming out of the production line that fulfil the product specifications.

Kesterites

Figure 13.1 shows the abundance in the Earth's crust for several elements. Is we can see, indium is a very rare element. As we have seen above, In is a crucial element of CIGS solar cells. Because of its scarcity, In might be *the* limiting step in the upscaling of the CIGS PV technology to future Terawatt scales. In addition, the current thin-film display industry depends on In as well, as ITO is integrated in many display screens.

As a consequence, other chalcogenic semiconductors are investigated that do not contain rare elements. A interesting class of materials are the *kesterites* which are *quarternary* or *pentary* semiconductors consisting of four or five elements, respectively. While the name giving mineral kesterite [$Cu_2(ZnFe)SnS_4$], where zinc and iron atoms can substitute each other, is not used as a semiconductor, kesterite without iron (Cu_2ZnSnS_4) is used. It also is known as *copper zinc tin sulfide* (CZTS) and is a I_2-II-IV-VI_4 semiconductor. Other kesterites are for example *copper zinc tin selenide* ($Cu_2ZnSnSe_4$, CZTSe), or ones using a mixture of sulphur and selenium, $Cu_2ZnSn(SSe)_4$ (CZTSS).

In difference to CIGS, CZTS is based on non-toxic and abundantly available elements. The current record efficiency is 12%. It is achieved with an CZTSS solar cells on lab-scale by IBM [38].

13.4.2 Cadmium telluride solar cells

In this section we will discuss the *cadmium telluride* (CdTe) technology, which currently is the thin-film technology with the lowest demonstrated cost per W_p. We start with discussing the physical properties of CdTe, which is a II-VI semiconductor because it consists of the II valence electron element cadmium (Cd) and the VI valence electron element tellurium (Te). Like the III-V semiconductors discussed in Section 13.2, CdTe forms a *zincblende* lattice structure where every Cd atom is bonded to four Te atoms and vice versa.

The band gap of CdTe is 1.44 eV, a value which is close to the optimal band gap for single junction solar cell. CdTe is a direct band gap material, consequently only a few micrometres of CdTe are required to absorb all the photons with an energy higher than the band gap energy. If the light-excited charge carriers should be efficiently collected at the contacts, their diffusion has to be in the order of the thickness.

CdTe can be *n*-doped by replacing the II-valence electron atom Cd with a III-valence electron atom like aluminium, gallium or indium. *n*-doping can be achieved as well by replacing a VI-valence tellurium atom with a VII-valence electron element like fluorine, chlorine, bromine and iodine atoms. The III- and VII-valence atoms act as shallow donors. Also a tellurium vacancy acts like a donor.

p-doping of CdTe can be achieved by replacing Cd with a I-valence electron atom like copper, silver or gold. It can be achieved as well by replacing Te atoms with V-valence electron elements such as nitrogen, phosphorus or arsenicum. These elements act as shallow acceptors. But also a Cd vacancy acts as an acceptor. In solar cells, *p*-doped CdTe is used. However, it is difficult to obtain CdTe with a high doping level.

Figure 13.23 (a) shows the structure of a typical CdTe solar cell. First, transparent front contact is deposited onto the glass superstrate. This can be *tin oxide* or *cadmium stannate*, which is a Cd-Sn-oxide alloy. On top of that the *n*-layer is deposited, which is a *cadmium sulfide* layer, similar to the *n*-uffer layer in CIGS solar cells (Section 13.4.1). Then, the *p*-type CdTe absorber layer is deposited with a typical thickness of a few micrometres. Making a good back contact on CdTe is rather challenging because the material properties of CdTe do not allow a large choice of acceptable metals. Heavily doping the contact area with a semiconductor material improves the contact qualities, however, achieving high doping levels in CdTe is problematic. Copper containing contacts have been used as back contacts, however, on the long term they may face instability problems due to diffusion of Cu through the CdTe layer up to the CdS buffer layer. Nowadays, stable *antimony telluride* layers in combination with molybdenum are used.

The band diagram of a CdTe solar cell is shown on Fig.

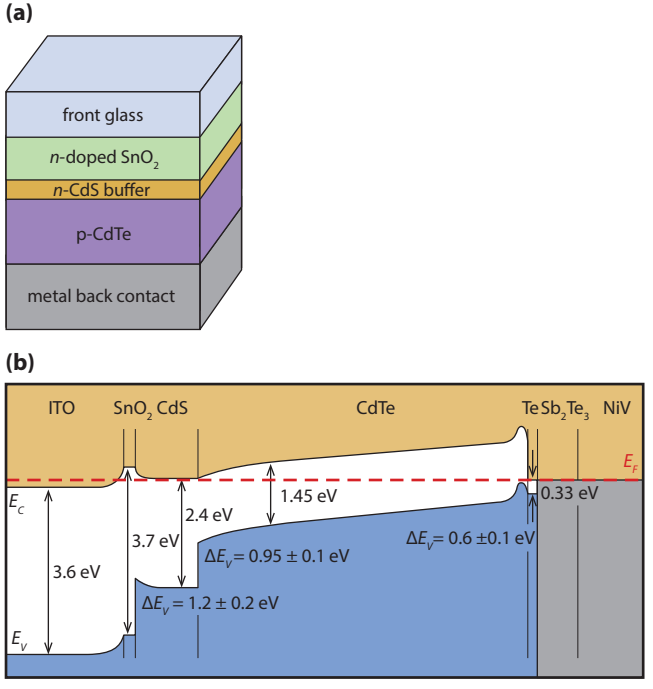

Figure 13.23: The (a) layer structure and (b) band diagram of a typical CIGS solar cell. Band diagram data taken from [71].

13.23 (b). The *p*-type semiconductor CdTe has a band gap of 1.44 eV whereas the *n*-type CdS has a band gap of 2.4 eV. Consequently, the junction is a heterojunction, similar to what we have seen for CIGS PV devices. The light excited minority electrons in the *p*-layer are separated at the heterojunction and collected at the TCO based front contact. The holes are collected at the back contact.

Manufacturing CdTe solar cells

Usually, the CdS/CdTe layers are processed using the *closed space sublimation method*. In this method, the source and the substrate are placed at a short distance from each other, ranging from a few millimetres up to several cm, under vacuum conditions.The source can either consist of granulates or powders of CdTe. Both, the source and the substrate are heated, where the source is kept at a higher temperature than the substrate. This temperature difference induces a driven force on the precursors, which are deposited on the substrate. Such, the bulk *p*-type CdTe is formed. In the process chamber, an inert carrier gas like argon or nitrogen can be used.

The US company First Solar Inc. is the leading company in the CdTe technology [72]. Like the German company Antec Solar, they use the closed space sublimation method. Another company producing CdTe

solar cells is the German Calyxo. First Solar is by far the largest CdTe manufacturer. In 2008, First Solar had an annual production rate of 500 MW. In 2006 and 2007 it already was one of the biggest solar module manufacturers in the world.

The record conversion efficiency of lab-scale solar CdTe solar cells is 19.6% [38] and was obtained by GE Global Research in 2013. The open circuit voltage of this cell is 857.3 mV, the short circuit current density is 28.59 mA/cm^2, and the FF is 80.0%. The current record module based on CdTe technology has a conversion efficiency of 16.1% at an area of 7200 cm^2.

The current cost of modules from First Solar product is in the order of 0.68 to 0.70 USD/W_p and is expected to drop to 0.40 USD/W_p, hence keeping the cost per W_p lower than that of modules based on crystalline silicon wafers.

An important aspect that needs to be addressed is the toxicity of cadmium. However, insoluble Cd compounds like CdTe and CdS are much less toxic than the elementary Cd. Nonetheless, it is very important to prevent cadmium entering into the ecosystem. It is an important question whether, CdTe modules could become a major source of Cd pollution. AS OF»», First Solar has an installed production capacity of 2 GW/year. With this capacity, about 2% of the total industrial Cd consumption would be taken by First Solar. Nevertheless, recycling schemes have been set up for installed CdTe solar modules. For example, First Solar has a recycling scheme in which a deposit of 5 dollar cents per W_p is included to cover the cost for the recycling at the end of the module lifetime.

Maybe an even bigger challenge for the CdTe technology is the supply with tellurium. Let us take another look at Fig. 13.1 that shows the abundance of the various elements in the Earth's crust. Tellurium is one of the rarest stable solid elements in the Earth's crust with an abundance of about 1 µg/kg, which is comparable to that of platinum [73]. Because of its scarcity, the supply with Te might be the limiting step to upscale the CdTe PV technology to future Terawatt scales. On the other hand, Tellurium as source material has only had a few users so far. Thus, for Te no dedicated mining has been explored so far. In addition, new supplies of tellurium-rich ores were found in Xinju, China. Therefore at this moment it remains unclear to which extent the CdTe PV technology might be limited by the supply with tellurium.

13.5 Organic Photovoltaics

So far we discussed inorganic thin-film semiconductor materials such as III-V semiconductors, amorphous and nanocrystalline silicon, CIGS and CdTe solar cells. In this section we will take a look at *organic photovoltaics*.

The used absorber materials are either *conductive organic polymers* or *organic molecules* that are based on carbon, which may form a cyclic, a-cyclic, linear or mixed compound structure.

Figure 13.24 shows some example of organic materials that can be used for PV applications: P3HT, Phtalocyanine, PCBM and Ruthenium Dye N3. All these materials can be considered as large conjugated systems, which means that carbon atoms in the chain have an alternating single or a double bond and every atom in the chain has a p-orbital available. In such conjugated compounds, the p-orbitals are delocalised, which means that they can form one big mixed orbital. Hence, the valence electron of the original p-orbital is shared over all the orbitals. A classical example would be the benzene molecule, which is a cyclic conjugated compound. As we can see in Fig. 13.25 (a), This molecule has 6 carbon atoms and six p-orbitals, which mix and from two circular orbitals that contain three electrons each. These electrons do not belong to one single atom but to a group of atoms.

In contrast, a methane molecule (CH_4) is tretrahedrally coordinated, which means that it has 4 equivalent sp^3 hydride bonds with a bond angle of 109.5°. An ethene (C_2H_4) molecule has three equivalent sp^2 hybrid bonds with a bond angle of 120° plus an electron in a p^z orbital. Two neighbouring p^z orbitals form a so-called π orbital, as illustrated in 13.25 (b).

In Chapter 6 we have discussed that two individual sp^3 hybrid orbitals can make an anti-bonding and a bonding state. The same is valid for the two p^z orbitals forming a molecular π orbital. They can make bonding and anti-bonding π states. Therefore, conjugated molecules can have similar properties as semiconductor materials.

At room temperature, most electrons are in the bonding state, which is also called the *highest occupied molecular orbital* (HUMO). The anti-bonding state can be considered as the *lowest unoccupied molecular orbital* (LUMO). As the conjugated molecules are getting longer, the HOMO and LUMO will broaden and act similar to valence and conduction band in conventional semiconductors. The energy difference between the HOMO and LUMO levels can be considered as the band gap of the polymer material.

To discuss whether an organic material is *p*-type or *n*-type, we first have to introduce the concept of the *vacuum level*, shown in Fig. 13.26 (a). The vacuum level is defined as the energy of a free stationary electron that is outside of any material, or in other words, *in vacuo*. This level often is used as the level of alignment for the energy levels of two different materials. The *ionisation energy* is the energy required to excite an electron from the valence band or HOMO to the vacuum state. The *electron affinity* is the energy obtained by moving an electron form the vacuum just outside the semicon-

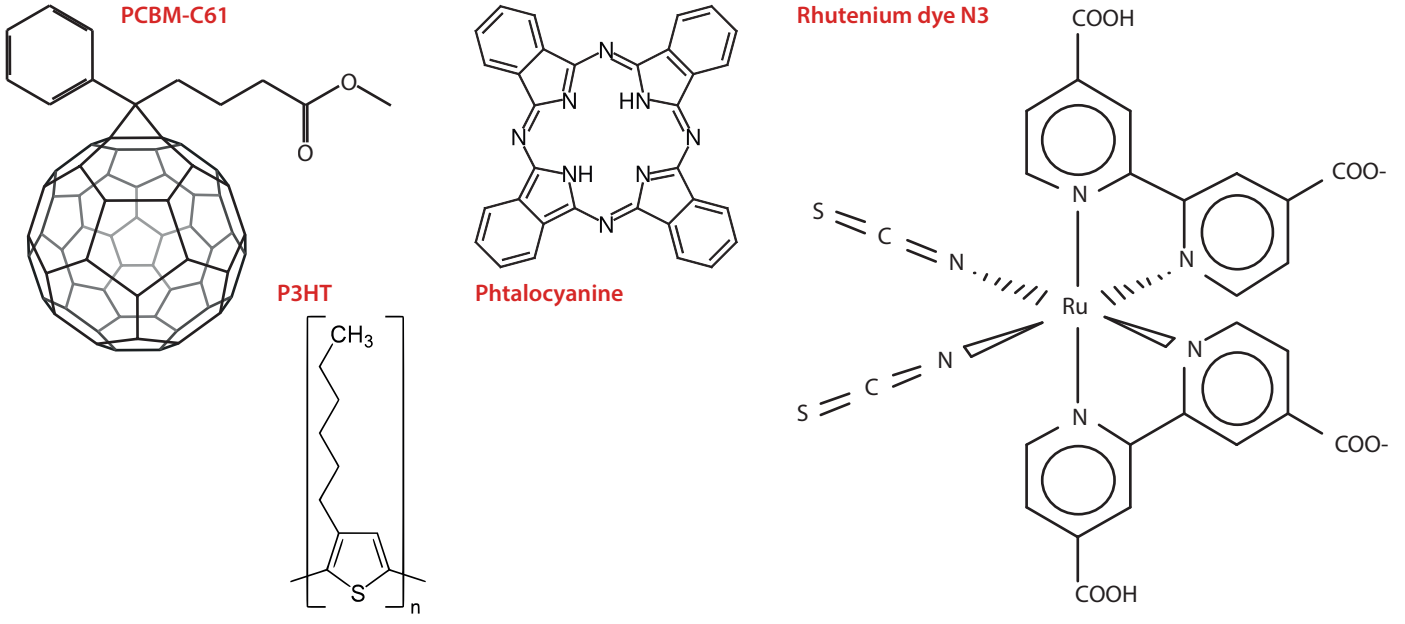

Figure 13.24: Some examples of organic molecules used for organic photovoltaics.

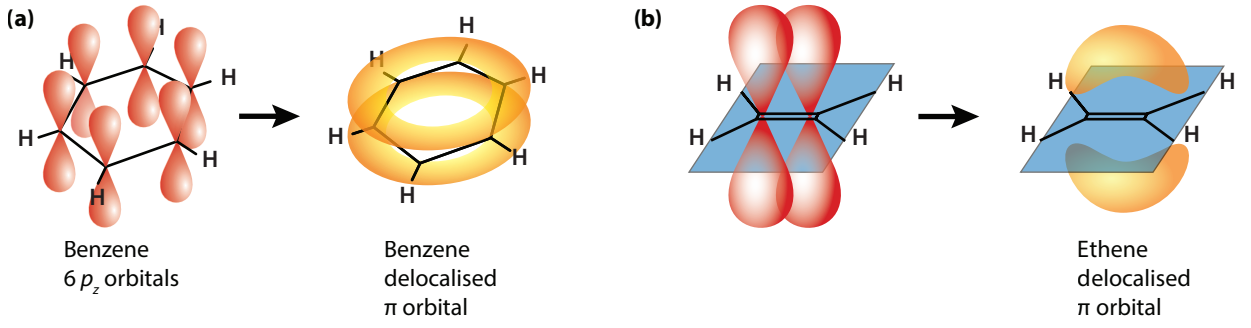

Figure 13.25: The chemical structure and π orbitals of (a) benzene and (b) ethene.

ductor or conjugated polymer to the bottom of the conduction band or LUMO. When a material has a low ionisation potential, it can release an electron out of the material relatively easy, *i.e.* it can act as an *electron donor*. On the other hand, when a material has a high electron affinity, it can easily accept an additional electron in the LUMO or conduction band, it thus acts as an *electron acceptor*.

As we have discussed earlier, in inorganic semiconductors an electron can be excited from the valence band to the conduction band leaving a hole in the valence band. Such an electron-hole pair is only weakly bound and both entities are easily separated and can diffuse away from each other. In organic materials this is not the case. Absorption of a photon with sufficient energy results in the creation of an *exciton*, illustrated in Fig. 13.26 (b) ❶. An exciton is an excited electron-hole pair that is still in bound state because of the mutual Coulomb forces between the particles. Such excitons can diffuse through the material, but they have a low life time in organic materials, recombining back to the ground state within a few nanoseconds. Hence, the diffusion length of such excitons is in the order of only 10 nm.

If an electron donor and an electron acceptor material are brought together, an interface is formed between those two. The HOMO and LUMO of both polymers can be aligned considering their energy levels with reference to the vacuum level, as illustrated in Fig. 13.26 (b). At the interface we see a difference in the HOMO

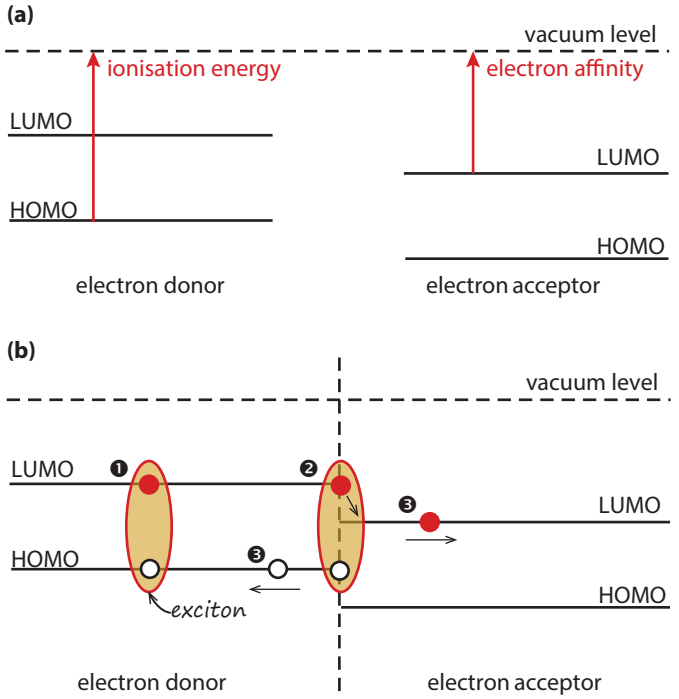

Figure 13.26: Illustrating (a) the energy levels in organic PV materials and (b) the separation of electrons and holes in an exciton at the acceptor-donor interface..

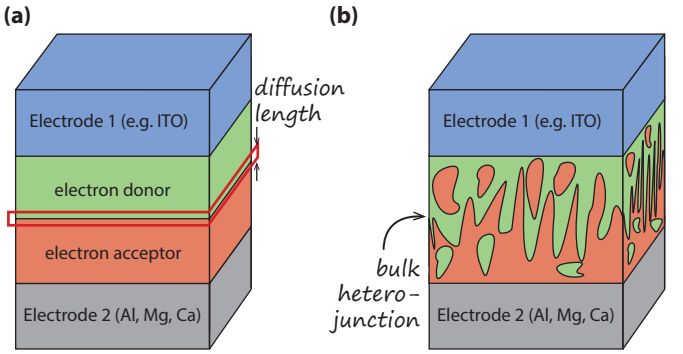

Figure 13.27: Illustrating (a) the layer structure of organic solar cells and (b) an organic solar cell with a *bulk heterojunction*.

and LUMO levels. Because of this difference, an electrostatic force exists between the two materials. If the materials are chosen such that the difference is large enough, these local electric fields can break up the exciton ❷. The electron then can be injected into the electron acceptor and a hole remains in the electron donor material ❸.

Figure 13.27 (a) shows the structure of a typical organic solar cell. Here, we consider a solar cell consisting of both organic acceptor-type and donor-type materials. Similar as for semiconductor materials, a heterojunction based on two different materials or conjugated compounds can be constructed. As mentioned before, the

typical diffusion length in the organic materials is only about 10 nm. Hence, the thickness of the solar cell in principle is strongly limited by the diffusion length, while it has to be at least 100 nm to absorb a sufficient fraction of the light. Therefore, organic solar cells are based on *bulk heterojunction* photovoltaic devices that is illustrated in Fig. 13.27 (b), where the electron-donor and the electron-acceptor materials are mixed together and form a *blend*. In this way, typical length scales in the order of the exciton diffusion lengths can be achieved. Hence, a large fraction of the excitons can reach an interface, where they are separated into an electron and a hole. The electrons move though the acceptor material to the electrode and the holes move through the donor material to be collected at the other electrode. Usually, the holes are usually collected at a TCO electrode, for example *indium tin oxide* (ITO). The electrons are collected at a metal back electrode.

The record organic solar cells are based on double junctions. Heliatek achieved a 12.0% solar cell efficiency on lab-scale, with unknown stability.

An advantage of organic solar cells is that they have low production costs. With chemical engineering the band gap can be tuned. Organic solar cell can be integrated into flexible substrates. Important disadvantages are a low efficiency, low stability and low strength compared to inorganic PV cells.

Although organic solar cells themselves can be cheaply, the need of highly expensive encapsulation materials to protec the organic materials from humidity, moisture and air limits the industrial application. To my knowledge, no company is producing organic PV modules at the moment. Konarka Technologies has been active in the past. Their small PV modules had efficiencies in the range of 3 up to 5% and worked for a couple of years only.

13.6 Hybrid organic-inorganic solar cells

In this section we discuss two hybrid concepts, where the junction consists of both inorganic and organic compounds. These concepts are *dye-sensitised solar cells* and *perovskite solar cells*.

13.6.1 Dye-sensitised solar cells

An alternative solar cell concept based on organic materials is the dye-sensitised solar cell (DSSC), which is a photo-electrochemical system. It contains *titanium dioxide* (TiO_2) nanoparticles, organic *dye* particles, an electrolyte and platinum contact. Figure 13.28 (a) shows an illustration of a DSSC. The dye sensitiser, which acts like an electron donor, is the photo-active component

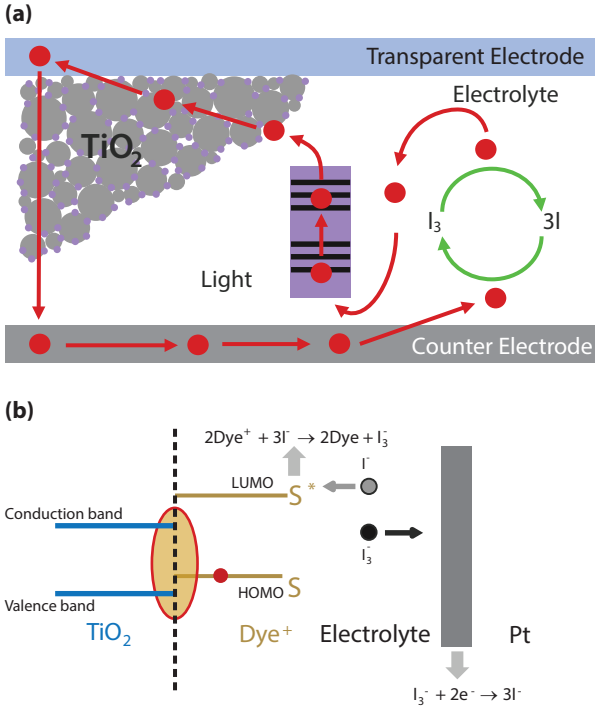

Figure 13.28: (a) A sketch of a dye sensitised solar cell and (b) the relevant energy levels in its operation.

of the solar cell. The second major component are the TiO_2 nanoparticles that act as the electron–acceptor. Similar to organic bulk heterojunction cells, the dye is mixed with TiO_2.

As a photoactive dye sensitiser, *ruthenium polypyridine* is used. The operation is illustrated in 13.28 (b). If a photon is absorbed by the ruthenium polypyridine, it can excite an electron from its ground state, the S state, to an excited state, referred to as S*. In this case the S can be considered as a HOMO and the S* state can be consider as LUMO. The S* energy level is above the energy level of the conduction band of the TiO_2. As a result, the light excited electrons are injected into the TiO_2 nanoparticles, while the dye-sensitiser molecule remains positively charged. The electrons move through the TiO_2 to the TCO based back contact in a diffusion based transport mechanism. Via the electric circuit, the electrons move to the front contact. The front contact is electrically connected to the dye via an *electrolyte*. Electrolytes are solutions or compounds that contain ionised entities that can conduct electricity. The typical electrolyte used for DSSC contains *iodine*. The positively charged oxidised dye molecule is neutralised by a negatively charge iodide. Three negatively charged iodides, neutralise two dye molecules and create one negatively charged tri-iodide. This negatively charged tri-iodide moves to the counter electrode where it is reduced using two electrons into three negatively charged iodine. To facilitate the chemical reactions, these *photoelectro-*

chemical cells require a *platinum* back contact.

The performance of a DSSC depends on the HOMO and LUMO level of the dye material, the Fermi level of TiO_2 nanoparticles and the *redox potential* of the iodide and triodide reactions. The current record efficiency of dye-sensitised PV devices on lab-scale is 11.9% and was achieved by Sharp [38].

The major advantage of DSSC are the low production costs. A disadvantage is the stability of the electrolyte under various weather conditions: At low temperatures the electrolyte can freeze, which stops the device from generating power and might even result in physical damage. High temperatures result in thermal expansion of the electrolyte, which make encapsulating modules more complicated. Another challenge is the high cost of the platinum electrodes, hence replacing the platinum with cheaper materials is a topic of ongoing research. Secondly, more stable and resistive electrolyte materials must be developed. Thirdly, research is done on improved dyes that enhance the spectral and band gap utilisation of the solar cells.

Yet, no dye-sensitised PV products are available commercially.

14 Third Generation Concepts

The term *third generation photovoltaics* refers to all novel approaches that aim to overcome the Shockley-Queisser (SQ) single bandgap limit, preferably at a low cost. The SQ limit that we discussed extensively in Chapter 10 is a thermodynamic approach to estimate the maximum efficiency of a single junction solar cell in dependence of the bandgap of its semiconductor material. For its derivation we assumed that the AM1.5 spectrum is incident on a solar cell. Further, we do not allow the solar cell to increase in temperature but we force it to keep the ambient temperature of 300 K. This means that all energy absorbed by the solar cell can escape the solar cell by either the generated current density or by radiative recombination of charge carriers. Under these assumptions the efficiency limit is around 33% in the band gap range from 1.0 eV up to 1.8 eV.

Now we are going to look at some very fundamental limitations of classical single-junction solar cells.

First, in single-junction solar cells only on band gap material is used. Hence, a large fraction of the energy of the most energetic photons is lost as heat as illustrated in Fig. 10.7.

Secondly, most solar cells concepts are based on an incident irradiance level of 1 sun. However, higher irradiance means more current generation and also higher voltage levels, resulting in a higher overall efficiency.

Thirdly, every photon only excites one electron in the conduction band creating only one electron-hole pair.

The energy of highly energetic photons could be utilised better if they would create more than one excited electron in the conduction band.

Fourthly, the photons with energies below the bandgap are not used. Hence, they do not result in charge carrier excitation, as illustrated in Fig. 10.7.

If these fundamental limitations could be solved, PV concepts with conversion efficiencies exceeding the Shockley-Queisser limit could be developed. In this chapter, we will discuss several third generation concepts: Multi-junction solar cells, concentrator photovoltaics, spectral up and down conversion, multi-exciton generation, intermediate band-gap solar cells and hot carrier solar cells. Note, that besides multi-junction and the concentrator approach, none of these concepts have resulted in high efficiency solar cells or even been demonstrated yet. These other concepts are still in fundamental research phase and it is not clear whether they will ever become a large scale PV technology [74].

14.1 Multi-junction solar cells

The first limitation discussed in the list above can be attacked by using *multi-junction* solar cells. Depending on the author, these solar cells are seen as part of the second or the third generation, and indeed we already briefly discussed them in Chapter 13 on thin-film solar cells when looking at the III-V technology (Section 13.2) and thin-film silicon technology (Section 13.3). For completeness, we here will summarise this technology.

In multi-junction cells, several cell materials with different bandgaps are combined in order to maximise the amount of the sun light that can be converted into electricity, as we already illustrated in Fig. 10.9. To realise this, two or more cells are stacked onto each other. The top cell has the highest bandgap, in order to absorb and convert the short wavelength (blue) light. Light with wavelengths longer than the bandgap-wavelength can traverse the top cell and be absorbed in the cells below with lower bandgaps. The bottom cell has the lowest bandgap to absorb the long wavelength (red and near infrared) light. In order to optimise the performance of multi-junction solar cells with two electrical terminals, matching the currents of all the subcells (current matching) is crucial. Multi-junction cells with more terminals do not have this restriction, but their production is more complicated.

In thin-film silicon tandem cells, an a-Si:H top cell is stacked onto a nc-Si:H bottom cell. In order to achieve current matching, the top cell is much thinner than the bottom cell. The cell can be further optimised by using an intermediate reflector between the top and the bottom cell in order to reflect the blue light back into the top cell while letting the red light pass to the bottom cell. The reported record efficiency of a-Si:H/nc-Si:H

tandem cells is 12.3% and for a a-Si:H/nc-Si:H/nc-Si:H it is 13.4% [38].

Multi-junction cells containing III-V semiconductors are at present the most efficient solar cells. The current world record efficiency is 44.4% for a triple-junction GaInP/GaAs/GaInNAs cell that is used in a concentrated PV-system [38]. Because of the high concentration of 302 the overall efficiency increases and hence we also tackle the second limitation mentioned in the table above. As a result the SQ limit can be exceeded by more than 10%.

14.2 Spectral conversion

Single-junction solar cells have the limitation that every photon can only generate one collected electron at a maximum. In theory, this limitation can be tackled by *spectral conversion*. The main idea of spectral conversion to use an additional layer consisting of a "magical" material such that the incident spectrum can be altered. Materials that are investigated for spectral conversion are organic dyes, quantum dots, lanthanide ions, and transition metal ion systems [75].

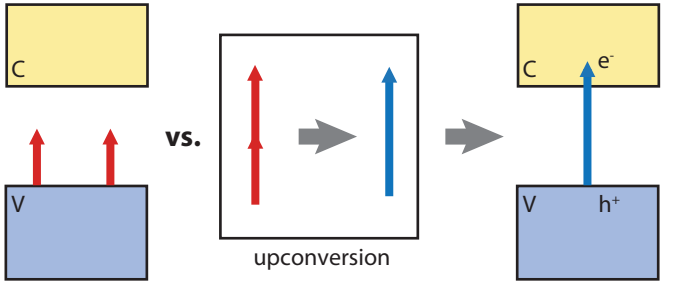

Figure 14.1: Illustrating the principle of spectral upconversion.

14.2.1 Spectral upconversion

In *spectral upconversion*, two or more low-energy photons excite electron-hole pairs in several steps, as illustrated in Fig. 14.1. If these electron-hole pairs recombine radiatively, they emit a photon with a higher energy. As a result, two or more low-energy photons are converted in one high-energy photon, which can be absorbed in the PV-active material.

The up-converting layers should be placed at the back of the solar cell as low-energy photons can pass through the solar cell absorber to the up-converting layer and can be converted there to high energy photons that are absorbed by the solar cell in a second pass [75]. In this way, also up-converting layers with

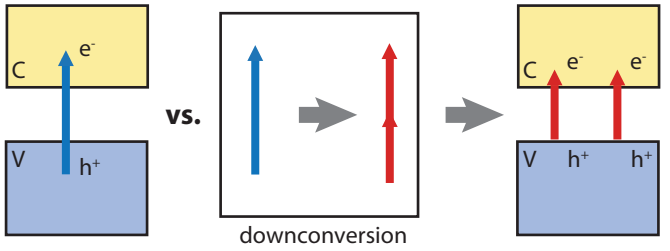

Figure 14.2: Illustrating the principle of spectral down-conversion.

low efficiencies become interesting. Due to parasitic absorption it is not advised to put these layers at the front of the solar cell.

14.2.2 Spectral downconversion

The idea of *spectral down-conversion* is to split one high-energy photon in multiple lower energetic photons [76], as illustrated in Fig 14.2. A high-energy photon is absorbed at the front of the solar cell and converted into at least two photons with lower energies. If the energy of the initial photon is $E_{\mathrm{ph}} > 2E_g$ and the energy the resulting two photons is still larger than that of the band gap of the absorber material, both photons can be absorbed and used for exciting charge carriers. As a result, a high energetic photon, for example in the blue, can result in two excited electrons in the blue part. In other words, the maximum theoretical EQE of 100% at the wavelength of the blue photon can be increased to 200%. If the photon had sufficient energy to be split into three photons with an energy higher than the band gap, a theoretical EQE of 300% could be obtained. In difference to upconversion, a down-converting layer has to be at the front of the solar cell, as highly energetic photons are always absorbed in the absorber layer. Hence, parasitic absorption might be a problem in this technology.

One possibility that is investigated for realising spectral down-conversion is to use so-called *quantum dots* (QDs). These are small spherical nanoparticles made of semiconductor materials with typical diameters of a few nanometers, as illustrated in Fig. 14.3 (a). These semiconductor particles still behave like a semiconductor material; however, due to so-called *quantum confinement* the band gap of the semiconductor quantum dots can be larger than that of the same semiconductor in a bulk configuration. The band gap of the QDs can be tuned by varying their size. The smaller the particles, the larger the band gap. This enables interesting opportunities for bandgap engineering, such as a multijunction solar cells based on junctions with different QDs of different sizes in every junction.

If QDs should be used for down-conversion, an ensemble of nanoparticles is embedded in a host mater-

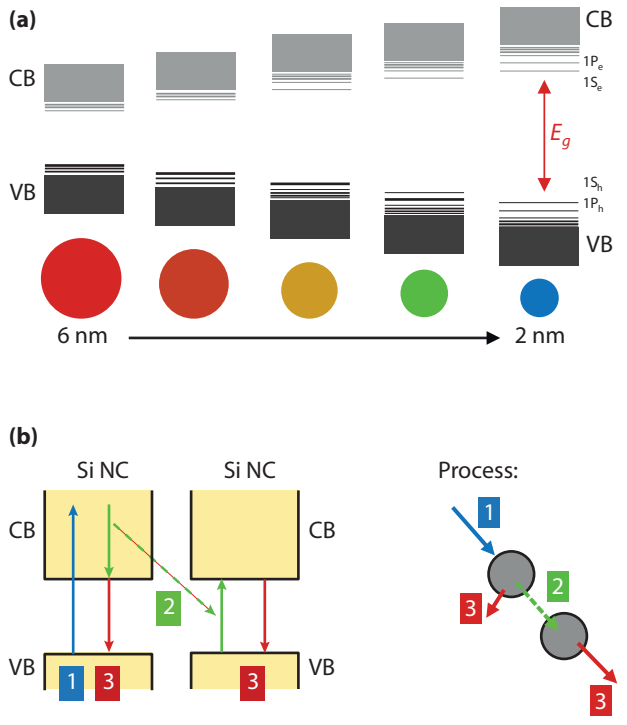

Figure 14.3: Illustrating (a) the effect of the quantum dot (QD) size on the bandgap and (b) spectral down-conversion with QDs.

ial, where the particles are in very close proximity of each other. Figure 14.3 (b) shows the electronic band gap diagram of two nanoparticles. Now, a high-energy photon is absorbed by one QD and hence one electron is excited into the conduction band of the particle. In difference to a bulk semiconductor, the excess energy of the photon is not necessarily lost as heat, but it can be transferred as a *quantised energy package* to a neighbouring quantum dot. Here a second electron is excited into conduction band of the second quantum dot. As a result, two electron-hole pairs have been created out of one photon. If non-radiative recombination mechanisms like Auger recombination and SRH recombination are sufficiently suppressed, the electron-hole pairs in both quantum dots can radiatively recombine such that each of the two QDs emits one reddish photon. In summary, one incident bluish photon is converted in to two reddish photons, which can be absorbed by a PV material.

Figure 14.4 shows some experimental results by Jurbergs *et al.* on down conversion based on silicon quantum dots in a narrow spectral range [77]. The horizontal axis represents the photon emission wavelength. At around 790 nm a down-conversion efficiency of 60% achieved. The EQE exceeds 100% in the blue region from 3.1 up to 3.4 eV. The major challenge is have QD layers with a spectral response exceeding 100% at lower photon energies, because the solar spectrum contains far more photons in that spectral range.

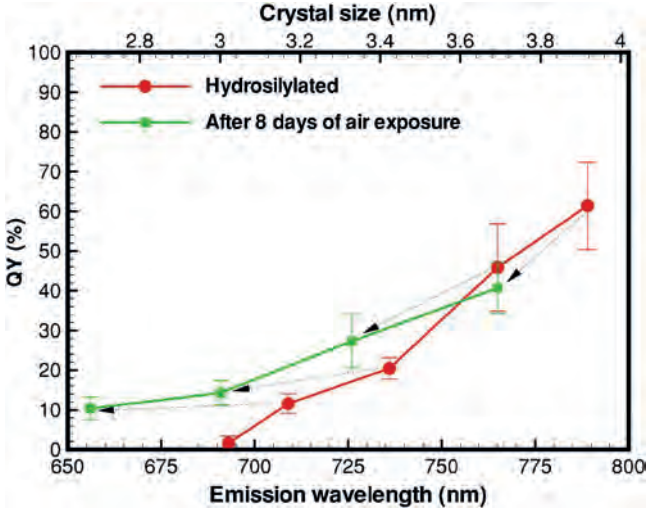

Figure 14.4: Down conversion at Si quantum dots [77].

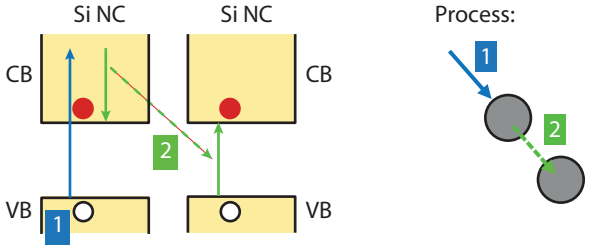

Figure 14.5: Illustrating multi exciton generation with quantum dots.

In practice, only small enhancements in efficiency due to up/down-converters have been reported [74].

14.3 Multi-exciton generation

Another approach to enhance the charge carrier excitation by a single energetic photon is called *multiple exciton generation* (MEG). Here, more than one electron-hole pairs is generated from high energy photons. In difference to spectral down-conversion here all the fundamental energy conversion steps occur in the PV-active layers.

Like down conversion, in principle MEG can be realised with quantum dots, as illustrated in Fig. 14.5. Again, in

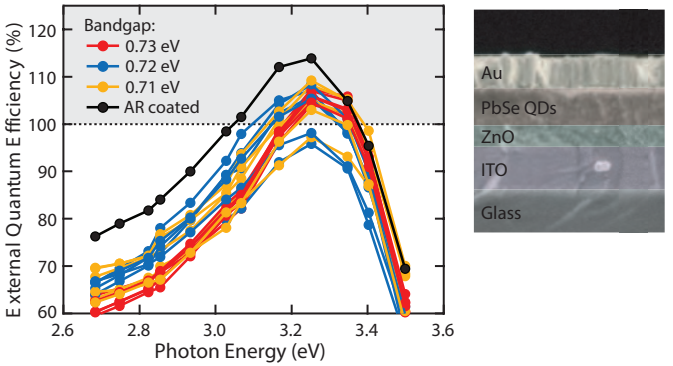

Figure 14.6: Multi-exciton generation using PbSe quantum dots with EQEs exceeding 100% [78].

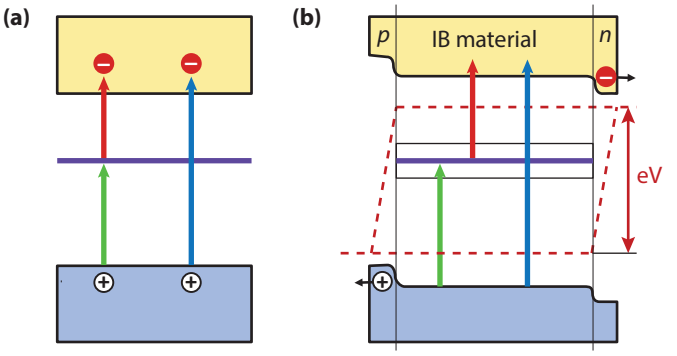

Figure 14.7: Band diagram (a) with the intermediate band depicted and (b) with the quasi-Fermi levels exceeding the energy of the low-energy photons.

one particle an electron is excited into the conduction band and the excess energy is transferred to a neighbouring QD, where a second electron is excited in to conduction band of the second quantum dot. However, here the charge carriers in the two electron-hole pairs are separated before they can recombine such that one incident photons results in more then one generated electrons [79]. Just as for spectral down-conversion, also here quantum efficiencies exceeding 100% are theoretically possible when one incident photon creates statistically more than charge carriers. Figure 14.6 shows that EQEs exceeding 100% can be achieved as demonstrated by Semonin *et al.* [78]. Here, the absorber layer consists PbSe quantum dots.

14.4 Intermediate band solar cells

The concept of *intermediate band solar cells* (IB) tries to tackle the problem photons with energies below the bandgap cannot be utilised for current generation. As shown in Fig. 14.7 (a), in intermediate band cells energy levels are created artificially in the bandgap of the absorber material [80]. As in conventional single-junction solar cells, photons with a sufficient energy can excite an electron from the valence band in to the conduction band. However, in difference to conventional semiconductors photons with energies below the bandgap can excite an electron from the valence band in to the inter-

mediate band. A second low-energy photon is required to excite the electron from the intermediate band into the conduction band. As illustrated in 14.7 (a), therefore two photons with small energies can result in quasi-Fermi level splitting larger than the energy of each of these photons.

Various studies have been performed on how intermediate band cells can be realised, as for example summarised in Reference [80]. In such solar cells, a layer with the intermediate states is placed in between *p*- and *n*-layers. For example quantum dots can be used to realise the intermediate states. Further, various bulk materials are studied for realising the intermediate states. One major problem of experimental IB cells is that the voltages are lower than the voltages of reference cells without IB structures. Yet, the efficiency loss due to the lower voltage cannot be overcome by sub-bandgap current density originating from the IB transitions.

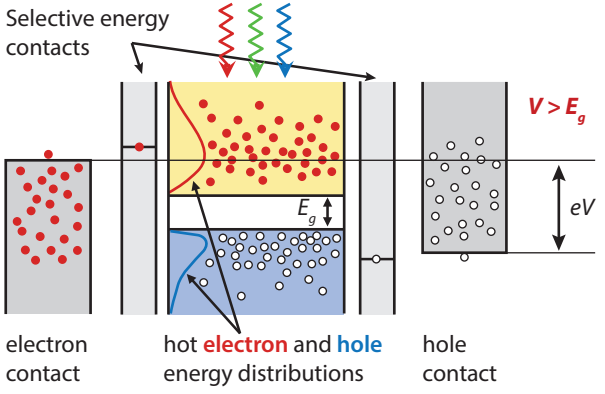

Figure 14.8: Illustrating the working principle of a hot carrier solar cell.

14.5 Hot carrier solar cells

The idea of *hot carrier solar cells* is to reduce the energy losses due to relaxation and hence thermalisation. As illustrated in Fig. 14.8, this should be achieved by collect electron-hole pairs of high energy photons just after light excitation before they have a chance to relax back

to the edges of the electronic bands. In the figure, the population of the charge carrier levels reflects the situation just after the excitation by the absorption of a photon. This distribution is not in thermal equilibrium as many electrons are excited into position further up in the conduction band and the holes are excited down to lower levels in the valence band. These charge carriers are called *hot* electrons and holes [22, 81].

It takes only a few picoseconds (10^{-12} s) for the hot charge carriers to relax back to the edges of the electronic bands. The idea of hot carrier cells it to collect the charge carriers as long as they are still *hot*. Hence, an energy larger than the band gap energy can be utilised per excited charge carrier and the average bandgap utilisation would exceed the bandgap.

The fundamental challenge is to collect the hot carriers before they relax back to the edges of the electronic bands. Such a concept would require *selective contacts*, which only select electrons above a particular energy level in the conduction band and holes below a certain energy level in the valence band, respectively. At the moment the main challenge is to increase the lifetime of the hot charge carriers, such that they have time to move from the absorber material to the selective contacts.

Part IV

PV Systems

15 Introduction to PV systems

15.1 Introduction

After we have discussed the fundamental scientific theories required for solar cells in Part II and have taken a look at modern PV technology in Part III, we now will use the gained knowledge to discuss complete PV systems. A PV system contains many different components besides the PV modules. For successfully planning a PV system it is crucial to understand the function of the different components and to know their major specifications. Further, it is important to know the effect on the location of the (expected) performance of a PV system.

15.2 Types of PV systems

PV systems can be very simple, consisting of just a PV module and load, as in the direct powering of a water pump motor, which only needs to operate when the sun shines. However, when for example a whole house should be powered, the system must be operational day and night. It also may have to feed both AC and DC loads, have reserve power and may even include a back-up generator. Depending on the system configuration, we can distinguish three main types of PV systems: stand-alone, grid-connected, and hybrid. The basic PV system principles and elements remain the same. Systems are adapted to meet particular requirements by varying the type and quantity of the basic elements. A

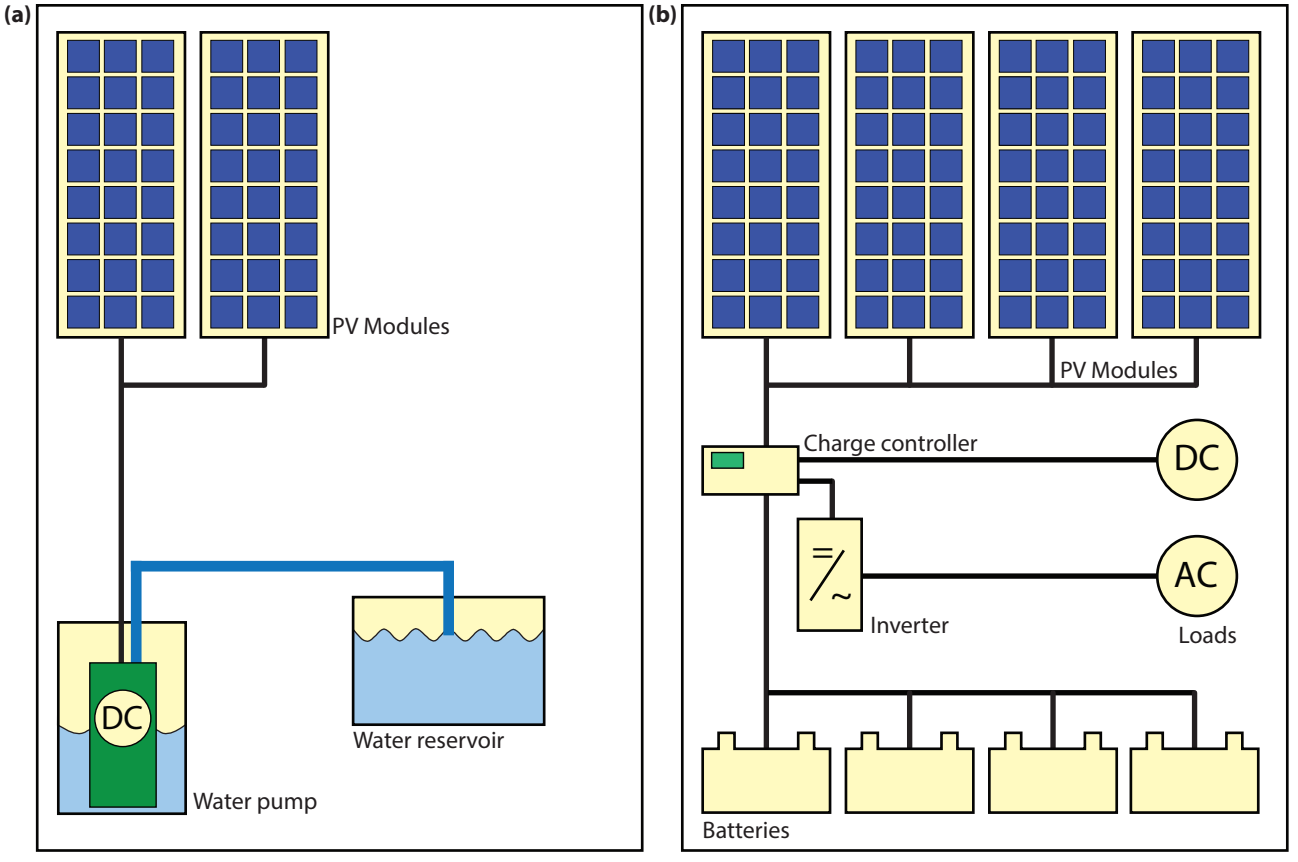

Figure 15.1: Schematic representation of (a) a simple DC PV system to power a water pump with no energy storage and (b) a complex PV system including batteries, power conditioners, and both DC and AC loads.

modular system design allows easy expansion, when power demands change.

15.2.1 Stand-alone systems

Stand-alone systems rely on solar power only. These systems can consist of the PV modules and a load only or they can include batteries for energy storage. When using batteries charge regulators are included, which switch off the PV modules when batteries are fully charged, and may switch off the load to prevent the batteries from being discharged below a certain limit. The batteries must have enough capacity to store the energy produced during the day to be used at night and during periods of poor weather. Figure 15.1 shows schematically examples of stand-alone systems; (a) a simple DC PV system without a battery and (b) a large PV system with both DC and AC loads.

15.2.2 Grid-connected systems

Grid-connected PV systems have become increasingly popular for building integrated applications. As illustrated in Fig. 15.2, they are connected to the grid via inverters, which convert the DC power into AC electricity. In small systems as they are installed in residential homes, the inverter is connected to the distribution

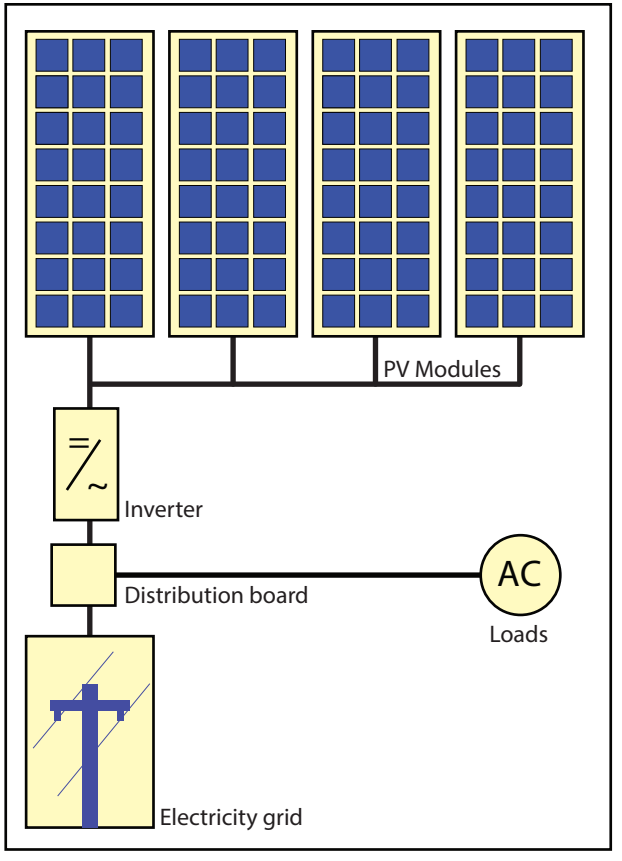

Figure 15.2: Schematic representation of a grid-connected PV system.

board, from where the PV-generated power is transferred into the electricity grid or to AC appliances in the house. These systems do not require batteries, since they are connected to the grid, which acts as a buffer into that an oversupply of PV electricity is transported while the grid also supplies the house with electricity in times of insufficient PV power generation.

Large PV fields act as power stations from that all the generated PV electricity is directly transported to the electricity grid. They can reach peak powers of up to several hundreds of MW_p. Figure 15.3 shows a 25.7 MW_p system installed in Germany.

Figure 15.3: The 25.7 MW_p Lauingen Energy Park in Bavarian Swabia, Germany [82].

15.2.3 Hybrid systems

Hybrid systems consist of combination of PV modules and a complementary method of electricity generation such as a diesel, gas or wind generator. A schematic of an hybrid system shown in Fig. 15.4. In order to optimise the different methods of electricity generation, hybrid systems typically require more sophisticated controls than stand-alone or grid-connected PV systems. For example, in the case of an PV/diesel system, the diesel engine must be started when the battery reaches a given discharge level and stopped again when battery reaches an adequate state of charge. The back-up generator can be used to recharge batteries only or to supply the load as well.

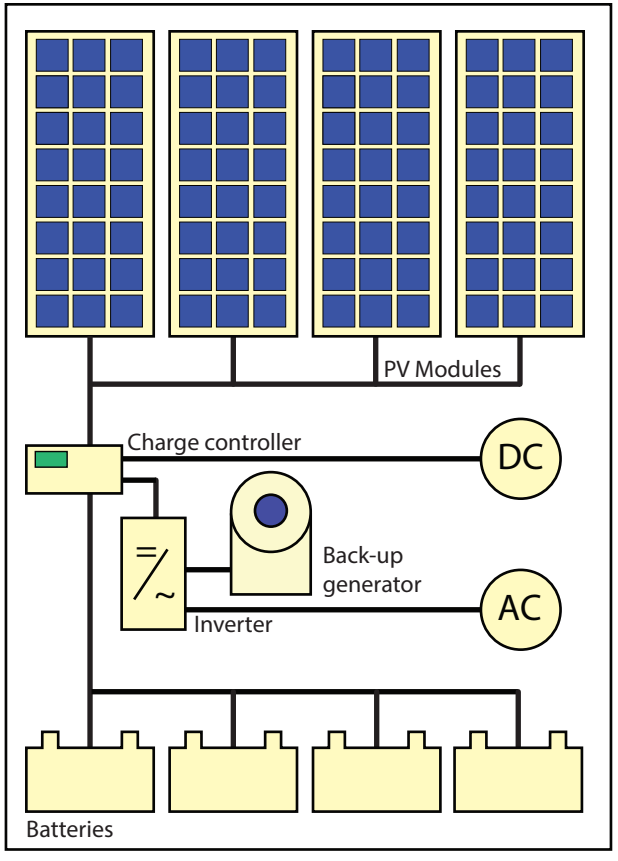

Figure 15.4: Schematic representation of a hybrid PV system that has a diesel generator as alternative electricity source..

15.3 Components of a PV system

As we have seen earlier in this book, a solar cell can convert the energy contained in the solar radiation into electrical energy. Due to the limited size of the solar cell it only delivers a limited amount of power under fixed current-voltage conditions that are not practical for most applications. In order to use solar electricity for practical devices, which require a particular voltage and/or current for their operation, a number of solar cells have to be connected together to form a *solar panel*, also called a *PV module*. For large-scale generation of solar electricity solar panels are connected together into a *solar array*.

Although, the solar panels are the heart of a *PV system*, many other components are required for a working system, that we already discussed very briefly above. Together, these components are called the *Balance of System* (BOS). Which components are required depends on whether the system is connected to the electricity grid or whether it is designed as a stand-alone system. The most important components belonging to the BOS are:

- A *mounting structure* is used to fix the modules and to direct them towards the sun.
- *Energy storage* is a vital part of stand-alone systems because it assures that the system can deliver electricity during the night and in periods of

bad wheather. Usually, *batteries* are used as energy-storage units.

- *DC-DC converters* are used to convert the module output, which will have a variable voltage depending on the time of the day and the weather conditions, to a fixed voltage output that *e. g.* can be used to charge a battery or that is used as input for an inverter in a grid-connected system.
- *Inverters* or *DC-AC converters* are used in grid-connected systems to convert the DC electricity originating from the PV modules into AC electricity that can be fed into the electricity grid.
- *Cables* are used to connect the different components of the PV system with each other and to the electrical load. It is important to choose cables of sufficient thickness in order to minimise resistive losses.

Even though not a part of the PV system itself, the *electric load*, *i.e.* all the electric appliances that are connected to it have to be taken into account during the planning phase. Further, it has to be considered whether the loads are AC or DC loads.

The different compontents of a PV system are schematically presented in Fig. 15.5 and will be discussed in detail in Chapter 17.

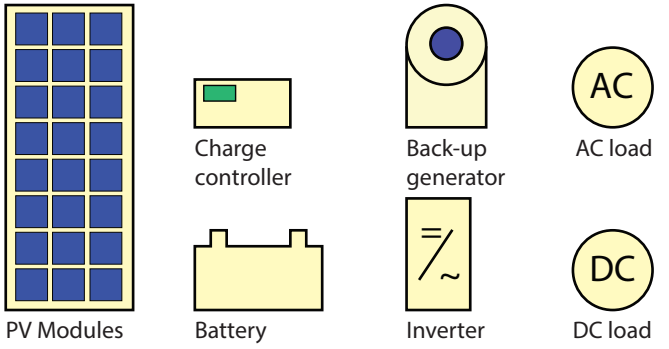

Figure 15.5: A schematic of the different components of a PV system.

16 Location issues

In Chapter 5 we discussed solar radiation on earth and introduced the AM1.5 spectrum, normalised to a total irradiance of 1000 W/m^2. This spectrum is used to evaluate the performance of solar cells and modules in laboratories and industry. The AM1.5 spectrum represents the solar irradiance if the centre of the solar disc is at an angle of 48.2° off the zenith (or 41.2° above the horizon).

Of course, the sun not always is at this position, but the position is dependent on the time of the day and the year and of course it is dependent on the location on earth. In this section we will discuss how to calculate the position of the sun at every location on earth at an arbitrary time and date. Furthermore we will discuss scattering of sunlight when it traverses the atmosphere and how this influences the direct and diffuse spectrum. We also will discuss the influence of the mounting angle and position of a PV module on the irradiance at the module.

16.1 The position of the sun

For planning a PV system it is crucial to know the position of the sun in the sky at the location of the solar system at a given time. In this section we explain how this position can be calculated.

Since celestial objects like the sun, the moon and the stars are very far away from the earth it is convenient to describe their motion projected on a sphere with ar-

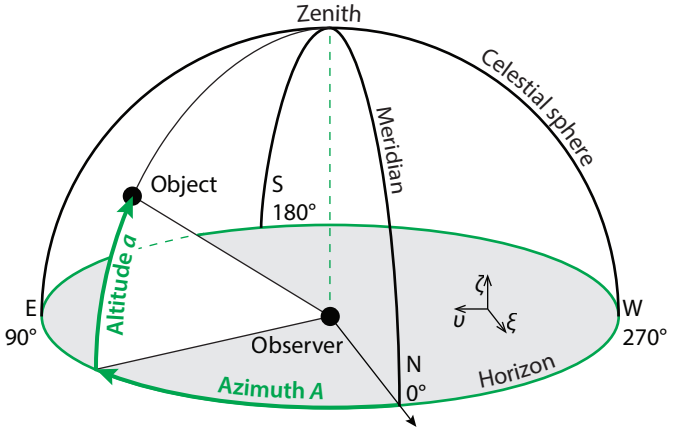

Figure 16.1: Illustrating the definition of the *altitude a* and the *azimuth A* in the horizontal coordinate system. Note that North is at the bottom of the figure.

bitrary radius and concentric to the earth. This sphere is called the *celestial sphere*. The position of every celestial object thus can be parameterised by two angles. For photovoltaic applications it is most convenient to use the *horizontal coordinate system*, where the horizon of the observer constitutes the *fundamental plane*. In this coordinate system, the position of the sun is expressed by two angles that are illustrated in Fig. 16.1: The *altitude a* that is the angular elevation of the centre of the solar disc above the horizontal plane. Its angular range is $a \in [-90°, 90°]$, where negative angles correspond to the object being below the horizon and thus not visible. The *azimuth A* that is the angle between the line of sight projected on the horizontal plane and due North. It is usually counted eastward, such that $A = 0°$, $90°$, $180°$, $270°$ correspond to due *North*, *East*, *South* and *West*, respectively. Its angular range is $A \in [0°, 360°]$. In a different convention also used by the PV community, due South corresponds to $0°$ and is counted westward, the angles then are in between $-180°$ and $180°$. Figure 16.1 also shows the *meridian*, which is great circle on the celestial sphere passing through the celestial North and South poles as well as the zenith.

Instead of using the spherical coordinates a and A, we also could use Cartesian coordinates, that we here call ξ (xi), υ (upsilon) and ζ (zeta) and that are also depicted in Fig. 16.1. The principal direction is parallel to ξ. The Cartesian coordinates are connected to the spherical

coordinates via

$$\begin{pmatrix} \xi \\ v \\ \zeta \end{pmatrix} = \begin{pmatrix} \cos a \cos A \\ \cos a \sin A \\ \sin a \end{pmatrix}. \qquad (16.1)$$

Note that $\xi^2 + v^2 + \zeta^2 = 1$ for all points on the celestial sphere.

Earth orbits the sun in an elliptic orbit at an average distance of about 150 million kilometres. Due to the elliptic orbit the speed of Earth is not constant. This is because of Kepler's second law that states that "A line joining a planet and the Sun sweeps out equal areas during equal intervals of time." On the celestial sphere the sun seems to move on a circular path with one revolution per year. This path is called the *ecliptic* and illustrated in Fig. 16.2. For describing the apparent movement of the sun on the celestial sphere is is convenient to use coordinates in that the ecliptic lies in the fundamental plane. These coordinates are called the *ecliptic coordinates*. As principal direction, the position of the sun at the spring (*vernal*) *equinox* (thus around 21 March) is used, which is indicated by the sign of Aries, ♈. As obvious from Fig. 16.2, ♈ lies both in the ecliptic plane as well as in the equatorial plane. The ecliptic coordinate system is sketched in Fig. 16.3 (a). The two angular coordinates are called the *ecliptic longitude* λ and the *ecliptic latitude* β. Note, that in this coordinate system the rotation of the earth around its axis is not taken into account.

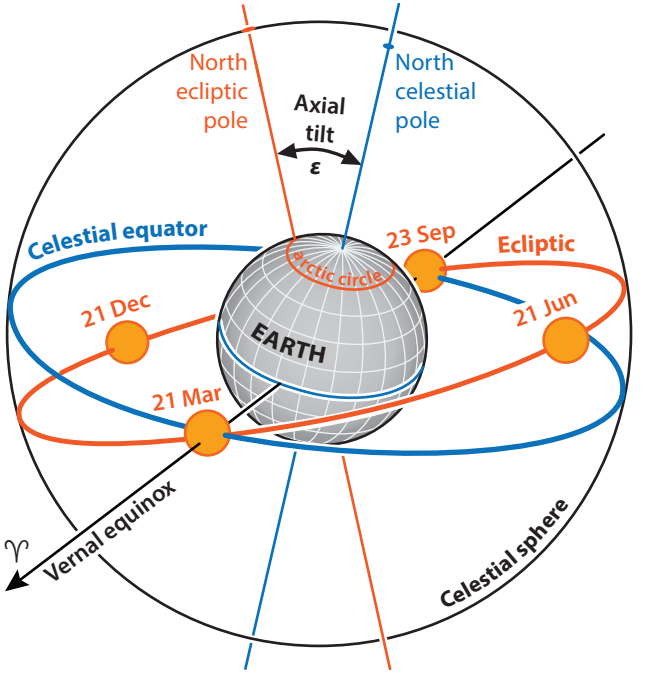

Figure 16.2: Illustrating the *ecliptic*, *i.e.* the apparent movement of the sun around earth. Further, the *celestial equator* and the direction of the *vernal equinox* are indicated. The sizes of Sun and Earth are not in scale.

In ecliptic coordinates approximate position the position of the sun can be expressed easily. The approximation presented here has an accuracy of about 1 arcminute within two centuries of 2000 and is published by the the *Astronomical Applications Departement* of the *U.S. Naval Observatory*.

To express the position of the sun we first have to express the time D elapsed since Greenwich noon, Terrestrial Time, on 1 January 2000, in days. For astronomic purposes it may be convenient to relate D to the Julian date JD via

$$D = \mathrm{JD} - 2451545.0. \tag{16.2}$$

The Julian Date[1] that is defined as the number of days since 1 January 4713 BC in a proleptic[2] Julian calendar or since 24 November 4717 BC in a proleptic Gregorian calendar.

Now, the *mean longitude* of the sun corrected to the aberration of the light is given by

$$q = 280.459^\circ + 0.98564736^\circ D \tag{16.3}$$

Because of the elliptic orbit of earth and hence a varying speed throughout the year, we have to correct with the so-called *mean anomaly* of the Sun,

$$g = 357.529^\circ + 0.98560028^\circ D. \tag{16.4}$$

[1] Calculating the Julian Date is implemented in `MatLab`.

[2] *Proleptic* means that a calendar is applied to dates before its introduction.

It may be convenient to normalise q and g to the range $[0^\circ, 360^\circ)$ by adding or subtracting multiples of 360°.

Now the ecliptic longitude of the sun is given by

$$\lambda_S = q + 1.915^\circ \sin g + 0.020^\circ \sin 2g. \tag{16.5}$$

The ecliptic latitude can be approximated by

$$\beta_S = 0. \tag{16.6}$$

For estimating the radiation it might also be convenient to approximate the distance of the Sun from the Earth. In astronomical units (AU) this is given by

$$R = 1.00014 - 0.01671 \cos g - 0.00014 \cos 2g. \tag{16.7}$$

As stated above, for PV applications it is convenient to use horizontal coordinates. We therefore have to transform from ecliptic coordinates into the horizontal coordinates. This is done via *three rotations* that are to be performed after each other consecutively:

First, we have to transform from ecliptic coordinates into equatorial coordinates. As illustrated in Fig. 16.2, the fundamental plane of these coordinates is tilted to the ecliptic with an angle ϵ,

$$\epsilon = 23.429^\circ - 0.00000036^\circ D \tag{16.8}$$

The principal direction is again given by the vernal equinox Υ. In Fig. 16.3 (b) the equatorial coordinate

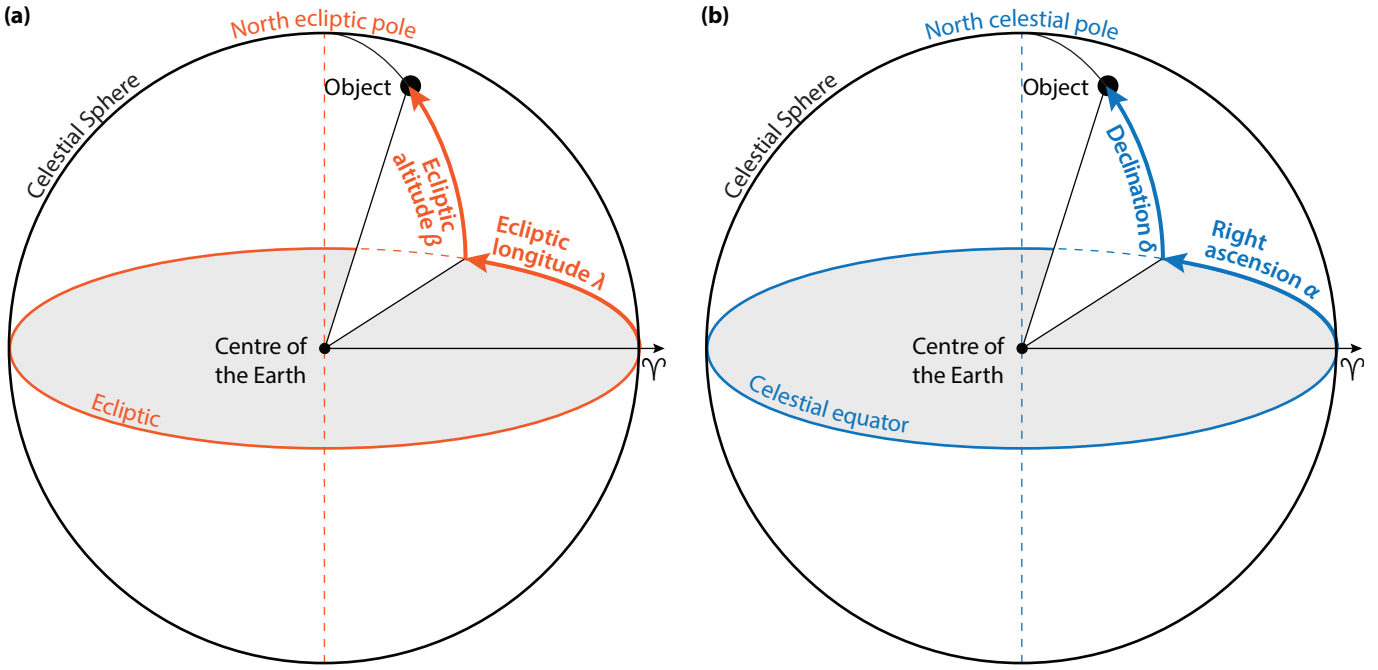

Figure 16.3: Illustrating (a) the *ecliptic coordinate system* and (b) the *equatorial coordinate system*.

system is sketched. The two coordinates are called the *right ascension* α and the *declination* δ. The transformation from ecliptic coordinates to equatorial coordinates is a rotation by the angle ϵ about the vernal equinox as rotational axis. Mathematically this is expressed by

$$\begin{pmatrix} \cos\delta\cos\alpha \\ \cos\delta\sin\alpha \\ \sin\delta \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\epsilon & -\sin\epsilon \\ 0 & \sin\epsilon & \cos\epsilon \end{pmatrix} \begin{pmatrix} \cos\beta\cos\lambda \\ \cos\beta\sin\lambda \\ \sin\beta \end{pmatrix}. \tag{16.9}$$

Secondly, we have to take the rotation of the earth around its axis into account. We do this by using the so-called *hour angle* h instead of the right ascension α. Those two angles are connected to each other via

$$h = \theta_L - \alpha, \tag{16.10}$$

where θ_L is the *local mean sidereal time*, *i.e.* the angle between the vernal equinox and the meridian. All these angles are illustrated in Fig. 16.4. A sidereal day is the duration between two passes of the vernal equinox through the meridian and it is slightly shorter than a solar day. We can understand this by realising that the earth has to rotate by 360° and approximately 360°/365.25 between two passes through the meridian. The duration of *mean* sidereal day is approximately 23 h, 56 m and 4 s.

For calculating θ_L we first have to determine the Greenwich Mean Sidereal Time (GMST), which is (approxim-

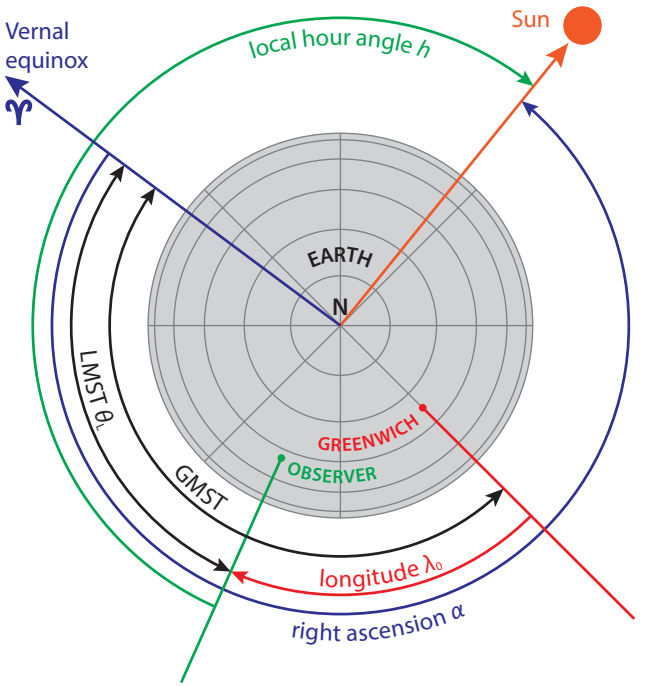

Figure 16.4: Illustrating the right ascension α, the local hour angle h, the Greenwich Mean Sidereal Time GMST and the local mean sidereal time θ_L.

ately) given by

$$\begin{aligned}\text{GMST} =& 18.697374558\,\text{h} \\ &+ 24.06570982441908\,\text{h} \cdot D \\ &+ 0.000026\,\text{h} \cdot T^2,\end{aligned} \tag{16.11}$$

where D is as defined above and T is the number of centuries past since Greenwich noon, Terrestrial Time, on 1 January 2000,

$$T = \frac{D}{36525}. \tag{16.12}$$

For many applications, the quadratic term may be omitted. GMST is given in hours and has to be normalised to the range [0 h, 24 h). We then can obtain the *local* mean sidereal time in degrees with

$$\theta_L = \text{GMST}\frac{15^\circ}{\text{hour}} + \lambda_0, \tag{16.13}$$

where λ_0 is the longitude of the observer.

We thus have to the following transform about the rotational axis of the earth,

$$\begin{pmatrix}\cos\delta\cos h\\ \cos\delta\sin h\\ \sin\delta\end{pmatrix} = \begin{pmatrix}\cos\theta_L & \sin\theta_L & 0\\ \sin\theta_L & -\cos\theta_L & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\cos\delta\cos\alpha\\ \cos\delta\sin\alpha\\ \sin\delta\end{pmatrix}. \tag{16.14}$$

Note that this transform is no rotation but a reflection at an angle of $\theta_L/2$. We thus now transformed the principal direction of the coordinate system from vernal equinox to the local mean sidereal time.

Thirdly we transform to the horizontal coordinate system by rotating with the latitude angle ϕ_0 of the observer. While the ecliptic and equatorial coordinate systems use the centre of the earth as origin, the horizontal coordinate system uses the actual position on the surface of Earth as origin. However, because of the distance of celestial objects in general and the Sun in particular being much larger than the radius of the Earth, we may neglect this translational shift of the origin of the coordinate systems.

The rotational axis of the third transform is the axis that is normal to both the principal direction θ_L and the rotational axis of the earth $(\delta = 90^\circ)$.

$$\begin{pmatrix}\xi'\\ v'\\ \zeta\end{pmatrix} = \begin{pmatrix}\sin\phi_0 & 0 & -\cos\phi_0\\ 0 & 1 & 0\\ \cos\phi_0 & 0 & \sin\phi_0\end{pmatrix}\begin{pmatrix}\cos\delta\cos h\\ \cos\delta\sin h\\ \sin\delta\end{pmatrix}. \tag{16.15}$$

However, here the directions ξ' and v' point to due South and West, respectively. We thus must apply $\xi' \to -\xi$ and $v' \to -v$ in order to get the directions such as they are defined in Fig. 16.1 (ξ and v pointing to due North and East, respectively). We do this with the matrix transform

$$\begin{pmatrix}\xi\\ v\\ \zeta\end{pmatrix} = \begin{pmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\xi'\\ v'\\ \zeta\end{pmatrix}. \tag{16.16}$$

By combining Eqs. (16.9), (16.14)–(16.16) we directly can transform from ecliptic to horizontal coordinates,

$$\begin{pmatrix}\xi\\ v\\ \zeta\end{pmatrix} = \begin{pmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\sin\phi_0 & 0 & -\cos\phi_0\\ 0 & 1 & 0\\ \cos\phi_0 & 0 & \sin\phi_0\end{pmatrix}\begin{pmatrix}\cos\theta_L & \sin\theta_L & 0\\ \sin\theta_L & -\cos\theta_L & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\epsilon & -\sin\epsilon\\ 0 & \sin\epsilon & \cos\epsilon\end{pmatrix}\begin{pmatrix}\cos\beta\cos\lambda\\ \cos\beta\sin\lambda\\ \sin\beta\end{pmatrix}. \quad (16.17)$$

Please note that matrix multiplications do not commute, *i.e.* the order in which the rotations are applied must not be altered. Now, we apply that the ecliptic latitude of the sun $\beta_S = 0$. By calculating Eq. (16.32) we find

$$\xi_S = \cos a_S \cos A_S = -\sin\phi_0\cos\theta_L\cos\lambda_S - (\sin\phi_0\sin\theta_L\cos\epsilon - \cos\phi_0\sin\epsilon)\sin\lambda_S, \quad (16.18a)$$

$$v_S = \cos a_S \sin A_S = -\sin\theta_L\cos\lambda_S + \cos\theta_L\cos\epsilon\sin\lambda_S, \quad (16.18b)$$

$$\zeta_S = \sin a_S = \cos\phi_0\cos\theta_L\cos\lambda_S + (\cos\phi_0\sin\theta_L\cos\epsilon + \sin\phi_0\sin\epsilon)\sin\lambda_S, \quad (16.18c)$$

where we also used the the relationship between Cartesian and spherical horizontal coordinates from Eq. (16.1). Dividing Eq. (16.18b) by Eq. (16.18a) and leaving Eq. (16.18c) unchanged leads to the final expressions for the solar position,

$$\tan A_S = \frac{v_S}{\xi_S} = \frac{-\sin\theta_L\cos\lambda_S + \cos\theta_L\cos\epsilon\sin\lambda_S}{-\sin\phi_0\cos\theta_L\cos\lambda_S - (\sin\phi_0\sin\theta_L\cos\epsilon - \cos\phi_0\sin\epsilon)\sin\lambda_S}, \quad (16.19a)$$

$$\sin a_S = \zeta_S = \cos\phi_0\cos\theta_L\cos\lambda_S + (\cos\phi_0\sin\theta_L\cos\epsilon + \sin\phi_0\sin\epsilon)\sin\lambda_S. \quad (16.19b)$$

A_S and a_S now can be derived by applying inverse trigonometric functions. While arcsin uniquely delivers an altitude in between $-90°$ and $90°$, applying arctan leads to ambiguities. For deriving an azimuth in between $0°$ and $360°$, we have to look in which quadrant is lying. Therefore we use ξ and v from Eqs. (16.18a) and (16.18b), respectively. We find

$$\xi > 0 \wedge v > 0 \Rightarrow A_S = \arctan f(\ldots), \quad (16.20a)$$

$$\xi < 0 \Rightarrow A_S = \arctan f(\ldots) + 180°, \quad (16.20b)$$

$$\xi > 0 \wedge v < 0 \Rightarrow A_S = \arctan f(\ldots) + 360°. \quad (16.20c)$$

$f(\ldots)$ denotes the function at the right hand side of Eq. (16.19a). Note that an altitude $a_S < 0°$ corresponds to the sun being below the horizon. This means that the Sun is not visible and no solar energy can be harvested.

The approximations presented on the previous pages are accurate within arcminutes for 200 centuries of 2000. Several years ago, NREL presented a much more complicated model, the so-called Solar Position Algorithm (ASP), with uncertainties of only $\pm 0.0003°$ in the period from 2000 BC to 6000 AD [83].

Example

As an example we will calculate the position of the Sun in ***Delft on 14 April 2014 at 11:00 local time****.*

For determining the solar position we need next to date and time (in UTC) the latitude and longitude. Since the time zone in Delft on 14 April is the CEST, the Central European Summer Time, the time difference with UTC is +2 hours, such that 11:00 CEST corresponds to 9:00 UTC. According to `Google Maps`*, the latitude and longitude of the* Markt *in the centre of Delft are given by*

$$\phi_0 = 52.01° \text{ N} = +52.01°,$$
$$\lambda_0 = \ \ 4.36° \text{ E} = +\ 4.36°.$$

For the calculation we first have to express date and time as the time elapsed since 1 January 2000 noon UTC.

$$D = 4 \cdot 366 + 10 \cdot 365 + 2 \cdot 31 + 28 + 13 - 0.5 + \tfrac{9}{24} = 5216.875.$$

Now we can calculate the mean longitude q and the mean anomaly g of the sun according to Eqs. (16.3) and (16.4),

$$q = 280.459° + 0.98564736° D = 22.4580712°,$$
$$g = 357.529° + 0.98560028° D = 99.28246073°,$$

where the values were normalised to $[0°, 360°)$. From Eq. (16.5) we thus obtain for the latitude of the Sun in ecliptic coordinates

$$\lambda_S = q + 1.915° \sin g + 0.020° \sin 2g = 24.34162696°.$$

For the axial tilt ϵ of the Earth we obtain from Eq. (16.8)

$$\epsilon = 23.429^\circ - 0.00000036^\circ D = 23.42712193^\circ.$$

The Greewich Mean Sidereal Time (GMST) is given by Eq. (16.11),

$$GMST = 18.697374558\,\mathrm{h} + 24.06570982441908\,\mathrm{h} \cdot D + 0.000026\,\mathrm{h} \cdot T^2 = 22.49731535\,\mathrm{h},$$

where we used $T = D/36525$ and normalised to $[0\,h,\ 24\,h)$. We then find for the local mean sidereal time θ_L

$$\theta_L = \mathrm{GMST}\frac{15^\circ}{\mathrm{hour}} + \lambda_0 = 341.8197303^\circ,$$

Now we have all variables required to calculate the solar position. From Eqs. (16.19) and (16.20) we thus find

$$\tan A_S = \frac{-\sin\theta_L\cos\lambda_S + \cos\theta_L\cos\epsilon\sin\lambda_S}{-\sin\phi_0\cos\theta_L\cos\lambda_S - (\sin\phi_0\sin\theta_L\cos\epsilon - \cos\phi_0\sin\epsilon)\sin\lambda_S} = -1.318180633.$$
$$\sin a_S = \cos\phi_0\cos\theta_L\cos\lambda_S + (\cos\phi_0\sin\theta_L\cos\epsilon + \sin\phi_0\sin\epsilon)\sin\lambda_S = 0.589415473,$$

which leads to the solar altitude $\boldsymbol{a_S}$**=36.1**$^\circ$ *and the solar azimuth* $\boldsymbol{A_S}$**=127.2**$^\circ$.

16.2 The sun path at different locations

In this section we discuss the solar paths throughout the year at several locations around the earth. Figures 16.5-16.8 shows four examples: Delft, the Netherlands ($\phi_0 = 52.01^\circ$ N), the North Cape, Norway ($\phi_0 = 71.17^\circ$ N), Cali, Colombia ($\phi_0 = 3.42^\circ$ N), and Sydney, Australia ($\phi_0 = 33.86^\circ$ S). Note that all the times are given in the apparent solar time (AST). While in Delft and on the North Cape, the Sun at noon always is South of the zenith, in Sydney it is always North. In Cali, close to the equator, the Sun is either South or North, depending on the time of the year. Since the North Cape is north of the arctic circle, the Sun does not set around 21 June. This phenomenon is called the *midnight sun*. On the other hand, the sun always stays below the horizon around 21 December - this is called the *polar night*.

16.3 The equation of time

Figure 16.9 shows the position of the sun throughout the year in Delft at 8:30, 12:00 and 15:30 mean solar time (MST), *i.e.* UTC $+ \lambda_0$, where the longitude λ_0 is expressed in hours. We see that the Sun not only changes from altitude but also from azimuth in the course of a year, such that it seems to run along the shape of an *Eight*. This closed curve is called the *analemma*.

The difference between the apparent solar time (AST), *i.e.* the timescale where the sun really is highest at noon every day, and mean solar time is described by the so-called equation of time, which is defined as

$$\mathrm{EoT} = \mathrm{AST} - \mathrm{MST}. \tag{16.21}$$

The equation of time is given as the difference between the mean longitude q, as defined in Eq. (16.3), and the right ascendent α_S of the sun in the equatorial coordinate system,

$$\mathrm{EoT}(D) = [q(D) - \alpha_S(D)]\,\frac{1}{15}\frac{\mathrm{hour}}{\mathrm{deg}}. \tag{16.22}$$

The right ascendent is connected to the ecliptic longitude of the sun λ_S, as given in Eq. (16.5) via

$$\tan\alpha_s = \cos\epsilon\,\tan\lambda_S, \tag{16.23}$$

and ϵ is the axial tilt as given in Eq. (16.8).

The equation of time has two major contributors: the anomaly due to the elliptic orbit of the Earth around the Sun and the Axial tilt of the rotational axis of the Earth with respect to the ecliptic. Both effects are shown in Fig. 16.10. In this figure, also the total EoT is shown, which is nearly the sum of the two contributors (The maximal deviation is less than a minute).

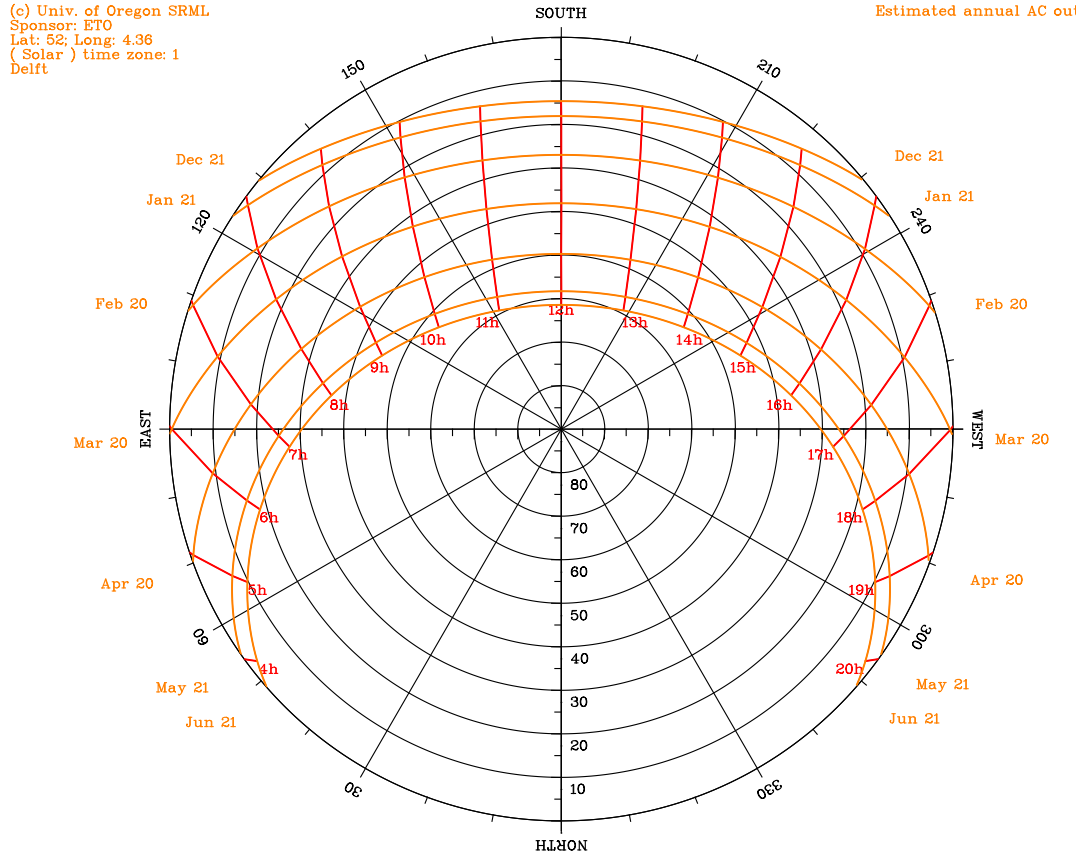

Figure 16.5: The sun path in apparent solar time in Delft, the Netherlands ($\phi_0 = 52.01°$ N). The sun path was calculated with the Sun path chart program by the *Solar Radiation Monitoring Lab.* of the *Univ. of Oregon* [84].

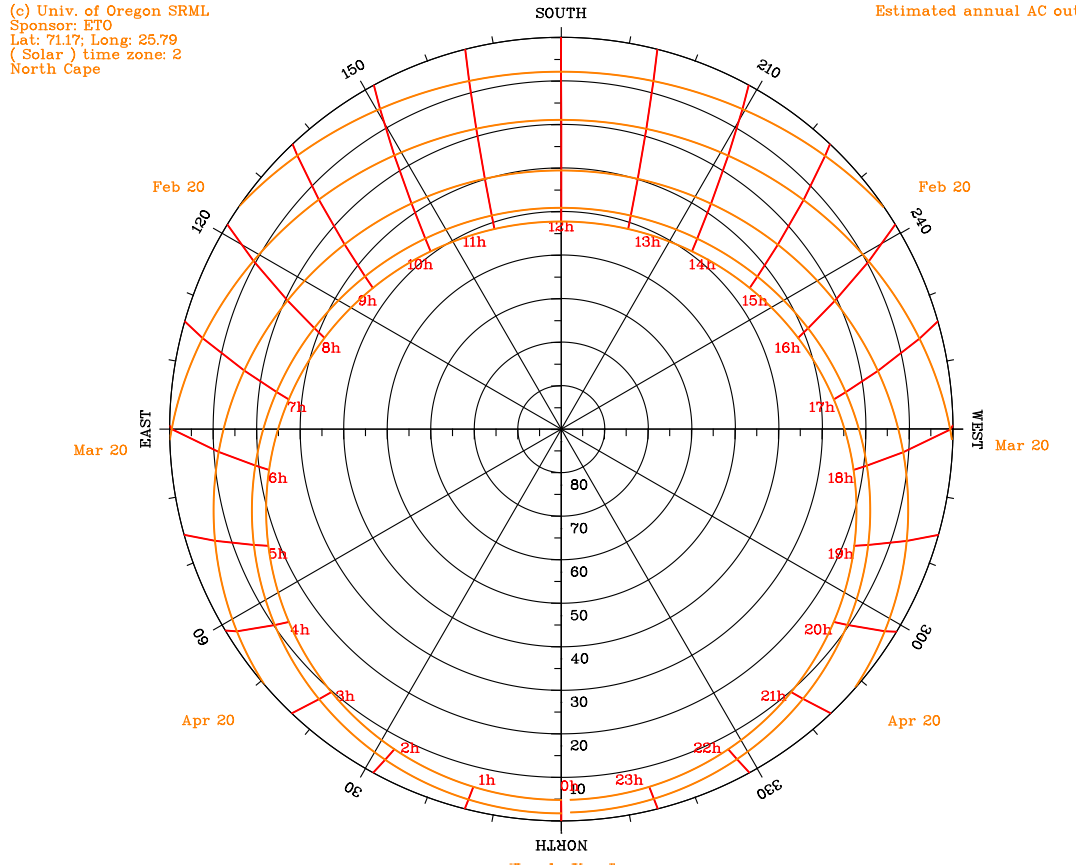

Figure 16.6: The sun path in apparent solar time on the North Cape, Norway ($\phi_0 = 71.17°$ N). The sun path was calculated with the Sun path chart program by the *Solar Radiation Monitoring Lab.* of the *Univ. of Oregon* [84].

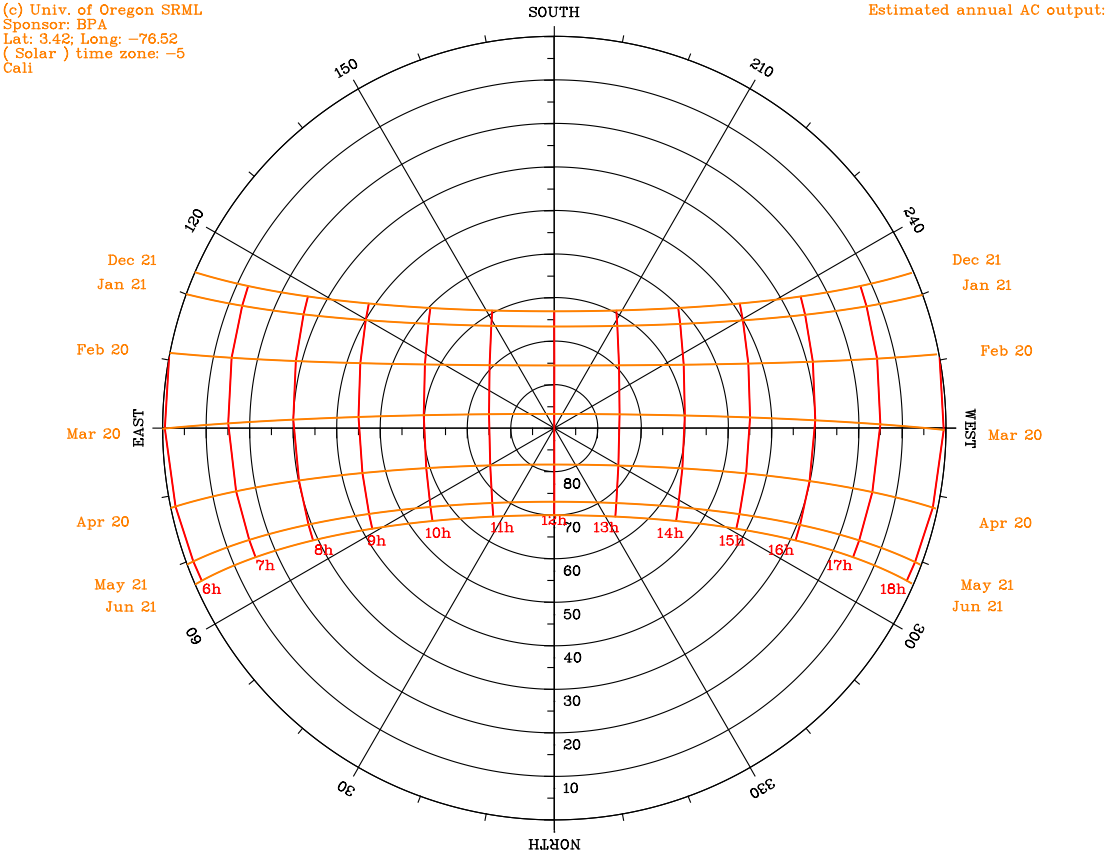

Figure 16.7: The sun path in apparent solar time in Cali, Colombia ($\phi_0 = 3.42°$ N). The sun path was calculated with the Sun path chart program by the *Solar Radiation Monitoring Lab.* of the *Univ. of Oregon* [84].

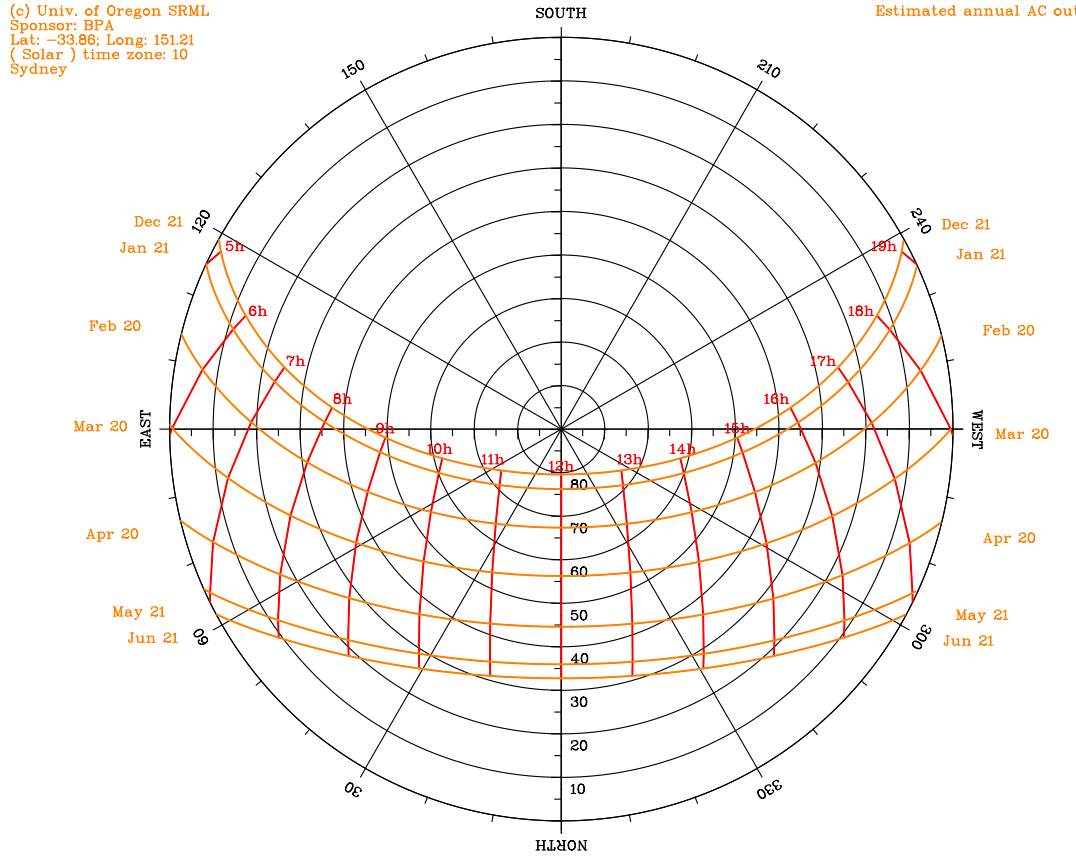

Figure 16.8: The sun path in apparent solar time in Sydney, Australia ($\phi_0 = 33.86°$ S). The sun path was calculated with the Sun path chart program by the *Solar Radiation Monitoring Lab.* of the *Univ. of Oregon* [84].

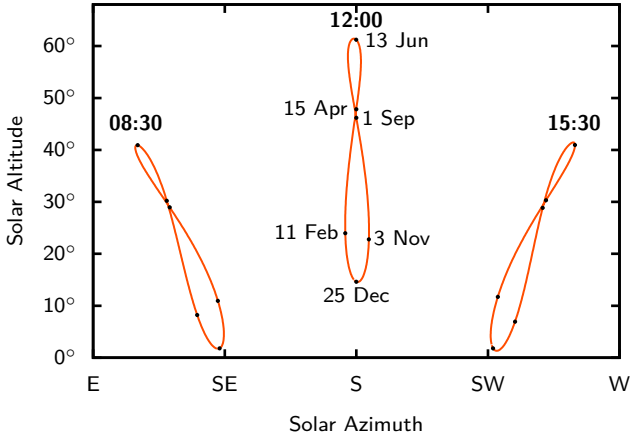

Figure 16.9: The analemma, *i.e.* the apparent curve of the sun throughout the year when observed at the same mean solar time every day. The analemma is shown for Delft (52° N latitude) at three points in time during the day.

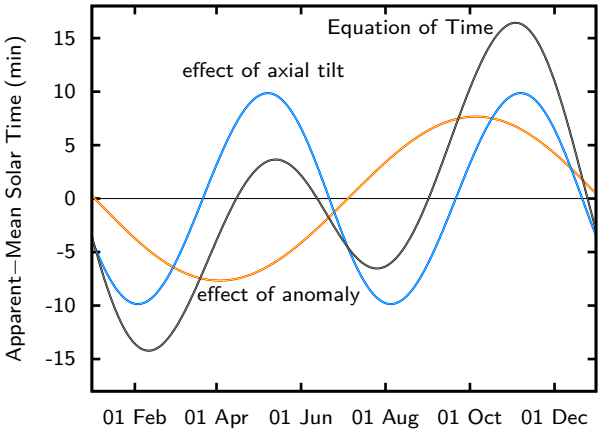

Figure 16.10: The effect of the anomaly of the terrestrial orbit and the axial tilt of the Earth rotation axis on the difference between apparent and mean solar time. The equation of time nearly is the sum of these two effects.

We see that the largest negative shift is on 11 February, where, the apparent noon is about 14 min 12 s prior to the mean solar noon. The largest positive shift is on 3 November, when the apparent solar noon is about 16 min 25 s past the mean solar noon. These points are also marked in Fig. 16.9. There, also the zeros of the EoT are shown, which are on 15 April, 13 June, 1 September, and 25 December.

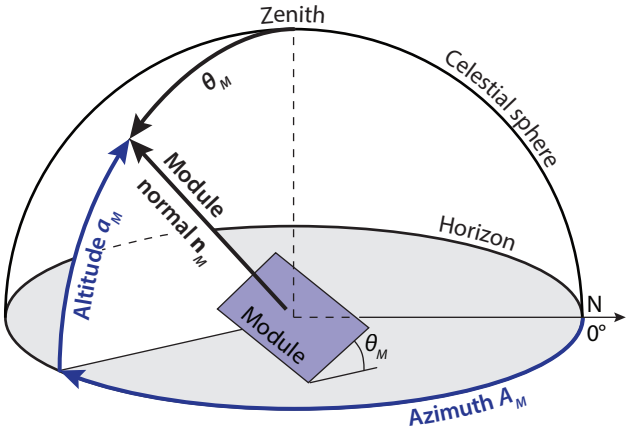

Figure 16.11: Illustrating the angles used to describe the orientation of a PV module installed on a horizontal plane.

16.4 Irradiance on a PV module

After having discussed how to calculate the position of the Sun everywhere on the Earth and having looked at several examples, it now is time to discuss the implications for the irradiance present on solar modules. For this discussion we assume that the solar module is mounted on a horizontal plane and that it is tilted under an angle θ_M, as illustrated in Fig. 16.11. The angle between the projection of the normal of the module onto the horizontal plane and due north is A_M. We then can describe the position of the module by the direction of the module normal in horizontal coordinates (A_M, a_M), where the altitude is given by $a_M = 90° - \theta$. Let now the sun be at the position (A_S, a_S). Then the direct irradiance on the module G_M is given by the equation

$$G_M^{\text{dir}} = I_e^{\text{dir}} \cos\gamma, \tag{16.24}$$

where $\gamma = \sphericalangle(A_M, a_M)(A_S, a_S)$ is the angle between the surface normal and the incident direction of the sunlight. Now we know that the scalar product of two unit vectors is equal to the cosine of the enclosed angle. Thus, we can write,

$$\cos\gamma = \mathbf{n}_M \cdot \mathbf{n}_S. \tag{16.25}$$

The normal vectors are given by

$$\mathbf{n}_M = \begin{pmatrix} \xi_M \\ \upsilon_M \\ \zeta_M \end{pmatrix} = \begin{pmatrix} \cos a_M \cos A_M \\ \cos a_M \sin A_M \\ \sin a_M \end{pmatrix}, \tag{16.26}$$

$$\mathbf{n}_S = \begin{pmatrix} \xi_S \\ \upsilon_S \\ \zeta_S \end{pmatrix} = \begin{pmatrix} \cos a_S \cos A_S \\ \cos a_S \sin A_S \\ \sin a_S, \end{pmatrix}. \tag{16.27}$$

where we used the relationship between Cartesian and spherical horizontal coordinates given in Eq. (16.1).

Hence, we find

$$\begin{aligned}\cos\gamma =& \mathbf{n}_M \cdot \mathbf{n}_S \\ =& \cos a_M \cos A_M \cos a_S \cos A_S \\ &+ \cos a_M \sin A_M \cos a_S \sin A_S + \sin a_M \sin a_S \\ =& \cos a_M \cos a_S \left(\cos A_M \cos A_S + \sin A_M \sin A_S\right) \\ &+ \sin a_M \sin a_S \\ =& \cos a_M \cos a_S \cos\left(A_M - A_S\right) + \sin a_M \sin a_S .\end{aligned} \tag{16.28}$$

Thus we obtain for the irradiance

$$\begin{aligned}G_M^{\mathrm{dir}} &= I_e^{\mathrm{dir}} \left[\cos a_M \cos a_S \cos\left(A_M - A_S\right) + \sin a_M \sin a_S\right] \\ &= I_e^{\mathrm{dir}} \left[\sin\theta \cos a_S \cos\left(A_M - A_S\right) + \cos\theta \sin a_S\right] .\end{aligned} \tag{16.29}$$

Note that this equation only holds when the sun is above the horizon ($a_S > 0$) and the azimuth of the sun is within $\pm 90^\circ$ of A_M, $A_S \in [A_S - 90^\circ, A_S + 90^\circ]$. Otherwise, $G_M^{\mathrm{dir}} = 0$.

If the azimuth of the solar position is the same as the azimuth of the module normal $A_M = A_S$, Eq. (16.29) becomes

$$\begin{aligned}G_M^{\mathrm{dir}} &= I_e^{\mathrm{dir}} \left[\cos a_M \cos a_S + \sin a_M \sin a_S\right] \\ &= I_e^{\mathrm{dir}} \cos\left(a_M - a_S\right) .\end{aligned} \tag{16.30}$$

When using the tilt angle $\theta = 90^\circ - a_M$ we find

$$G_M^{\mathrm{dir}} = I_e^{\mathrm{dir}} \sin\left(\theta + a_S\right) . \tag{16.31}$$

Modules mounted on a tilted roof

When a module is mounted on a horizontal plane, it is easy to determine its normal $\mathbf{n}_M$. However, when a module is to be mounted on an arbitrarily tilted roof things become more complicated. We thus will derive how to calculate the normal of the module in horizontal coordinates when the coordinates with respect to the roof are given. In fact, we thus have to transform the module normal from the *roof coordinate system* to the horizontal coordinate system.

As illustrated in Fig. 16.12 (a), the orientation of the roof in horizontal coordinates is characterised by the azimuth A_R and the altitude a_R of its normal $\mathbf{n}_R$. The module is installed on the roof, and its orientation with respect to the roof is best described in the roof coordinate system, where the fundamental plane is parallel to the roof and the principal direction is along the gradient of the roof, as illustrated in Fig. 16.12 (b). In this system, the module normal is given by the azimuth ϕ_M and the altitude is given by δ_M. The coordinate transform itself is transformed by combining two rotations:

First, we rotate with the angle $90^\circ - a_R$ around the axis that is perpendicular to both $\mathbf{n}_R$ and the gradient direction of the roof. Secondly, we rotate with the angle $A_R + 180^\circ$ along the zenith. We thus obtain

$$\begin{pmatrix}\xi_M\\ \upsilon_M\\ \zeta_M\end{pmatrix} = \begin{pmatrix}\cos a_M \cos A_M\\ \cos a_M \sin A_M\\ \sin a_M\end{pmatrix} = \begin{pmatrix}-\cos A_R & \sin A_R & 0\\ -\sin A_R & -\cos A_R & 0\\ 0 & 0 & 1\end{pmatrix} \begin{pmatrix}\sin a_R & 0 & -\cos a_R\\ 0 & 1 & 0\\ \cos a_R & 0 & \sin a_R\end{pmatrix} \begin{pmatrix}\cos\delta_M \cos\phi_M\\ \cos\delta_M \sin\phi_M\\ \sin\delta_M\end{pmatrix}. \tag{16.32}$$

The coordinates of the module in the horizontal coordinate system then are given by

$$\xi_M = \cos a_M \cos A_M = -\cos A_R \sin a_R \cos\delta_M \cos\phi_M + \sin A_R \cos\delta_M \sin\phi_M + \cos A_R \cos a_R \sin\delta_M, \tag{16.33a}$$
$$\upsilon_M = \cos a_M \sin A_M = -\sin A_R \sin a_R \cos\delta_M \cos\phi_M - \cos A_R \cos\delta_M \sin\phi_M + \sin A_R \cos a_R \sin\delta_M, \tag{16.33b}$$
$$\zeta_M = \sin a_M = \cos a_R \cos\delta_M \cos\phi_M + \sin a_R \sin\delta_M. \tag{16.33c}$$

Dividing Eq. (16.33b) by Eq. (16.33a) and leaving Eq. (16.33c) unchanged leads to the final expressions for the module orientation in horizontal coordinates,

$$\tan A_M = \frac{-\sin A_R \sin a_R \cos\delta_M \cos\phi_M - \cos A_R \cos\delta_M \sin\phi_M + \sin A_R \cos a_R \sin\delta_M}{-\cos A_R \sin a_R \cos\delta_M \cos\phi_M + \sin A_R \cos\delta_M \sin\phi_M + \cos A_R \cos a_R \sin\delta_M}, \tag{16.34a}$$
$$\sin a_M = \cos a_R \cos\delta_M \cos\phi_M + \sin a_R \sin\delta_M. \tag{16.34b}$$

Finally, the cosine of the angle between the module orientation and the solar position is given by

$$\begin{aligned}\cos\gamma = \mathbf{n}_M \cdot \mathbf{n}_S = {} & \cos a_S \cos(A_R - A_S)\left(\cos a_R \sin\delta_M - \sin a_R \cos\delta_M \cos\phi_M\right)\\ & + \cos a_S \sin(A_R - A_S)\cos\delta_M \sin\phi_M + \sin a_S \left(\cos a_R \cos\delta_M \cos\phi_M + \sin a_R \sin\delta_M\right).\end{aligned} \tag{16.35}$$

We will try to understand these results by discussing easy examples:

First, we look at a roof that faces eastward and has a tilt angle θ_R. Then, $a_R = 90^\circ - \theta_R$ and $A_R = 90^\circ$. On this roof a solar module is installed under a tilting angle θ_M with respect to the roof. The modules are mounted parallel to the gradient of the roof. We thus have $\delta_M = 90^\circ - \theta_M$ and $\phi_M = 270^\circ$. From Eqs. (16.34) we thus obtain

$$\begin{aligned} \sin a_M &= \cos a_R \cos \delta_M \cdot 0 + \sin a_R \sin \delta_M. \\ &= \sin a_R \sin \delta_M \end{aligned} \tag{16.36a}$$

$$\begin{aligned} \tan A_M &= \frac{-0 - 0 + 1 \cdot \cos a_R \sin \delta_M}{0 + \sin A_R \cos \delta_M \sin \cdot 1 + 0} \\ &= \cos a_R \tan \delta_M. \end{aligned} \tag{16.36b}$$

In the second example, the roof is facing southwards and tilted under an angle θ_R. We thus have $a_R = 90^\circ - \theta_R$ and $A_R = 180\circ$. Now the module tilted under an angle θ_M with respect to the roof and mounted perpendicular to the gradient of the roof. Hence, $\delta_M = 90^\circ - \theta_M$ and $\phi_M = 180^\circ$. Using Eqs. (16.34) we find

$$\begin{aligned} \sin a_M &= \cos a_R \cos \delta_M \cdot (-1) + \sin a_R \sin \delta_M. \\ &= \cos(a_R + eta a_M) = \sin(a_R - \theta_M). \end{aligned} \tag{16.37a}$$

$$\tan A_M = \frac{-0 - 0 + 0}{-\sin a_R \cos \delta_M + 0 + \cos A_R \sin \delta_M} = 0. \tag{16.37b}$$

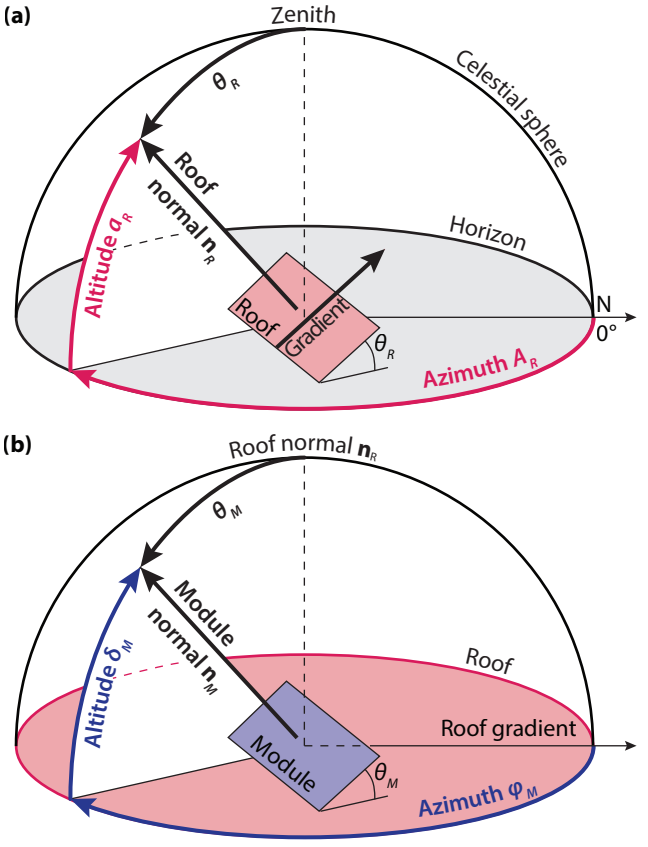

Figure 16.12: Illustrating the angles used to describe (a) the orientation of a roof on a horizontal plane and (b) the orientation of a module mounted on a roof.

16.5 Direct and diffuse irradiance

As sunlight traverses the atmosphere, it is partially scattered, leading to an attenuation of the *direct beam* component. On the other hand,the scattered light also partially will arrive at on the earths surface as *diffuse light*. For PV applications it is important to be able to estimate the strength of the direct and diffuse components.

First, we discuss a simple model that allows to estimate the irradiance on a *cloudless sky* in dependence of the *air mass* and hence the altitude of the sun. As we have seen in section 5.5, the air mass is defined as

$$\mathrm{AM} = \frac{1}{\cos\theta} = \frac{1}{\sin a_S}, \tag{16.38}$$

where we used that the angle between the sun and the zenith θ is connected to the solar altitude via $\theta = 90^\circ - a_S$. This equation, however, does not take the curvature of the earth into account. If the curvature is taken into account, we find [85]

$$\mathrm{AM}(a_S) = \frac{1}{\sin a_S + 0.50572(6.07995 + a_S)^{-1.6364}}. \tag{16.39}$$

To estimate the direct irradiance at a certain solar altitude a_S and altitude of the observer h, we can use the following empirical equation [86]

$$I_e^{\mathrm{dir}} = I_e^0 \left[(1 - ch) \cdot 0.7^{\left(\mathrm{AM}^{0.678}\right)} + ch\right], \tag{16.40}$$

with the constant $c = 0.14$. The solar constant is given as $I_e^0 = 1361\ \mathrm{Wm}^{-2}$. In a first approximation, the diffuse irradiance is about 10% of the direct irradiance. For the global irradiance we hence obtain [87]

$$I_e^{\mathrm{global}} \approx 1.1 \cdot I_e^{\mathrm{dir}}. \tag{16.41}$$

A more accurate model was developed in the framework of the *European Solar Radiation Atlas* [88]. In that model the direct irradiance for clear sky is given by

$$I_e^{\mathrm{dir}} = I_e^0 \varepsilon \exp\left[-0.8662\, T_L(\mathrm{AM2})\, m\, \delta_R(m)\right]. \tag{16.42}$$

I_0 is the *solar constant* that takes a value of $1361\ \mathrm{W/m^2}$. The factor ε allows to correct for deviations of the sun-earth distance from its mean value. a_S is the solar altitude angle. $T_L(\mathrm{AM2})$ is the *Linke turbidity factor* with that the haziness of the atmosphere is taken into account. In this equation its value at an Air Mass 2 is used. m is the relative optical air mass, and finally $\delta_R(m)$ is the integral Rayleigh optical thickness. The different components can be evaluated as follows:

The *correction factor* ε is given by

$$\varepsilon = \frac{I_e(R)}{I_e^0} = \frac{R^2}{\mathrm{AU}^2}. \tag{16.43}$$

The distance between earth and sun as a multiple of astronomic units (AU) is given in Eq. (16.7). We thus

obtain

$$\varepsilon = (1.00014 - 0.01671\cos g - 0.00014\cos 2g)^2, \quad (16.44)$$

which leads to annual variations of about $\pm 3.3\%$.

The *Linke Turbidity factor* approximates absorption and scattering in the atmosphere and takes both absorption by water vapour and scattering by aerosol particles into account. It is a unit-less number and typically takes values between 2 for very clear skies and 7 for heavily polluted skies.

The relative optical air mass m expresses the ratio of the optical path length of the solar beam through the atmosphere to the optical path through a standard atmosphere at sea level with the Sun at zenith. In can be approximated as a function of the solar altitude a_S by

$$m(a_S) = \frac{\exp(-z/z_h)}{\sin a_S + 0.50572(a_S + 6.07995)^{-1.6364}}. \quad (16.45)$$

Here, z is the site elevation and z_h is the scale height of the Rayleigh atmosphere near the Earth surface, given by 8434.5 m.

Finally, the Rayleigh optical thickness $\delta_R(m)$ is given by

$$\frac{1}{\delta_R(m)} = 6.62960 + 1.75130\,m - 0.12020\,m^2 + 0.00650\,m^3 - 0.00013\,m^4. \quad (16.46)$$

In their paper, Rigollier *et al.* also take the effect of of refraction at very low altitudes into account [88]. This, however is not relevant for our application.

They also present and expression for the diffuse irradiance of the light, which is given by

$$I_e^{\mathrm{dif}} = I_e^0 \varepsilon T_{rd}[T_L(\mathrm{AM2})] F_d[a_S,\, T_L(\mathrm{AM2})], \quad (16.47)$$

where T_{rd} is the diffuse transmission function at zenith, which is a second-order polynomial of $T_L(\mathrm{AM2})$ T_{rd} has typical values in between 0.05 for very clear skies and 0.22 for a very turbid atmosphere. F_d is a diffuse angular function, given as a second-order polynomial of $\sin a_S$. For more details we refer to the paper [88].

16.6 Shadowing

Shadowing had to be kept in mind when planning a PV system that consists of several rows of PV modules, which are placed behind each other. In this section we will determine how far behind the module the shadow reaches in dependence of the solar position, the module orientation and the length l of the module. Figure 16.13 shows the important notions that we use in this derivation.

For the determination we look at a module that is tilted at an angle θ_M. Its normal angle has an azimuth A_M.

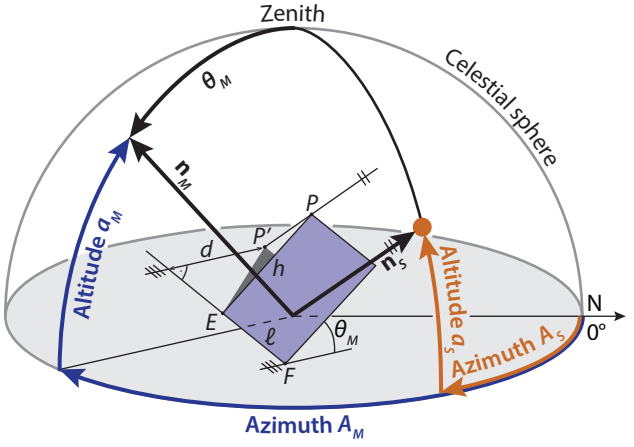

Figure 16.13: Derivation of the length d of a shadow behind a PV module with ground length l and height h. The module has the normal $\mathbf{n}_M$ while the position of the sun is given by the direction $\mathbf{n}_S$. The length of the shadow d is given as the distance between the line connecting the lower module corners E and F and the projection of the upper corner P on the ground (P').

This module touches the ground at two corner points that we call E and F. Without loss of generality, we may assume that E is at the origin of our horizontal coordinate system, $E = (0, 0, 0)$. Further, P is the corner point of top the module lying above E. The length of the module, *i.e.* the distance between E and P is l, $\overline{EP} = l$. The shadow of this point on the horizontal plane, we denote by P'. Then, the *length of the shadow* d is defined as the shortest distance between P' and the line g, which connects E with F.

For the determination of d we first must derive P, and then P'. The normal of the module $\mathbf{n}_M$ is given by

$$\mathbf{n}_M = \begin{pmatrix} \cos a_M \cos A_M \\ \cos a_M \sin A_M \\ \sin a_M \end{pmatrix} = \begin{pmatrix} \sin \theta_M \cos A_M \\ \sin \theta_M \sin A_M \\ \cos \theta_M \end{pmatrix}. \quad (16.48)$$

The direction vector $\mathbf{r}$ of the line g connecting E with F is then given as

$$\mathbf{r} = \begin{pmatrix} \sin A_M \\ -\cos A_M \\ 0 \end{pmatrix}. \quad (16.49)$$

Then, we can calculate the direction vector $\mathbf{h}$ of the line

that connects E with P with the vector product

$$\begin{aligned}\mathbf{h} &= -\mathbf{n}_M \times \mathbf{r} \\ &= -\begin{pmatrix}\sin\theta_M \cos A_M \\ \sin\theta_M \sin A_M \\ \cos\theta_M\end{pmatrix} \times \begin{pmatrix}\sin A_M \\ -\cos A_M \\ 0\end{pmatrix} \\ &= \begin{pmatrix}-\cos\theta_M \cos A_M \\ -\cos\theta_M \sin A_M \\ \sin\theta_M\end{pmatrix},\end{aligned} \tag{16.50}$$

where the $-$ sign is because of the fact that the horizontal coordinate system is left-hand. Since $\mathbf{h}$ has length 1, *i.e.* it is a unit vector, we easily can derive the position of the point P with

$$P = E + l \cdot \mathbf{h} = l \begin{pmatrix}-\cos\theta_M \cos A_M \\ -\cos\theta_M \sin A_M \\ \sin\theta_M\end{pmatrix}. \tag{16.51}$$

For calculating the position of P' we define the line s, which goes through P and points towards the sun,

$$s(t) = P + t \cdot \mathbf{n}_S. \tag{16.52}$$

When the position of the sun is described by its altitude a_S and azimuth A_S, we find

$$s(t) = l \begin{pmatrix}-\cos\theta_M \cos A_M \\ -\cos\theta_M \sin A_M \\ \sin\theta_M\end{pmatrix} + t \begin{pmatrix}\cos a_S \cos A_S \\ \cos a_S \sin A_S \\ \sin a_S,\end{pmatrix} \tag{16.53}$$

We find the shadow P' of the point P as the *intersection* of the line s with the horizontal plane, $z = 0$,

$$l \sin\theta_M + t \sin a_S = 0. \tag{16.54}$$

Hence,

$$t = -l \frac{\sin\theta_M}{\sin a_S}. \tag{16.55}$$

The coordinates of $P' = (P'_x, P'_y, 0)$ then are given as

$$P'_x = -l\left(\cos\theta_M \cos A_M + \sin\theta_M \cot a_S \cos A_S\right), \tag{16.56a}$$

$$P'_y = -l\left(\cos\theta_M \sin A_M + \sin\theta_M \cot a_S \sin A_S\right). \tag{16.56b}$$

As stated already earlier, the length of the shadow d is given as the shortest distance between P' and the line g connecting E and F. Let g' be the line through P' that is perpendicular to g, $g \perp g'$. Since $E = (0, 0, 0)$ we find for g and g'

$$g(u) = \quad u \cdot \mathbf{r}, \tag{16.57a}$$

$$g'(v) = P' + v \cdot \mathbf{r}', \tag{16.57b}$$

where the direction vector $\mathbf{r}'$ is given as

$$\mathbf{r}' = \begin{pmatrix}\cos A_M \\ \sin A_M \\ 0\end{pmatrix}. \tag{16.58}$$

d is the distance between P' and the intersection of g

with g'. At this intersection we have

$$g(u) = g'(v),$$
$$u \begin{pmatrix} \sin A_M \\ -\cos A_M \\ 0 \end{pmatrix} = P' + v \begin{pmatrix} \cos A_M \\ \sin A_M \\ 0 \end{pmatrix}$$

$$u \sin A_M = P'_x + v \cos A_M$$
$$-u \cos A_M = P'_y + v \sin A_M$$

$$u \sin A_M \cos A_M = P'_x \cos A_M + v \cos^2 A_M$$
$$-u \cos A_M \sin A_M = P'_y \sin A_M + v \sin^2 A_M$$

By adding the last two equations we find

$$P'_x \cos A_M + v \cos^2 A_M + P'_y \sin A_M + v \sin^2 A_M = 0, \tag{16.59}$$

and hence

$$v = -P'_x \cos A_M - P'_y \sin A_M. \tag{16.60}$$

Using Eqs. (16.56), we derive

$$\begin{aligned} v =& l[(\cos\theta_M \cos A_M + \sin\theta_M \cot a_S \cos A_S) \cos A_M \\ &+ (\cos\theta_M \sin A_M + \sin\theta_M \cot a_S \sin A_S) \sin A_M] \\ =& l(\cos\theta_M \cos^2 A_M + \sin\theta_M \cot a_S \cos A_S \cos A_M \\ &+ \cos\theta_M \sin^2 A_M + \sin\theta_M \cot a_S \sin A_S \sin A_M). \end{aligned} \tag{16.61}$$

Because the direction vector $\mathbf{r}'$ of the line g' is a unit vector, the length of the shadow d is equal to v. Therefore we obtain from Eq. (16.61) with some trigonometric operations

$$d = l\left[\cos\theta_M + \sin\theta_M \cot a_S \cos(A_M - A_S)\right]. \tag{16.62}$$

As a rule of thumb, d should be at least three times l, $d > 3l$.

Example

A PV system should be installed on a flat roof in Naples (Italy). The area of the roof that can be utilized for installing the PV system is 10×10 m^2. *The roof is oriented such that the sides are parallel to the East-West and North-South directions, respectively.*

The owner of the roof decides to use Yingli PANDA 60 *modules with dimensions of* $1650 \times 990 \times 40$ mm^3. *The modules are installed facing south with a tilt of* $30°$.

He wants to install ***as many modules as possible*** *under the condition that on the shortest day of the year* ***no mutual shading must occur for the duration of 6 hours***.

Should the modules be mounted with the long or short side touching the ground? How many modules can be mounted in this case?

Answer: *The shortest day of course is 21 December. The solar position on this day at 9:00 and 15:00 is*

Time	*Altitude* (°)	*Azimuth* (°)
9:00	13.59	138.55
15:00	13.13	222.17

Because of the Equation of Time, *the sun is not at its highest point at exactly 12:00 noon. We see, that the solar altitude at 15:00 is just slightly lower than at 9:00. Thus, when using 9:00 for calculating the length of the shadow, the duration without mutual shading will be slightly shorter than 6 hours. Thus, we use the position at 15:00 for the calculation.*

The length of the shadow can be calculated with equation (16.62)

$$d = l\left[\cos\theta_M + \sin\theta_M \cot a_S \cos(A_M - A_S)\right].$$

We have θ_M=30°, A_M=180°, a_S=13.13°, *and* A_S=222.17°.

If the module is mounted to the ground on the short side, we have l=1650 mm. *Hence, we find* d=4050 mm. *The area directly beneath the module at the last row is*

$$d' = l\cos\theta_M = 1429 \text{ mm}.$$

Thus, we can mount ***three rows*** *behind each other, because*

$$2d + d' = 2 \cdot 4050 + 1429 = 9529 \text{ mm}, \qquad (16.63)$$

which is less than 10 m. *In one row fit* 10 *modules because* $10 \cdot 990 = 9900$ mm. *Thus, we can place* 30 *modules.*

If the module is mounted to the ground on the long side, we have l=990 mm. *Hence, we find* d=2430 mm. *The area directly beneath the module at the last row is*

$$d' = l\cos\theta_M = 857 \text{ mm}.$$

Thus, we can mount ***four rows*** *behind each other, because*

$$3d + d' = 3 \cdot 2430 + 857 = 8147 \text{ mm},$$

which is lower than 10 m. *In one row fit 6 modules because* $6 \cdot 1650 = 9900$ mm. *Thus, we can place 24 modules.*

17 Components of PV Systems

17.1 PV modules

In this section we will discuss *PV modules* (or solar modules), their fabrication and how to to determine their performance.

Before we start with the actual treatment of PV modules, we briefly want to introduce different terms. Figure 17.1 (a) shows a crystalline *solar cell*, which we discussed in Chapter 12. For the moment we will consider only modules that are made from this type of solar cells. A *PV module*, is a larger device in which many solar cells are connected, as illustrated in Fig. 17.1 (b). The names PV module and solar module are often used interchangeably. A *solar panel*, as illustrated in Fig. 17.1 (c), consists of several PV modules that are electrically connected and mounted on a supporting structure. Finally, a *PV array* consists of several solar panels. An example of such an array is shown in Fig. 17.1 (d). This array consists of two strings of two solar panels each, where *string* means that these panels are connected *in series*.

17.1.1 Series and parallel connections in PV modules

If we make a solar module out of an ensemble of solar cells, we can connect the solar cells in different ways: first, we can connect them in a *series connec-*

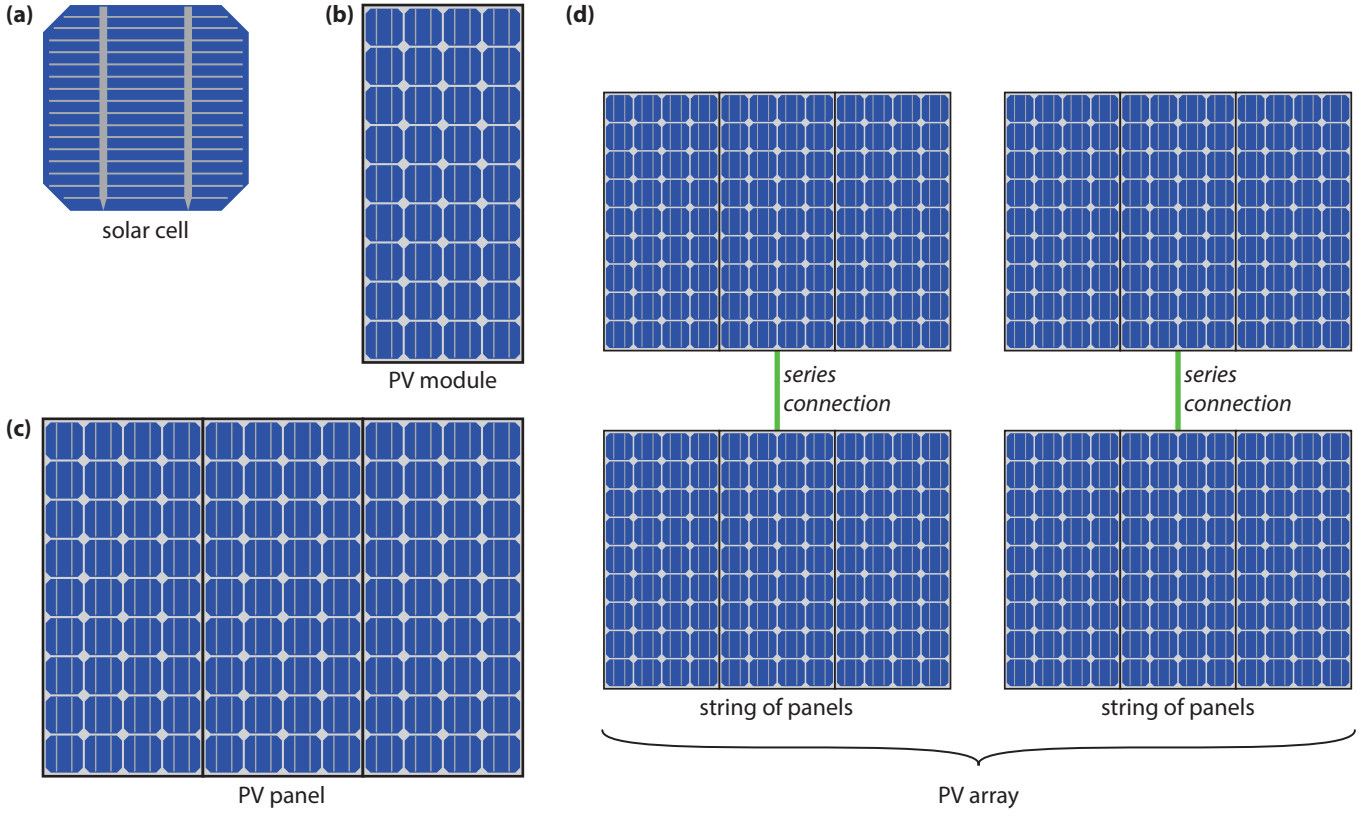

Figure 17.1: Illustrating (a) a solar cell, (b) a PV module, (c) a solar panel, and (d) a PV array.

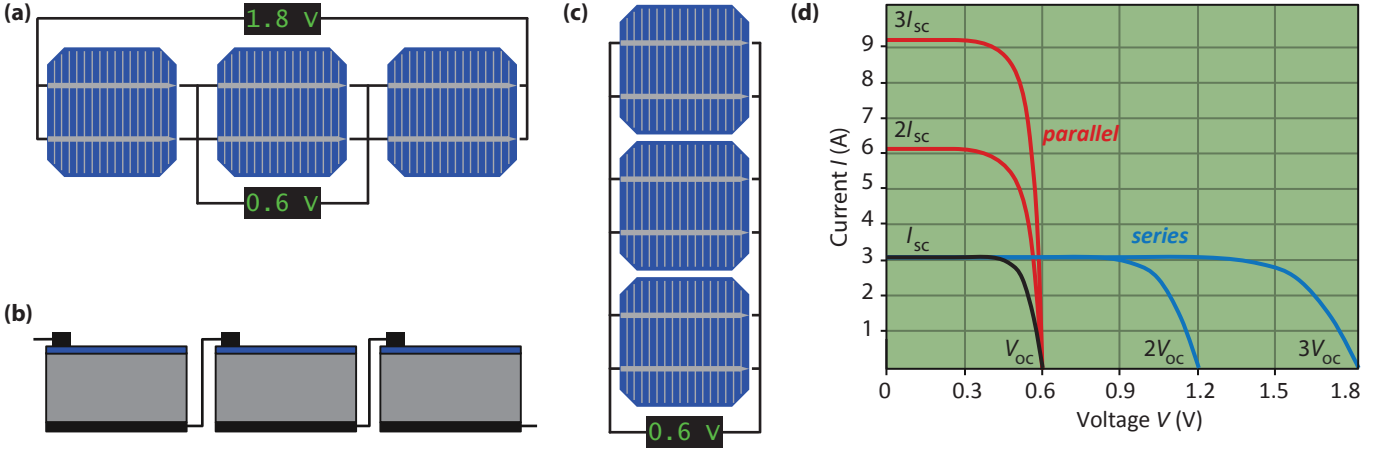

Figure 17.2: Illustrating (a) a series connection of three solar cells and (b) realisation of such a series connection for cells with a classical front metal grid. (c) Illustrating a parallel connection of three solar cells. (d) *I-V* curves of solar cells connected in series and parallel.

tion as shown in Fig. 17.2 (a). In a series connection the voltages add up. For example, if the open circuit voltage of one cell is equal to 0.6 V, a string of three cells will deliver an open circuit voltage of 1.8 V. For solar cells with a classical front metal grid, a series connection can be established by connecting the bus bars at the front side with the back contact of the neighbouring cell, as illustrated in Fig. 17.2 (b). For series connected cells, the current does not add up but is determined by the photocurrent in each solar cell. Hence, the total current in a string of solar cells is equal to the current generated by one single solar cell.

Figure Fig. 17.2 (d) shows the *I-V* curve of solar cells connected in series. If we connect two solar cells in series, the voltages add up while the current stays the same. The resulting open circuit voltage is two times that of the single cell. If we connect three solar cells in series, the open circuit voltage becomes three times as large, whereas the current still is that of one single solar cell.

Secondly, we can connect solar cells in *parallel* as illustrated in Fig. 17.2 (c), which shows three solar cells connected in parallel. If cells are connected in parallel, the voltage is the same over all solar cells, while the currents of the solar cells add up. If we connect *e.g.* three cells in parallel, the current becomes three times as large, while the voltage is the same as for a single cell, as illustrated in Fig. 17.2 (d).

The reader may have noticed that we used *I-V* curves, *i.e.* the *current-voltage* characteristics, in the previous paragraphs. This is different to Parts II and III, where we used *I-V* curves instead, *i.e.* the *current density - voltage* characteristics. The reason for this switch from J to I is that on module level, the total current that the module can generate is of higher interest than the current density. As the area of a module is a constant, the shapes of the *I-V* and *J-V* curves of a module are similar.

For a total module, therefore the voltage and current output can be partially tuned via the arrangements of the solar cell connections. Figure 17.3 (a) shows a typical PV module that contains 36 solar cells connected in series. If a single junction solar cell would have a short circuit current of 5 A, and an open circuit voltage of 0.6 V, the total module would have an output of $V_{oc} = 36 \cdot 0.6\,V = 21.6\,V$ and $I_{sc} = 5\,A$. However, if two strings of 18 series-connected cells are connected in parallel, as illustrated in Fig. 17.3 (b), the output of the module will be $V_{oc} = 18 \cdot 0.6\,V = 10.8\,V$ and $I_{sc} = 2 \times 5\,A = 10\,A$. In general, for the *I-V* characteristics of a module consisting of m identical cells in series and n identical cells in parallel the voltage multiplies by a factor m while the current multiplies by a factor n. Modern PV modules often contain 60 (10×6), 72 (9×8) or 96 (12×8) solar cells that are usually all connected in series in order to minimise resistive losses.

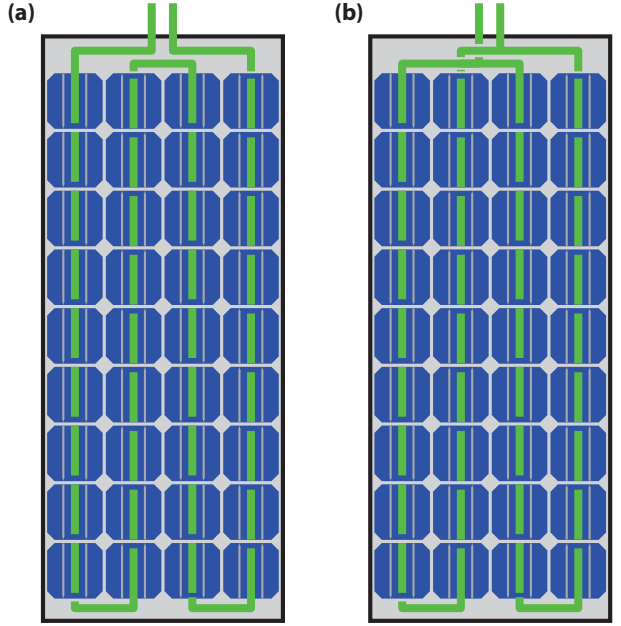

Figure 17.3: Illustrating a PV module consisting (a) of a string of 36 solar cells connected in series and (b) of two strings of 18 solar cells each that are connected in parallel.

17.1.2 PV module parameters

In a nutshell, for a PV module a set of parameters can be defined, similar than for solar cells. The most common parameters are the *open circuit voltage* V_{oc}, the *short circuit current* I_{sc} and the *module fill factor* FF_M. On module level, we have to distinguish between the *aperture area efficiency* and the *module efficiency*. The aperture area is defined as the area of the PV-active parts only. The total module area is given as the aperture area plus the dead area consisting of the interconnections and the edges of the module. Clearly, the aperture area efficiency is larger than the module efficiency.

Determining the the efficiency and the fill factor of a PV module is less straight-forward than determining voltage and current. In an ideal world with perfectly matched solar cells and no losses, one would expect that the efficiency and fill factor at both the module and cell levels to be the same. This is not the case in real life. As mentioned above. The cells are connected with each other using interconnects that induce resistive losses. Further, there might be small mismatches in the interconnected cells. For example, if $m \times n$ cells are interconnected, the cell with the lowest current in a string of m cells in series determines the module current. Similarly, the string with the lowest voltage in the n strings that are connected in parallel dictates the module voltage. The reason for mismatch between individual cells are inhomogeneities that occur during

the production process. Hence, in practice PV module perform a little less than what one would expect from ideally matched and interconnected solar cells. This loss in performance translates to a lower fill factor and efficiency at module level. If the illumination across the module is not constant or if the module is heats up non-uniformly, the module performance reduces even further.

Often, differences between cell and module performance are mentioned in datasheets that are provide by the module manufacturers. For example, the datasheet of a Sanyo HIT-N240SE10 module gives a cell level efficiency of 21.6%, but a module level efficiency of only 19%. Despite all the technological advancements being made at solar cell level for improving the efficiency, still a lot must be done at the PV systems level to ensure a healthy PV yield. For the performance of a PV system, not only the module performance is important, but also the yield of the PV system.

17.1.3 Partial shading and bypass diodes

PV modules have so-called *bypass diodes* integrated. To understand the reason for using such diodes, we have to consider modules in real-life conditions, where they can be *partially shaded*, as illustrated in Fig. 17.4 (a). The shade can be from an object nearby, like a tree, a chimney or a neighbouring building. It also can be caused by a leaf that has fallen from a tree. Partial shading can have significant consequences for the output of the solar module. To understand this, we consider the situation in which one solar cell in the module shaded for a large part shaded. For simplicity, we assume that all six cells are connected in series. This means that the current generated in the shaded cell is significantly reduced. In a series connection the current is limited by the cell that generates the lowest current, this cell thus dictates the maximum current flowing through the module.

In Fig. 17.4 (b) the theoretical I-V curve of the five unshaded solar cells and the shaded solar cell is shown. If the cells are connected to a constant load R, the voltage across the module is dropping due to the lower current generated. However, since the five unshaded solar cells are forced to produce high voltages, they act like a reverse bias source on the shaded solar cell. The dashed line in Fig. 17.4 (b) represents the reverse bias load put on the shaded cell, which is the I-V curve of the five cells, reflected across the vertical axis equal to 0 V. Hence, the shaded solar cell does not generate energy, but starts to dissipate energy and heats up. The temperature can increase to such a critical level, that the encapsulation material cracks, or other materials wear out. Further, high temperatures generally lead to a decrease of the PV output as well.

These problems occurring from partial shading can be

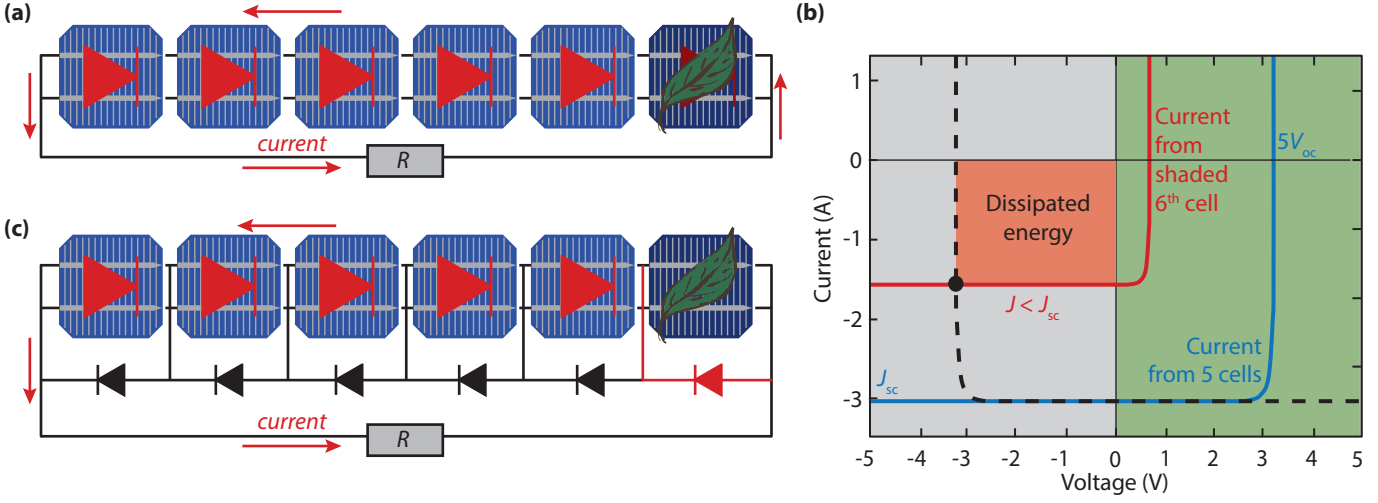

Figure 17.4: Illustrating (a) string of six solar cells of which one is partially shaded, which (b) has dramatic effects on the *I-V* curve of this string. (c) Bypass diodes can solve the problem of partial shading.

prevented by including bypass diodes in the module, as illustrated in 17.4 (c). As discussed in Chapter 8, a diode blocks the current when it is under negative voltage, but conducts a current when it is under positive voltage. If no cell is shaded, no current is flowing through the bypass diodes. However, if one cell is (partially) shaded, the bypass diode starts to pass current through because of the biasing from the other cells. As a result current can flow around the shaded cell and the module can still produce the current equal to that of a unshaded single solar cell.

For cells that are connected in parallel, partial shading is less of a problem, because the currents generated in the others cells do not need to travel through the shaded cell. However, a module consisting of 36 cells in parallel have very high currents (above 100 A) combined with a very low voltage (approx. 0.6 V). This combination would lead to very high resistive losses in the cables; further an inverter that has only 0.6 V as input will not be very efficient, as we will see in Section 17.3. Therefore, combining the cells in series and using bypass diodes is much better an option to do.

17.1.4 Fabrication of PV modules

As discussed in the subsection 17.1.5, a PV module must withstand various influences in order to survive a lifetime of 25 years or even longer. In order to ensure

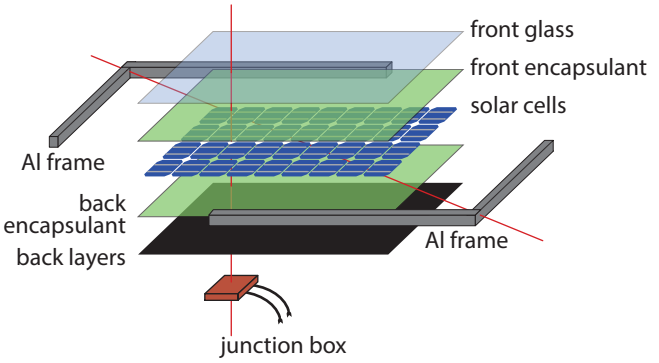

Figure 17.5: The components of a typical c-Si PV module.

a long lifetime, the components of that a PV module is built must be well chosen. Fig. 17.5 shows the typical components of a usual crystalline silicon PV module. Of course, the layer stack may consist of different materials dependent on the manufacturer. The major components are [89]:

- *Soda-lime* glass with a thickness of several millimetres, which provides mechanical stability while being transparent for the incident light. It is important the glass has a low iron content because iron leads to absorption of light in the glass which can lead to losses. Further, the glass must be *tempered* in order to increase its resistance to impacts.

- The solar cells are sandwiched in between two layers of *encapsulants*. The most common material is *ethylene-vinyl-acetate* (EVA), which is a thermoplastic polymer (plastic). This means that it goes into shape when it is heated but that these changes are reversible.
- The *back layer* acts as a barrier against humidity and other stresses. Depending on the manufacturer, it can be another glass plate or a composite polymer sheet. A material combination that is often used is PVF-polyester-PVF, where PVF stands for *polyvinyl fluoride*, which is often known by its brand name *Tedlar*®. PVF has a low permeability for vapours and is very resistive against weathering. A typical polyester is *polyethylene terephthalate* (PET)
- A *frame* usually made from aluminium is put around the whole module in order to enhance the mechanical stability.
- A *junction box* usually is placed at the back of the module. In it the electrical connections to the solar cell are connected with the wires that are used to connect the module to the other components of the PV system.

One of the most important steps during module production is *laminating*, which we briefly will explain for the case that EVA is used as encapsulant [89]. For lamination, the whole stack consisting of front glass, the encapsulants, the interconnected solar cells, and the back layer are brought together in a laminator, which is heated above the melting point of EVA, which is around 120°C. This process is performed *in vacuo* in order to ensure that air, moisture and other gasses are removed from within the module stack. After some minutes, when the EVA is molten, pressure is applied and the temperature is increased to about 150°C. Now the *curing* process starts, *i.e.* a curing agent, which is present in the EVA layer, starts to cross-link the EVA chains, which means that transverse bonds between the EVA molecules are formed. As a result, EVA has elastomeric, rubberlike properties.

The choice of the layers that light traverses before entering the solar cell is also very important from an optical point of view. If this layers have an increasing refractive index, they act as antireflective coating and thus can enhance the amount of light that is in-coupled in the solar cell and finally absorbed, which increases the current produced by the solar cell.

17.1.5 Lifetime testing of PV Modules

The typical lifetime of PV systems is about 25 years. In these as little maintenance as possible should be required on the system components, especially the PV modules are required to be maintenance free. Furthermore, degradation in the different components of that

the module is made should be little: manufacturers typically guarantee a power between 80% and 90% of the initial power after 25 years. During the lifetime of 25 years or more, PV modules are exposed to various external stress from various sources [90]:

- *temperature* changes between night and day as well as between winter and summer,
- *mechanical stress* for example from wind, snow and hail,
- stress by agents transported via the *atmosphere* such as dust, sand, salty mist and other agents,
- *moisture* originating from rain, dew and frost,
- *humidity* originating from the atmosphere,
- *irradiance* consisting of direct and indirect irradiance from the sun; mainly the highly-energetic UV radiation is challenging for many materials.

Before PV modules are brought to the market, they are usually tested extensively in order to assure their stability against these various stresses. The required tests are extensively defined in the standards *IEC 61215* for modules based on crystalline silicon solar cells and in *IEC 61646* for thin-film modules. Since the modules cannot be tested during a period of 25 years, *accelerated stress testing* must be performed. The required tests are [91]:

- *Thermal cycles* for studying whether thermal stress leads to broken interconnects, broken cells, electrical bond failure, adhesion of the junction box,...
- *Damp heat* testing to see whether the modules suffer from corrosion, delamination, loss of adhesion and elasticity of the encapsulant, adhesion of the junction box,...
- *Humidity freeze* testing in order to test delamination, adhesion of the junction box, ...
- *UV testing*, because UV light can lead to delamination, loss of adhesion and elasticity of the encapsulant, ground fault due to backsheet degradation. Mainly, UV light can lead to a discoloration of the encapsulant and back sheet, which means that they get yellow. This can lead to losses in the amount of light that reaches the solar cells.
- *Static mechanical loads* in order to test whether strong winds or heavy snow loads lead to structural failures, broken glass, broken interconnect ribbons or broken cells.
- *Dynamic mechanical load*, which can lead to roken glass, broken interconnect ribbons or broken cells.
- *Hot spot* testing in order to see whether hot spots due to shunts in cells or inadequate bypass diode protection are present.
- *Hail testing* to see whether the module can handle the mechanical stress induced by hail.

- *Bypass diode thermal testing* to study whether overheating of these diodes causes degradation of the encapsulant, backsheet or the junction box.
- *Salt spray* testing to see whether salt that is present in salty mist or that is used in salty water for snow and ice removal leads to corrosion of PV module components.

How these tests are to be performed is defined in other standards, for example *IEC 61345* for UV testing and *IV 61701* for salt-mist corrosion testing Usually these tests are carried out by organisations like TÜV Rheinland. Refining the test requirements and understanding which accelerated tests are required to guarantee a lifetime of 25 years and more is subject to ongoing research and development.

17.1.6 Thin-film modules

Making thin-film modules is very different from making modules from c-Si solar cells. While for c-Si technology producing solar cells and producing PV modules are two distinct steps, in thin-film technology producing cells and modules cannot be separated from each other. To illustrate this we look at a PV module where the thin-films are deposited in *superstrate configuration* on glass, as illustrated in Fig. 17.6. For making such a module, a transparent front contact, a stack of (photo)active layers that also contain one or more semiconductor junctions, and a metallic back contact are deposited onto each other. In industrial production, the glass plates on that these layers are deposited can be very large, with sizes significantly exceeding 1×1 m^2.

Such a stack of layers deposited onto a large glass plate in principle forms one very large solar cell that will produce a very high current. Since all the current would have to be transported across the front and back contacts, which are very thin, resistive losses in the module is even a bigger problem than for c-Si modules. Therefore, the module is produced such that it consists of many very narrow cells of about 1 cm width and the length being equal to the module length. These cells then are connected in series across the width of the module. On the very left and right of the module metallic busbars collect the current and conduct it to the bottom of the module where they are connected with external cables.

The series connection is established with *laser scribing*. In total, *three* laser scribes are required for separating two cells from each other and establishing a series connection between them. The first laser scribe, called P1, is performed after the transparent front contact is deposited, as shown in Fig. 17.6 (a). The wavelength of the laser is such that the laser light is absorbed in the front contact and the material is evaporated, leaving a "gap" in the front contact, as shown in Fig. 17.6 (b).

Then the photoactive layers are deposited onto the front contact and also fill the gaps. Then, the second laser scribe, called P2, is performed, as illustrated in Fig. 17.6 (c). The laser wavelength has to be chosen such that it is not absorbed in the transparent front contact but in the absorber layer. For example, if the absorber consists of amorphous silicon, green laser light can be used. The P2 scribe leaves a gap in the absorber layer, as illustrated in Fig. 17.6 (d). The next step is the deposition of the metallic back contact that also fills the P2 gap. Finally, the third laser scribe (P3) is performed as illustrated in Fig. 17.6 (e). The wavelength for this scribe has to be chosen such that it is neither absorbed in the front contact nor in the absorber stack, so it is, for example, infrared. The P3 scribe shoots a gap into the back contact, as shown in Fig. 17.6 (f).

To understand the action of the laser scribes, we take a look at Fig. 17.6 (f): the P1 scribe filled with absorber material forms an barrier, since the absorber is orders of magnitudes less conductive than the transparent front contact. Similar, the P3 scribes forms an insulating gap in the metallic back contact. However, the P2 scribe that also is filled with metal forms a highly conducting connection between the front and back contacts – here the actual series connection is performed.

For example, for making CIGS solar cells, first the molybdenum back contact is deposited on top of the glass substrate and the cell areas are defined by P1 laser

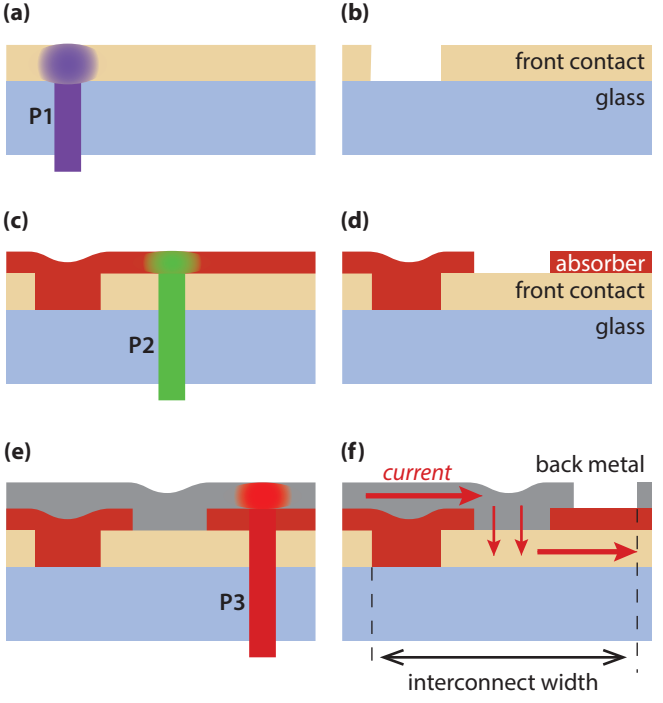

Figure 17.6: Schematic of creating an interconnect in thin-film module. (Explanation given in the text).

scribes. Then the CIGS p layer and the CdS n layer are deposited including a P2 laser scribe step. Finally the intrinsic and n-doped zinc oxide is deposited, followed by a final P3 laser scribe step. Now the front TCO electrode is connected with the Molybdenum back contact of the next solar cell.

The performance of such an interconnect established via laser scribes and hence the total module performance is determined by several things. First, the P2 scribe has to be highly conductive, This means that it has to be wide enough and that there must be no barrier at the interface between the front contact and the metal of the P2 scribe. Further, the P1 and P3 scribes must perform good barriers to effectively separate the cells from each other. Thirdly, the region between the P1 and P3 scribes does not contribute to the the current generated by the module. Therefore, the ratio between this width and the total cell width (including the scribes) has to be as small as possible. Another issue is the fact that the three laser scribes are performed in different steps of production and thus often in different machines. Further, the distance between the scribes might be different at the different processes when they are performed at different temperatures. This, aligning the glass plates in all the production steps is extremely important for manufacturing high-quality thin-film modules.

The production steps and also the exact processing of the laser scribes is of course dependent on which thin-film technology is used and even on the manufacturer itself. However, the basic principles and the action behind these processes is valid in general.

One advantage of thin-film PV technology is that they can be deposited onto flexible substrates. For example, the Dutch company HyET Solar developed a technology, where thin-film silicon layers are deposited onto a temporary aluminium substrate [92]. After the solar cell layers are encapsulated on the back side, the temporary substrate is etched away, and the front side is encapsulated. This results in a very low weight flexible substrate, which can be integrated for example in curved roof top elements. A very big advantage is that such very light modules can be installed on simple roof top constructions that only can handle little ballast. On such roofs, heave PV panels with glass cannot be installed. Further, if such flexible modules are directly integrated into roofing elements, installation costs can be reduces significantly. Often, installation costs are the largest contributor to the non-modular costs of a PV system. Currently, only thin-film silicon technologies have demonstrated flexible modules with reasonable efficiencies.

17.1.7 Some examples

Table 17.1 shows some parameters of PV modules using different PV technologies:

Table 17.1: Specifications of different PV modules.

			SunPower X21-345	Avancis PowerMax140	Kaneka U-EA120	First Solar FS-392
Technology			c-Si	CIS	a-Si/nc-Si	CdTe
Rated power	P_{mpp}	(W_p)	345	140	114	92.5
Rated current	I_{mpp}	(A)	6.02	2.98	2.18	1.94
Rated voltage	V_{mpp}	(V)	57.3	47	55	47.7
Short circuit current	I_{sc}	(A)	6.39	3.31	2.6	2.11
Open circuit voltage	V_{oc}	(V)	68.2	61.5	71	60.5
Dimensions		(m×m)	1.56×1.05	1.6×0.67	1.2×1.00	1.2×0.60
Max warranty on P_{mpp}		(years)	25	25	25	25

- A SunPower module made based on monocrystalline silicon solar cells,
- an Avancis module based on copper indium diselenide (CIS) technology,
- a Kaneka amorphous silicon /nanocrystalline silicon tandem (a-Si:H/nc-Si:H) module, and
- a module of First Solar based on cadmium telluride (CdTe) technology.

17.2 Maximum power point tracking

In this section we discuss the concept of *Maximum power point tracking* (MPPT). This concept is very unique to the field of PV Systems, and hence brings a very special application of power electronics to the field of photovoltaics. The concepts discussed in this section are equally valid for cells, modules, and arrays, although MPPT usually is employed at PV module/array level.

As discussed earlier, the behaviour of an illuminated solar cell can be characterised by an *I-V* curve. Interconnecting several solar cells in series or in parallel merely increases the overall voltage and/or current, but

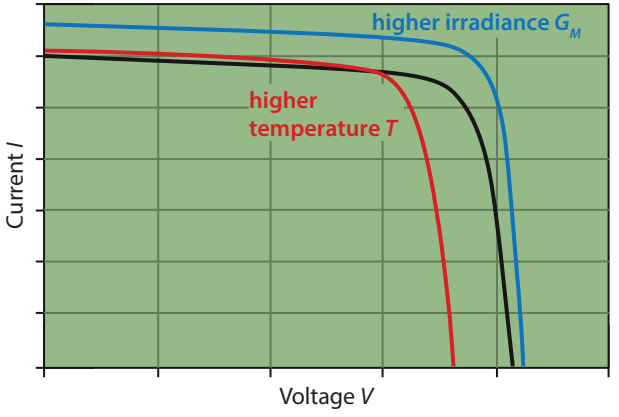

Figure 17.7: Effect of increased temperature T or irradiance G_M on the I-V curve.

does not change the shape of the I-V curve. Therefore, for understanding the concept of MPPT, it is sufficient to consider the I-V curve of a solar cell. The I-V curve is dependent on the module temperature on the irradiance, as we will discuss in detail in Section 18.3. For example, an increasing irradiance leads to an increased current and slightly increased voltage, as illustrated in Fig. 17.7. The same figure shows that an increasing temperature has a detrimental effect on the voltage.

Now we take a look at the concept of the *operating point*, which is the defined as the particular voltage and current, at that the PV module operates at any given point in time. For a given irradiance and temperature, the operating point corresponds to a unique (I, V) pair which lies onto the I-V curve. The power output at this operating point is given by

$$P = I \cdot V. \tag{17.1}$$

The operating point (I, V) corresponds to a point on the power-voltage (P-V) curve, shown in Fig. 17.8. For generating the highest power output at a given irradiance and temperature, the operating point should such correspond to the maximum of the (P-V) curve, which is called the *maximum power point* (MPP).

If a PV module (or array) is directly connected to an electrical load, the operating point is dictated by that load. For getting the maximal power out of the module, it thus is imperative to force the module to operate at the maximum power point. The simplest way of forcing the module to operate at the MPP, is either to force the voltage of the PV module to be that at the MPP (called V_{mpp}) or to regulate the current to be that of the MPP (called I_{mpp}).

However, the MPP is dependent on the ambient conditions. If the irradiance or temperature change, the I-V and the P-V characteristics will change as well and hence the position of the MPP will shift. Therefore, changes in the I-V curve have to be tracked continuously such that the operating point can be adjusted to

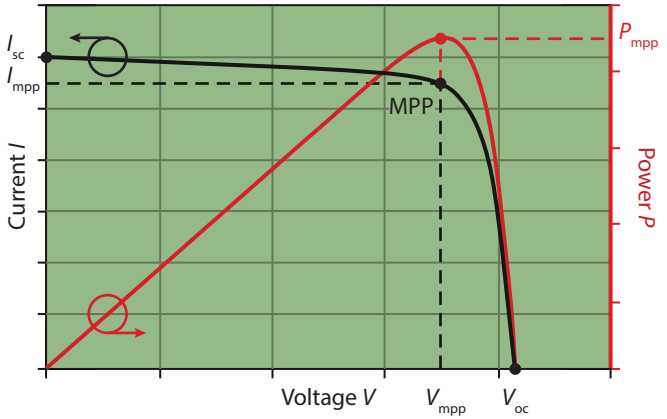

Figure 17.8: A generic I-V curve and the associated P-V curve. The maximum power point (MPP) is indicated.

be at the MPP after changes of the ambient conditions.

This process is called *Maximum Power Point Tracking* or MPPT. The devices that perform this process are called *MPP trackers*. We can distinguish between two categories of MPP tracking:

- *Indirect* MPP tracking, for example performed with the *Fractional Open Circuit Voltage* method.
- *Direct* MPP tracking, for example performed with the *Perturb and Observe* method or the the *Incremental Conductance* method.

All the MPPT algorithms that we discuss in this section are based on finding the and tuning the voltage until V_{MPP} is found. Other algorithms, which are not discussed in this section, work with the power instead and aim to find I_{MPP}.

17.2.1 Indirect MPPT

First, we discuss *indirect* MPP Tracking, where simple assumptions are made for estimating the MPP based on a few measurements.

Fixed voltage method

For example, in the *fixed voltage* method (also called *constant voltage method*), the operating voltage of the

solar module is adjusted only on a seasonal basis. This model is based on the assumption that for the same level of irradiance higher MPP voltages are expected during winter than during summer. It is obvious that this method is not very accurate. It works best at locations with minimal irradiance fluctuations between different days.

Fractional open circuit voltage method

One of the most common *indirect* MPPT techniques is the *fractional open circuit voltage* method. This method exploits the fact that – in a very good approximation – the V_{mpp} is given by

$$V_{\text{mpp}} = k \cdot V_{\text{oc}}, \tag{17.2}$$

where k is a constant. For crystalline silicon, k usually takes values in between 0.7 and 0.8. In general, k of course is dependent on the type of solar cells. As changes in the open circuit voltage can be easily tracked, changes in the V_{mpp} can be easily estimated just by multiplying with k. This method thus can be implemented easily. However, there are also certain drawbacks.

First, using a constant factor k only allows to roughly estimate the position of the MPP. Therefore, the operating point usually will not be exactly on the MPP but in its proximity, with is called the *MPP region*. Secondly, every time the system needs to respond to a change in illumination conditions, the V_{oc} must be measured. For this measurement, the PV module needs to be disconnected from the load for a short while, which will lead to a reduced total output of the PV system. The more often the V_{oc} is determined, the larger the loss in output will be. This drawback can be overcome my slightly modifying the method. For this modification a pilot PV cell is required, which is highly matched with the rest of the cells in the module. The pilot cell receives the same irradiance as the rest of the PV module, and a measurement of the pilot PV cell's V_{oc} also gives an accurate representation of that of the PV module, hence it can be used for estimating V_{mpp}. Therefore, the operating point of the module can be adjusted without needing to disconnect the PV module.

17.2.2 Direct MPPT

Now we discuss *direct* MPP tracking, which is more involved than indirect MPPT, because current, voltage or power measurements are required. Further, the system must response more accurately and faster than in indirect MPPT. We shall look at a couple of the most popular kind of algorithms.

Table 17.2: A summary of the possible options in the P&O algorithm.

Prior Perturbation	Change in Power	Next Perturbation
Positive	Positive	Positive
Positive	Negative	Negative
Negative	Positive	Negative
Negative	Negative	Positive

Perturb and observe (P&O) algorithm

The first algorithm that we discuss is the *Perturb and Observe* (P&O) algorithm, which also is known as "hill climbing" algorithm. In this algorithm, a perturbation is provided to the voltage at that the module is currently driven. This perturbation in voltage will lead to a change in the power output. If an increasing voltage leads to an increasing in power, the operating point is at a lower voltage than the MPP, and hence further voltage perturbation towards higher voltages is required to reach the MPP. In contrast, if an increasing voltage leads to a decreasing power, further perturbation towards lower voltages is required in order to reach the MPP. Hence, the algorithm will converge towards the MPP over several perturbations. This principle is summarised in Table 17.2.

A problem with this algorithm is that the operat-

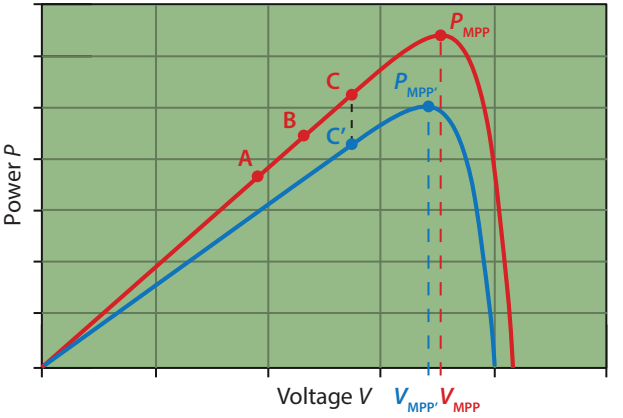

Figure 17.9: The *perturb&observe* algorithm struggles from rapidly changing illumination conditions.

ing point is never steady at the MPP but meandering around the MPP. If very small perturbation steps are used around the MPP, this meandering, however, can be minimised. Additionally, the P&O algorithm struggles from rapidly changing illuminations. For example, if the illumination (and hence the irradiance) changes in between two sampling instants in the process of convergence, then the algorithm essentially fails in its convergence efforts, as illustrated in Fig. 17.9 In the latest perturbation, the algorithm has determined that the MPP lies to the at a higher voltage than of point B, and hence the next step is a perturbation to converge towards the MPP accordingly. If the illumination was constant, it would end up at C and the algorithm would conclude that the MPP is at still higher voltages, which is correct. However, as the illumination changes rapidly before the next perturbation, the next perturbation shifts the operating point to C′ instead to C, such that

$$P_{\mathrm{C}'} < P_{\mathrm{B}} \tag{17.3}$$

While the MPP still lies to the right of C′, the P&O algorithm thinks that it is on the left of C′. This wrong assumption is detrimental to the speed of convergence of the P&O algorithm, which is one of the critical figures of merit for MPPT techniques. Thus, drastic changes in weather conditions severely affect the efficacy of the P&O algorithm's.

Incremental conductance method

Next, we look at the *Incremental Conductance* Method. The *conductance* G[1] of an electrical component is defined as

$$G = \frac{I}{V} \tag{17.4}$$

At the MPP, the slope of the P-V curve is zero, hence

$$\frac{\mathrm{d}P}{\mathrm{d}V} = 0. \tag{17.5}$$

We can write

$$\frac{\mathrm{d}P}{\mathrm{d}V} = \frac{\mathrm{d}(IV)}{\mathrm{d}V} = I + V\frac{\mathrm{d}I}{\mathrm{d}V}. \tag{17.6}$$

If the sampling steps are small enough, the approximation

$$\frac{\mathrm{d}I}{\mathrm{d}V} \approx \frac{\Delta I}{\Delta V} \tag{17.7}$$

can be used. We call $\Delta I/\Delta V$ the *incremental conductance* and I/V the *constanteneous conductance*. Hence, we have

$$\frac{\Delta I}{\Delta V} = -\frac{I}{V} \quad \text{if} \quad V = V_{\mathrm{mpp}}, \tag{17.8a}$$

$$\frac{\Delta I}{\Delta V} > -\frac{I}{V} \quad \text{if} \quad V < V_{\mathrm{mpp}}, \tag{17.8b}$$

$$\frac{\Delta I}{\Delta V} < -\frac{I}{V} \quad \text{if} \quad V > V_{\mathrm{mpp}}. \tag{17.8c}$$

[1] The electrical conductance G must not be confused with the irradiance incident on the module G_M.

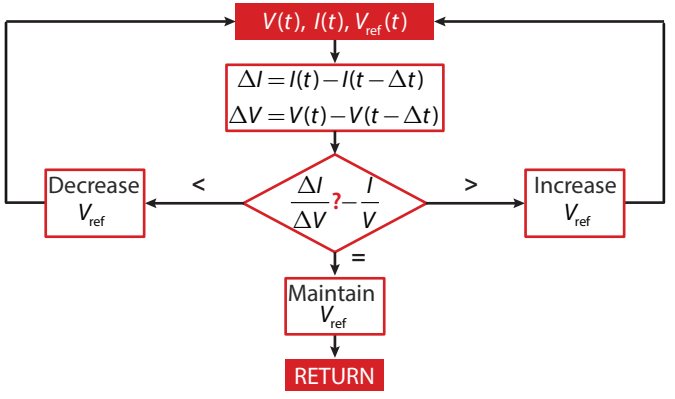

Figure 17.10: A conceptual flowchart of the *incremental conductance* algorithm.

These relationships are exploited by the incremental conductance algorithm.

Figure 17.10 shows a conceptual flowchart. Note that this flowchart is not exhaustive. While both, the instantaneous voltage and current are the observable parameters, the instantaneous voltage is also the controllable parameter. V_{ref} is the voltage value forced on the PV module by the MPPT device. It is the latest approximation of the V_{mpp}. For any change of the operating point, the algorithm compares the instantaneous with the incremental conductance values. If the incremental conductance is larger than the negative of the instantaneous conductance, the current operating point is to the left of the MPP; consequently, V_{ref} must be incremented. In contrast, if the incremental conductance is lower than the negative of the instantaneous conductance, the current operating point is to the left of the MPP and the is consequently decremented. This process is iterated until the incremental conductance is the same as the negative instantaneous conductance, in which case $V_{\text{ref}} = V_{\text{mpp}}$.

The incremental conductance algorithm can be more efficient than the P&O algorithm as it does not meander around the MPP under steady state conditions. Further, small sampling intervals make it less susceptible to changing illumination conditions. However, under conditions that are strongly varying and under partial shading, the incremental conductance method might also become less efficient. The major drawback of this algorithm is the complexity of its hardware implementation. Not only currents *and* voltages must be measured, but also the instantaneous and incremental conductances must be calculated and compared. How such a hardware design can look like however is beyond the scope of this book.

17.2.3 Some remarks

While a MPPT is used to find the MPP by changing the voltage, it does not perform changes of the operating voltage. This is usually done by a DC-DC converter

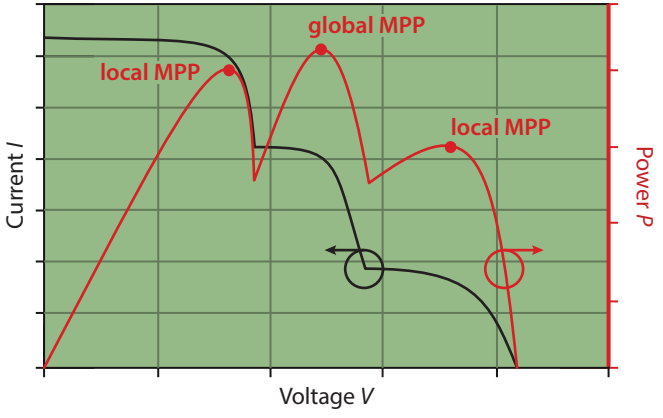

Figure 17.11: The *P*-*V* curve of partially shaded system that exhibits several local maxima.

that will be discussed in section 17.3.2.

In modern PV systems, the MPPT is often implemented within other system components like the inverters or charge controllers. The list of techniques presented in this section not exhaustive, we just discussed the most common ones. The development of more advanced MPPT techniques is going on rapidly and many scientific papers as well as patents are published in this area. Furthermore, manufacturers usually use proprietary techniques.

Up to now we only looked at situations the total *I*-*V* curve is similar to that of a single cell. Let us now consider a system that is partially shaded, as illustrated in Fig. 17.11. Then, the *P*-*V* curve will have different local maxima. Depending on the used MPPT algorithm, it is not sure at all that the algorithm finds the global maximum. Different companies use proprietary solutions to tackle this issue. Alternatively, each string can be connected to a separate MPPT. There are inverters that have connections for several strings (usually two).

17.3 Photovoltaic Converters

A core technology associated with PV systems is the power electronic converter. An ideal PV converter should draw the maximum power from the PV panel and supply it to the load side. In case of grid connected systems, this should be done with the minimum harmonic content in the current and at a power factor close to unity. For stand-alone systems the output voltage should also be regulated to the desired value. In this section a short review of different topologies often associated with PV systems is given. The semiconductor switches in the following are assumed to be ideal.

17.3.1 System configurations

Before digging into details about different converter topologies used for power conversion in PV systems, a general overview of different system architecture will be presented. The system architecture determines how PV modules are interconnected and how the interface with the grid is established. Which of these system architectures will be employed in a particular PV plant depends on many factors such as environment of the plant (whether the plant is situated in an urban environment or at an open area), scalability, costs etc. Figure 17.12 an overview of different system architectures is given. The main advantages and disadvantages of the different architectures are discussed below.

In general solar inverters should have the following characteristics [93]:

- Inverters should be highly efficient because the owner of the solar system requires the absolute maximum possible generated energy to be delivered to the grid/load.
- Special demands regarding the potential between solar generator and earth (depending on the solar module type).
- Special safety features like active islanding detection capability.
- Low limits for harmonics of the line currents. This requirement is enforced by law in most countries since the harmonic limits of both sources and loads connected to the grid are regulated.
- Special demands on *electromagnetic interference* (EMI), which are regulated by law in most countries. The goal of these minimise the unwanted influence of EMI on other equipment in the vicinity or connected to the same supply. Think for example of the influence of a mobile phone on an old radio.
- In many instances the solar system is to be installed outdoors and inverters should adhere to certain specification regarding temperature and humidity conditions, *e.g.* IP 54.
- Design for high ambient temperatures.
- Designed for 20 years operation under harsh environmental conditions.
- Silent operation (no audible noise).

We have to distinguish between *single-phase* and *three-phase inverters*. For low powers, as they are common in small residential PV systems, single-phase inverters are used. They are connected to one phase of the grid. For higher powers, three-phase inverters are used that are connected to all phases of the grid. If a high power would be delivered to one phase, the currents flowing across the three phases would become very asymmetric

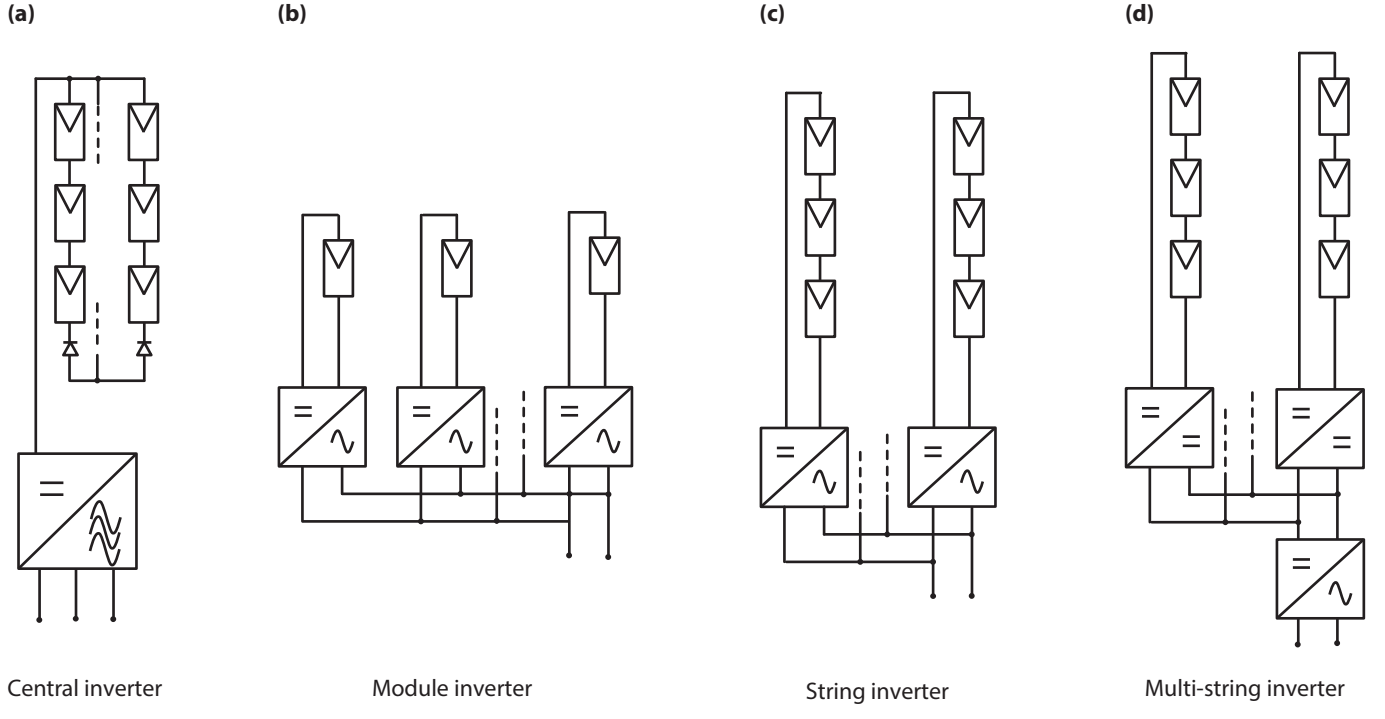

Figure 17.12: Different system architectures employed in PV systems.

leading to several problems in the electricity grid.

Note that the term *inverter* is often used for two different things: First, it is used for the actual inverter, which is the electronic building block that performs the DC-to-AC conversion, as described in section 17.3.3. Secondly, the term inverter also is used for the total unit produced by manufacturers, that nowadays usually contains, an MPP tracker, a DC-DC converter, an DC-AC converter and possibly also a charge controller of also a battery is connected.

Central inverters

This is the simplest architecture employed in PV systems. Here, PV modules are connected in strings leading to an increased system voltage. Many strings are then connected in parallel forming a PV array, which is connected to one central inverter. The inverter performs maximum power point tracking and power conversion as shown in Fig. 17.12 (a), where a three-phase inverter is depicted. This configuration is mostly employed in very-large scale PV production, with central inverter usually being DC to three phase [94].

Many different inverter topologies are utilised as a three phase inverter. Sometimes they are organised as a single DC to three phase unit but sometimes as a three separate DC-to-AC single phase units working with a phase displacement of 120 degrees each. Having all the PV modes connected in a single array in such a centralised configuration offers the lowest specific cost (cost per kWp_p of installed power). Since central inverters only use a few components, they are very reliable what makes them the preferred option in large scale PV power plants.

In spite of their simplicity and low specific cost, central inverters suffer from the following disadvantages:

1. Due to the layout of the system, a large amount of power is carried over considerable distances using DC wiring. This can cause safety issues because fault DC currents are difficult to interrupt. Special precaution measures must to be taken such as ticker insulation on the DC cabling and special circuit breakers, which can increase the costs.

2. All strings operate at the same maximum power point, which leads to mismatch losses in the modules. This is significant disadvantage. Mismatch losses increase even more with ageing and with partial shading of sections of the array. Mismatch between the different strings may significantly reduce the overall system output.

3. Low flexibility and expandability of the system. Due to the high ratings a system is normally designed as a unit and hence difficult to extend. In other words the system design is not very flexible.

4. Power losses in the string diodes, which are put in

series with each string to prevent current circulation inside strings.

Module Integrated or module oriented inverters

A very different architecture is that of the *module integrated inverters*, as shown in 17.12 (b). These inverters operate directly at one or several PV modules and have power ratings of several hundreds of watts. Because of the low voltage rating of the PV module, these inverters require often require a two stage power conversion. In a first stage boosts the DC voltage is boosted to the required value while it is inverted to AC in the second stage. Often, a high frequency transformer is incorporated providing full galvanic isolation, which enhances the system flexibility even further. As their name suggests, these inverters are usually integrated with PV panel (so called 'AC PV panels'). In this way the highest flexibility and the expandability of system is obtained. One of the most distinguishing features of this system is the "plug and play" characteristic, which allows to build a complete (and readily expandable) PV system at a low investment cost. Another advantage of these inverters, is minimisation of the mismatch losses that can occur because of non-optimal MPPT.

All these advantages come at certain expenses. Because these inverters are be mounted on a PV module, they must operate in harsh environment such as high temperature and large daily and seasonal temperature variations. Also, the specific are the highest of all the inverter topologies. Many topologies for module integrated inverters have been proposed, with some of them being already implemented in commercially available inverters.

String Inverters

String inverter, as illustrated in 17.12 (c), combine the advantages of central and module integrated inverter concepts with little tradeoffs. A number of PV modules that are connected in series form a PV string with a power rating of up to 5 kWp and with the open circuit voltage of up to 1 kV. Now, a number of smaller inverters can be used to connect the PV system to the grid.

One disadvantage of the topology is the fact that the high DC voltage requires special consideration, similarly as this already was the case for the central inverter architecture. Here, this issue is even more important because string inverters are usually being installed in households or on office buildings, without designated support structure or increased safety requirements. In general a qualified electrician is needed to perform the interconnections between the modules and the inverter. The protection of the system also requires special consideration, with emphasis on proper DC cabling.

Although partial shading of the string will influence the overall efficiency of the system, each string can independently be operated at its MPP. Also, because no strings are connected in parallel, there is no need for series diodes as in the case of PV arrays with many parallel strings. This reduces losses associated with these diodes. However, it still is a risk that within a string hot-spot occurs because of unequal current and power sharing inside the string.

Multi String Inverters

The Multi String inverter concept, illustrated in Fig. 17.12 (d), has been developed to combine the advantage of the higher energy yield of a string inverter with the lower costs of a central inverter [95]. Lower power DC-DC converters are connected to individual PV strings each having its own MPP tracking, which independently optimises the energy output from each PV string. To expand the system within a certain power range only a new string with a DC-DC converter has to be included. All DC-DC converters are connected via a DC bus through a central inverter to the grid.

Team Concept

Aside of the four system architecture already described, many other concepts are also present in literature.

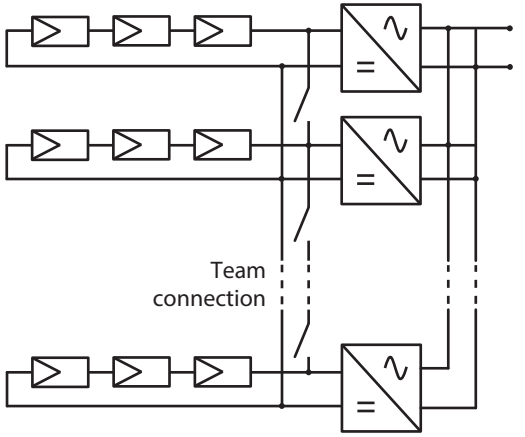

Figure 17.13: The team concept of inverters.

However, these concepts are less widely utilised than the ones already presented. One of the alternative concepts is so called team concept, which combines the string technology with the master-slave concept. A combination of several string inverters working with the team concept is shown in Figure 17.13. At very low irradiation the complete PV array is connected to a single inverter. This reduces the overall losses as any power electronic converter is designed such that it has maximum efficiency near full load. With increasing solar radiation more inverters are being connected dividing the PV array into smaller units until every string inverter operates close to its rated power. In this mode every string operates independently with its own MPP tracker. At low solar radiation the inverters are controlled in a master-slave fashion.

In literature many other concepts are discussed. For further reading on this subject, we refer for example to Ref. [96].

17.3.2 DC-DC converters

DC-DC converters fulfil multiple purposes. In an inverter, DC power is transformed into AC power. The DC input voltage of the inverter often is constant while the output voltage of the modules at MPP is not. Therefore a DC-DC converter is used to transform the variable voltage from the panels into stable voltage used by the DC-AC inverter. Additionally, as already stated in section 17.2, the MPP Tracker controls the operating point of the modules, but it cannot set it. This also is done by the DC-DC converter. Further, in a stand-alone system the MPP voltage of the modules might differ from that required by the batteries and the load. Also here, a DC-DC converter is useful. Three topologies are used for DC-DC converters: *buck*, *boost*, and buck-boost converters. They are described below.

Step-Down (Buck) Converter

Figure 17.14 (a) illustrates the simplest version of a buck DC-DC converter. The unfiltered output voltage waveform of such a converter operated with pulse-width-modulation (PWM) is shown in Fig. 17.14 (b). If the switch is *on*, the input voltage V_d is applied to the load. When the switch is *off*, the voltage across the load is zero. From the figure we see that the average DC output voltage is denoted as V_o. From the unfiltered voltage, the average output voltage is given as

$$V_o = \frac{1}{T_s}\int_0^{T_s} v_o(t)\,\mathrm{d}t = \frac{1}{T_s}\left(t_{\mathrm{on}}V_d + t_{\mathrm{off}}\cdot 0\right) = \frac{t_{\mathrm{on}}}{T_s}V_d. \tag{17.9}$$

The different variables are defined in Fig. 17.14 (b). To simplify the discussion we define a new term, the duty

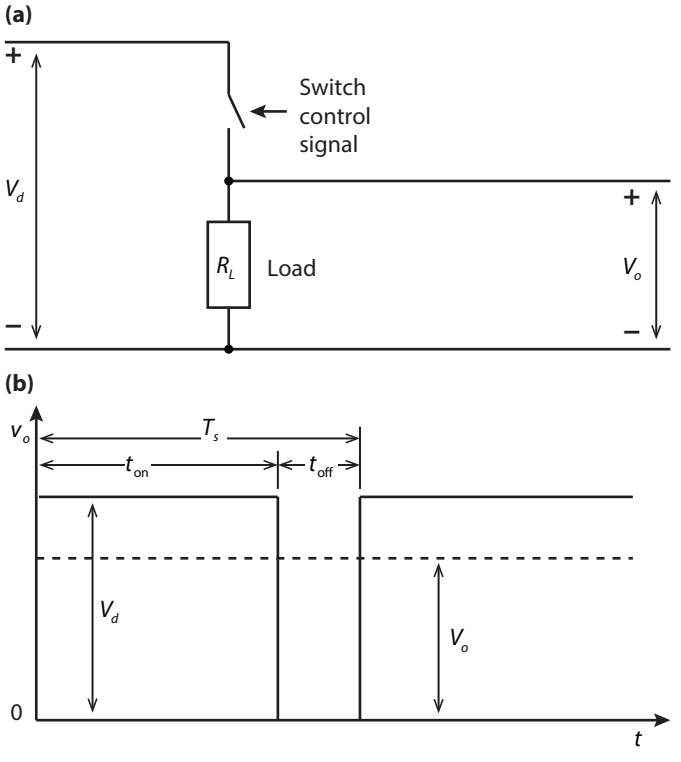

Figure 17.14: (a) A basic buck converter without any filters and (b) the unfiltered switched waveform generated by this converter.

cycle D, as

$$D := \frac{t_{\text{on}}}{T_s} \tag{17.10}$$

and hence

$$V_o = D \cdot V_d. \tag{17.11}$$

In general the output voltage with such a high harmonic content is undesirable, and some filtering is required. Figure 17.15 (a) shows a more complex model of a step-down converter that has output filters included and supplies a purely resistive load. As filter elements an inductor L and a capacitor C are used. The relation between the input and the output voltages, as given in Equ. 17.11, is valid in continuous conduction mode, *i.e.* when the current through the inductor never reaches zero value but flows continuously. We can change the ratio between the voltages on the input and output side by changing the duty cycle D. A detailed discussion about different modes of operation of a buck converter can be found in Reference [97].

In *steady-state operation* the time integral of the voltage across the inductor v_L taken during one switching cycle is equal to zero. If this is not the case, the circuit is not in steady state. Thus, in steady state, we obtain the following *inductor volt-second balance*:

$$\int_0^{T_s} v_L \, \mathrm{d}t = \int_0^{t_{\text{on}}} v_L \, \mathrm{d}t + \int_{t_{\text{on}}}^{T_s} v_L \, \mathrm{d}t = 0. \tag{17.12}$$

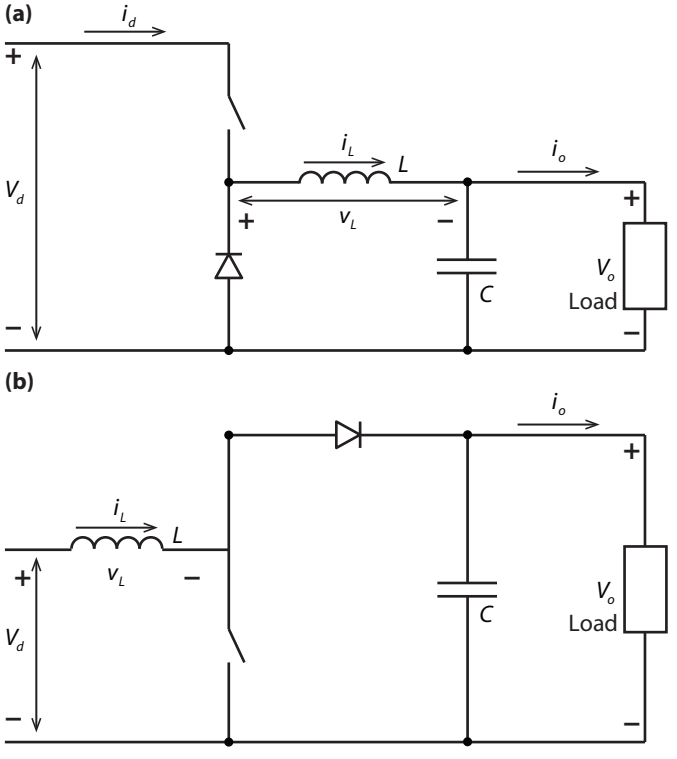

Figure 17.15: (a) A buck converter with filters and (b) a boost converter.

Solving this equation leads to

$$V_o = DV_d, \tag{17.13}$$

which is the same result as in Eq. 17.11.

Step-Up (Boost) Converter

In a boost converter, illustrated in Fig. 17.15 (b), an input DC voltage V_d is boosted to a higher DC voltage V_o. By applying the inductor volt-second balance across the inductor as explained in Eq. 17.12, we find

$$V_d t_{\text{on}} + (V_d - V_0) t_{\text{off}} = 0. \tag{17.14}$$

Using the definition for the duty cycle we find

$$\frac{V_0}{V_d} = \frac{1}{1-D}. \tag{17.15}$$

The above relation is valid in the continuous conduction mode. The principle of operation is that energy stored in the inductor (during the switch is *on*) is later released against higher voltage V_o. In this way the energy is transferred from lower voltage (solar cell voltage) to the higher voltage (load voltage).

Buck-Boost Converter

In a *buck-boost converter* the output voltage can be both higher or lower than the input voltage. The simpli-

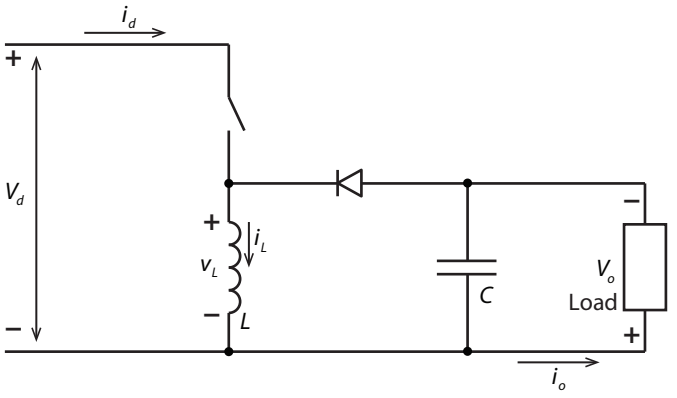

Figure 17.16: A buck-boost converter.

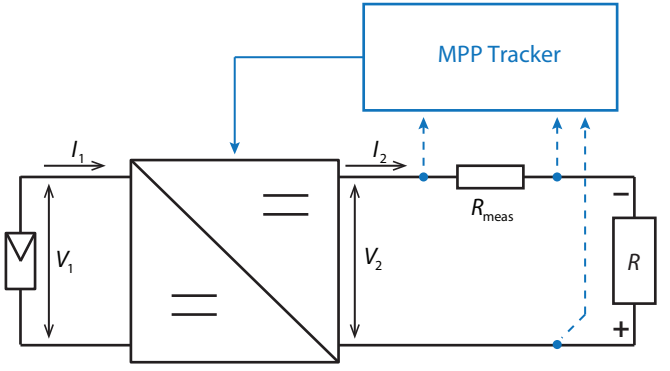

Figure 17.17: A combination of a unit performing an MPPT algorithm and a DC-DC converter (adapted from [98]).

fied schematic of a buck-boost converter is depicted in Fig. 17.16. Using inductor volt second balance as in Eq. 17.12, we find

$$V_d t_{\text{on}} + (-V_o) t_{\text{off}} = 0 \tag{17.16}$$

and hence

$$\frac{V_o}{V_d} = \frac{D}{1-D} \tag{17.17}$$

in the continuous conduction mode.

The above described topologies are only the most basic DC-DC converter topologies. The interested reader can find more in-depth information in Reference [97].

MPP Tracking

In section 17.2 we extensively discussed Maximum Power Point Tracking. Or – more specific we discussed different algorithms that are used for performing MPPT. In these algorithms, usually the operating point of the module is set such that its power output becomes maximal. However, the MPPT algorithm itself cannot actually adjust the voltage or current of the operating point. For this purpose a DC-DC converter is needed. Figure 17.17 shows such a combination of the unit performing the MPPT and DC-DC converter. As illustrated in the figure, this MPPT unit measures the voltage or the cur-

rent on the load side and can very those by adapting the duty cycle of the DC-DC converter. In this illustration, current and voltage on the load side are measured, but they also can be measured on the PV side.

Example

Assume a PV module has its MPP at $V_{\mathrm{PV}} = 17\,\mathrm{V}$ *and* $I_{\mathrm{PV}} = 6\,\mathrm{A}$ *at a given level of solar irradiance. The module has to power a load with a resistance* $R_L = 10\,\Omega$*. Calculate the duty cycle of the DC-DC converter, if a buck-boost converter is used.*

The maximum power from the module is $P_{\mathrm{MPP}} = V_{\mathrm{MPP}} \cdot I_{\mathrm{MPP}} = 102\,\mathrm{W}$*. If this power should be dissipated at the resistor, we have to use the relation*

$$P_R = U_R^2 / R$$

and hence find for the voltage at the resistor

$$V_R = \sqrt{P_{\mathrm{MPP}} R_L} = 31.94\,\mathrm{V}$$

Using Eq. (17.17),

$$\frac{V_o}{V_d} = \frac{D}{1-D}$$

with $V_o = V_R$ *and* $V_d = V_{\mathrm{PV}}$ *we find* $D = 0.65$*.*

17.3.3 DC-AC converters (inverters)

Earlier in this section we have discussed different architectures that are used for the power conversion in PV systems. Further, we looked at DC-DC converters that are mainly used in combination with MPPTs in order to push the variable output from the PV modules to a level of constant voltage.

As nowadays most appliances are designed for the standard AC grids, for most PV systems a DC-AC converter is required. As already stated earlier, the term inverter is used both for the DC-AC converter and the combination of all the components that form the actual power converter.

The H-bridge inverter

Figure 17.18 shows a very simple example of a so called H-bridge or full-bridge inverter. On the left, the DC input is situated. The load (or in our case the AC output) is situated in between four switches. During usual operation we can distinguish between three situations:

A **All four switches open**: No current flows across the load.

B **S_1 and S_4 closed, S_2 and S_3 open**: Now a current is flowing to through the load, where $+$ is connected

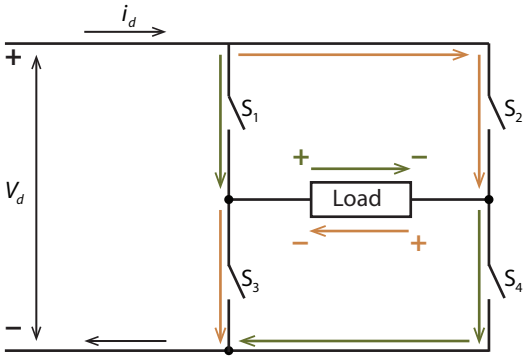

Figure 17.18: A simple representation of an H-bridge.

to the left- and − is connected to the right-hand side of the load.

C S_1 and S_4 open, S_2 and S_3 closed: Now a current is flowing to through the load, where − is connected to the left- and + is connected to the right-hand side of the load.

WARNING It must be assured that S_1 and S_2 never are open at the same time because this would lead to short-circuiting. The same is true for S_3 and S_4.

From this list we see that the H-bridge configuration allows to put the load on three different levels, which are $+V_d$, 0, and $-V_d$. Continuously switching between positive and negative voltages is exactly what is happening in AC. In the easiest operation mode, the H-bridge switches between situations B and C continuously, which will provide a square wave. Note that there will always be a *dead time* in the order of µs in order to prevent short-circuiting. While such a square wave might be useful for some applications, it is not suited at all for grid-connected installations. The reason for this are harmonic distortions. To understand this we look at a Fourier transform of a square wave, which is given as

$$V_{\text{square}}(t) = \frac{4}{\pi}\left[\sin(2\pi\nu t) + \frac{1}{3}\sin(6\pi\nu t) + \frac{1}{5}\sin(10\pi\nu t) + \cdots\right]. \tag{17.18}$$

Thus, such a square wave contains not only the principal sine function with frequency ν, but also all the higher harmonics with frequencies 3ν, 5ν, and so on. These higher harmonics can lead to distortions of the electricity grid and thus must be reduced as much as possible. One principle how to do this is *pulse-width modulation* (PWM) that we already discussed in the section on DC-DC conversion (17.3.2). In this configuration, each leg (the one via S_1 and S_4 and the one via S_2 and S_3) in fact acts as a buck converter.

If a low-pass filter consisting of capacitors and inductors is used, as in Fig. 17.19, the high frequency components are filtered out and hence a very smooth sine

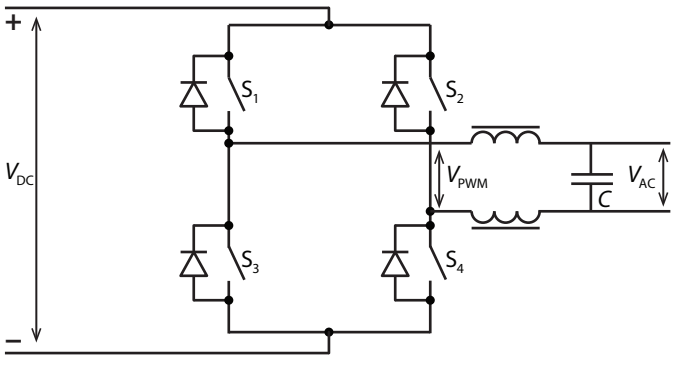

Figure 17.19: Illustration of an H-bridge containing a low-pass filter for removing the high-frequency components of the signal.

curve can be obtained that complies to the regulations for grid-connected systems. Figure 17.20 shows the unfiltered PWM output and the filtered sine output.

Note that in Fig. 17.19 diodes are connected in parallel to the switches. The reason for placing these diodes is the following: If the switch goes from closed to open very fast, no current can flow through the inductor anymore, meaning that the change of current flowing through the inductor is very high. This induces a voltage given by

$$V_{\text{induced}} = -L\frac{\mathrm{d}I}{\mathrm{d}t}, \tag{17.19}$$

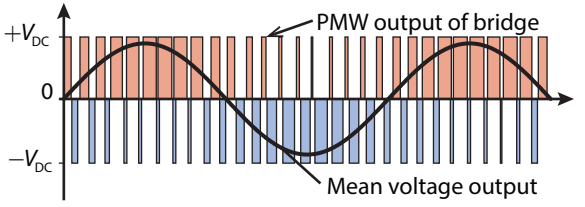

Figure 17.20: The unfiltered PWM signal and the sine signal that is obtained with a low-pass filter.

which is the higher the faster the current changes. As high induced voltages will damage the electric circuitry, they must be prevented. The diode ensures that current also can flow after the switch opens. Hence current is flowing always and no high induced voltages appear.

Since this configuration is grid-connected, we easily can determine the required DC input voltage, which must be minimal the peak voltage of the AC voltage. For an effective AC voltage of 240 V, as used in large parts of the world, the peak voltage is $V_{\text{peak}} = 1.1\sqrt{2} \cdot 240\ \text{V} = 373\ \text{V}$, where we also took a 10% tolerance into account.

If the PV array delivers a lower voltage, thus a boost converter would be required prior in addition to the inverter. Alternatively, a transformer can be used that transforms a lower-voltage AC signal to the 240 V AC signal. Such a system is sketched in Fig. 17.23. This system has the advantage of galvanic separation with all the advantages discussed below, but on the other hand

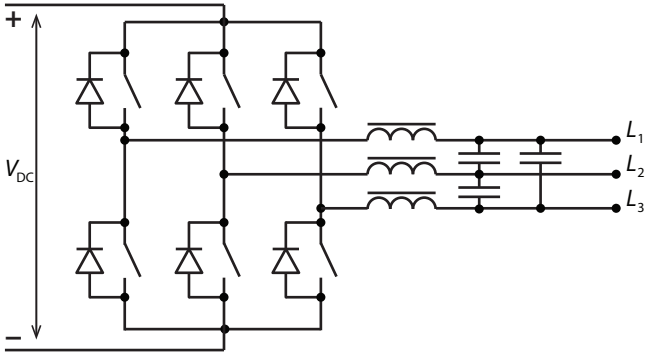

Figure 17.21: Illustrating a three-phase inverter.

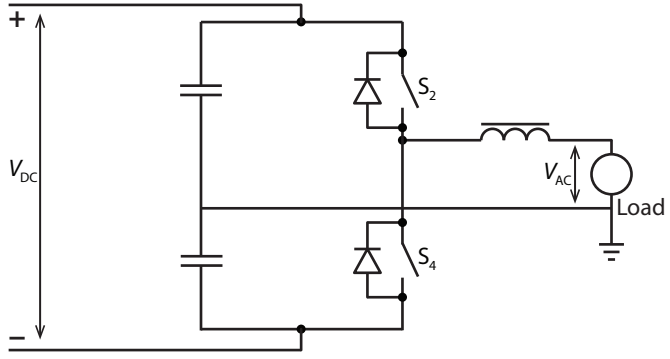

Figure 17.22: Illustrating a half-bridge inverter.

the transformer reduces the overall efficiency.

Figure 17.21 shows a three-phase inverter. As it can be seen, it is very similar to the single-phase inverter discussed above, just it contains three legs of that each creates a sine-wave output with a phase shift of 120° between them.

Half-bridge inverters

A simpler inverter topology, the so-called half-bridge inverter, is shown in Fig. 17.22. In contrast to the H-bridge configuration, here, two switches are replaced by capacitors and the midpoint in-between the two capacitors is directly connected to the ground. The half-bridge configuration is much simpler than the H-bridge configuration, but it has some drawbacks.

The main drawback is requirement for a high DC link voltage, which needs to be two times higher than that of a full bridge inverter. For a effective AC voltage of 240 V this would be 746 V. Because the topology provides two-levels in the output voltage (in difference to three levels for the full bridge configuration, a higher current ripple is present in the output filter inductor. Hence, larger value of the output filters are required.

17.3.4 Some remarks

Switches

All the DC-DC transformers and DC-AC transformers discussed above contain switches. Traditionally, in *line-commutated* inverters, *thyristors* are used as switches. A thyristor is a electronic component consistent of pnpn layers. It thus contains 3 p-n junctions. One disadvantage is that thyristors cannot be turned off, but only turned on. Thus, one has to wait for the next zero pass of the grid signal [98]. The current flow thus is rectangular which leads to a very high harmonic content requiring additional filters in order to make the output compatible with the electricity grid. Nowadays, thyristors only are used for inverters with a power of 100 kW and higher.

All other inverters are *self-commutated* inverters that generate an output with very little harmonic content as described above and in Fig. 17.20. The switches there are fully-controllable such that pulse-width modulation becomes possible. As switches, GTOs (gate turn-off thyristors), IGBTs (insulated-gate bipolar transistors) or MOSFETs (metal-oxide-semiconductor field-effect transistors) are used. More information can be found for example in Reference [20].

Overall configuration

Figure 17.23 shows an example of a transformer-less absorber as it could be sold for household systems. Besides the actual DC-AC converter, which is realised as an H-bridge, it also contains a DC-DC boost converter and an MPPT that uses the voltage and current measured on the PV side as input. The system sketched also contains several electronic components for increased safety: a residual current measurement to detect leakage currents above a certain threshold and to shut-down the inverter if these currents appear. Further, a grid-monitoring unit to prevent islanding (see below) [98].

Potential induced degradation (PID)

As mentioned earlier, in PV systems that have a transformer-less inverter, no galvanic separation between the DC- and the AC-parts of the system is given. Because of this lack of galvanic isolation, a potential of -500 V or more between the PV modules and the ground can occur, which can lead to potential induced degradation (PID). Many thin-film contain a TCO front contact that is deposited in superstrate configuration on the glass top plate. Positively charged sodium ions than can travel into the TCO because of this potential. This leads to corrosion and consequently to performance loss of the module. Also, for crystalline mod-

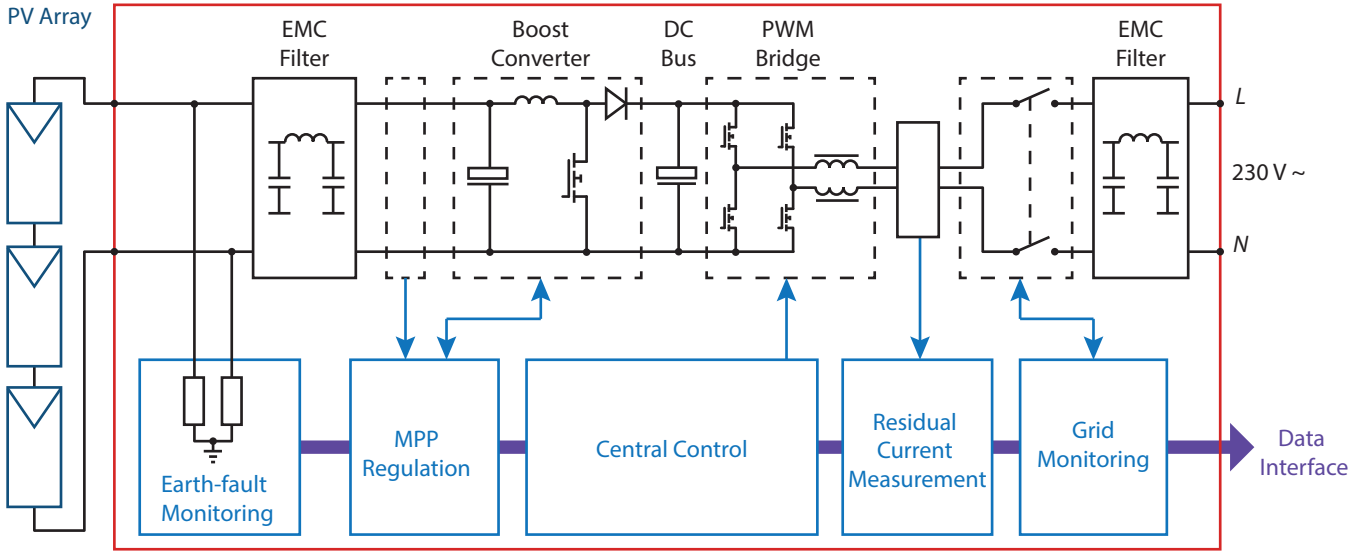

Figure 17.23: Illustrating an example of a transformer-less inverter unit as it is sold for residential PV systems. As switches, MOSFETs are used. (Adapted from [98]).

ules, PID can be a problem [98]. Therefore, for systems containing thin-film modules the inverter must have a transformer.

Islanding

A potential danger of grid-connected systems is *islanding*. Imagine that a potential PV system is installed in a street where the electricity grid is shut-down in order to do maintenance work on the electricity cables. If it is a sunny day, the PV system will produce power and – without protection – would deliver the power to the grid. The electricity worker thus can be in danger. This phenomenon is called islanding and due to its danger it must be prevented.

The inverter therefore must be able to detect, when the electricity grid is shut-down. If this is the case, also the inverter must stop delivering power to the grid.

Efficiency of power converters

For planning a PV system it is very important to know the efficiency of the power converters. This efficiency of DC-DC and DC-AC converters is defined as

$$\eta = \frac{P_o}{P_o + P_d}, \tag{17.20}$$

i.e. the fraction of the output power P_o to the sum of P_o with the dissipated power P_d. For estimating the efficiency we thus must estimate the dissipated (lost) power, which can be seen as a sum of several components,

$$P_d = P_L + P_{\text{switch}} + \text{other losses}, \tag{17.21}$$

i.e. the power lost in the inductor P_L and the power lost in the switch. P_L is given as

$$P_L = I_{L,\,\text{rms}}^2 R_L, \tag{17.22}$$

where $I_{L,\,\text{rms}}$ is the mean current flowing through the inductor. The losses in the switch strongly depend on the type of switch that is used. Other losses are for example resistive losses in the circuitry in-between the switches.

For a complete inverter unit it is convenient to define the efficiency as

$$\eta_{\text{inv}} = \frac{P_{\text{AC}}}{P_{\text{DC}}}, \tag{17.23}$$

which is the ratio of the output AC power to the DC input power. Figure 17.24 shows the efficiency of a commercially available inverter for different input voltages. As we can see, in general the lower the output power, the less efficient the inverter. This efficiency characteristic must be taken into account when planning a PV system. Further, the efficiency is lower if the input voltage deviates from the nominal value.

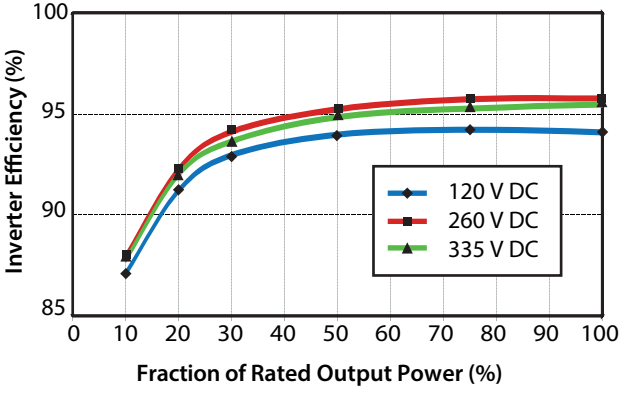

Figure 17.24: The power-dependent efficiency for several input voltages of a *Fronius Galvo 1.5-1 208-240* inverter at 240 V AC. The nominal input voltage is 260 V DC. (Data taken from Ref. [99]).

17.4 Batteries

In this section we discuss a vital component not only of PV systems but of renewable energy systems in general. *Energy storage* is very important at both small and large scales in order to tackle the intermittency of renewable energy sources. In the case of PV systems, the intermittency of the electricity generation is of two kinds: first, *diurnal* fluctuations, *i.e.* the difference of irradiance during the 24 hour period. Secondly, the *seasonal* fluctuations, *i.e.* the difference of irradiance between the summer and winter months. There are several technological options for realising storage of energy. Therefore it is important to make an optimal choice.

Fig. 17.25 shows a *Ragone* chart, where the *specific energy* is plotted against the *peak power density*. Because it uses a double-logarithmic chart storage technologies with very different storage properties can be compared in one plot. Solar energy application require a high energy density and – depending on the application – also a reasonably high power density. For example, we cannot use capacitors because of their very poor energy density. For short term to medium term storage, the most common storage technology of course is the *battery*. Batteries have both the right energy density and power density to meet the daily storage demand in small and medium-size PV systems. However, the seasonal storage problem at large scales is yet to be solved.

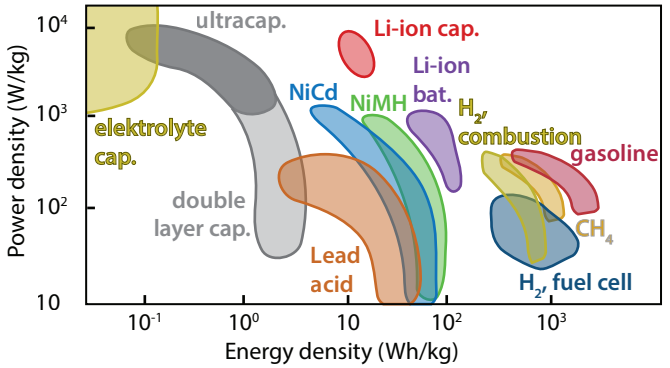

Figure 17.25: A *Ragone* chart of different energy storage methods. Capacitors are indicated with "cap".

The ease of implementation and efficiency of the batteries is still superior to that of other technologies, like pumping water to higher levels, compressed air energy storage, conversion to hydrogen, flywheels, and others. Therefore we will focus on battery technology as a viable storage option for PV systems.

Batteries are electrochemical devices that convert chemical energy into electrical energy. We can distinguish between primary or secondary batteries. Primary batteries convert chemical energy to electrical energy *irreversibly*. For example, *zinc carbon* and *alkaline* batteries are primary batteries.

Secondary batteries or *rechargeable batteries*, as they are more commonly called, convert chemical energy to electrical energy *reversibly*. This means that they can be recharged when an over-potential is used. In other words, excess electrical energy is stored in these secondary batteries in the form of chemical energy. Typical examples for rechargeable batteries are *lead acid* or *lithium ion* batteries. For PV systems only secondary batteries are of interest.

17.4.1 Types of batteries

Lead-acid batteries are the oldest and most mature technology available. They will be discussed in detail later in this section.

Other examples are *nickel-metal hydride* (NiMH) and *nickel cadmium* (NiCd) batteries. NiMH batteries have a high energy density, which is comparable to that of lithium-ion batteries (discussed below). However, NiMH batteries suffer from a high rate of self discharge. On the other hand, NiCd batteries have much lower energy density than lithium ion batteries. Furthermore, because of the toxicity of cadmium, NiCd batteries are widely banned in the European Union for consumer use. Additionally, NiCd batteries suffer from what is called the *memory effect*: the batteries loose their usable energy capacity if they are repeatedly charged after only being partially discharged. These disadvant-

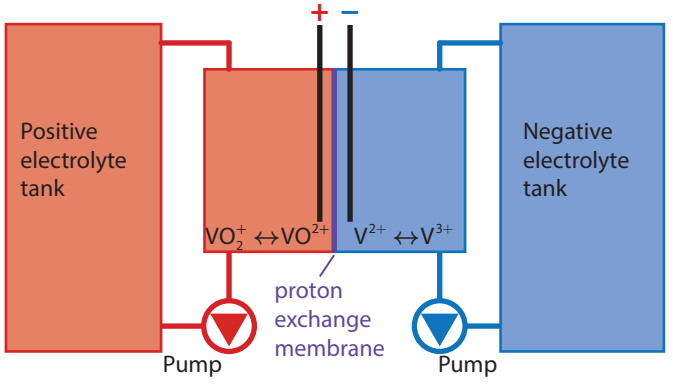

Figure 17.26: Schematic on a *vanadium* redox-flow battery that employs vanadium ions.

ages make NiMH and NiCD batteries unsuitable candidates for PV storage systems.

Lithium ion batteries (LIB) and *lithium-ion polymer* batteries, which are often referred to as lithium polymer (LiPo) batteries, have been heavily investigated in recent years. Their high energy density already has made them the favourite technology for light-weight storage applications for example in mobile telephones. However, these technologies still suffer from high costs and low maturity.

The last and the most recent category of batteries that we will discuss in this treatise are *redox-flow* batteries.

Lead-acid batteries and LIB, the two main storage options for PV systems, are similar in the sense that their electrodes undergo chemical conversion during charging and discharging, which makes their electrodes to degenerate with time, leading to inevitable "ageing" of the battery. In contrast, redox flow batteries combine properties of both batteries and fuel cells, as illustrated in Fig. 17.26. Two liquids, a *positive electrolyte* and a *negative electrolyte* are brought together only separated by a membrane, which only is permeable for protons. The cell thus can be charged and discharged without the reactants being mixed, which in principle prevents the liquids from ageing. The chemical energy in a redox flow battery is stored in its 2 electrolytes, which are stored in two separate tanks. Since it is easy to make the tanks larger, the maximal energy that can be stored in such a battery thus is not restricted. Further, the maximal output power can easily be increased by increasing the area of the membrane, for example by using more cells at the same time. The major disadvantages is that such a battery system requires additional components such as pumps, which makes them more complicated than other types of batteries.

Figure 17.27 shows a Ragone plot as it is used to compare different battery technologies. In contrast to Fig. 17.25, here the the gravimetric energy density and volumetric energy density are plotted against each other. The *gravimetric energy density* is the amount of energy stored per mass of the battery; it typically is meas-

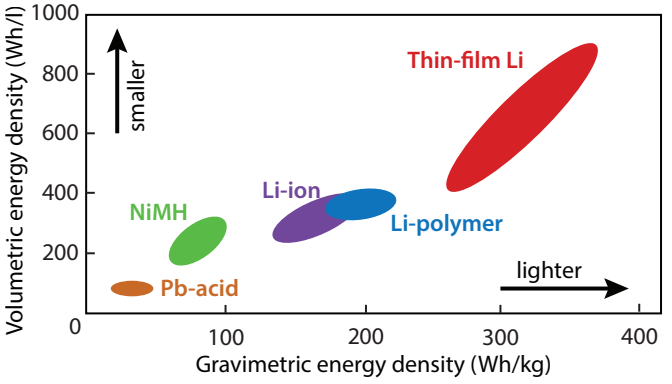

Figure 17.27: A Ragone chart for comparing different secondary battery technologies with each other.

ured in Wh/kg. The volumetric energy density is the amount of energy stored per volume of battery; it is given in Wh/l. The higher the gravimetric energy density, the lighter the battery can be. The higher the volumetric energy density, the smaller the battery can be.

Figure 17.27 shows that lead acid batteries have both the lowest volumetric and gravimetric energy densities among the different battery technologies. Lithium ion batteries show ideal material properties for using them as storage devices. Redox-flow batteries are very promising. However, both LIB and redox-flow batteries are still in a development phase which makes these technologies still very expensive. Thus, because of to the unequalled maturity and hence low cost of the lead-acid batteries, they are still the storage technology of choice for PV systems.

Figure 17.28 shows a sketch of a lead-acid battery. A typical battery is composed of several individual cells, of which each has a nominal cell voltage around 2 V. Different methods of assembly are used. In the *block assembly*, the individual cells share the housing and are interconnected internally. For example, to get the typical battery voltage of 12 V, six cells are connected in series. As the name suggests, lead acid batteries use an acidic electrolyte. More specifically they use diluted sulphuric acid H_2SO_4. Two plates of opposite polarity are inserted in the electrolyte solution, which act as the electrodes. The electrodes contain grid shaped lead carrier and porous active material. This porous active material has a sponge-like structure, which provides sufficient surface area for the electrochemical reaction. The active mass in the negative electrode is lead, while in the positive electrode lead dioxide (PbO_2) is used.

When the battery is discharged, electrons flow from the negative to the positive electrode through the external circuit, causing a chemical reaction between the plates and the electrolyte. This forward reaction also depletes the electrolyte, affecting its *state of charge* (SoC). When the battery is recharged, the flow of electrons is reversed, as the external circuit does not have a load, but

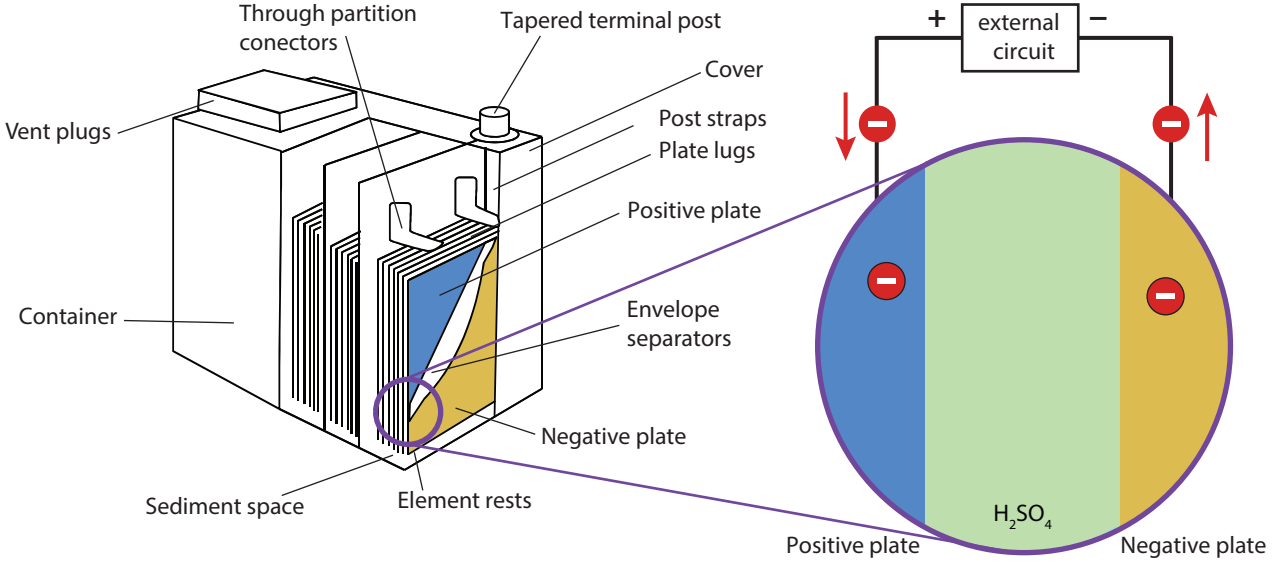

Figure 17.28: Schematic of a lead-acid battery.

a source with a voltage higher than that of the battery enables the reverse reaction. In the PV systems, this source is nothing but the PV module or array providing clean solar power.

17.4.2 Battery parameters

Let us now discuss some parameters that are used to characterise batteries.

Voltage

First, we will discuss the *voltage* rating of the battery. The voltage at that the battery is rated is the *nominal voltage* at which the battery is supposed to operate. The so called *solar batteries* or lead acid grid plate batteries are usually rated at 12 V, 24 V or 48 V.

Capacity

When talking about batteries, the term *capacity* refers to the amount of charge that the battery can deliver at the rated voltage. The capacity is directly proportional to the amount of electrode material in the battery. This explains why a small cell has a lower capacity than a large cell that is based on the same chemistry, even though the open circuit voltage across the cell will be the same for both the cells. Thus, the voltage of the cell is more chemistry based, while the capacity is more based on the quantity of the active materials used.

The capacity C_{bat} is measured in ampere-hours (Ah). Note that charge usually is measured in *coulomb* (C). As the electric current is defined as the rate of flow of electric charge, Ah is another unit of charge. Since 1 C = 1 As, 1 Ah = 3600 C. For batteries, Ah is the more convenient unit, because in the field of electricity the amount of energy usually is measured in watt-hours (Wh). The energy capacity of a battery is simply given by multiplying the rated battery voltage measured in Volt by the battery capacity measured in Amp-hours,

$$E_{\text{bat}} = C_{\text{bat}} V, \tag{17.24}$$

which results in the battery energy capacity in Watt-hours.

C-rate

A brand new battery with 10 Ah capacity theoretically can deliver 1 A current for 10 hours at room temperature. Of course, in practice this is seldom the case due to several factors. Therefore, the *C-rate* is used, which is a measure of the rate of discharge of the battery relative to its capacity. It is defined as the multiple of the current over the discharge current that the battery can sustain over one hour. For example, a C-rate of 1 for a 10

Ah battery corresponds to a discharge current of 10 A over 1 hour. A C-rate of 2 for the same battery would correspond to a discharge current of 20 A over half an hour. Similarly, a C-rate of 0.5 implies a discharge current of 5 A over 2 hours. In general, it can be said that a C-rate of n corresponds to the battery getting fully discharged in $1/n$ hours, irrespective of the battery capacity.

Battery efficiency

For designing PV systems it is very importent to know the *efficiency* of the storage system. For storage systems, usually the *round-trip efficiency* is used, which is given as the ratio of total storage system input to the total storage system output,

$$\eta_{\text{bat}} = \frac{E_{\text{out}}}{E_{\text{in}}} 100\%. \tag{17.25}$$

For example, if 10 kWh is pumped into the storage system during charging, but only 8 kWh can be retrieved during discharging, the round trip efficiency of the storage system is 80%. The round-trip efficiency of batteries can be broken down into two efficiencies: first, the *voltaic efficiency*, which is the ratio of the average discharging voltage to the average charging voltage,

$$\eta_V = \frac{V_{\text{discharge}}}{V_{\text{charge}}} 100\%. \tag{17.26}$$

This efficiency covers the fact that the charging voltage is always a little above the rated voltage in order to drive the reverse chemical (charging) reaction in the battery.

Secondly, we have the *coulombic efficiency* (or Faraday efficiency), which is defined as the ratio of the total charge got out of the battery to the total charge put into the battery over a full charge cycle,

$$\eta_C = \frac{Q_{\text{discharge}}}{Q_{\text{charge}}} 100\%. \tag{17.27}$$

The *battery efficiency* then is defined as the product of these two efficiencies,

$$\eta_{\text{bat}} = \eta_V \cdot \eta_C = \frac{V_{\text{discharge}}}{V_{\text{charge}}} \frac{Q_{\text{discharge}}}{Q_{\text{charge}}} 100\%. \tag{17.28}$$

When comparing different storage devices, usually this round-trip efficiency is considered. It includes all the effects of the different chemical and electrical non-idealities occurring in the battery.

State of charge and depth of discharge

At another important battery parameter is the *State of Charge* (SoC), which is defined as the percentage of the battery capacity available for discharge,

$$\text{SoC} = \frac{E_{\text{bat}}}{C_{\text{bat}} V}. \tag{17.29}$$

Thus, a 10 Ah rated battery that has been drained by 2 Ah is said to have a SoC of 80%. Also the *Depth of Discharge* (DoD) is an important parameter. It is defined as the percentage of the battery capacity that has been discharged,

$$\mathrm{DoD} = \frac{C_{\mathrm{bat}} V - E_{\mathrm{bat}}}{C_{\mathrm{bat}} V}. \tag{17.30}$$

For example, a 10 Ah battery that has been drained by 2 Ah has a DoD of 20%. The SoC and the DoD are complimentary to each other.

Cycle lifetime

The *cycle lifetime* is a very important parameter. It is defined as the number of charging and discharging cycles after that the battery capacity drops below 80% of the nominal value. Usually, the cycle lifetime is specified by the battery manufacturer as an absolute number. However, stating the battery lifetime as a single number is a oversimplification because the different battery parameters discussed so far are not only related to each other but are also dependent on the temperature.

Figure 17.29 (a) shows the cycle lifetime as a function of the DoD for different temperatures. Clearly, colder operating temperatures mean longer cycle lifetimes. Furthermore, the cycle lifetime depends strongly on the

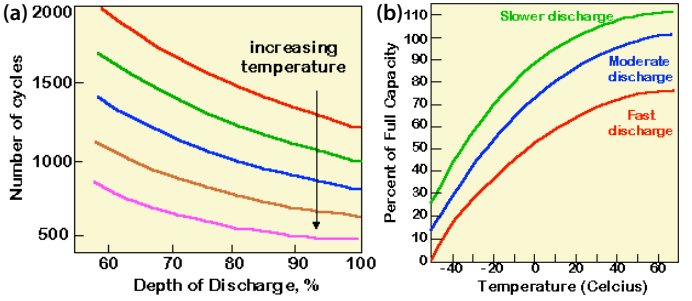

Figure 17.29: (a) Qualitative illustration of the cycle lifetime of a lead-acid battery in dependence of the DoD and temperature [100]. (b) Effect of the temperature on the battery capacity [100].

DoD. The smaller the DoD, the higher the cycle lifetime. Thus, that the battery will last longer if the average DoD can be reduced during the lifetime of the battery. Also, *battery overheating* should be strictly controlled. Overheating can because of overcharging and subsequent over-voltage of the lead acid battery. To prevent this, charge controllers are used that we address in the next section.

Temperature effects

While the battery lifetime is increased at lower temperatures, one more effect must to be considered. The

temperature affects the battery capacity during regular use too. As seen from Fig. 17.29 (b), the lower the temperature, the lower the battery capacity. This is because, the chemicals in the battery are more active at higher temperatures, and the increased chemical activity leads to increased battery capacity. It is even possible to reach an above rated battery capacity at high temperatures. However, such high temperatures are severely detrimental to the battery health.

Ageing

The major cause for ageing of the battery is *sulphation*. If the battery is insufficiently recharged after being discharged, sulphate crystals start to grow, which cannot be completely transformed back into lead or lead oxide. Thus the battery slowly loses its active material mass and hence its discharge capacity. *Corrosion* of the lead grid at the electrode is another common ageing mechanism. Corrosion leads to increased grid resistance due to high positive potentials. Further, the electrolyte can *dry out*. At high charging voltages, gassing can occur, which results in the loss of water. Thus, demineralised water should be used to refill the battery from time to time.

17.4.3 Charge controllers

After having discussed different types of batteries and different battery characteristics, it is now time to discuss *charge controllers*. Charge controllers are used in PV systems that use batteries, which are stand-alone systems in most cases. As we have seen before, it is very important to charge and discharge batteries at the right voltage and current levels in order to ensure a long battery lifetime. A battery is an electrochemical device that requires a small over-potential to be charged. However, batteries have strict voltage limits, which are necessary for their optimal functioning. Further, the amount of current sent to the battery by the PV array and the current flowing through the battery while being discharged have to be within well-defined limits for proper functioning of the battery. We have seen before that lead acid batteries suffer from overcharge and over-discharge. On the other hand, the PV array responds dynamically to ambient conditions like irradiance, temperature and other factors like shading. Thus, directly coupling the battery to the PV array and the loads is detrimental to the battery lifetime.

Therefore a device is needed that controls the currents flowing between the battery, the PV array and the load and that ensures that the electrical parameters present at the battery are kept within the specifications given by the battery manufacturer. These tasks are done by a *charge controller*, that nowadays has several different

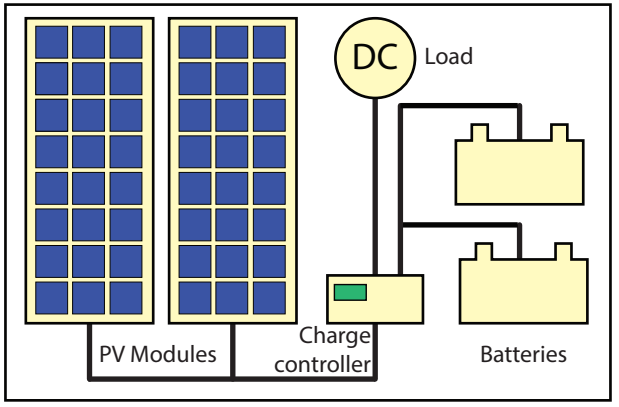

Figure 17.30: Illustrating the position of the *charge controller* in a generic PV system with batteries.

functionalities, which also depend on the manufacturer. We will discuss the most important functionalities. A schematic of its location in a PV system is shown in Fig. 17.30.

When the sun is shining at peak hours during summer, the generated PV power excesses the load. The excess energy is sent to the battery. When the battery is fully charged, and the PV array is still connected to the battery, the battery might overcharge, which can cause several problems like gas formation, capacity loss or overheating. Here, the charge controller plays a vital role by de-coupling the PV array from the battery. Similarly, during severe winter days at low irradiance, the load exceeds the power generated by the PV array, such that the battery is heavily discharged. Over-discharging the battery has a detrimental effect on the cycle lifetime, as discussed above. The charge controller prevents the battery from being over-discharged by disconnecting the battery from the load.

For optimal performance, the battery voltage has to be within specified limits. The charge controller can help in maintaining an allowed voltage range in order to ensure a healthy operation. Further, the PV array will have its V_{mpp} at different levels, based on the temperature and irradiance conditions. Hence, the charge controller needs to perform appropriate voltage regulation to ensure the battery operates in the specified voltage range, while the PV array is operating at the MPP. Modern charge controllers often have an MPP tracker integrated.

As we have seen above, certain C-rates are used as battery specifications. The higher the charge/discharge rates, the lower the coulombic efficiency of the battery. The optimal charge rates, as specified by the manufacturer, can be reached by manipulating the current flowing into the battery. A charge controller that contains a proper current regulation is also able to control the C-rates. Finally, the charge controller can impose the limits on the maximal currents flowing into and from the battery.

If no *blocking diodes* are used, it is even possible that the battery can "load" the PV array, when the PV array is operating at a very low voltage. This means that the battery will impose a forward bias on the PV modules and make them consume the battery power, which leads to heating up the solar cells. Traditionally, blocking diodes are used at the PV panel or string level to prevent this *back discharge* of the battery through the PV array. However, this function is also easily integrated in the charge controller.

We distinguish between *series* and *shunt* controllers, as illustrated in 17.31. In a series controller, overcharging is prevented by disconnecting the PV array until a particular voltage drop is detected, at which point the array is connected to the battery again. On the other hand, in a parallel or shunt controller, overcharging is prevented by short-circuiting the PV array. This means that the PV modules work under short circuit mode, and that no current flows into the battery. These topologies also ensure over-discharge protection using power switches for the load connection, which are appropriately controlled by the algorithms implemented into the charge controller algorithm. As charge controllers are a necessary component of of stand-alone PV systems, they should have a very high efficiency.

As we have seen above, temperature plays a crucial role in the functioning of the battery. Not only does temperature affect the lifespan of the battery, but it also changes its electrical parameters significantly. Thus, modern charge controllers have a temperature sensor included, which allows the charge controller to adjust the electrical parameters of the battery, like the operating voltage, to the temperature. The charge controller thus keeps the operating range of the battery within the optimal range of voltages. The charge controller is usually kept in close proximity to the battery, such that the operating temperature of the battery is close to that of the charge controller. However, when the battery is heavily loaded, the battery might heat up, leading to differences between the temperature expected by the charge controller and the actual temperature of the battery. Therefore, high end charge controllers also take temperature effects due to high currents into account.

17.5 Cables

The overall performance of PV systems also is strongly dependent on the correct choice of the cables. We therefore will discuss how to choose suitable cables. But we start our discussion with *color conventions*.

PV systems usually contain DC and AC parts. For correctly installing a PV system, it is important to know the color conventions. For *DC cables*,

- **red** is used for connecting the +-contacts of the different system components with each other while

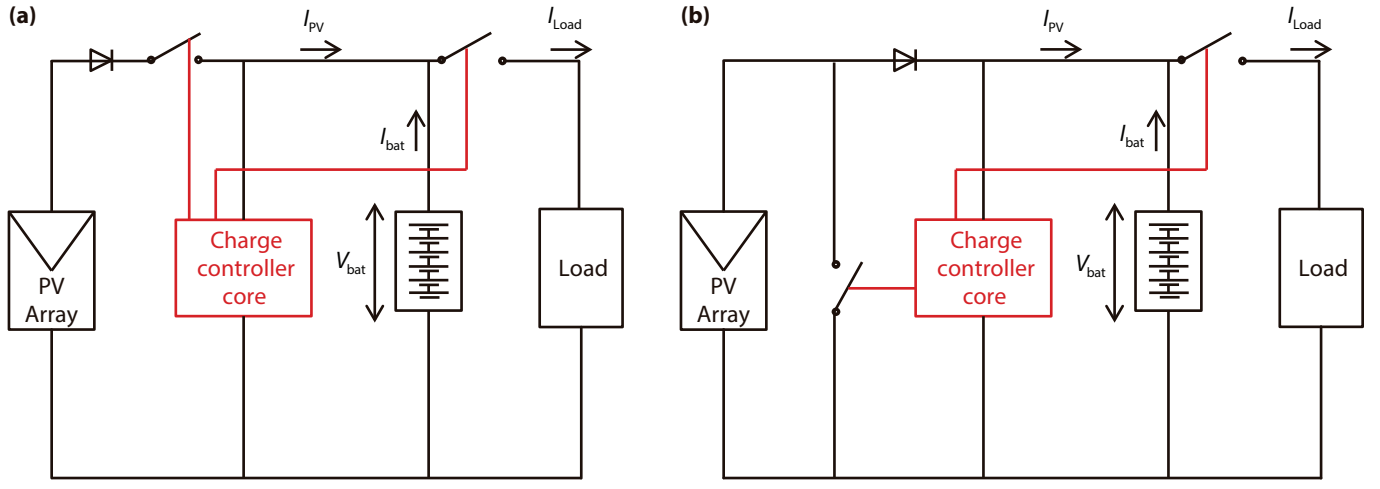

Figure 17.31: Basic wiring scheme of (a) a series and (b) a shunt charge controller.

- **black** is used for interconnecting the $-$-*contacts*.

For *AC wiring*, different colour conventions are used around the world.

- For example, in the *European Union*,
  blue is used for neutral,
  green-yellow is used for the protective earth and
  brown (or another color) is used for the phase.

- In the *United States* and *Canada*,
  silver is used for neutral,
  green-yellow, **green** or a **bare** conductor is used for the protective earth
  and **black** (or another color) is used for the *phase*.

- In India and Pakistan,
  black is used for *neutral*,
  green is used for the protective earth and
  blue, **red**, or yellow is used for the *phase*.

Therefore it is very important to check the standards of the country in that the PV system is going to be installed.

The cables have to be chosen such that *resistive losses* are minimal. For estimating these losses, we look at a very simple system that is illustrated in Fig. 17.32 The system consists of a power source and a load with resistance R_L. Also the cables have a resistance R_{cable}, which also is sketched. The power loss at the cables is given

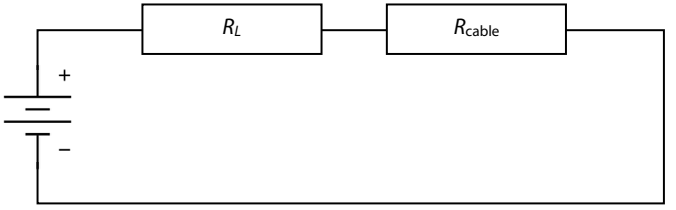

Figure 17.32: Illustrating a circuit with a load R_L and cable resistance R_{cable}. The connections drawn in between the components are lossless.

as

$$P_{\text{cable}} = I \cdot \Delta V_{\text{cable}}, \tag{17.31}$$

where ΔV_{cable} is the voltage drop across the cable, which is given as

$$\Delta V_{\text{cable}} = V \frac{R_{\text{cable}}}{R_L + R_{\text{cable}}}. \tag{17.32}$$

Using

$$V = I(R_L + R_{\text{cable}}) \tag{17.33}$$

we find

$$P_{\text{cable}} = I^2 R_{\text{cable}}. \tag{17.34}$$

Hence, as the current doubles, four times as much heat will be dissipated at the cables. It now is obvious why modern modules have connected all cells in series.

Let us now calculate the resistance of a cable with length ℓ and cross section A. It is clear, that if ℓ is doubled, also R_{cable} doubles. In contrast, if A doubles, R_{cable} decreases to half. The resistance thus is given by

$$R_{\text{cable}} = \rho \frac{\ell}{A} = \frac{1}{\sigma} \frac{\ell}{A}, \tag{17.35}$$

where ρ is the *specific resistance* or *resistivity* and σ is the *specific conductance* or *conductivity*. If both, ℓ and A are given in metres, their units are $[\rho] = \Omega \cdot \text{m}$ and $[\sigma] = S/\text{m}$, where S denotes the unit for conductivity, which is *Siemens*.

For electrical cables it is convenient to have ℓ in metres and A in mm^2. When using this convention, we find the following units for ρ and σ:

$$1\,\Omega \cdot \text{m} = 1\,\Omega \frac{\text{m}^2}{\text{m}} = 1\,\Omega \frac{10^6\,\text{mm}^2}{\text{m}} = 10^6\,\Omega \frac{\text{mm}^2}{\text{m}},$$
$$1\,\frac{S}{\text{m}} = 1\,S \frac{\text{m}}{\text{m}^2} = 1\,\Omega \frac{\text{m}}{10^6\,\text{mm}^2} = 10^{-6}\,\Omega \frac{\text{m}}{\text{mm}^2}.$$

The most widely used metals used for electrical cables are *copper* and *aluminium*. Their resistances and conductivities are

$$\rho_{\text{Cu}} = 1.68 \cdot 10^{-8}\,\Omega \cdot \text{m} = 1.68 \cdot 10^{-2}\,\Omega \frac{\text{mm}^2}{\text{m}},$$
$$\sigma_{\text{Cu}} = 5.96 \cdot 10^7\,\frac{\text{S}}{\text{m}} = 59.6\,\text{S} \frac{\text{m}}{\text{mm}^2},$$
$$\rho_{\text{Al}} = 2.82 \cdot 10^{-8}\,\Omega \cdot \text{m} = 2.82 \cdot 10^{-2}\,\Omega \frac{\text{mm}^2}{\text{m}},$$
$$\sigma_{\text{Al}} = 3.55 \cdot 10^7\,\frac{\text{S}}{\text{m}} = 35.5\,\text{S} \frac{\text{m}}{\text{mm}^2}.$$

Usual thicknesses for cables are 0.75 mm^2, 1.5 mm^2, 2.5 mm^2, 4 mm^2, 6 mm^2, 10 mm^2, 16 mm^2, 25 mm^2, 35 mm^2, *et cetera*. Since DC circuits are driven at lower voltages than AC currents, the currents are higher, requiring thicker cables.

18
PV System Design

In the previous chapters we discussed how to estimate the irradiance on a PV module in dependence of the position on Earth, date and time (Chapter 16). Further we introduced and discussed all the different components of PV Systems in Chapter 17.

In this chapter, we will put the things together and use them such that we can design complete PV systems. Of course, PV system designs can be made with different levels of complexity. For a first approximation the STC performance of the PV modules and the performance of the other components (like the inverter) at STC conditions and the number of *Equivalent Sun Hours* (ESH) at the location of the PV system are sufficient. The notion of ESH will be discussed below. In a more detailed approach, performance changes of the modules and the other components due to changing irradiance and weather conditions are taken into account. Since these performance changes can be quite high, they can alter the optimal system design considerably.

There are two main paradigms for designing PV systems. First, the system can be designed such that the generated power and the loads, *i.e.* the consumed power, match. A second way to design a PV system is to base the design on economics. We must distinguish between *grid-connected* and *off-grid* systems. As we will see, grid-connected systems have very different demands than off-grid systems.

This chapter is organised as follows: First, as an example, we will discuss a design of a simple off-grid

system in Section 18.1. After that we will take a more detailed look on *load profiles* in Section 18.2. In Section 18.3 and on how weather and irradiance conditions affect the performance of PV modules and BOS components, mainly inverters. Finally, in Sections 18.4 and 18.5 we learn how to design grid-connected and off-grid systems, respectively.

18.1 A simple approach for designing off-grid systems

In this section, we will design a simple off-grid system, as depicted in Fig. 18.1. The design presented here is based on very simple assumptions and does not take any weather-dependent performance changes into account. Nonetheless, we will see the major steps that are necessary for designing a system. Such a simple design can be performed in a six step plan:

1. Determine the total load current and operational time
2. Add system losses
3. Determine the solar irradiation in daily equivalent sun hours (ESH)
4. Determine total solar array current requirements

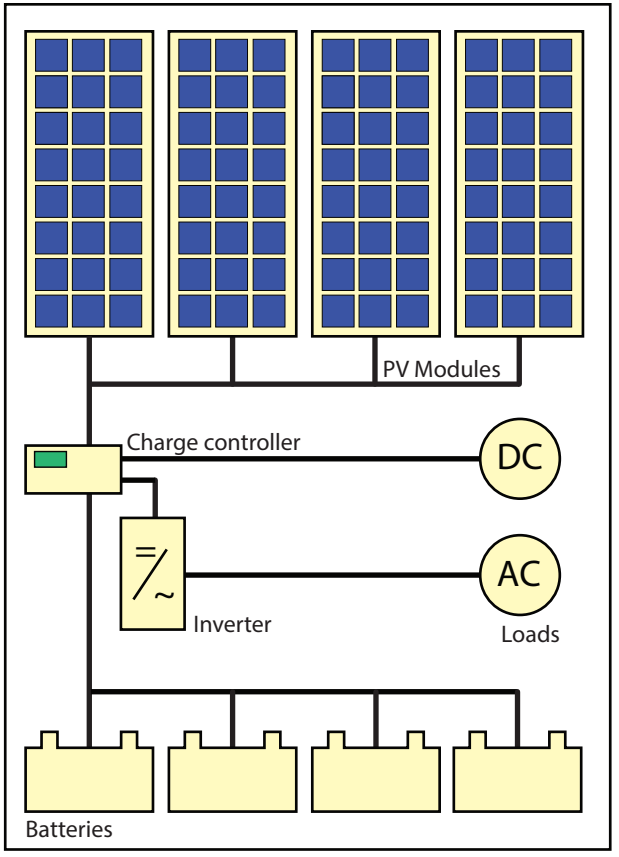

Figure 18.1: Illustrating a simple off-grid PV system with AC and DC loads (see also Fig. 15.1 b).

5. Determine optimum module arrangement for solar array
6. Determine battery size for recommended reserve time

1. Determine the total load current and operational time

Before starting determining the current requirements of loads of a PV system one has to decide the nominal operational voltage of the PV system. Usual nominal voltages are 12 V or 24 V. When knowing the voltage, the next step is to express the daily energy requirements of loads in terms of current and average operational time expressed in Ampere-hours [Ah]. In case of DC loads the daily energy [Wh] requirement is calculated by multiplying the power rating [W] of an individual appliance with the average daily operational time [h]. Dividing the Wh by the nominal PV system operational voltage, the required Ah of the appliance is obtained.

Example

A 12 V PV system has two DC appliances A and B requiring 15 and 20 W respectively. The average operational time per day is 6 hours for device A and 3 hours for device B. The daily energy requirements of the devices expressed in Ah are calculated as follows:

Device A: $15\,\text{W} \cdot 6\,\text{h} = 90\,\text{Wh}$

Device B: $20\,\text{W} \cdot 3\,\text{h} = 60\,\text{Wh}$

Total: $90\,\text{Wh} + 60\,\text{Wh} = 150\,\text{Wh}$

$150\,\text{Wh}/12\,\text{V} = 12.5\,\text{Ah}$

In case of AC loads the energy use has to be expressed as a DC energy requirement since PV modules generate DC electricity. The DC equivalent of the energy use of an AC load is determined by dividing the AC load energy use by the efficiency of the inverter, which typically can be assumed to be 85%. By dividing the DC energy requirement by the nominal PV system voltage the Ah is determined.

Example

An AC computer (device C) and TV set (device D) are connected to the PV system. The computer, which has rated power 40 W, runs 2 hours per day and the TV set with rated power 60 W is 3 hours per day in operation. The daily energy requirements of the devices expressed in DC Ah are calculated as follows:

Device C: $40\,\mathrm{W} \cdot 2\,\mathrm{h} = 80\,\mathrm{Wh}$

Device D: $60\,\mathrm{W} \cdot 3\,\mathrm{h} = 180\,\mathrm{Wh}$

Total: $80\,\mathrm{Wh} + 180\,\mathrm{Wh} = 260\,\mathrm{Wh}$

DC requirement: $260\,\mathrm{Wh}/0.85 = 306\,\mathrm{Wh}$

$306\,\mathrm{Wh}/12\,\mathrm{V} = 25.5\,\mathrm{Ah}$

2. Add system losses

Some components of the PV system, such as charge regulators and batteries require energy to perform their functions. We denote the use of energy by the system components as system energy losses. Therefore, the total energy requirements of loads, which were determined in step 1, are increase with 20 to 30% in order to compensate for the system losses.

Example

The total DC requirements of loads plus the system losses (20%) *are determined as follows:*

$(12.5\,\mathrm{Ah} + 25.5\,\mathrm{Ah}) \cdot 1.2 = 45.6\,\mathrm{Ah}$

3. Determine the solar irradiation in daily equivalent sun hours (EHS)

How much energy a PV module delivers depends on several factors, such as local weather conditions, seasonal changes, and installation of modules. PV modules should be installed under the optimal *tilt-angle* in order to achieve best year-round performance. It is also important to know whether a PV system is expected to be used all-year round or only during a certain period of a year. The power output during winter is much less than the yearly average and in the summer months the power output will be above the average.

In the PV community, 1 equivalent sun means a solar irradiance of 1000 $\mathrm{W/m^2}$. This value corresponds to the standard, at which the performance of solar cells and modules is determined. The rated parameters of modules (see for example Table 17.1) are determined at solar irradiance of 1 sun.

When solar irradiation data are available for a particular location than the equivalent sun hours can be determined. For example, in the Netherlands the average annual solar irradiation is 1000 $\mathrm{kWh/m^2}$. According to the AM1.5 spectrum, 1 sun delivers 1000 $\mathrm{W/m^2}$ = 1 $\mathrm{kW/m^2}$. Hence, the Dutch average annual solar irradiation can be expressed as

$$\frac{1000\,\mathrm{kWh} \cdot \mathrm{m}^{-2}}{1\,\mathrm{kW} \cdot \mathrm{m}^{-2}} = 1000\,\mathrm{h},$$

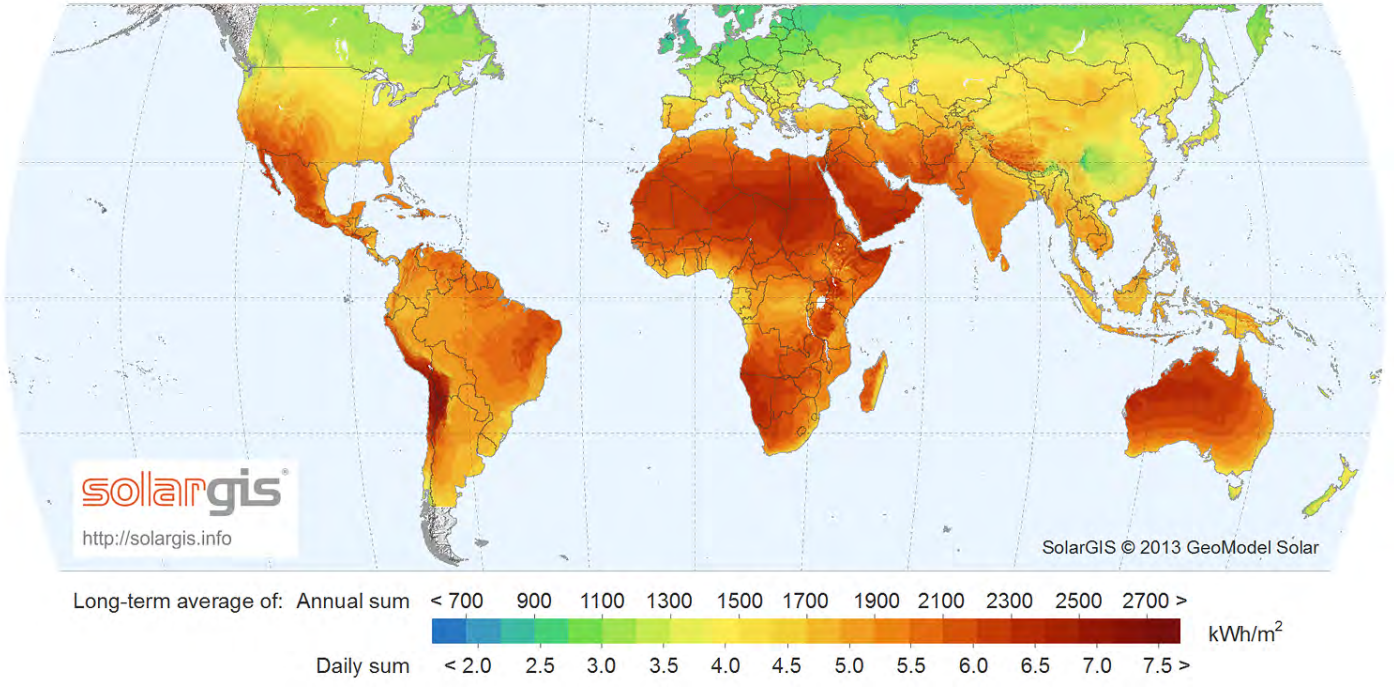

Figure 18.2: The average global horizontal insolation of the World given in kWh/year and PSH/day [101]. Note that this map only shows latitudes in between 60° N and 60° S.

i.e. 1000 equivalent sun hours, which corresponds to 2.8 h per day. Figures 18.2 and 18.3 show the *global horizontal irradiation* given in units of daily equivalent sun hours (ESH) and in kWh/year. Note that these figures show horizontal values, meaning that a PV module would get this irradiance if it would be flat. For a tilted module, the annual irradiance can be considerably higher, as we have seen in Chapter 16.

Figure 18.3: The average global horizontal insolation of the Netherlands given in kWh/year and PSH/day [101].

4. Determine total solar array current requirements

The current that has to be generated by the solar array is determined by dividing the total DC energy requirement of the PV system including loads and system losses (calculated in step 2 and expressed in Ah) by the daily equivalent sun hours (determined in step 3).

Example

The total DC requirements of loads plus the system losses are 45.6 Ah. *The daily EHS for the Netherlands is about 3 hours. The required total current generated by the solar array is* 45.6 Ah / 3 h = 15.2 A.

5. Determine optimum module arrangement for solar array

Usually, PV manufacturers produce modules in a whole series of different output powers. In the optimum arrangement of modules the required total solar array current (as determined in step 4) is obtained with the minimum number of modules. Modules can be either connected in series or in parallel to form an array. When modules are connected in series, the nominal voltage of the PV system is increased, while the parallel connection of modules results in a higher current.

The required number of modules in parallel is calculated by dividing the total current required from the solar array (determined in step 4) by the current generated by module at peak power (rated current in the specification sheet). The number of modules in series is determined by dividing the nominal PV system voltage with the nominal module voltage (in the specification sheet under configuration). The total number of modules is the product of the number of modules required in parallel and the number required in series.

Example

The required total current generated by the solar array is 15.2 A. *We have Shell SM50-H modules available. The specification of these modules is given in Table 17.1. The rated current of a module is* 3.15 A. *The number of modules in parallel is* $15.2\,\text{A} / 3.15\,\text{A} = 4.8 < 5$ *modules. The nominal voltage of the PV system is* 12V *and the nominal module voltage is* 12 V. *The required number of modules in series thus is* $12\,\text{V} / 12\,\text{V} = 1$ *module. Therefore, the total number of modules in the array is* $5 \cdot 1 = 5$ *modules.*

6. Determine battery size for recommended reserve time

Batteries are a major component of stand-alone PV systems. The batteries provide load operation at night or in combination with the PV modules during periods of limited sunlight. For a safe operation of the PV system one has to anticipate periods with cloudy weather and plan a reserve energy capacity stored in the batteries. This reserve capacity is referred to as *PV system autonomy*, which means the period of time that the system is not dependent on energy generated by PV modules. It is given in days. The system autonomy depends on the type of loads. For critical loads such as telecommunications components the autonomy can be 10 days and more, for residential use it is usually 5 days or less. The capacity [Ah] of the batteries is calculated by multiplying the daily total DC energy requirement of the PV system including loads and system losses (calculated in step 2 and expressed in Ah) by the number of days of recommended reserve time. In order to prolong

Table 18.1: Worksheet for designing a simple off-grid PV system based on rough assumptions.

Daily DC loads requirements			
DC load	W ×	h =	Wh
Total DC loads energy use:			

Daily AC loads requirements			
DC load	W ×	h =	Wh
Total AC loads energy use:			
/0.85 = **DC energy requirement**			

1	Daily DC energy use (DC loads)		
1	Daily DC energy use (AC loads)	+	
1	Daily DC energy use (all loads)	=	
	PV system nominal voltage	/	
	Daily Ah requirements (all loads)	=	
2	Add PV system losses	×	
	Daily Ah requirements (system)	=	
3	Design EHS	/	
4	Total solar array current	=	
5	**Select module type**		
	Module rated current	/	
	Number of modules in parallel	=	
	PV system nominal voltage		
	Modules nominal voltage	/	
	Number of modules in series	=	
	Number of modules in parallel	×	
	Total number of modules	=	
6	**Determine battery capacity**		
	Daily Ah requirements (system)		
	Recommended reserve time	×	
	Usable battery capacity	/	
	Minimum battery capacity	=	

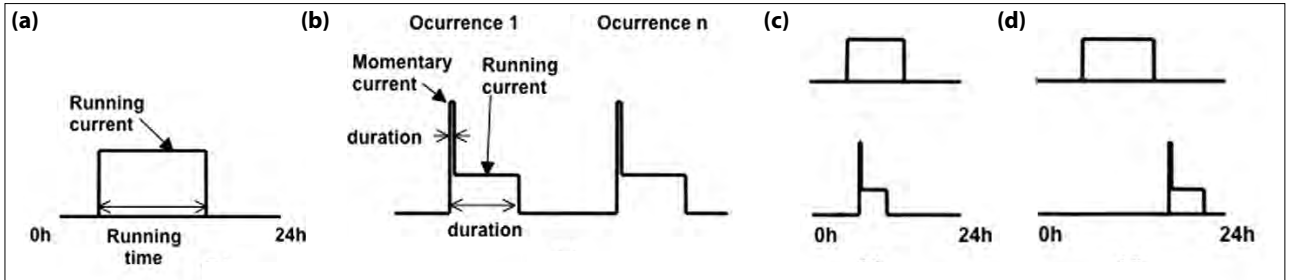

Figure 18.4: Illustrating different load profiles.

the life of the battery it is recommended to operate the battery using only 80% of its capacity. Therefore, the minimal capacity of the batteries is determined by dividing the required capacity by a factor of 0.8.

Example

The total DC requirements of loads plus the system losses are 45.6 Ah. *The recommended reserve time capacity for the installation side in the Netherlands is 5 days. Battery capacity required by the system is* $45.6\,\text{Ah} \cdot 5 = 228\,\text{Ah}$. *The minimal battery capacity for a safe operation therefore is* $228\,\text{Ah}/0.8 = 285\,\text{Ah}$.

Designing a simple PV system as described in this section can be carried out using a worksheet as in Table 18.1, where the PV system design rules are summarised.

18.2 Load profiles

Now we take a look at the load profile. Figure 18.4 illustrates different shapes that loads can have. (a) A simple load draws a constant amount of power for a certain time. (b) However, the consumed power does not need to be constant but can also show peaks that correspond to switching electrical appliances on or off. A household of course has several different loads that (c) can be switched on at the same time (coincident) or (d) at different times (non-coincident).

Analysing load profiles can be performed with increasing complexity and hence accuracy. The simplest method is to determine the loads on a 24-hour basis. To do this, an arbitrary day can be taken and the electricity consumption can be monitored. However, several loads do not fit in such a scheme. Several examples for

this already were treated in Section 18.1. For example, washing machines and dishwashers do not fit in a 24-hour scheme because typically they are only used several times in a week. Additionally, several loads are seasonal in nature, for example, air conditioning or heating, in case this is performed with a heat pump. Therefore it is advisable to look at load profiles for a whole year.

The total energy consumed in a year is given by

$$E_L^Y = \int_{\text{year}} P_L(t)\,\mathrm{d}t. \tag{18.1}$$

It is expressed in kWh/year.

18.3 Meteorological effects

Standard Test Conditions (STC) of Photovoltaic (PV) modules are generally not representative of the real working conditions of a solar module. For example, high levels of incident irradiation, may cause the temperature of a module to rise many degrees above the STC temperature of 25°C, therefore lowering the module performances. For example, in the in a climate such as the one of the Netherlands, real operating conditions for PV systems correspond to relatively low levels of irradiance combined with a cold and windy weather. In order to effectively estimate real time energy production from a given set of ambient parameters, it is essential to develop an accurate thermal model of the solar module, which evaluates the effects of the major meteorological parameters on the final temperature of a solar module.

In this section, we will develop an accurate thermal model for estimating the PV module working temperature as a function of meteorological parameters will be developed. The model is based on a detailed energy balance between the module itself and the surrounding environment. Both the installed configuration of the array together with external parameters such as direct incident solar irradiance on the panels, wind speed and cloud cover will be taken into account. From the calculated equilibrium temperature, the overall efficiency of the PV array will be calculated by separately assessing the temperature and irradiance effect on the efficiency at STC.

18.3.1 Simplified thermal model for a photovoltaic array

The temperature strongly influences the performances of a PV module [102]. While the level of incident irradiation can be easily determined by measuring it with a pyranometer, the temperature reached inside the cell is much harder evaluate. In order to give an estimate of the average operating temperature of the module, olar manufacturers provide, together with rated per-

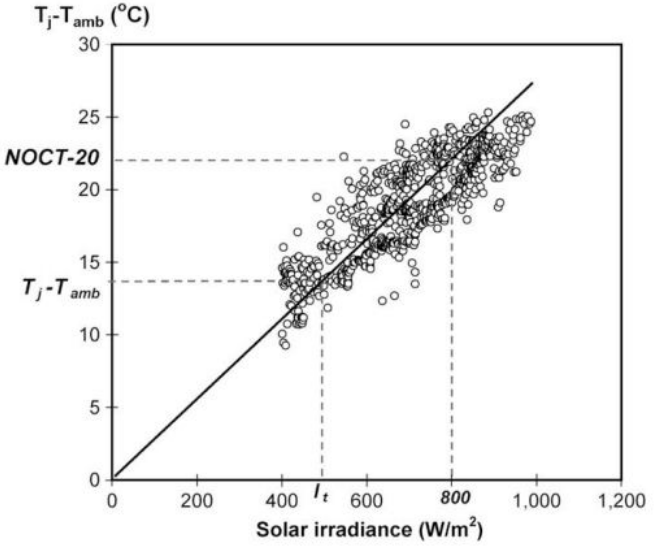

Figure 18.5: Rise in temperature above the ambient level with respect of increasing solar irradiance [103].

formances at STC, also the so called *Nominal Operating Cell Temperature* (NOCT). This value corresponds to the performances of a cell under an irradiance level of 800 W/m^2, ambient temperature of 20°C and an external wind speed of 1 m/s [102].

Simplified steady state models use a linear relationship between the solar irradiance G_M and the difference between the module temperature and the ambient temperature $(T_M - T_a)$, where the NOCT is used as a reference point, [103]

$$T_M = T_a + \frac{T_{\text{NOCT}} - 20^\circ}{800} G_M. \tag{18.2}$$

Equation (18.2) is however too simplistic and thus leads to significant errors in the modules temperature evaluations. The reason is that it does not take environmental conditions such as wind speed and mounting configuration of the array into account [103].

In order to take into mounting configuration of the module into account, the Installed Nominal Operating Conditions Temperature (INOCT) has been defined [104]. This value is described as the cell temperature of an installed array at NOCT conditions. Its value can therefore be obtained from the NOCT and the mounting configuration. The evaluation of how the INOCT varies with the module mounting configuration has been experimentally determined by measuring the NOCT at various mounting heights. The results can be found in literature and are here summarised in the Table 18.2.

Evaluating the influence of external meteorological parameters on the the module temperature is more complex and has been approached by developing a detailed thermal energy balance between the module and the surroundings. Here, the module is assumed to be a

Table 18.2: Derivation of the INOCT from the NOCT for various mounting configurations [104].

Rack Mount	INOCT = NOCT − 3°C
Direct Mount	INOCT = NOCT + 18°C
Standoff	INOCT = NOCT + X

where X is given by

W (inch)	X (°C)
1	11
3	2
6	− 1

single uniform mass at temperature T_M. The three types of heat transfer between the module itself with the surroundings are *conduction*, *convection* and *radiation*. The contributions considered in the model are:

- Heat received from the Sun in the form of insolation φG_M, where φ is the absorptivity of the module.
- Convective heat exchange with surrounding air from the front and rear side of the module

  $$h_c(T_M - T_a),$$

  where h_c denotes the overall convective heat transfer coefficient of the module.
- Radiative heat exchange between the upper module surface and the sky

  $$\epsilon_{\text{top}}\sigma\left(T_M^4 - T_{\text{sky}}^4\right),$$

  where $\epsilon_{\text{top}} = 0.84$ is the emissivity of the module front glass and σ is the Stefan-Boltzmann constant as defined in Eq. (5.20). and between the rear surface and the ground

  $$\epsilon_{\text{back}}\sigma\left(T_M^4 - T_{\text{gr.}}^4\right),$$

  where the emissivitiy of the back is assumed to be $\epsilon_{\text{back}} = 0.89$.
- Conductive heat transfer between the module and the mounting structure. We neglect this contribution due to the small area of the contact points [105].

By separately considering each of the above mentioned contributions, we write down the balance for the heat transfer, [104]

$$\begin{aligned} mc\frac{\mathrm{d}T_M}{\mathrm{d}t} =& \varphi G_M - h_c(T_M - T_a) - \epsilon_{back}\sigma\left(T_M^4 - T_{\text{gr.}}^4\right) \\ & - \epsilon_{\text{top}}\sigma\left(T_M^4 - T_{\text{sky}}^4\right). \end{aligned} \tag{18.3}$$

Before providing a solution for this differential equation, it is appropriate to remark the fact that we are considering the entire module as a uniform piece at a

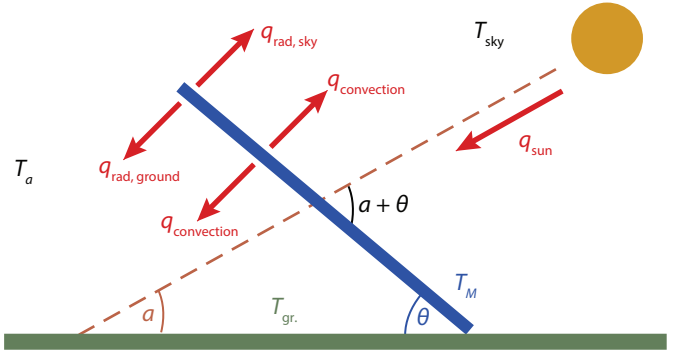

Figure 18.6: Representation of heat exchange between a tilted module surface and the surroundings.

temperature T_M. However, this is not entirely realistic since modules are made of various layers of different materials surrounding the actual solar cells. It is the purpose of this section to evaluate the temperature of the inner cell, which is the place where absorption of solar radiation effectively place. This temperature will be to some extent higher than the surface module temperature T_M due to the heat produced in the cell due to light absorption. The approximation of considering the temperature uniform throughout the module layers however is justified because of the relatively low thickness of the active cell together with the low heat capacity of the cell material compared to the other layers [105]. This results into a very low thermal resistance of the cell to heat flow and therefore justifies the uniform temperature approximation.

In addition to this, a steady state approach will here be considered, meaning that the module temperature will not change over time for each of the 10-minute time steps. In reality, the temperature follows an exponential decay lagging behind variations in irradiation level. It is defined as *Time Constant* of a module, which is 'the time it takes for the module to reach 63% of the total change in temperature resulting from a step change in irradiance' [105]. Time constants for PV modules are generally of the order of approximately 7 minutes. For time steps greater that the Time Constant, as it is in our case, the module can be approximated as being in a steady state condition. In light of this assumption, the term on the left hand side of Eq. (18.3) vanishes.

It is now possible to proceed with the solution of the thermal energy balance equation. The formula can be linearised by noticing that

$$\left(a^4 - b^4\right) = \left(a^2 + b^2\right)(a+b)(a-b). \tag{18.4}$$

Since

$$\left(T_M^2 + T_{\text{sky}}^2\right)(T_M + T_{\text{sky}})$$

changes less than 5% for a 10°C variation in T_M, we can consider this term to be constant when T_M varies [104]. Therefore the energy balance can be simplified

becoming linear with respect to T_M. By defining

$$h_{r,\,\mathrm{sky}} = \epsilon_{\mathrm{top}}\sigma\left(T_M^2 + T_{\mathrm{sky}}^2\right)(T_M + T_{\mathrm{sky}}), \tag{18.5a}$$

$$h_{r,\,\mathrm{gr.}} = \epsilon_{\mathrm{back}}\sigma\left(T_M^2 + T_{\mathrm{gr.}}^2\right)(T_M + T_{\mathrm{gr.}}), \tag{18.5b}$$

we can rewrite Eq. (18.3) and find

$$\begin{aligned}\varphi G_M - h_c(T_M - T_a) - h_{r,\,\mathrm{sky}}(T_M - T_{\mathrm{sky}}) \\ - h_{r,\,\mathrm{gr.}}(T_M - T_{\mathrm{gr.}}) = 0.\end{aligned} \tag{18.6}$$

By rearranging the terms, the formula can be explicitly expressed as a function of T_M,

$$T_M = \frac{\varphi G + h_c T_a + h_{r,\,\mathrm{sky}} T_{\mathrm{sky}} + h_{r,\,\mathrm{gr.}} T_{\mathrm{gr.}}}{h_c + h_{r,\,\mathrm{sky}} + h_{r,\,\mathrm{gr.}}} \tag{18.7}$$

However, since $h_{r,\,\mathrm{gr.}}$ and $h_{r,\,\mathrm{sky}}$ are also function of the module temperature, the equation needs to be solved iteratively: an initial module temperature is assigned and by $h_{r,\,\mathrm{gr.}}$ and $h_{r,\,\mathrm{sky}}$ are updated each iterations. A nearly exact solution can be obtained after 5 iterations.

Before solving iteratively Eq. (18.7) there are still many unknown variables that need to be determined, which we will do in the following sections.

18.3.2 Calculating the convective transfer coefficients

Convection is a form of energy transfer from one place to another caused by the movement of a fluid. Convective heat transfer can be either *free* or *forced* depending on the cause of the fluid motion.

In free convection, heat transfer is caused by temperature differences which affect the density of the fluid itself. Air starts circulating due to difference in buoyancy between hot (less dense) fluid and cold (denser) fluid. A circular motion is therefore initiated with rising hot fluid and sinking cold fluid. Free convection only takes place in a gravitational field [106].

Forced convection, on the other hand, is caused by a fluid flow due to external forces which therefore enhance the convective heat exchange. The heat transfer depends very much on whether the induced flow over a solid surface is laminar or turbulent. In the case of turbulent flow, an increased heat transfer is expected with respect to the laminar situation. This fact is due to an increased heat transport across the main direction of the flow. On the contrary, in laminar flow regime, only conduction is responsible for transport in the cross direction. For this reason, forced convection is always studied separately in the laminar and turbulent regime [106]. The overall convection transfer is made of the two relative contribution for free and forced compon-

ents. Mixed convective coefficient can be obtained by taking the cube root of the cubes of the forced and convective coefficients according to the equation [102]

$$h^3_{\text{mixed}} = h^3_{\text{forced}} + h^3_{\text{free}}. \tag{18.8}$$

Since the convective heat transfer coefficients will be different on the top and rear surface of the module, determining the total heat transfer coefficient has to be decoupled between the top h_c^T and rear h_c^B surfaces. The overall heat transfer will eventually be determined by the sum of the two components.

Convective heat transfer on the top surface

Convective heat transfer has to be distinguished in free and forced components. For the forced component we further have to distinguish between laminar and turbulent flow. We obtain for the laminar and turbulent convective heat transfer coefficients

$$h^{\text{lam.}}_{\text{forced}} = \frac{0.86\,\text{Re}^{-0.5}}{\text{Pr}^{0.67}}\rho c_{\text{air}} w, \tag{18.9a}$$

$$h^{\text{turb.}}_{\text{forced}} = \frac{0.028\,\text{Re}^{-0.2}}{\text{Pr}^{0.4}}\rho c_{\text{air}} w. \tag{18.9b}$$

Re is the *Reynolds number* that expresses the ratio of of the inertial forces to viscous forces,

$$\text{Re} = \frac{w D_h}{\nu}, \tag{18.10}$$

where w is the *wind speed* at the height of the PV array, D_h is the *hydraulic diameter* of the module, which is used as relevant length scale, and ν is the *kinematic viscosity* of air. Pr is the *Prandtl number* which is the ratio of the momentum diffusivity to the thermal diffusivity. It is considered to be 0.71 for air. Finally, ρ and c_{air} are the density and heat capacity of air, respectively. The hydraulic diameter of a rectangle of length L and width W, and thus of the PV module, is given as

$$D_h = \frac{2LW}{L+W}. \tag{18.11}$$

Figure 18.7 shows the variation of the forced heat transfer coefficient at various wind speeds. We notice an overall increase in the heat transfer with increasing wind speeds due to the increased force transfer component. There are two different regions in the graph which represent the laminar and turbulent regimes, respectively. The laminar flow extends till around 3 m/s and is characterised by a lower convective heat exchange compared to the turbulent regime.

In a good approximation, $h_{\text{forced}}$and w are proportional

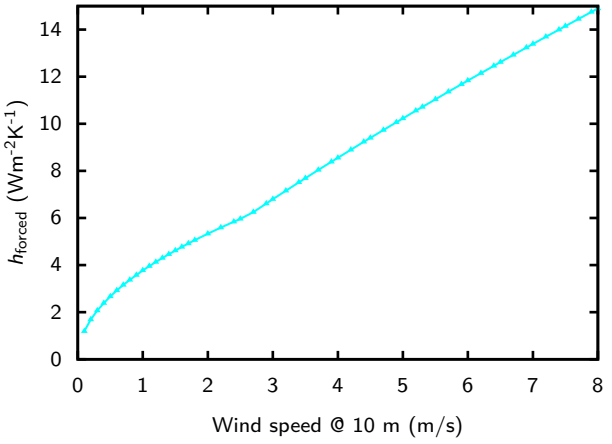

Figure 18.7: Forced convective heat transfer coefficient with increasing wind speeds

to each other and we may write

$$h_{\text{forced}}^{\text{lam.}} \simeq w^{0.5}, \tag{18.12a}$$

$$h_{\text{forced}}^{\text{turb.}} \simeq w^{0.8}. \tag{18.12b}$$

To determine the free heat transfer coefficient we only need to utilise the dimensionless *Nusselt number* Nu, which expresses the ratio between the convective and conductive heat transfer [104]

$$\text{Nu} = \frac{h_{\text{free}} D_h}{k} = 0.21(\text{Gr} \cdot \text{Pr})^{0.32}, \tag{18.13}$$

where k is the heat conductivity of air and Gr is the *Grasshoff number* that is the ratio between the buoyancy and viscous forces,

$$\text{Gr} = \frac{g\beta(T - T_a)D_h^3}{\nu^2}. \tag{18.14}$$

Here, g is the *acceleration due to gravity on Earth* and β is the volumetric thermal expansion coefficient of air, which can be approximated to be $\beta = 1/T$.

With the value of both free and forced coefficient we can calculate the total mixed heat convective mass transfer coefficient using Eq. (18.8).

$$h_{\text{mixed}} = h_c^T = \sqrt[3]{h_{\text{forced}}^3 + h_{\text{free}}^3}. \tag{18.15}$$

Convective heat transfer coefficient on the rear surface of the module.

Convection on the back side of the module will be lower than on the top because of the mounting structure and the relative vicinity to the ground. For example, a rack mount configuration, which is approximately installed at 1 m height, will achieve a larger heat exchange rate than a standoff mounted array that is mounted 20 cm above the ground. We model the effect of the different mounting configurations by scaling the convection coefficient obtained for the top of the module. We determine the scaling factor by performing an energy balance at the INOCT conditions [104],

$$\varphi G_M - h_c^T(T_{\text{INOCT}} - T_a) - h_{r,\,\text{sky}}(T_{\text{INOCT}} - T_{\text{sky}}) = h_c^B(T_{\text{INOCT}} - T_a) + h_{r,\,\text{gr.}}(T_{\text{INOCT}} - T_{\text{gr.}}) \tag{18.16}$$

We define R as the ratio of the actual to the ideal heat loss from the back side,

$$R = \frac{h_c^B(T_{\text{INOCT}} - T_a) + \epsilon_{\text{back}}\sigma(T_{\text{INOCT}}^4 - T_{\text{gr.}}^4)}{h_c^T(T_{\text{INOCT}} - T_a) + \epsilon_{\text{top}}\sigma(T_{\text{INOCT}}^4 - T_{\text{sky}}^4)}. \tag{18.17}$$

Substituting this into Eq. (18.16) at INOCT conditions yields

$$R = \frac{\varphi G_M - h_c^T(T_{\text{INOCT}} - T_a) - \epsilon_{\text{top}}\sigma(T_{\text{INOCT}}^4 - T_{\text{sky}}^4)}{h_c^T(T_{\text{INOCT}} - T_a) + \epsilon_{\text{top}}\sigma(T_{\text{INOCT}}^4 - T_{\text{sky}}^4)}. \tag{18.18}$$

The back side convection is therefore given by

$$h_c^B = R \cdot h_c^T. \tag{18.19}$$

We therefore find the overal convective heat transfer coefficient to be

$$h_c = h_c^T + h_c^B. \tag{18.20}$$

18.3.3 Other parameters

Sky temperature evaluation

The sky temperature can be expressed as a function of the measured ambient temperature, humidity, cloud cover and cloud elevation [104]. On a cloudy day, usually when the cloud cover is above 6 okta[1], the sky temperature will approach the ambient temperature, $T_{\text{sky}} = T_a$ [105]. However, on a clear day the sky temperature can drop below T_a and can be estimated by

$$T_{\text{sky}} = 0.0552 \cdot T_a^{3/2}. \tag{18.21}$$

Wind speed at module height evaluation

Since the anemometer used for the evaluation of the wind speed is at a higher height than the module array,

[1] Okta is a measure for the cloud cover, where 0 is clear sky and 8 is a completely cloudy sky.

the real wind speed experienced by the module will be scaled down with

$$w = w_r \left(\frac{y_M}{y_r} \right)^{\frac{1}{5}}, \tag{18.22}$$

where y_M and y_r denote the module and anemometer heights, respectively [104].

The elevation factor of 1/5 is determined by the landscape surrounding the installation, which is supposed to be open country.

Absorptivity and emissivity of the module

In the thermal model developed, absorptivity is defined as the fraction of the incident radiation that is converted into thermal energy in the module. This value is linked to the reflectivity R and efficiency of the module by the equation [104]

$$\varphi = (1 - R)(1 - \eta). \tag{18.23}$$

Typical value for reflectivity of solar modules are 0.1 [104]. As far as the emissivity is concerned, a value of 0.84 has been used for the front glass surface and 0.89 for the back surface [107].

Now that all the unknown variables have been determined it is possible to the make iterations to calculate the final temperature of the module as a function of the level of irradiance, wind speed and ambient temperature.

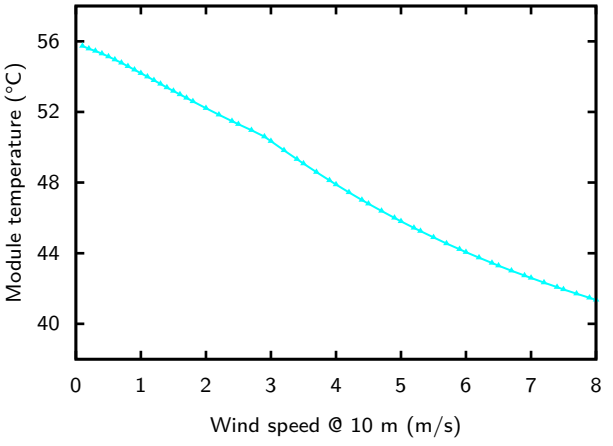

Figure 18.8: Influence of the wind speed on the temperature of a solar module at a fixed radiation of 1000 W/m^2 and ambient temperature of 25°C on a clear day.

18.3.4 Evaluation of the thermal model

In the previous section, all the variables that are required to solve the thermal model have been derived. Therefore, the thermal model now can be solved via it-

eration.

Figure 18.8 below shows the effect of the wind speed on the module temperature at a fixed level of irradiance (1000 W/m^2) and an ambient temperature of 25 °C on a clear day. It is obvious that the module temperature is significantly above the value of 25 °C for all wind-speeds. Thus, STC are not representative for real operating cell conditions. Two different regions are visible in the graph corresponding to laminar and turbulent flow, respectively.

In Fig. 18.9 the thermal model and the NOCT model presented in Eq. (18.2) are compared to each other in dependence of the incident radiation for 1 m/s wind speed. We see that for 1000 W/m^2 the NOCT model predicts a module temperature of 54.3°C regardless of the wind speed, which is much above the values obtained with the thermal model. The contribution of free convective and radiative heat exchange lower the module temperature with respect to the NOCT model. The difference between the two models is more pronounced at low levels of radiation, but for irradiances higher than 1000 W/m^2 the two curves start diverging again. This can be understood when we consider that the NOCT model has been developed as a linearisation of the temperature/irradiation dependency around the NOCT conditions [103]. The accuracy of the model therefore decreases with the increasing distance from the NOCT.

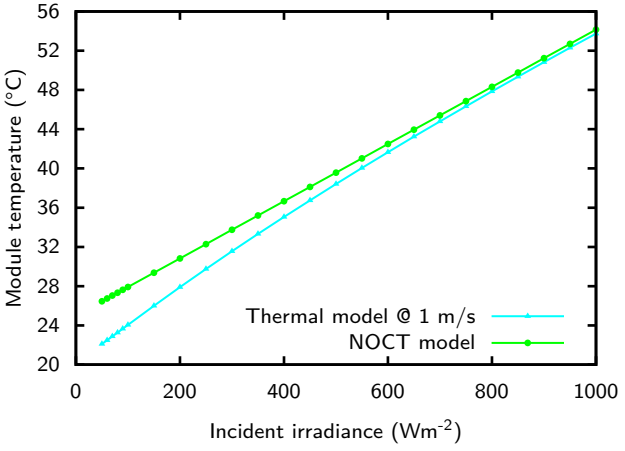

Figure 18.9: The module temperature calculated with the NOCT and the thermal model in dependence of the incident irradiance for 1 m/s wind speed.

18.3.5 Effect of temperature on the solar cell performance

The effect of a module temperature deviating from the 25°C of STC is expressed by the temperature coefficients that given on the data sheet provided by solar manufacturers. When knowing the temperature coefficient of a certain parameter, its value at a certain temperature T_M can be estimated with

$$V_{oc}(T_M, G_{STC}) = V_{oc} + \frac{\partial V_{oc}}{\partial T}(\text{STC})(T_M - T_{STC}), \tag{18.24}$$

$$I_{sc}(T_M, G_{STC}) = I_{sc} + \frac{\partial I_{sc}}{\partial T}(\text{STC})(T_M - T_{STC}), \tag{18.25}$$

$$P_{mpp}(T_M, G_{STC}) = P_{mpp} + \frac{\partial P_{mpp}}{\partial T}(\text{STC})(T_M - T_{STC}), \tag{18.26}$$

$$\eta(T_M, G_{STC}) = \frac{P_{mpp}(T_M, G_{STC})}{G_{STC}}. \tag{18.27}$$

If the efficiency temperature coefficient $\partial\eta/\partial T$ is not given in the datasheet, it can be obtained by rearranging

$$\eta(T_M, G_{STC}) = \eta(\text{STC}) + \frac{\partial \eta}{\partial T}(\text{STC})(T_M - 25°\text{C}). \tag{18.28}$$

An increase in the solar cell temperature will shift the *I-V* curve as shown in Fig. 18.10. The slight increase

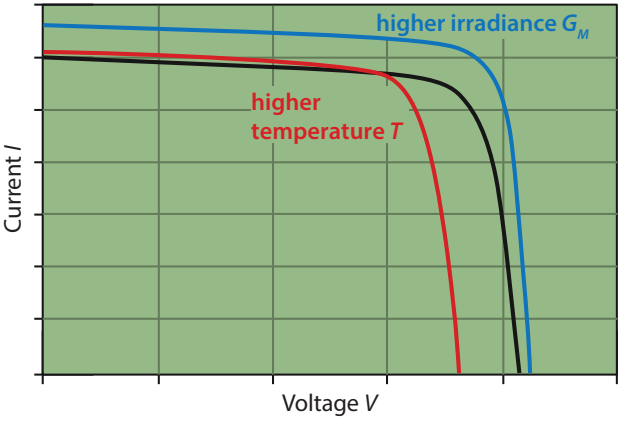

Figure 18.10: Effect of a temperature increase on the *I-V* solar cell characteristic.

in the short circuit current at higher temperatures is completely outweighed by the decrease in open circuit voltage. The overall effect is of a general linear decrease in the maximum achievable power and therefore a decrease in the system efficiency and fill factor. This effect is due to an increase of the intrinsic carrier concentration at higher temperatures which in turns leads to an increase of the reverse saturation current I_0, which represents a measure of the *leakage* of carriers across the solar cell junctions, as we have seen in Chapter 8. The exponential dependence of I_0 from the temperature is the main cause of the linear reduction of $V_{oc}$with the temperature,

$$V_{oc} = \frac{kT}{e} \ln\left(\frac{I_{sc}}{I_0}\right). \tag{18.29}$$

On the other hand, the slight increase in the generated current is due to a moderate increase in the photogenerated current resulting from an increased number of thermally generated carriers. The overall reduction of power at high temperature shows that cold and sunny climates are the best environment where to place a solar system.

18.3.6 Effect of light intensity on the solar cell performance

Intuitively, performances of a solar cell decrease considerably with decreasing light intensity incident on the module with respect to STC. The evaluation of the extent of this reduction is however less straightforward than for the case of the temperature since solar manufacturer often do not explicitly provide a reduction factor of the efficiency at every light intensity level.

By definition the efficiency is given by

$$\eta = \frac{I_{sc} V_{oc} \mathrm{FF}}{G_M} \tag{18.30}$$

The maximum variation of the FF for light intensity between 1 and 1000 W/m^2 is about 2% for CdTe, 5% for a-Si:H, 22% for poly-cristalline silicon, and 23% for mono-crystalline silicon [108].

The short circuit current of a solar cell is directly proportional to the incoming radiation,

$$I_{sc} \simeq \lambda G_M, \tag{18.31}$$

where λ is simply a constant of proportionality. By expressing V_{oc} as in Eq. (18.29), the efficiency can be written as

$$\eta \simeq \mathrm{FF}\,\lambda \frac{kT}{e} (\ln G_M + \ln \lambda - \ln I_0). \tag{18.32}$$

By defining

$$a = \mathrm{FF}\,\lambda\frac{kT}{e}, \tag{18.33a}$$

$$b = \mathrm{FF}\,\lambda\frac{kT}{e}(\ln\lambda - \ln I_0), \tag{18.33b}$$

the efficiency can be finally written as

$$\eta(25^\circ\mathrm{C},\, G_M) = a\ln G_M + b \tag{18.34}$$

The values of the coefficients a and b are device specific parameters and are rarely given by the manufacturer. The overall trend of the efficiency is represented by a straight line on a logarithmic scale [108].

From this model the values of I_{sc}, V_{oc} and the efficiency at a irradiance level G_M can be determined from the STC as follows

$$V_{oc}(25^\circ\mathrm{C},\, G_M) = V_{oc}(\mathrm{STC})\frac{\ln G_M}{\ln G_{\mathrm{STC}}}, \tag{18.35}$$

$$I_{sc}(25^\circ\mathrm{C},\, G_M) = I_{sc}(\mathrm{STC})\frac{G_M}{G_{\mathrm{STC}}}, \tag{18.36}$$

$$P_{MPP}(25^\circ\mathrm{C},\, G_M) = \mathrm{FF}\,V_{oc}(25^\circ\mathrm{C},\, G_M)I_{sc}(25^\circ\mathrm{C},\, G_M), \tag{18.37}$$

$$\eta(25^\circ\mathrm{C},\, G_M) = \frac{P_{MPP}(25^\circ\mathrm{C},\, G_M)}{A_M G_M}, \tag{18.38}$$

where A_M is the module area.

18.3.7 Overall module performance

By combining the two effects of temperature and light intensity, the final efficiency of the module at every level of irradiance and temperature can be determined as [109]

$$\eta(T_M,\, G_M) = \eta(25^\circ\mathrm{C},\, G_M)\,[1 + \kappa(T_M - 25^\circ\mathrm{C})]\,, \tag{18.39}$$

where

$$\kappa = \frac{\partial\eta}{\partial T}\frac{1}{\eta(\mathrm{SCT})}. \tag{18.40}$$

Typical values for κ are $-0.0025/^\circ$C for CdTe, $-0.0030/^\circ$C for CIS, and $-0.0035/^\circ$C for c-Si [109].

All the other parameters such as I_{sc}, V_{oc}, and P_{mpp} can also be evaluated at every level of irradiance and module temperature by simply adapting Eq. (18.39) to the corresponding coefficients and parameters.

In Fig. 18.11 the difference between $\eta(25^\circ\mathrm{C},\, G_M)$ and $\eta(T_M,\, G_M)$, which takes both irradiance G_M and module temperature T_M into account, is shown. The graph has been derived by keeping the wind speed constant at a value of 1 m/s and by determining T_M iterating the heat thermal model at each level of irradiance for a constant ambient temperature of 25°C. At low levels of irradiance no effects of the module temperature are observed. At higher level of incident light intensity the difference between the two curves becomes more pronounced. since only $\eta_{\mathrm{mpp}}(G_M,\, T_M)$ takes into account

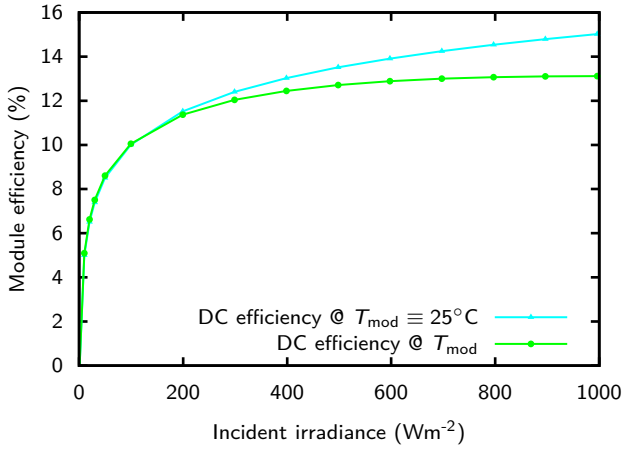

Figure 18.11: Comparison of $\eta(25°C, G_M)$ and $\eta(T_M, G_M)$ in dependence of the light irradiance at a wind speed of 1 m/s and ambient temperature of $T_a = 25°C$ for a c-Si solar cell.

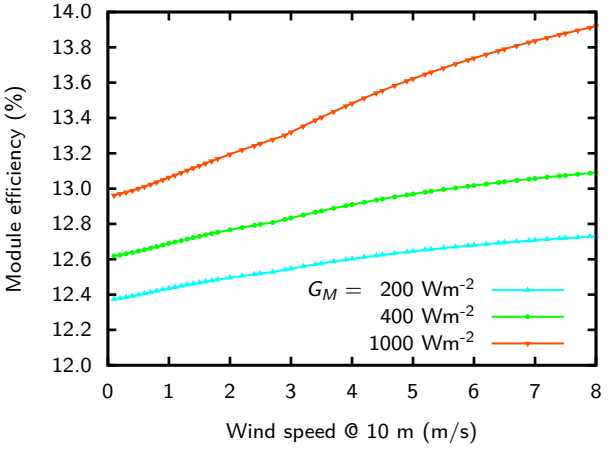

Figure 18.12: Overall efficiency of the module with increasing wind speed at various irradiance levels.

the marked reduction in efficiency resulting from the rising T_M.

Figure 18.12 shows the overall efficiency at various light intensities in dependence of the wind speed. Once more, the beneficial effect of turbulent motion in cooling the module is reflected by an increase of efficiency with the wind speed.

Finally, the power output of a module before the BOS is given by

$$P_{\mathrm{DC}} = \eta(G_M, T_M) \cdot G_M \cdot A_M, \tag{18.41}$$

and the power output at STC is given by

$$P_{\mathrm{STC}} = \eta(G_{\mathrm{STC}}, 25°\mathrm{C}) \cdot G_{\mathrm{STC}} \cdot A_M, \tag{18.42}$$

The *energy yield* of the DC side then is defined as

$$Y_{\mathrm{DC}} = \frac{P_{\mathrm{DC}}}{P_{\mathrm{STC}}} \cdot 100\% = \frac{\eta(G_M, T_M) \cdot G_M}{\eta(G_{\mathrm{STC}}, 25°\mathrm{C}) \cdot G_{\mathrm{STC}}} \cdot 100\%. \tag{18.43}$$

18.4 Designing grid-connected PV-system

Now we can use all the things learned in this and the other chapters for actually designing a PV system. For designing the PV system we use the *energy balance* approach, meaning that we design the system such that the generated energy and the consumed energy during one year match. Of course, there are also other ways of designing systems for example based on economic arguments.

For the energy balance we first need to calculate the annual load, which already happened in section 18.2. The energy yield at the DC side is given by

$$E^Y_{\mathrm{DC}} = A_{\mathrm{tot}} \int_{\mathrm{year}} G_M(t)\eta(t)\,\mathrm{d}t, \tag{18.44}$$

where A_{tot} is the total module area. It is related to the area of one module A_M via

$$A_{\mathrm{tot}} = N_T \cdot A_M, \tag{18.45}$$

where N_T is the number of modules. The energy balance now writes as

$$E^Y_{\mathrm{DC}} = E^Y_L \cdot \mathrm{SF}, \tag{18.46}$$

where SF is a *sizing factor* that usually is assumed to be 1.1. We therefore can calculate the required number of modules,

$$N_T = \left\lceil \frac{E^Y_L \cdot \mathrm{SF}}{A_M \cdot \int_{\mathrm{year}} G_M(t)\eta(t)\,\mathrm{d}t} \right\rceil, \tag{18.47}$$

where $\lceil x \rceil$ denotes the ceiling function, *i.e.* the lowest integer that is greater or equal than x.

Now it is important to decide how many modules are to be connected in *series* (N_S) and in *parallel* (N_P). Of course,

$$N_T = N_S \cdot N_P. \tag{18.48}$$

Such a PV array hence consists of P strings of S modules each. The N_T determined in Eq. (18.86) not necessarily needs to be an even number. For example, if $N_T = 11$, you might want to choose for $N_T = 12$ panels, because you can install them as $S \times P$= 12×1, 6×2, 4×3, 3×4, 2×6 or 1×12 strings. In principle, it is

preferable to connect as many modules as possible in series since then the currents on the DC side and hence the cable losses stay low. Further, thinner cables can be chosen. However, many modern inverters can connect two or even more individual strings, which can be important if the installation contains two or more areas with different shading, for example on two different sides of a roof. Further, it can be chosen to connect different strings to different inverters as well, such that the system uses several string inverters. N_S and hence the voltage at the inverter is also restricted by the chosen inverter typ — or, if seen the other way round, the inverter has to be chosen such that its operating voltage fits well to the string voltage.

In a conservative assumption, the power on the DC side at STC now is given as

$$P_{\mathrm{DC}}^{\mathrm{STC}} = N_T \times P_{\mathrm{MPP}}^{\mathrm{STC}}. \tag{18.49}$$

The inverter must be chosen such that its maximal power $P_{\mathrm{DC,\,max}}^{\mathrm{inv}}$ is above the maximal PV output,

$$P_{\mathrm{DC,\,max}}^{\mathrm{inv}} > P_{\mathrm{DC}}^{\mathrm{STC}}. \tag{18.50}$$

Further, the nominal DC power of the inverter should be approximately equal to the PV power,

$$P_{\mathrm{DC}_0} \approx P_{\mathrm{DC}}^{\mathrm{STC}}. \tag{18.51}$$

Usually, for $P_{\mathrm{DC}_0} < 5$ kWp, *single-phase* inverters are used while for $P_{\mathrm{DC}_0} > 5$ *three-phase* inverters are advised.

Also the inverter efficiency is dependent on the input power and voltage. A model discussing the inverter efficiency is presented below.

18.4.1 Inverter efficiency

As we already discussed in Chapter 17, modern *inverters* fulfil two major functions: First, *maximum power point tracking* (MPPT), and secondly, the actual inverter function, *ie* converting the incoming direct current (DC) to alternating current (AC) that can be fed into the electricity grid.

In order not to waste electricity produced by the PV array, an inverter should always work as close as possible at its maximum achievable efficiency. However, the inverter efficiency of an inverter strongly depends on the DC input voltage as well as the total DC input power as well as on the DC input voltage of the system. The inverter efficiency η_{inv} with respect to the input DC power at various DC voltage level is usually given at the data sheet, at least for some values. Often, only the *peak efficiency* is given as a single value.

Weighted Efficiencies

A more reliable way of expressing the inverter efficiency in a single number is to use *weighted efficiencies*,

which combine the inverter efficiencies over a wide range of solar resource regimes [110]. Two different weighted efficiencies are commonly used. First, the *European Efficiency*, which represents a low-insolation climate such as in central Europe, and the *California Energy Commission* (CEC) efficiency, which represents the PV system performance in high-insolation regions such as in the southwest of the United States [110]. They are given by

$$\eta_{\text{Euro}} = 0.03\,\eta_{\ 5\%} + 0.06\,\eta_{10\%} + 0.13\,\eta_{20\%} + 0.10\,\eta_{30\%} + 0.48\,\eta_{50\%} + 0.20\,\eta_{100\%}, \tag{18.52a}$$

$$\eta_{\text{CEC}} = 0.04\,\eta_{10\%} + 0.05\,\eta_{20\%} + 0.12\,\eta_{30\%} + 0.21\,\eta_{50\%} + 0.53\,\eta_{75\%} + 0.05\,\eta_{100\%}, \tag{18.52b}$$

where $\eta_{x\%}$ denotes the efficiency at x% of nominal power of the inverter. Note that the CEC efficiency contains a 75% value that is not present in the European efficiency.

Even though the weighted efficiencies represents a more accurate approximation of the effective yearly working performance of the inverter compared to the mere peak efficiency, it still only is an approximation of the average performance of a system in the European climate.

If a better estimate of the *real time energy yield* of an extended PV system is needed, a more accurate representation of the instantaneous inverter performance at every level of input power and voltage must be developed, for example the model discussed below.

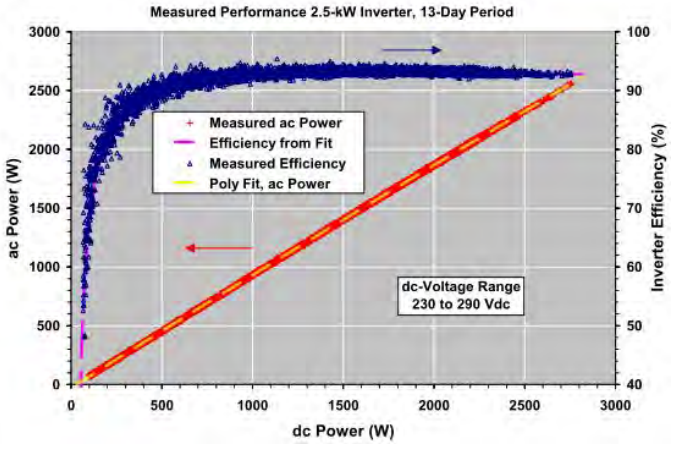

Figure 18.13: Field test results for a 2.5 kW Solectria PVI2500 inverter recorded during a period of 13 days for system operation at Sandia Laboratories [111].

Sandia National Laboratories (SNL) model

Due to a lack of detailed data from inverters manufacturers, many research institutes around the world have published extended data, which are publicly available online. These data present efficiency curves for a large

range of inverters as a function of a several characteristic parameters. One database that can be used is the one provided by the Sandia National Laboratories [112].

Figure 18.13 shows an example of an inverter efficiency curve CITE. For the graph, field test measurements were taken during a period of 13 days with changing weather condition. While the relationship between the AC and DC power appears to be linear at first, a closer look to the graph reveals that this is not entirely the case, as shown in Fig. 18.14. The power consumption of the inverter itself together with the electrical characteristics of the switching modes and circuits at different power levels results in a degree of non linearity between AC and DC power at a given DC voltage level. Assuming that the inverter efficiency is a constant value throughout the whole DC power range is equivalent to assuming a linear relationship between DC and AC power, which has been shown not to be the case.

The dependency of the inverter efficiency on the DC input voltage is very complex phenomenon that still lacks a full physical explanation [113]. The differences between the different inverter types can partially be explained with different types of switches used. Figure 18.15 shows the voltage dependent inverter efficiency for different inverter types. These curves were determined with the Sandia model that is described in the next paragraph.

In the following, we will introduce a mathematical

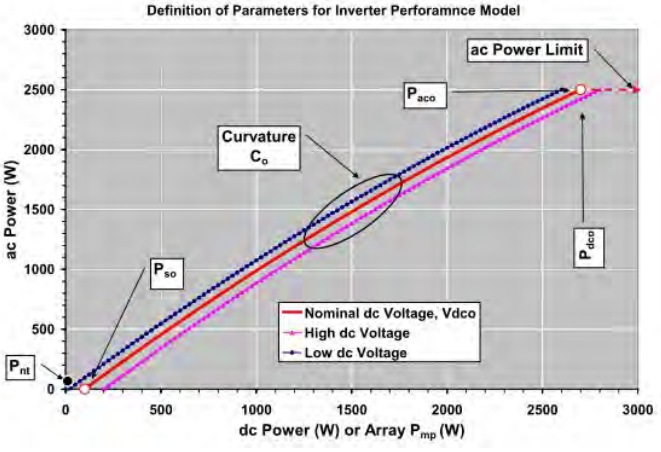

Figure 18.14: A closer look on to the relationship between AC output and DC input power for an inverter [111].

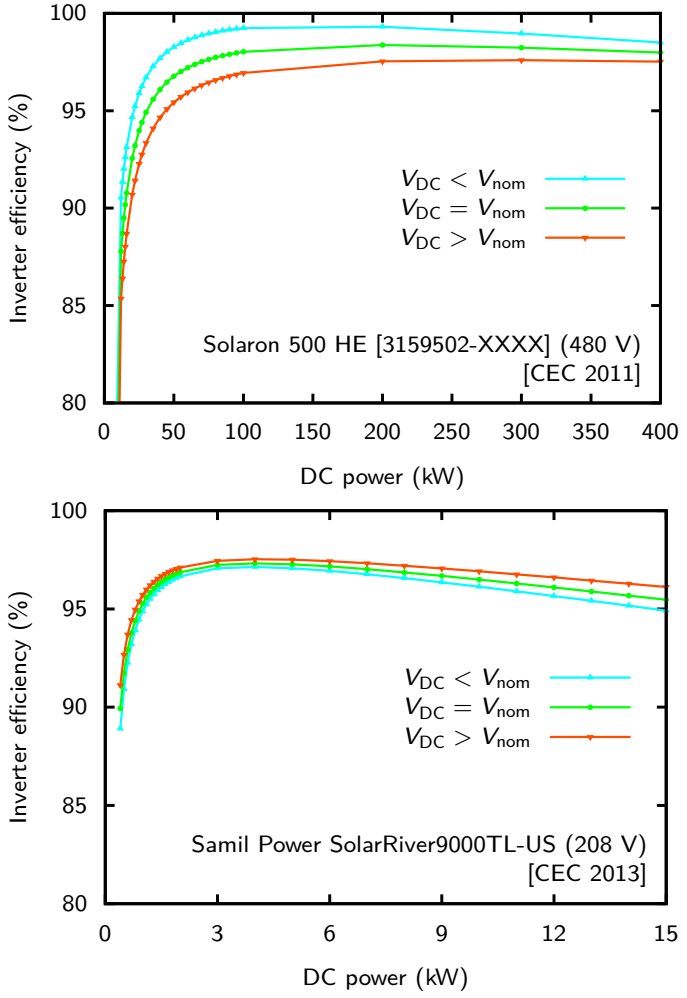

Figure 18.15: Variation in the inverter efficiency with DC input voltage for two different inverters.

model developed at Sandia Research Institute that has been chosen due to its accuracy of both the effect of DC input power and voltage on the AC power output. In this model the relationship between P_{AC} and P_{DC} is given by [111]

$$P_{\mathrm{AC}} = \left[\frac{P_{\mathrm{AC}_0}}{A-B} - C(A-B)\right] \cdot (P_{\mathrm{DC}} - B) - C(P_{\mathrm{DC}} - B)^2, \tag{18.53}$$

where the coefficients A, B, and C are given by

$$A = P_{DC_0}\left[1 + C_1\left(V_{\mathrm{DC}} - V_{\mathrm{DC}_0}\right)\right], \tag{18.54a}$$

$$B = P_{S_0}\left[1 + C_2\left(V_{\mathrm{DC}} - V_{\mathrm{DC}_0}\right)\right], \tag{18.54b}$$

$$C = C_0\left[1 + C_3\left(V_{\mathrm{DC}} - V_{\mathrm{DC}_0}\right)\right]. \tag{18.54c}$$

Several of the parameters are depicted in Fig. 18.14. The parameters are [111]:

P_{AC}: AC-power output from inverter based on input power and voltage, (W).

P_{DC}: DC-power input to inverter, typically assumed to be equal to the PV array maximum power, (W).

V_{DC}: DC-voltage input, typically assumed to be equal to the PV array maximum power voltage, (V).

P_{AC_0}: Maximum AC-power "rating" for inverter at reference or nominal operating condition, assumed to be an upper limit value, (W).

P_{DC_0}: DC-power level at which the ac-power rating is achieved at the reference operating condition, (W).

V_{DC_0}: DC-voltage level at which the ac-power rating is achieved at the reference operating condition, (V).

P_{S_0}: DC-power required to start the inversion process, or self-consumption by inverter, strongly influences inverter efficiency at low power levels, (W).

C_0: Parameter defining the curvature (parabolic) of the relationship between AC-power and DC-power at the reference operating condition, default value of zero gives a linear relationship, (1/W).

C_i: Empirical coefficient allowing P_{DC_0} to vary linearly with DC-voltage input, default value is zero, ($i = 1, 2, 3$, 1/V).

This model takes the following losses into account [113]:

- Self consumption of the inverter. This value corresponds to the DC power required to start the inversion process.
- Losses proportional to the output power due to fixed voltage drops in semiconductors and switching losses.
- Ohmic losses

The accuracy of the model depends on the data available for determining the performance parameters. An initial estimate can be performed using the little information provided by the manufacturer. Using all the required parameters will provide a model with an error of approximately 0.1% between the modeled and measured inverter efficiency [113].

All the parameters required for handling Eq. (18.53) are given in the Sandia laboratory database where a full list of a wide range of inverters with nominal powers from 200 W up to 1 MW are covered [112]. This model therefore uses the instantaneous value of P_{DC} and V_{DC} produced by the entire PV array to evaluate the AC power output and inverter efficiency.

Maximum Power Point Tracker and additional losses

The efficiency of the MPPT has not been included explicitely in the Sandia performance model. This is because the efficiency of most inverters ranges between 98% and nearly 100% at every level of input power and the voltage provided is within the accepted minimum and maximum window for the MPPT to function correctly. A decrease of 1% in the system performance has therefore been used to take the MPPT losses into account.

An an additional decrease of 3% in the system performance can be considered to cover losses caused by mismatch between modules (-1.5%), ohmic cable losses (-0.5%) and soiling (-1%) [109], if the cable losses are not determined via the method described in section 17.5.

18.4.2 Performance Analysis

Now we will put all things together that we discussed earlier in this section. The *instantaneous power output* on the AC side can be described with

$$P_{\mathrm{AC}}(t) = A_M\, G_M(t)\, \eta_M(t)\, \eta_{\mathrm{inverter}}(t)\, \eta_{\mathrm{MPPT}}(t)\, \eta_{\mathrm{other}}. \tag{18.55}$$

The system efficiency then is given as

$$\eta_{\mathrm{system}}(t) = \frac{P_{\mathrm{AC}}(t)}{A_{\mathrm{tot}}\, G_M(t)} \cdot 100\%, \tag{18.56}$$

which leads us to the *instantaneous AC-side yield* (also known as *performance ratio*),

$$Y_{\mathrm{AC}}(t) = \frac{P_{\mathrm{AC}}(t)}{P_{\mathrm{STC}}} \cdot 100\%. \tag{18.57}$$

Then, the *yearly energy yield at the AC side* can be calculated with

$$E^Y_{\mathrm{AC}} = \int_{\mathrm{year}} P_{\mathrm{AC}}(t)\, \mathrm{d}t; \tag{18.58}$$

it is given in Wh/year. Another important parameter is the *annual efficiency* of the system

$$\eta^Y_{\mathrm{system}} = \frac{E^Y_{\mathrm{AC}}}{E^Y_{i,\,\mathrm{sys}}} \cdot 100\%, \tag{18.59}$$

where $E^Y_{i,\,\mathrm{sys}}$ is the solar energy incident on the PV system throughout the year. It can be calculated with

$$E^Y_{i,\,\mathrm{sys}} = A_{\mathrm{tot}} \int_{\mathrm{year}} G_M(t)\, \mathrm{d}t. \tag{18.60}$$

The last parameter we look at is the *yearly electricity yield*

$$Y_E = \frac{E^Y_{\mathrm{AC}}}{N_S N_P \cdot P_{\mathrm{STC}}}, \tag{18.61}$$

which is given by Wh/(year kWp).

At the end of the design phase it is important to check whether the system really fulfils the requirements. If the yearly energy yield exceeds the annual load, the system is well designed. Otherwise, another iteration has to be done in order to scale up the system. However, as stated earlier, for a grid-connected system it also can be a choice not the require the whole load to by covered by PV electricity.

18.5 Designing off-grid PV systems

In this last section of this chapter on PV system design we take a closer look on the design of off-grid PV systems (also called stand-alone systems). Choosing a good design is more critical for off-grid systems than for grid-connected systems. The reason for this is that

off-grid systems cannot fall back on the electricity grid, which increases the requirements on the reliability of the off-grid system.

A major component of off-grid systems is the *storage component*, which can store energy in times when the PV modules generate more electricity than required and it can deliver energy to the electric appliances when the electricity generated by the PV modules in not sufficient. A major design parameter for off-grid systems is the required number of autonomous days, *i.e.* the number of days a fully charged storage must be able to deliver energy to the system until discharged.

The sizing of both the PV array and the storage component, usually a *battery bank*, is interconnected. Here, two parameters arise: first, E_{fail}, which is the energy required by the electric load that cannot be delivered by the PV system, for example if the batteries are emptied after several cloudy days. Secondly, E_{dump}, which is the energy produced by the PV array that neither is used for driving a load nor is stored in the battery, for example, if the batteries already are full after a number of sunny days. Now, we can define the *Loss of Load Probability*,

$$\text{LLP} = \frac{E_{\text{fail}}}{\int_{\text{year}} P_L(t)\,\mathrm{d}t}. \tag{18.62}$$

Of course, the lower LLP, the more stable and reliable the PV system. Table 18.3 shows recommended LLPs for different applications.

Table 18.3: Recommended Loss of Load Probability (LLP) for some Applications

Application	Recommended LLP
Domestic illumination	10^{-2}
Appliances	10^{-1}
Telecommunications	10^{-4}

Figure 18.16 (a) shows a schematic of an off-grid system with all the required components, *i.e.* the PV array, a maximum power point tracker (MPPT), a charge controller (CC), a discharge controller (DC), a battery bank (BB), an inverter, and the load. The CC prevents the BB from being overcharged by the PV system, while the DC prevents the battery to be discharged below the minimal allowed SoC. In the simplest case, they (dis)connect the PV system and load from the battery by switches. For optimal charging, however, nowadays often CC with pulse-width modulation (PWM) are used. By far not all off-grid systems contain an MPPT, thus it is represented with a dashed line. However, if an MPPT is present it usually is delivered in one unit together with the CC. Usually, the DC and inverter are combined in one *battery inverter unit*. Especially at larger systems the inverter currents may become very large. For example, if a 2400 W load is present for a short time, this means 100 A on the DC side in a 24 V system. Therefore the battery inverter usually is directly connected to the BB with thick cables. In very small

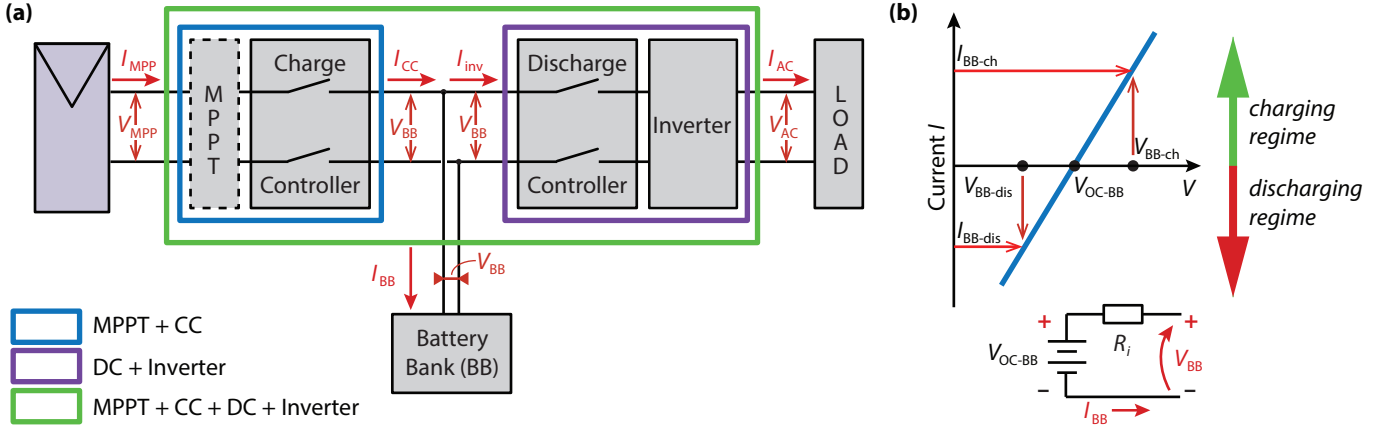

Figure 18.16: (a) PV topology of an off-grid system. (b) Simplified *I*-*V* curve and schematic of the battery bank.

systems with maximal powers of several hundreds of Watt, the CC and DC may be combined in one unit.

18.5.1 A closer look at the battery

The battery bank is the workhorse of any off-grid system, because it is its stable power source. It is thus very important to understand how the battery bank will act in the PV system. Therefore, we will discuss it in more detail in this section.

When we want to understand how a battery works we have to consider *the net effect of all the forces that try to charge or discharge the battery*. In Fig. 18.16 (a) the most important currents and voltages that we need and the following discussion are depicted. In Fig. 18.16 (b) a schematic and an *I-V* curve of an idealised battery bank is shown. As we can see, this battery bank is considered to consist of an voltage source with a total voltage $V_{\text{OC-BB}}$ and an internal resistance R_i. In reality, $V_{\text{OC-BB}}$ will not be constant, but a function of the state of charge, the ambient temperature, and others,

$$V_{\text{OC-BB}} = f(\text{SoC}, T, \ldots). \tag{18.63}$$

In the following discussion, however, we will neglect this dependence and assume $V_{\text{OC-BB}}$ to be constant. Because of R_i, the voltage V_{BB} will differ from $V_{\text{OC-BB}}$ and also is dependent on the current I_{BB} flowing through the battery, as seen by the equation

$$I_{\text{BB}} = \frac{1}{R_i}\left(V_{\text{BB}} - V_{\text{OC-BB}}\right). \tag{18.64}$$

Note that we use the convention, where I_{BB} is positive when the battery is charged and negative when it is discharged. Let us now derive an expression for the voltage of the battery bank (V_{BB}) as a function of the other PV system parameters. We start with the power on the left and right-hand sides of the MPPT,

$$\eta_{\text{MPP}} I_{\text{MPP}} V_{\text{MPP}} = I_{\text{CC}} V_{\text{BB}} =: \beta, \tag{18.65}$$

and hence

$$I_{\text{CC}} = \frac{\eta_{\text{MPP}} I_{\text{MPP}} V_{\text{MPP}}}{V_{\text{BB}}} = \frac{\beta}{V_{\text{BB}}}. \tag{18.66}$$

Consequently, $\beta = 0$ means that the PV system is not active, for example during night.

In a similar manner, we look at the power on the left and right hand side of the inverter:

$$P_L = I_{\text{AC}} V_{\text{AC}} = \eta_{\text{inv}} I_{\text{inv}} V_{\text{BB}} =: \eta_{\text{Inv}} \alpha, \tag{18.67}$$

and consequently

$$I_{\text{inv}} = \frac{P_L}{\eta_{\text{inv}} V_{\text{BB}}} = \frac{\alpha}{V_{\text{BB}}}. \tag{18.68}$$

Combining Eqs. (18.66) and (18.68), we find

$$I_{\text{BB}} = I_{\text{CC}} - I_{\text{inv}} = \frac{\beta - \alpha}{V_{\text{BB}}}. \tag{18.69}$$

Clearly, $\alpha = 0$ indicates that no load is present.

Combining Eqs. (18.64) and (18.69), multiplying with $R_i V_{BB}$ and rearranging leads to a quadratic equation,

$$V_{BB}^2 - V_{OC\text{-}BB} V_{BB} - R_i(\beta - \alpha) = 0. \tag{18.70}$$

with the solutions

$$V_{BB}^{\pm} = \frac{V_{OC\text{-}BB}}{2} \pm \sqrt{\left(\frac{V_{OC\text{-}BB}}{2}\right)^2 + R_i(\beta - \alpha)}. \tag{18.71}$$

The correct solution is the "+" solution as we can check with

$$V_{BB}^{+}(\beta - \alpha = 0) = V_{OC\text{-}BB}. \tag{18.72}$$

Hence, the final solution is

$$V_{BB} = \frac{V_{OC\text{-}BB}}{2} + \sqrt{\left(\frac{V_{OC\text{-}BB}}{2}\right)^2 + R_i(\beta - \alpha)}. \tag{18.73}$$

Let us now take a short look on this solution. If the battery is charged, I_{CC} is higher than I_{inv} and hence $(\beta - \alpha)$ and I_{BB} are positive. From Eq. (18.73) it follows that in this case V_{BB} is higher than $V_{OC\text{-}BB}$. On the other hand, if the battery is discharged, I_{CC} is lower than I_{inv} and hence $(\beta - \alpha)$ and I_{BB} are negative. Therefore, V_{BB} will be lower than $V_{OC\text{-}BB}$.

Note that only the net current $I_{BB} = I_{CC}$ flows in or out of the battery. Therefore, only this current determines the power loss in the battery,

$$P_{BB}(\text{loss}) = I_{BB}^2 R_i. \tag{18.74}$$

This power always is lost, irrespective of the sign of I_{BB}.

Table 18.4: Recommended number of autonomous days d_A at several latitudes.

Latitude (°)	Recommended d_A
0-30	5-6
20-50	10-12
50-60	15

18.5.2 Designing a system with energy balance

Now we discuss how to design a PV system based on the principle of *energy balance*. For determining the adequate components, the analysis for the load side and the PV side is performed separately. Let us begin with the *load side*. We can determine the annual load E_L^Y as discussed in section 18.2 and Eq. (18.1),

$$E_L^Y = \int_{\text{year}} P_L(t)\,\mathrm{d}t. \tag{18.75}$$

From that we can estimate the *average daily load* with

$$E_L^D = \frac{1}{365} E_L^Y. \tag{18.76}$$

Next, we have to choose an adequate number of days of autonomy d_A. Some values given in table 18.4. Now, we can calculate the required energy of the battery bank,

$$E_{\mathrm{BB}} = d_A \frac{E_L^D \cdot \mathrm{SF_{bat}}}{\mathrm{DoD_{max}}}, \tag{18.77}$$

where $\mathrm{SF_{bat}}$ is the sizing factor of the battery, which is similar to the sizing factor for the PV array already defined in section 18.4. $\mathrm{DoD_{max}}$ is the maximally allowed depth of discharge of the batteries. The rated energy of the chosen batteries is

$$E_{\mathrm{bat}} = V_{\text{OC-bat}} C_{\mathrm{bat}}, \tag{18.78}$$

where C_{bat} is the battery capacity (unit Ampere-hours). Hence, the required number of batteries is

$$N_{\mathrm{bat}} = \left\lceil \frac{E_{\mathrm{BB}}}{E_{\mathrm{bat}}} \right\rceil. \tag{18.79}$$

Similarly as for grid-connected systems we now need to choose a suitable inverter. Therefore we must consider the maximal load power $P_L^{\max}$. This we can do, for example, by looking at the appliance with the maximal power consumption, or by adding up the power of all appliances. It may be more beneficial to choose a system design, such that not all appliances can be used at the same time. The final decision of course is up to the designer of the system. Similar as for grid-connected systems, the inverter must fulfil several requirements: First, its maximally allowed power output must exceed the maximal power required by the appliances,

$$P_{\mathrm{DC,max}}^{\mathrm{inv}} > P_L^{\max}. \tag{18.80}$$

Secondly, the nominal power of the inverter should be approximately equal to the maximal load power,

$$P_{\mathrm{DC},0} \approx P_L^{\max}. \tag{18.81}$$

Thirdly, the nominal inverter input voltage should be approximately equal to the nominal voltage of the battery back,

$$V_{\mathrm{DC,inv}} \approx V_{\text{OC-BB}}. \tag{18.82}$$

For a more detailed analysis of the inverter performance the Sandia model can be used that was discussed in section 18.4.

Typical voltages for the battery bank are 12 V, 24 V, 48 V or 96 V. It can be adjusted by the number of batteries that are connected in series,

$$N_{\mathrm{bat}}^S = \frac{V_{\text{OC-BB}}}{V_{\text{OC-bat}}}. \tag{18.83}$$

From that we also can determine the number of batteries that must be connected in parallel,

$$N_{\text{bat}}^{P} = \left\lceil \frac{N_{\text{bat}}}{N_{\text{bat}}^{S}} \right\rceil . \tag{18.84}$$

Now, after designing the load side is completed with choosing inverter and batteries, we look at the PV side of the system. Sizing the PV array is very similar to the procedure used for grid-connected systems in section 18.4. The energy balance can be written down as The energy balance now writes as

$$E_{\text{DC}}^{Y} = E_{L}^{Y} \cdot \text{SF}, \tag{18.85}$$

where SF is a *sizing factor* that usually is assumed to be 1.1. We therefore can calculate the required number of modules,

$$N_T = \left\lceil \frac{E_L^Y \cdot \text{SF}}{A_M \cdot \int_{\text{year}} G_M(t)\eta(t)\,\mathrm{d}t} \right\rceil . \tag{18.86}$$

For minimising losses, the MPP voltage of the PV array and the nominal voltages of the inverter and the battery pack should be approximately equal, since otherwise the losses of the DC-DC converter that is included in the MPPT-CC unit will be higher. The number of PV modules that are connected in series in the PV array is given by

$$N_S = \left\lfloor \frac{V_{\text{OC-BB}}}{\overline{V}_{\text{mod-MPP}}} \right\rfloor , \tag{18.87}$$

where $\overline{V}_{\text{MPP-mod}}$ denotes the annual average of the MPP voltage of the PV modules. Of course, the maximally allowed input voltage of the MPPT-CC unit must not be exceeded by the PV array,

$$V_{\text{MPP}} \geq N_S \cdot V_{\text{mod-MPP}}^{\max}. \tag{18.88}$$

The number of required parallel PV strings is given by

$$N_P = \left\lceil \frac{N_T}{N_S} \right\rceil . \tag{18.89}$$

18.5.3 Performance Analysis

Like for grid-connected systems, also for off-grid systems, a performance analysis should be done to evaluate the chosen design. For this analysis, it is useful to use an algorithm that can simulate the performance of the PV system throughout the year. Conceptually, this algorithm can look as follows

1. Set a starting state of charge (SoC) of the battery bank.
2. Then calculate the SoC of the battery throughout the year for time steps Δt,
   - Depending on the actual load and PV array output, determine the battery current $I_{\text{BB}}(t)$.
   - Determine the actualised SoC.

Of course, the function of the charge controller, *i.e.* its switching behaviour, must be accurately mimicked by the algorithm.

3. Determine when the system cannot deliver the required load and hence E_{fail}.
4. Now, calculate the Loss of Load Probability of the system (LLP).
5. Finally, determine the yearly energy yield on the AC-side of the system, E_{AC}^{Y}.

Mathematically, the annual energy yield on the AC side can be expressed by

$$E_{\text{AC}}^{Y} = \int_{\text{year}} P_L(t)\,\mathrm{d}t - E_{\text{fail}} + \Delta E_{\text{BB}}. \tag{18.90}$$

In contrast to the expression used for determining the AC yield for grid-connected systems in Eq. (18.58), this equation contains two additional components: E_{fail}, the energy that is required by the load but cannot be delivered and ΔE_{BB} is the difference in energy stored in the battery bank between the beginning and end of the year,

$$\Delta E_{\text{BB}} = \left(\text{SoC}_{\text{end}} - \text{SoC}_{\text{beginning}}\right) C_{\text{bat}} V_{\text{BB}}. \tag{18.91}$$

If $\Delta E_{\text{BB}} < 0$, the system might not be sustainable. On the other hand, if it is > 0, E_{dump} will increase in the following, if the average meteorological conditions are unchanged.

Because of $E_{\text{fail}} > 0$,

$$E_{\text{AC}}^{Y} < \int_{\text{year}} P_L(t)\,\mathrm{d}t. \tag{18.92}$$

As already stated earlier, the Loss of Load Probability is given by

$$\text{LLP} = \frac{E_{\text{fail}}}{\int_{\text{year}} P_L(t)\,\mathrm{d}t}. \tag{18.93}$$

For evaluating the design, it is very important to look at LLP:

- LLP acceptable:
  - E_{dump} low: The system design is OK.
  - E_{dump} high: Resize the PV array.
- LLP not acceptable:
  - Increase the size of the PV array.
  - Increase the capacity of the battery bank.

19 PV System Economics and Ecology

19.1 PV System Economy

We conclude our discussions on PV systems with looking on several important topics on the economics of PV systems. Note that the economics of PV can be discussed at several levels, such as the consumer level, the manufacturing level, the level of PV installers, and the technology level where PV is compared to other electricity generation technologies on the scale of the electricity grid.

We will start this discussion with the definition of the *payback time*, which in finance is defined as the amount of time required to recover the cost of an investment. It can be calculated with

$$\text{payback time} = \frac{\text{initial investment}}{\text{annual return}}. \tag{19.1}$$

Translated to the consumer level, the payback time is the time it takes to recover the initial investment of the PV system as the system continuously reduces the electricity bill. Please note that the *financial* payback time is different from the *energy payback time* that we will discuss in Section 19.2 Let us look at an example:

Example

Let us assume that family Smith have installed a PV system with a power of 1 kW_p on their rooftop. The initial investment

was €8000. Family Smith has an annual electricity bill of €2000. The installation of the PV system leads to an average annual reduction of the electricity bill of €800. As a part of their consumed electricity is provided by their PV system, the electricity bill is therefore constantly reduced. Hence, the average annual return on their PV system is €800. As a consequence, the Smith's have earned the final investment back after 5 years, the payback time *hence is 5 years.*

The payback time is strongly influenced by the annual solar radiation on the PV system. As we have seen in Chapter 16, this is dependent on the orientation of the PV modules and on the location of the PV systems. In general we can say that the sunnier the location, the greater the PV yield and the shorter the payback time. Another factor that influences the payback time is the grid electricity costs: the higher these costs, the shorter the payback time. Finally, the payback time also is strongly dependent on the initial costs of the PV system.

In practice, often more factors must be taken into account than in the simple example above. This will increase the complexity of calculating the payback time. For instance, if we are considering a significant period of time, also the change of the value of money has to be taken into account, which is due to inflation. For example today €1000 will have a different *purchasing power* than in ten years time. Another factor that should be considered are policies regarding renewable energy. For example, subsidies and feed-in tariffs can affect the initial investments and savings.

Let us briefly discuss the concept of *feed-in tariffs*. At the consumer level, the feed-in tariff is the price at which a consumer can sell renewable electricity to the electricity provider. We distinguish between two kinds of feed-in tariffs, gross and net. *Gross* feed-in tariffs are paid for all the electricity the panels produce, irrespective of the consumer's electricity consumption. In contrast, *net* feed-in tariffs promise a higher rate for the surplus electricity fed into the grid after domestic use of the consumers is subtracted. For implementing feed-in tariff schemes at consumer level, the facility of *net metering is pivotal*.

Another very important concept is the *Levelized Cost of Electricity* (LCoE), which is defined as the cost per kWh of electricity produced by a power generation facility. It is usually used to compare the lifetime costs of different electricity generation technologies. To be able to estimate the effective price per kWH, the concept of LCoE allocates the costs of an energy plant across its full lifecycle. It is somehow similar to averaging the upfront costs of production over a long period of time. Depending on the number of variables that are to be taken into account, calculating the LCoE can become very com-

plex. In a simple case the LCoE can be determined with

$$\mathrm{LCoE} = \frac{\sum_{t=1}^{n} \frac{I_t + M_t + F_t}{(1+r)^t}}{\sum_{t=1}^{n} \frac{E_t}{(1+r)^t}}. \qquad (19.2)$$

The sums expand across the whole lifetime of the system n, every year accounts for one summand. I_t are the investment expenditures in the year t, M_t are the operational and maintenance expenditures in the year t, and F_t are the fuel expenditures in the year t. Of course, for PV $F_t \equiv 0$. Further, E_t is the electricity yield in the year t. Finally, r is the discount rate which is a factor used to discount future costs and translating them into the present value.

Figure 19.1 shows the LCoE for different methods of electricity generation. We see that the LCoE of wind energy is lowest, it is even below the LCoE of fossil fuel and nuclear based electricity. The LCoE of PV generated electricity is still slightly above that of the non-renewable technologies.

Depending on the location of a PC system and the initial investment required for the PV system, the LCoE for PV can vary a lot between different projects. Additionally, the discount rate r used for the calculation will strongly influence the LCoE.

For the electricity supplier, the LCoE is a valuable indicator of the cost competitiveness of a certain energy technology. It is also a good indicator for determining

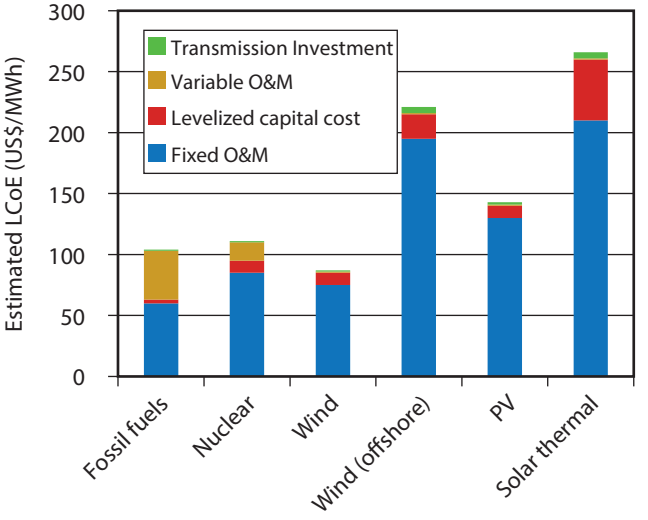

Figure 19.1: The levelised cost of electricity for different electricity generation technologies.

the electricity price: for making profit this price must be above the LCoE. Surely, the supplier cannot determine the electricity price independently, as it is strongly influenced by policy factors such as feed-in tariffs, subsidies and other incentives.

The final concept that we discuss is that of *grid parity*. For solar electricity this is the situation at which it can be generated at an LCoE that is equal to the electricity price from the conventional electric grid. Of course, if the PV electricity price is below the grid price, the situation becomes even better,

$$\mathrm{LCOE}_{\mathrm{PV}} \leq \mathrm{LCoE}_{\mathrm{conventional}}. \tag{19.3}$$

In principle, the concept of grid parity can be generalised to the other renewable technologies as well. However, there is one significant difference between PV and other renewable technologies such as wind and hydro electricity. Wind and hydro electricity installations usually only can be financed by companies but are no option for a single consumer. In contrast, PV can be scaled down to the level of a single module, such that a house owner can become an electricity producer with his small scalable PV installation on his roof. As a consequence, the consumer offering PV electricity to the grid now is effectively competing with the retail grid price, which conventionally includes other costs like transmission, distribution, and so on. As for the consumer these costs are no part of the PV electricity price, PV grid parity for retail grid prices can be reached faster, as shown in Fig. 19.2.

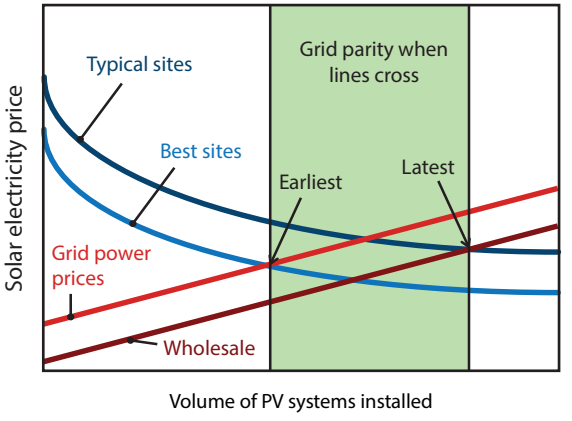

Figure 19.2: Grid parity for PV systems [114].

The graph is showing the volume of installed PV systems versus the price of PV electricity. The installed volume can be directly correlated with time, as the past decade has seen the implemented PV volume rise tremendously. As capital costs decline with increasing volumes, the price of PV generated electricity is expected to decrease in the future. On the other hand, the price of fossil fuels is expected to rice because of increasing scarcity and cost linked to the right to emit *CO2* emissions into the atmosphere costs. These trends

will lead to increasing prices for electricity generated with combusting fossil fuels. Grid parity is reached when the prices of PV electricity and grid electricity are crossing. Naturally, with incentives and subsidies grid parity can be reached even faster, however grid parity reached with public financial support often is disregarded. Therefore it is said that true grid parity is reached when the PV electricity prices fall below the grid prices without any subsidies. Real grid parity for example already is reached in the sunny country of Spain.

To conclude, grid parity is a very useful concept to indicate the feasibility of a renewable energy technology. The closer a technology is to grid parity, the easier it can be integrated in the electricity mix. With the advancements in technology and the maturity of manufacturing processes, grid parity for solar is expected to be reached at many locations around the World in the next years.

19.2 PV System Ecology

Besides discussing the economics of PV systems, it also is very important to consider their ecological and environmental aspects. The main reason for that is that the aim of photovoltaics is to generate electricity without any considerable effect on the environment. It is therefore very important to check the ecological aspects of the different PV technologies. In this section we will discuss different concepts to quantify the environmental impact of PV systems.

The concept of the *carbon footprint* estimates the emissions of CO_2 caused by manufacturing the PV modules and compares them with the reduction of CO_2 emissions due to the electricity generated with PV instead of combusting fossile fuels. A more analytical approach is to look at the total energy required to produce either the PV modules or all the components of a PV system. As the different PV technologies vary considerably in the required production processes, the energy consumption for producing 1 kW_p varies considerably between the different technologies. If a complete *life cycle assessment* (LCA)is performed, it is tried to trace the energy and carbon footprints of the PV panels throughout their lifetime. Therefore LCA also is known as *cradle-to-grave analysis*.

We now are going to introduce several indicators that are used to judge the different ecological aspects. The *Energy Yield Ratio* is defined as the ratio of the total energy yield of a PV module or system throughout its lifetime with all the energy that has to be invested in the PV system in that time. This invested energy not only contains the energy for producing the components, transporting them to the location and installing them but also the energy that is required to recycle the differ-

ent components at the end of their lifecycle.

As the energy required for producing a PV system depends strongly on the PV technology and also on the quality of the panels, the energy yield ratio for the different technologies varies a lot. While the energy yield ratio for PV modules can be as large as 10 to 15, PV systems usually have a lower ratio because of the energy invested in the components other than the modules.

A very important concept is the *Energy Payback Time*, which is defined as the total required energy investment over the lifetime divided by the average annual energy yield of the system,

$$\text{Energy Payback Time} = \frac{\text{totally invested energy}}{\text{average annual energy yield}}. \tag{19.4}$$

Note that the energy payback time is different from the economic payback time introduced in Section 19.1.

The energy payback time of typical PV systems is between 1 and 7 years and it also depends on location issues such as the orientation of the PV array as well as the solar irradiance throughout the year.

Figure 19.3 shows the specific energy required for producing PV modules with different technologies. As we can see, the differences between the technologies are large. The specific energy required for producing thin-film modules from materials such as amorphous silicon,

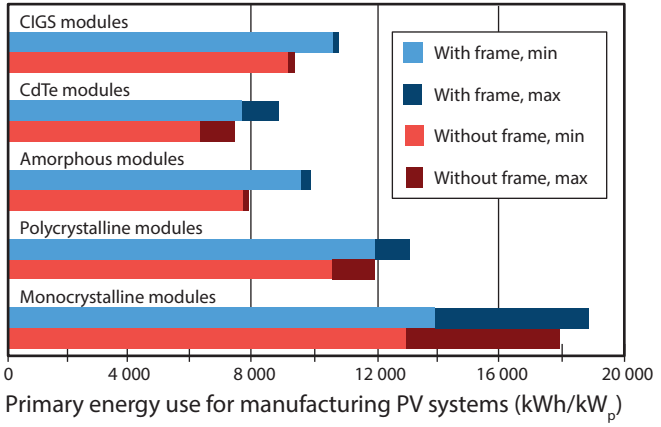

Figure 19.3: Specific energy used to produce PV modules of different technologies (Data from [115]).

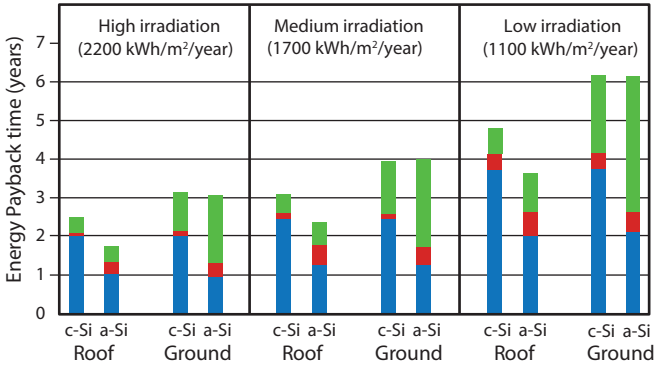

Figure 19.4: The energy payback time for the different components constituting PV systems (Data from [114]).

cadmium telluride and CIGS is significantly below that of modules made from polycrystalline and monocrystalline silicon, where the specific energy can reach values up to 12000-18000 kWh/kW$_p$. Because of further improvements in the module efficiency and the manufacturing process we however may expect that that the specific energy follows a decreasing trend.

For a PV system it is more difficult to allocate the energy that was used for its production, as all the components constituting the balance of system have to be taken into account as well. For example, for components like batteries and inverters the technologies and manufacturing processes may vary a lot between the different products available on the market. Nonetheless, studies were carried out that estimated the energy required by whole PV systems. Generally, as we can see in Fig. 19.4 it is found that the energy required for the BOS is significantly below that used for manufacturing the modules. In the figure, amorphous and crystalline silicon modules are compared. As expected, we see that the energy payback time in regions with high solar irradiance have significantly shorter energy payback times than regions with low irradiance. While a-Si:H based modules have a shorter energy payback time than c-Si based modules, the energy payback time for the module frame and the BOS of a-Si:H based systems can be significantly higher than that of c-Si based systems. This can be explained with the lower efficiency of a-Si:H that increases the required framing material per W$_p$.

The irradiance strongly influences the energy payback time and varies between two years (high irradiance) and six years (low irradiance). Roof-mounted systems always have a shorter energy payback time than systems mounted on the ground, mainly because of the BOS that is more energy extensive for ground-mounted systems.

No matter which PV technology is chosen, the energy payback time always is far below the expected system lifetime, which usually is between 25 and 30 years. For the PV systems discussed in Fig. 19.4, the energy yield

ratio is between 4 and 10. Hence, the energy invested in the PV system is paid back several times throughout the life cycle of the PV system. The urban legend that PV modules require more energy to be produced than they will ever produced thus is not backed by any data. In contrast, the net energy produced is much larger than the energy required for PV production.

However, a lot of work still needs to be done and can be done. Some studies indicate that the energy required for producing PV modules can be reduced by up to 80%. Further, as the amount of installed PV systems becomes larger and larger, recycling of the components at the end of the lifecycle becomes very important. For example, the European Union introduced already several directives that induced recycling schemes for c-Si based PV modules.

The last environmental issue that we want to mention is pollution caused by the production of PV modules. As many sometimes toxic chemicals are required for producing PV modules, this can be a serious threat to the environment. Therefore it is very important to have strong legislation in order to prevent pollution of the surroundings of PV factories. Especially in countries with weak environmental legislation pollution can be a severe problem that affects the environment and people living in the surroundings negatively.

Part V

Alternative Methods of Solar Energy Utilisation

20 Solar Thermal Energy

Solar thermal energy is an application of solar energy that is very different from photovoltaics. In difference to photovoltaics, where we used electrodynamics and solid state physics for explaining the underlying principles, solar thermal energy is mainly based on the laws of thermodynamics. In this chapter we give a brief introduction to that field. After introducing some basics in Section 20.1, we will discuss Solar Thermal Heating in Section 20.2 and Concentrated Solar (electric) Power (CSP) in Section 20.3.

20.1 Solar thermal basics

We start this section with the definition of *heat*, which sometimes also is called *thermal energy*. The molecules of a body with a temperature different from 0 K exhibit a disordered movement. The kinetic energy of this movement is called *heat*. The average of this kinetic energy is related linearly to the temperature of the body.[1] Usually, we denote heat with the symbol Q. As it is a form of energy, its unit is Joule (J).

If two bodies with different temperatures are brought together, heat will flow from the hotter to the cooler

[1] The interested reader will find more details in every textbook on physics or thermodynamics. This definition is taken from Ref. [116].

body and as a result the cooler body will be heated. Dependent on its physical properties and temperature, this heat can absorbed in the cooler body in two forms, sensible heat and latent heat.

Sensible heat is that form of heat that results in changes in temperature. It is given as

$$Q = mC_p(T_2 - T_1), \tag{20.1}$$

where Q is the amount of heat that is absorbed by the body, m is its mass, C_p is its *heat capacity* and $(T_2 - T_1)$ is the temperature difference. On the other hand, if a body absorbs or releases *latent heat*, the temperature stays constant but the phase changes. This happens for example when ice is melting: When its temperature is equivalent to its melting point, heat that is absorbed by the ice will not result in increasing temperature but in transformation into from the solid to the liquid phase, which is water. Mathematically, this is expressed as

$$Q = mL, \tag{20.2}$$

where L is the *specific latent heat*.

The two forms of heat are illustrated in Fig. 20.1, which shows what happens when a body absorbs heat. In the beginning, the body is solid and has temperature T_1. It then heats up and the heat is stored as solid sensible heat. When its melting point T^* is achieved, its temperature will not increase any more but the phase will change from solid to liquid. After everything is molten,

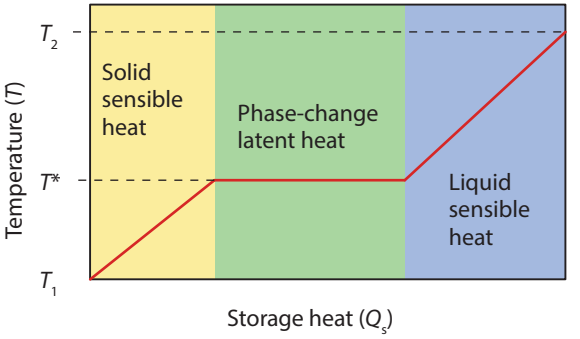

Figure 20.1: Illustrating the difference of sensible and latent heat.

the liquid will heat up further, the heat now is stored as liquid sensible heat.

Now we will take a look at the three basic mechanisms of heat transfer: conduction, convection and radiation.

Conduction is the transfer of heat in a medium due to a temperature gradient. For a better understanding we look at the heating of a house during winter, as illustrated in Fig. 20.2. The inside of the house is warm at 20°C, but the outside is cold with a temperature of 0°C. Because of this heat gradient, heat will be transferred from the inside through the wall to the outside. Mathematically, heat conduction can be described by

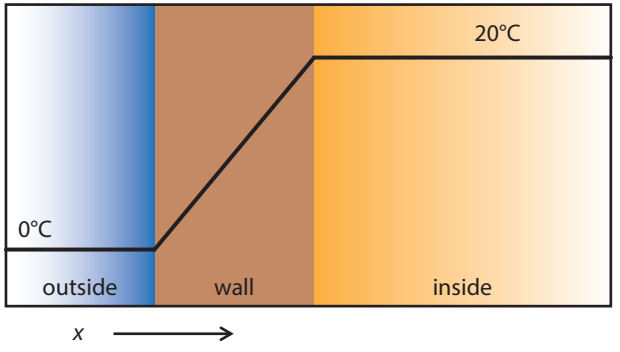

Figure 20.2: Illustrating conductive heat transfer through a wall.

the *Fourier law*,

$$\frac{\mathrm{d}Q_{\mathrm{cond}}}{\mathrm{d}t} = -kA\frac{\mathrm{d}T}{\mathrm{d}x}, \tag{20.3}$$

where $\mathrm{d}Q/\mathrm{d}t$ is the heat flow in W, k is the thermal conductivity of the wall [given in W/(Km)], A is the contact area given in m^2 and $\mathrm{d}T/\mathrm{d}x$ is the temperature gradient in the x direction given in K/m. If we assume the wall to be made from a uniform material, we can assume that $\mathrm{d}T/\mathrm{d}x$ is constant throughout the wall and we can replace it by $\Delta T/\Delta x$, where Δx is the thickness of the wall.

Convection is the second possible mechanism for heat transfer. It is the transfer of heat by the movement of a fluid. We distinguish between two forms of convection, as sketched in Fig. 20.3: In *forced convection* the movement of the fluid is caused by external variables, while in *natural convection* the movement of the fluid is caused by density differences due to temperature gradients. In both cases, the heat transfer from a medium of temperature T_1 into a fluid of temperature T_2 can be described by *Newton's law*

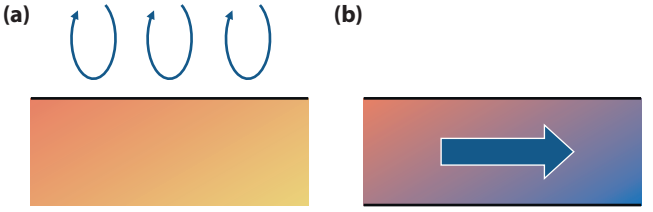

Figure 20.3: Illustrating (a) natural and (b) forced conductive heat between a solid and a surrounding fluid.

$$\frac{\mathrm{d}Q_{\mathrm{conv}}}{\mathrm{d}t} = -hA(T_1 - T_2), \tag{20.4}$$

where h is the *heat transfer coefficient* which is given in $\mathrm{W/(m^2K)}$. We will not discuss the calculation of h in detail but we want to mention that it depends on various factors such as the velocity of the fluid, the shape of the surface or the kind of flow that is present, *i.e.* whether it is laminar or turbulent flow.

The third heat transfer mechanism is *radiative heat trans-*

fer, which is the most important mechanism of heat transfer for solar thermal systems. As we already discussed in Chapter 5, thermal radiation is electromagnetic radiation propagated through space at the speed of light. Thermal radiation is emitted by bodies depending on their temperature. This is caused when excited electrons fall back on their ground level and emit a photon and hence electromagnetic radiation.

As discussed in Chapter 5, a *black body* is an idealised concept of a body, which is a perfect absorber of radiation, independent on the wavelength or direction of the incident light. Further, it is a perfect emitter of thermal radiation. The Stefan-Boltzmann law [Eq. (20.5)] describes the total irradiant emittance of a black body of temperature T,

$$M_e^{BB}(T) = \sigma T^4, \tag{20.5}$$

where M is given in W/m^2 and $\sigma \approx 5.670 \cdot 10^{-8}\,\mathrm{W}/(\mathrm{m}^2\mathrm{K}^4)$ is the Stefan-Boltzmann constant. In nature, no black bodies exist. However we can describe them as so-called *grey bodies*. The energy emitted by a grey body still can be described by Plank's law [Eq. (5.18a)] when it is multiplied with a wavelength dependent emission coefficient $\epsilon(\lambda)$. For a black body we would have $\epsilon(\lambda) \equiv 1$.

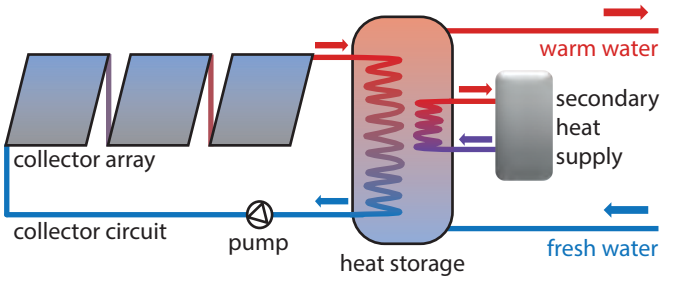

Figure 20.5: Illustrating the main components of a solar water heating system.

20.2 Solar thermal heating

About half of the world's energy consumption is in the form of *heat*. About two thirds of the heat demand is covered by coal, oil, and natural gas, as we can see in Fig. 20.4 (a). As shown in Fig. 20.4 (b), heat is mainly used in the industrial sector for facilitating for example chemical processes and in the residential sector for heating and warm water supply.

Figure 20.4 (c) shows the total energy demand of a typical household in the United States. We see that space heating and water heating represent 59% of the total energy consumption. If also the demand for cooling is taken into account, about two third of the energy consumption are related to the use of heat.

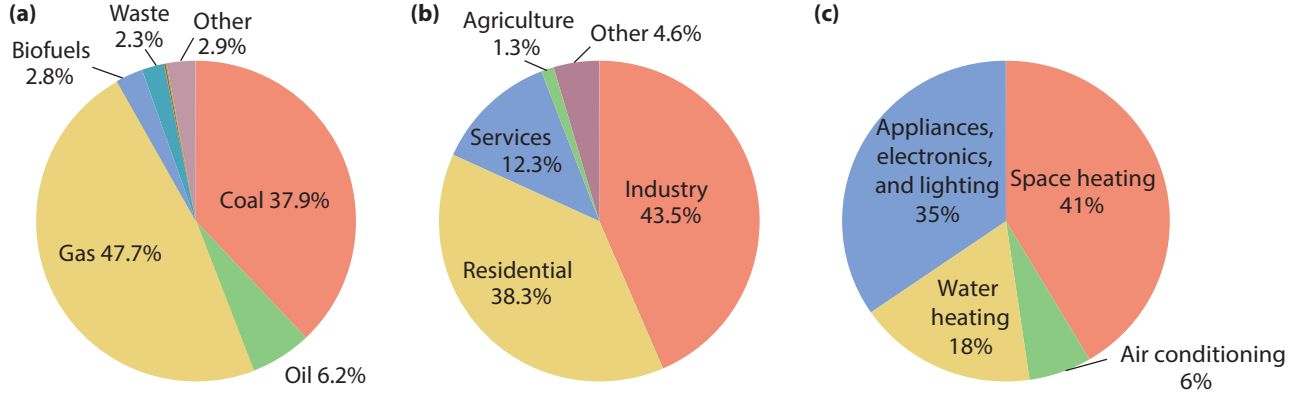

Figure 20.4: (a) The used primary energy suppliers for the generation of heat and (a) the demand of heat by sector [9]. (c) Energy consumption of U.S. homes [117]. All data is for 2009.

The residential demand for heat can at least partially covered with a *solar water heater*, which is a combination of a solar collector array, an energy transfer system and a storage tank, as sketched in Fig. 20.5. The main part of a solar water heater is the *collector array*, which absorbs solar radiation and converts it into heat. This heat is absorbed by a heat transfer fluid that passes through the collector. The heat can be stored or used directly. The amount of the hot water produced by a solar water heater throughout a year depends on the type and size of the solar collector array, the size of the water storage, the amount of sunshine available at the site and the seasonal hot water demand pattern.

There are several ways to classify solar water heating systems. The first way is by the fluid heated in the collector. When the fluid used in the application is the same that is heated in the collector it is called a direct or open loop. In contrast, when the fluid heated in the collector goes to a heat exchanger to heat up the utility fluid, it is called an indirect or closed loop.

The second way to classify the systems is by the way the heat transfer fluid is transported. This can either be passive, where no pumps are required, or by forced circulation, using a pump. The passive solar water heating systems, uses natural convection to transport the fluid from the collector to the storage. This happens because the density of the fluid drops when the temperature increases, such that the fluid rises from the bottom to the top of the collector – this is the same as natural convection that we discussed in Section 20.1. The advantage of passive systems is that they do not require any pumps or controllers, which makes them very reliable and durable. However, depending on the quality of the used water, pipes can be clogged, which considerably reduced the flow rate.

On the other hand, active systems like the one sketched in Fig. 20.5 require pumps that force the fluid to circulate from the collector to the storage tank and the rest of the circuit. These systems usually are more expensive than passive systems. However, they have the advantage that the flow rates can be tuned more easily.

20.2.1 Solar Thermal Collectors

Now we will take a closer look at the solar collector, in which the working fluid is heated by the solar radiation. The collector determines how efficiently the the incident light is used. It usually consists of a black surface, called the absorber, and a transparent cover. The absorber is able to absorb most of the incident energy from the sun, Q_{sun}, raising its temperature and transferring that heat to a working fluid. Hence, the absorber can be cooled and the heat can be transferred elsewhere.

Not all of Q_{sun} can be used, as there are some losses,

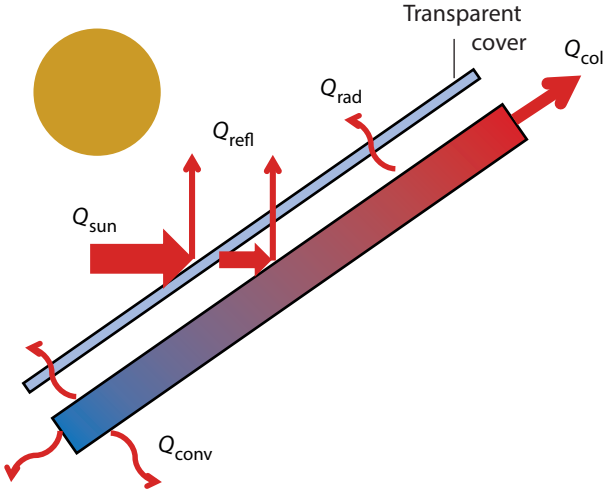

Figure 20.6: Illustrating the major energy fluxes in a covered solar collector.

as illustrated in Fig. 20.6. A part Q_{refl} is lost as reflection either in the encapsulation or in the absorber itself. Other losses are related to the heat exchanged with the surrounding air by the convection mechanism, Q_{conv} and radiation from the hot absorber, Q_{rad}. When we put all these energies in an energy balance, we find for the heat Q_{col} that can be collected by the collector

$$Q_{col} = Q_{sun} - Q_{refl} - Q_{conv} - Q_{rad}. \qquad (20.6)$$

The efficiency of the collector depends mainly on two factors: the extent to which the sunlight is converted into heat by the absorber and the heat losses to the surroundings. It will therefore depend on the weather conditions and the characteristics of the collector itself. To reduce losses, insulation from the surroundings is important, especially when the temperature difference between the absorber and the ambient is high.

Usually, collectors are classified in three categories: uncovered, covered and vacuum, as shown in Fig. 20.7. *Uncovered collectors* do not have a transparent cover, so the sun strikes directly the absorber surface, hence the reflection losses are minimised. This collector type only is used for applications where the temperature differences between the absorber and the surroundings are small, for example for swimming pools. *Covered collectors* are covered by a transparent material, providing extra insulation but also increasing the reflection losses. These collectors are used for absorber temperatures of up to 100°C. Finally, in *vacuum collectors* the absorber is

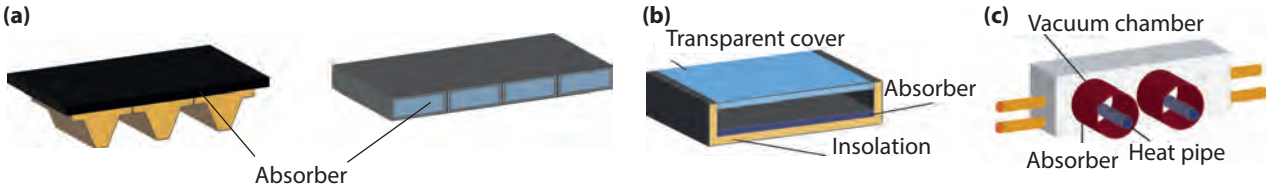

Figure 20.7: Illustrating (a) an uncovered, (b) a covered and (c) a vacuum collector.

encapsulated in vacuum tubes. In that case, little heat is lost to the surroundings. The manufacturing process of these collectors in more complicated and expensive, but the collector can be used for high temperature applications since the convection losses to the surroundings are much lower than for the other types.

Modern solar collectors often use light trapping techniques just as solar cells. For example, transparent conducting oxide layers at the front window, that we discussed in Section 13.1, can be used. If they have the plasma frequency in the infrared, almost all the solar radiation can enter the window, but the long-wavelength infrared radiation originating from the hot parts in the collector cannot leave the collector as it is reflected back at the window.

Collectors also can be classified by their shape. We can distinguish between flat-plate collectors and concentrating collectors. *Flat-plate collectors* consist of flat absorbers that are oriented towards the sun. They can deliver moderate temperatures, up to around 100°C. They use both direct and diffuse solar radiation and no tracking systems are required. Their main applications are solar water heating, heating of building, air conditioning and industrial processes heat. In contrast, *concentrating collectors* are suited for systems that require a higher temperature than achievable with flat collectors. The performance of concentrating collectors can be optimised by decreasing the area of heat loss. This is done by placing an optical device in between the source of radiation and the energy absorbing surface. because of this optical device the absorber will be smaller and hence will have a lower heat loss compared to a flat-plate collector at the same absorber temperature. One disadvantage of concentrator systems is that they require a tracking system to maximise the incident radiation at all times. This increases the cost and leads to additional maintenance requirements.

Like for PV systems,[2] also for solar heat collector arrays it is important to decide whether the collectors should be connected in series or in parallel. Connecting collectors in parallel means that all collectors have the same input temperature, while for collectors connected in series the outlet temperature of one collector is the input temperature of the next collector. Most commercial and industrial systems require a large number of collectors to satisfy the heating demand. Therefore most of the time, a combination of collectors in series and in parallel is used. Parallel flow is used more frequently because it is inherently balanced and minimises the pressure drop. In the end, the choice of series or parallel arrangement will ultimately depend on the temperature required by the application the system.

20.2.2 Heat storage

Now we will discuss the heat storage, which is an extremely important component as it has an enormous influence on the overall system cost, performance and reliability. Its design affects the other basic elements such as the collector or the thermal distribution system. The task of the storage is twofold. First, it improves the utilisation of the collected solar energy. Secondly, it improves the system efficiency by preventing the fluid flowing through the collectors from quickly reaching high temperatures.

Several storage technologies are available; some of them can even be combined to cover daily and seasonal fluctuations. In general, heat can be stored in liquids, solids or phase?change materials, abbreviated as PCM. Water is the most frequently used storage medium for liquid systems, because it is inexpensive, non-toxic, and it has a high specific heat capacity. In addition, the energy can be transported by the storage water itself, without the need for additional heat exchangers.

The usable energy stored in a water tank can be calculated with

$$Q_{\text{stored}} = V\rho C_p \Delta T, \tag{20.7}$$

where V is the volume of the tank, ρ is the density of water, C_p is the specific heat capacity of water and ΔT is the temperature range of operation. The lower temperature limit is often set by external boundaries, such as the temperature of the cold water, or by specific process requirements. The upper limit may be determined by the process, the vapour pressure of the liquid or the heat loss of the water storage. For example, for residential water heating systems the maximally allowed temperature is set to 80°C because at higher temperatures calcium carbonate will be released from the water, clogging the warm water tubes [118].

[2] See Section 17.1.

The heat loss of the tank, $\dot{Q}_{\text{loss}}$, can be determined with

$$\dot{Q}_{\text{loss}} = UA\Delta T, \tag{20.8}$$

where A is the area of the heat storage tank. U is the global heat exchange coefficient and is a measure for the quality of the insulation. Usually it varies between 2 and 10 W/K. Further, U is also a function of the different media between which the heat exchange takes place.

The same principles can be applied to small and big storage systems. Small water energy storage can cover daily fluctuations and is usually in the form of water tanks with volumes from several hundreds of litres up to several thousands of litres. Large storage systems can be used as a seasonal storage. Often they are realised as underground reservoirs.

Another type of energy storage are so-called *packed beds*, which are based on heat storage in solids. They use the heat capacity of a bed of loosely packed particulate material to store the heat. A fluid, usually air, is circulated through the bed to add or remove energy. A variety of solids can be used, rock being the most common. In operation, flow is maintained through the bed in one direction during addition of heat and in the opposite direction during removal. The bed is in general heated during the day with hot air from the collector. In the evening and during night energy is removed with air temperatures around 20°C flowing upward.

Another way to store heat in buildings is to use the solids of that the walls and roofs are built as a storage material. A case of particular interest is the *collector-storage wall*, which is arranged such that the solar radiation is transmitted through a glazing and absorbed in one side of the wall. As a result, the temperature of the wall increases and the energy can be transferred from the wall to the room by radiation and convection. Some of these walls are vented to transfer more energy to the room via forced convection.

The last method to store heat that we will discuss is to use phase change materials (PCM). While in all the storage methods discussed so far energy is stored as sensible heat, in PCM storage takes place as *latent heat*, which is used for a phase change, without any change in the material temperature. PCM used for energy storage must have a high specific latent heat L, such that a large amount of energy can be stored. In addition, the phase change must be reversible, and survive cycles without significant degradation. The stored amount of heat can be calculated with

$$Q_{\text{stored}} = m\left[C_s(T^* - T_1) + L + C_l(T_2 - T^*)\right], \tag{20.9}$$

where the different temperatures are also sketched in Fig. 20.1. We consider the specific heat capacity in the solid phase, C_s, to be constant from the initial temperature T_1 up to the phase change temperature T^*, the latent heat of the material L and the specific heat capacity in the liquid phase C_l from T^* up to the final temper-

ature T_2. Materials that are commonly used s PCM are molten salts, such as Na_2SO_4, $CaCl_2$ or $MgCl_2$. PCM storage in general is used for high temperature applications.

20.2.3 System design

Often, at least at larger latitudes with large differences between summer and winter, solar thermal systems include a boiler as a backup. Its main function is to provide the necessary energy when the solar power is not sufficient. It is basically a normal heater that adds the remaining heat needed to achieve the desired temperature in the storage tank. As energy source, usually natural gas, oil, or biomass is used. Sometimes, the additional heat also is provided with electricity or a heat pump.

Up to now we discussed how to collect and store the energy, but we did not consider yet how to transport it from the collector to the storage system. The transport of heat is done with a *collector circuit*, which usually transports heat using either a liquid or gas. If a liquid is used, it is important that it neither should freeze nor boil, even at the most extreme operating conditions. Further, the medium should have a large specific heat capacity, a low viscosity, and it should be non-toxic, cheap and abundant. The most common fluids are water, oils or air.

As mentioned above, the flux can be caused either naturally by the temperature gradients, forced by a pump, or by a heat pipe, in which the fluid is allowed to boil and condense again. The optimal choice will depend on the specific system design. Further, heat losses in the collector circuit must be taken into account, especially when the pipes are very long. During the planning phase it therefore is important to minimise the circuit length.

In systems with a pump often a *controller* is present that regulates the fluxes of fluid through the collector, the storage and the boiler in order to assure the desired temperature of the. This can for example be done by calculating the optimal flux in dependence of the fluid temperature at the collectors and in the storage such that the heat transfer is maximised.

20.2.4 Solar air conditioning

Another interesting application is *solar air conditioning* (also called *solar cooling*), which seems a bit contradictory on first sight. Before we start with the actual discussion on solar air conditioning, we briefly recap the principle of a compressor-driven air conditioning system, which is sketched in Fig. 20.8. Similar to every *heat pump*, the task of an air conditioning system is to transport heat from a cool reservoir B to a warmer reservoir A. Such a system has typically four components:

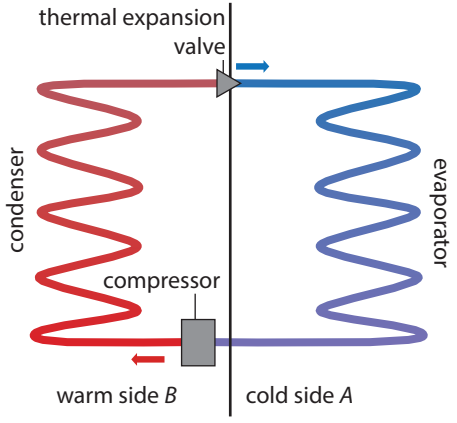

Figure 20.8: A compressor-driven cooling circuit.

a compressor, a condenser, a thermal expansion valve (throttle) and an evaporator.

In the compressor (1), vaporised coolant is compressed to a higher pressure. Under release of its latent heat to the surroundings, it is condensed to its liquid state in the condenser (2). Then the coolant passes a thermal expansion valve (3), such that its pressure on the other side of this valve is so low that it can evaporated again. For the evaporation, it must absorb latent heat from the cool reservoir A. Then, the coolant containing the latent heat from A is compressed again. We see that this cooling circuit utilises the latent heat stored and released during phase changes.

Naturally, heat only flows from warm to cold reservoires. For reversing the heat flow, as it happens in a heat engine, additional energy must be added to the system. This happens in the compressor and traditionally is supplied as electric energy. Probably the simplest option to realise solar cooling is to generate this electricity by a conventional PV system or a solar thermal power system as discussed in Section 20.3. Further, it is also possible to drive the compressor with mechanic energy obtained from solar energy directly with a Rankine cycle, which is discussed in 20.3.

The choice of the coolant is based on the temperature regime in which the air conditioner is supposed to operate. However, also their effect on the environment has to be taken into account. For example the use, *chlorofluorocarbons* that were widely used as coolants was strongly banned in the 1980s because of their destructive effect on the ozone layer [119].

Another concept that that directly can be driven by solar heat is that of an *absorption cooling machine*. In difference to compressor-driven cooling, no mechanical energy is required for absorption cooling, but the cooling process is directly driven by heat that can be supplied by solar collectors. Instead of one coolant, here *two substances*are used: a *refrigerant* and an *absorbant*. Often, ammonia (NH_3) is used as refrigerant, while water is used as an absorbant. Under atmospheric pressure,

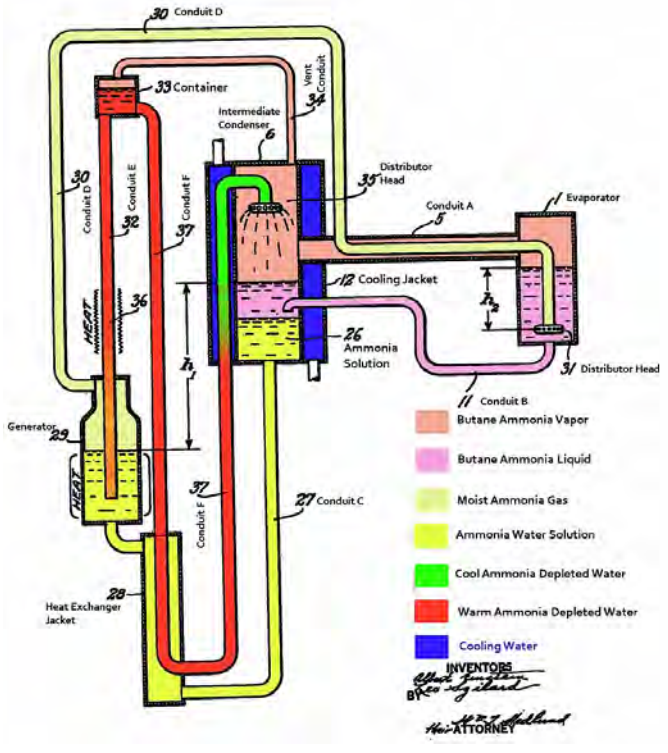

Figure 20.9: Illustrating an absorption cooling circuit as invented by Einstein and Szilard [120]. (Annotations by P. Brandon Malloy.)

ammonia has a boiling point of -33°C. Further, it is very soluble in water, a property that is utilised in the cooling process. Figure 20.9 sketches the cooling circuit of an absorption cooler. In the boiler (1), a mixture of both substances is heated. Because of its low boiling point the refrigerant will leave the mixture in the gas phases. Absorbant that is still present in this gas is separated (2) and brought back into the boiler. Then, the refrigerant is then is condensed in a condenser (3). Then, the condensed refrigerant is brought to a thermal expansion valve (4). Similar to the compressor circuit, latent heat is used to evaporate the expanded refrigerant, such that cooling takes place. In the last step, the refrigerant is absorbed by the absorbant and the circuit starts again. Instead of using a throttle valve to expand the refrigerant, it also can be mixed with a third gas, for example hydrogen (H_2), such that the partial pressure of the refrigerant is reduced allowing it to evaporate. In that version, the whole system operates at one pressure.

The last option of cooling that we discuss is that of solar *desiccant cooling*, illustrated in Fig. 20.10. In such a system, air air is dried when passing through a desiccant, like silica. Then water is sprayed into the air. As it evaporates, the air is cooled and humidified, such that it guarantees optimal interior conditions. Air that leaves the interior is heated up using a solar collector. The hot air streams through the desiccant. Hence, the desiccant is tried and can be reused for adsorbing humidity.

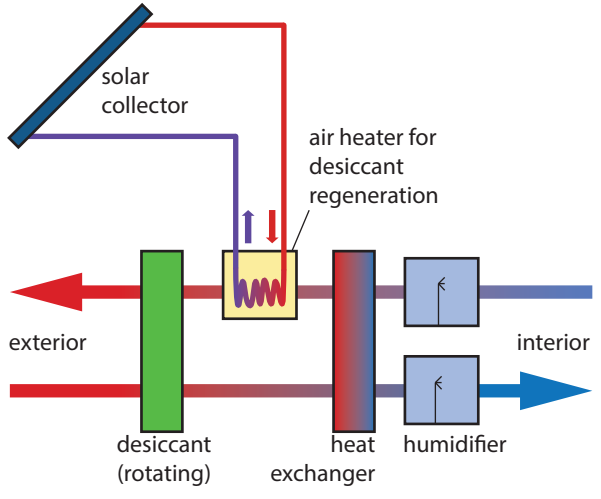

Figure 20.10: Desiccant cooling.

20.3 Concentrated solar power (CSP)

In this section we take a closer look at *concentrated solar power* (CSP), where high temperature fluids are used in steam turbines to produce electricity. Much of the early attention on CSP systems was on small-scale applications, mainly for water pumping. However, since about 1985 several large-scale power systems with a power output up to 80 MW have been built.

The CSP technology is especially interesting for desert regions, where almost all the solar radiation is incident as direct radiation. As illustrated in Fig. 20.11, these systems consist basically a collector, where the solar energy is absorbed, a storage system, usually water or phase-change storage, a boiler that acts as a heat exchanger between the operational fluids of the collector and the heat engine, and the heat engine itself, which converts the thermal energy into mechanical energy. In large CSP plants, the heat engine is a *steam turbine*, where the energy stored in the hot steam is partially converted to rotational (*mechanic*) energy as the steam expands along the turbine. This thermodynamic process is called *Rankine cycle*, called after the Scottish engineer William John Macquorn Rankine (1820-1872).

The mechanical energy can be further used to drive an electrical generator. The collectors are built as concen-

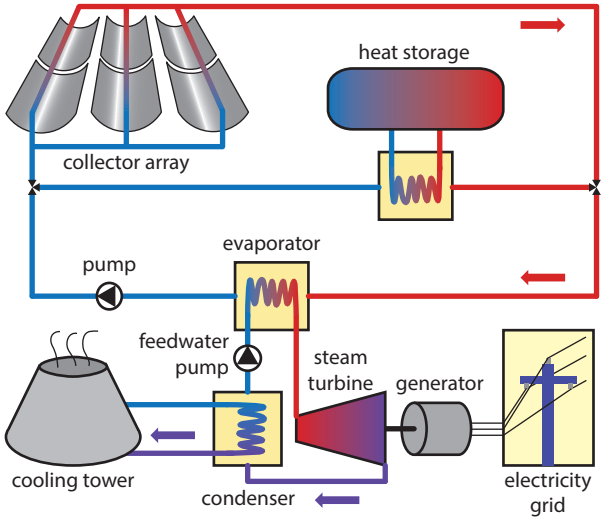

Figure 20.11: Sketching a solar thermal electric power system.

trator systems, in order to reach the high temperatures of several hundreds of degree Celsius, which are required to operate the heat engine.

One problem of CSP systems is that the efficiency of the collector diminishes as its operating temperature rises, while the efficiency of the engine increases with temperature as we already have seen when discussing the thermodynamic efficiency limit in Section 10.1. Hence, a compromise between the two has to be found when choosing the operating temperature. Current systems have efficiencies up to about 30% [121],

Now we will take a closer look at the solar concentrator system, illustrated in Fig. 20.12. First concepts already were developed by the ancient scientist Archimedes (287 BC - 212 BC); Leonardo da Vinci (1452 - 1519) designed some concentrators. Different types of concentrators produce different peak temperatures and hence thermodynamic efficiencies, due to the different ways of tracking the sun and focusing light. Innovations in this field are leading to more and more energy efficient and cost effective systems.

The most common concentrator type with about 96% of all installed CSP installations is the *parabolic trough* [126], which consists of a linear parabolic reflector that concentrates the light onto an absorber tube located in the middle of the parabolic mirror, in which the working fluid is located. The fluid is heated to 150 to 350 degrees Celsius, and then used in a heat engine. This

(a) Parabolic trough

(b) Fresnel concentrator

(c) Dish Stirling

(d) Solar power tower

Figure 20.12: Different types of concentrators used in CSP systems [122–125].

technology is far developed, well-known examples are the Nevada Solar One power plant and the power plant that the Plataforma Solar de Almeria (PSA) in Spain.

Another concentrator concept is that of *Fresnel concentrators*, where flat mirrors are used, as illustrated in 20.12. Flat mirrors allow more reflection in the same amount of space as parabolic mirrors, reflect more sunlight, and are much cheaper.

Other important concentrator systems is the *dish Stirling* or dish engine system. It consists of a parabolic reflector that concentrates light to the focal point, where the working fluid absorbs the energy and is heated up to about 500°C. The heat is then converted to mechanical energy using a *Stirling engine*. With about 31.25% efficiency demonstrated at Sandia National Laboratories, this design has currently the highest demonstrated efficiency [121].

The last concentrator type is that of *solar power tower plants.* Here, an array of dual-axis tracking reflectors, commonly named *heliostats*, are arranged around a tower. There the concentrated sunlight hits a central receiver, which contains the working fluid, which can be heated to 500 or even 1000°C. Examples for solar power towers are Solar One and Solar Two in the Unites Stars and the Eureka project in Spain [126].

CSP systems can be combined with heat storage, such as phase-change materials discussed in Section 20.2.

This allows CSP systems to operate during day *and* night.

21 Solar fuels

In this chapter we discuss how to store solar energy in the form of chemical energy in so-called *solar fuels*. As we have seen in previous chapters, we are able to convert solar energy to electricity with efficiencies of up to 44%. Also, we are able to convert solar energy easily in heat, which we can use for preparing warm water, heating and even cooling.

The biggest obstacle towards a economy that is driven by renewable energy is not the generation of sustainable energy sources but the storage the renewable energy. As we have seen in Chapter 17, storing electricity with batteries is very difficult. While day-night storage that makes solar electricity generated during daytime available in the night is easily done, seasonal storage that allows to store electricity generated in the summer until winter would require very large battery systems that usually are not feasible. This problem becomes more severe in regions that are far away from the equator such that the differences in solar irradiance between summer and winter are significant. The same also is true for solar heat that was discussed in Chapter 20: while a hot water storage tank with a volume of several hundreds of litres is sufficient as day-night storage, seasonal storage that would allow to use solar-heated water throughout the year would require storage tanks with a volume of tens of thousands litres. This tank also would need to be heavily isolated. If we manage to store solar energy as chemical energy, we do not have this problem as chemicals can be easily stored.

Let us take another look at the Ragone chart in Fig.

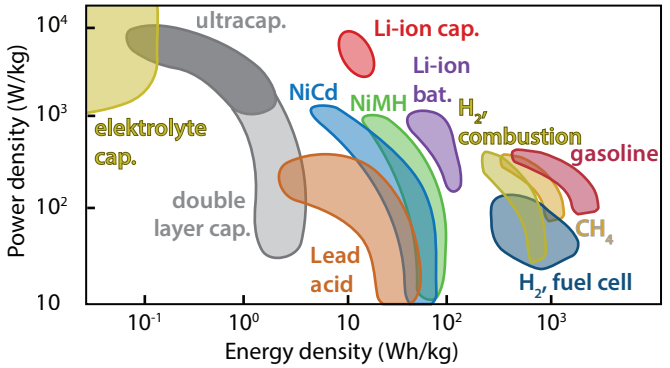

Figure 21.1: A *Ragone* chart of different energy storage methods. Capacitors are indicated with "cap".

21.1 that we already analysed in Section 17.4. This chart shows the amount of energy per mass stored in a certain storage technology versus the specific power provided by the technology. As we know, by far the most used form of stored energy are fossil fuels. As we can see from the graph, fossil fuels have high energy density properties, are stable and reliable. But we also know that fossil fuels are depleting and their use is detrimental to our environment. Hydrogen (H_2) has an even better gravimetric energy density compared to fossil fuels, but it is a very light gas, so the volumetric energy density, *i.e.* the energy per unit volume, is much lower. This is one of the reasons it has not been widely used until now. When we compare batteries and hydrogen as energy storage for solar electricity, we see that batteries have a lower energy density. Further, they require a much higher initial investment than hydrogen. Hence, it seems an interesting option to use hydrogen as a storage material, *i.e.* as a solar fuel.

As we see in this chapter, we can produce hydrogen using solar energy by utilising electrochemistry. In nature, *photosynthesis* takes place, where sunlight is used to convert carbon dioxide and water into oxygen and sugars, which can be seen as nature has been converting the energy from the sun into chemical energy for a long time by converting carbon dioxide and water into oxygen and sugars. Here, we try to mimic nature with inorganic semiconductor materials that re able to split a water molecule into oxygen and hydrogen using the energy of sunlight. This process sometimes is referred to as *artificial photosynthesis*.

For storing solar energy as chemical energy in the form of hydrogen, *water splitting* can be used,

$$2H_2O \rightarrow 2H_2 + O_2. \tag{21.1}$$

The energy required for this reaction is given by the *Gibbs free energy* and it has a value of $G = 237.2$ kJ/mol. In *solar* water splitting this energy is provided by the sun.

Another product that seems very promising for energy storage is *methane* (CH_4), as it is easier to store and

has less hazardous problems than H_2. It can be produced from hydrogen and carbon dioxide (CO_2) with the *Fischer-Tropsch process*. In this process, first the hydrogen is combined with carbon dioxide in a *water gas shift reaction* to obtain carbon monoxide and water in the form of steam,

$$H_2 + CO_2 \rightarrow 2H_2O + CO. \tag{21.2}$$

This gas mixture, known as *synthesis gas*, can be then refined to finally obtain methane,

$$CO + 3H_2 \rightarrow CH_4 + H_2O. \tag{21.3}$$

When the stored energy is required, the methane can be burned in a combustion reaction,

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O \tag{21.4}$$

that will give water and carbon dioxide as byproducts. The water can be reused for the electrolysis and the carbon dioxide can be used for the water gas shift reaction, hence, the cycle is closed. Each reaction in the cycle has a certain efficiency; the overall efficiency can be obtained by combining all those efficiencies.

Many different methods can be used for water splitting. In this chapter we will discuss two methods: *electrolysis of water* and *photoelectrochemical water splitting* (PEC).

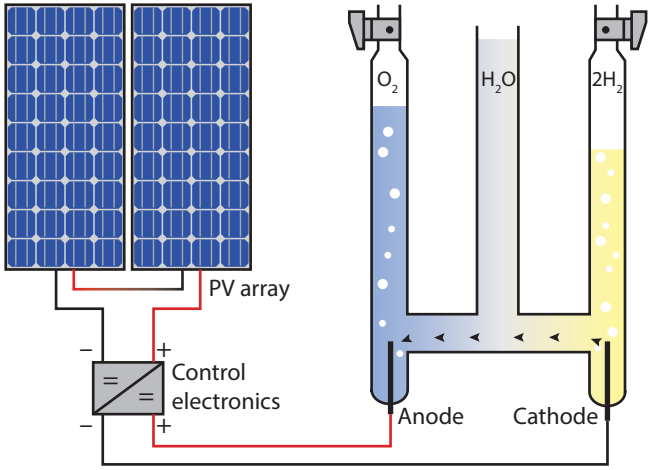

Figure 21.2: Illustrating a lab-scale Hofmann voltameter that is used to split water. In the sketch the voltameter is connected to a PV system via some control electronics.

21.1 Electrolysis of water

Figure 21.2 shows a typical lab-scale setup for the hydrolysis of water. It is called a *Hofmann voltameter* after the German chemist August Wilhelm von Hofmann (1818-1892). It contains of three upright cylinders that are joined at the bottom. While electrodes are placed in the left and the right cylinder, the central cylinder is used to refill the device with water. For starting the electrolysis, a voltage has to be applied between the two electrodes.

Water splitting is a *reduction-oxidation* reaction, which is commonly abbreviated as *redox reaction*. In such redox reactions, the reaction happens due to the exchange of electrons between atoms or molecules. They can be divided in two half reactions, the *oxidation* and the *reduction*. For water splitting these half reactions look as follows: During the oxidation, which happens at the *anode*, water is split giving oxygen and protons to the solutions and electrons to the anode,

$$H_2O \rightarrow \tfrac{1}{2}O_2 + 2H^+ + 2e^-. \tag{21.5}$$

In the reduction, electrons originating from the *cathode* react with the protons to hydrogen,

$$2H^+ + 2e^- \rightarrow H_2. \tag{21.6}$$

The total reaction therefore can be written as

$$H_2O \rightarrow \tfrac{1}{2}O_2 + H_2. \tag{21.7}$$

Each half reaction has an associated potential and the sum of the potentials of each half reaction is the potential for the whole redox reaction. Potentials always are defined with respect to a reference, *i.e.* a zero point. For redox reactions, the zero is defined as the hydrogen half reaction that happens at the cathode. For calculating the potential at the anode, *ie* of the oxygen production reaction, we can apply the equation

$$V^0 = \frac{G}{nF}, \tag{21.8}$$

where G is the Gibbs free energy already mentioned above, n is the number of electrons transferred between the anode and the cathode for the production of one H_2 molecule, and $F = 96\,485$ C/mol is the *Faraday constant* that tells us the amount of charge per mole of electrons. Using $G = 237.2$ kJ/mol and $n = 2$, as two electrons are transferred per H_2 molecule, we find $V^0 = 1.23$ V.

In reality, an applied voltage of 1.23 V is not enough to drive the redox reaction because of some voltage losses. These voltage losses are mainly because of the activation energy at the electrodes as well as Ohmic losses in the cables, electrodes and also in the solution. For increasing the conductivity of the water, a base or an acid can be added. For the classic Hofmann voltameter, a small amount of sulphuric acid (H_2SO_4) is used.

The potential difference between V_0 and the potential used in the real device is called *overpotential* ΔV. The

overpotential due to the activation energy required at the electrodes is strongly related to the electrode material and the gas that is produced at that electrode. For hydrogen production at the cathode, *platinized (black) platinum* is one of the best materials with an overpotential of -0.07 V. For the oxygen production at the anode, *nickel* with an overpotential of $+0.56$ V is very well suited. Combining these two voltages and the overpotentials we can understand that the typical overpotential is usually around 0.8V.

It is important to realise that all the energy consumed by the hydrolysis process that is related to the overpotential is a loss that will lead to heating up of the hydrolysis device. Only a fraction

$$\eta_{\mathrm{WS}} = \frac{V^0}{V^0 + \Delta V} \tag{21.9}$$

of the supplied energy is actually stored in hydrogen. It is clear that it is a very important task for science and industry to reduce the required overpotential. Further, the current requirement for platinum electrodes makes hydrolysis not feasible from an economic point of view.

For an overpotential of 0.8 V, we would get an efficiency of around 60%. If we would use a good PV system with an overall efficiency of $\eta_{\mathrm{PV}} = 18\%$ we would have a total maximal *solar-to-hydrogen efficiency* of

$$\eta_{\mathrm{STH}} = \eta_{\mathrm{PV}}\eta_{WS} = 10.9\%. \tag{21.10}$$

21.2 Photoelectrochemical (PEC) water splitting

Hydrogen can also be produced utilising solar energy by using a *photoelectrode*, which uses light to produce an electrochemical reaction, in which water is split into oxygen and hydrogen. In this process, the photons reach the surface of the photoelectrode, which is made of a photoactive semiconductor. As in any other semiconductor, photons with an energy equal or larger than the semiconductor band gap energy will create an electron-hole pair. The electrons and holes will be separated by an electric field, and both will be used in the two half reactions involved in the overall water splitting process. To generate the required electrical field, we need a voltage source, for example a solar cell. The photoelectrode can be either an anode or a cathode.

If the solar-splitting device is made of a solar cell that is placed behind the photoanode, the solar cell will receive the light transmitted through the photoanode. This light creates another electron-hole pair and electric field that brings the electrons to the photoanode and the holes to the photocathode with enough potential to drive the redox reaction in the electrolyte. As a result, the water molecule is split into oxygen and hydrogen.

When the photoelectrode is made from an *n*-type semiconductor it acts as electron donor; if it is *p*-type it

acts as electron acceptor. As an acceptor, the material will attract more holes to the interface, which will enhance the reduction half reaction. Hence it acts as the photocathode and produces hydrogen. If the semiconductor is *n*-type, it acts as the photoanode. Electrons are moved to the interface by the internal electric field, and those electrons are involved in the oxidation half reaction, such that oxygen is produced.

The semiconductor material has to fulfil several requirements. First, it has to absorb the light that is incident on its surface. Secondly, charge carrier transport inside the material and separation into the two electrodes must be efficient. Thirdly, a bandgap of 1.23 eV is not enough to drive the reaction because of the overpotential already discussed above. It has been estimated that materials with an energy band gap close to 2.1 eV have the potential to split water. Fourthly, the energy levels of the material have to be adequate to couple with the energy needed for the reaction. In particular, the energy levels of the reactions have to be located somewhere in the energy band gap of the semiconductor, which is called the favourable position. To further enhance the reaction, a catalyst may be added to the semiconductor surface. Fifthly, on the practical side, it is important that the used materials are photochemically stable and relatively cheap. From these criteria we can conclude that the main technical challenges to be addressed are light absorption, the separation of charges and the catalysis of the reaction.

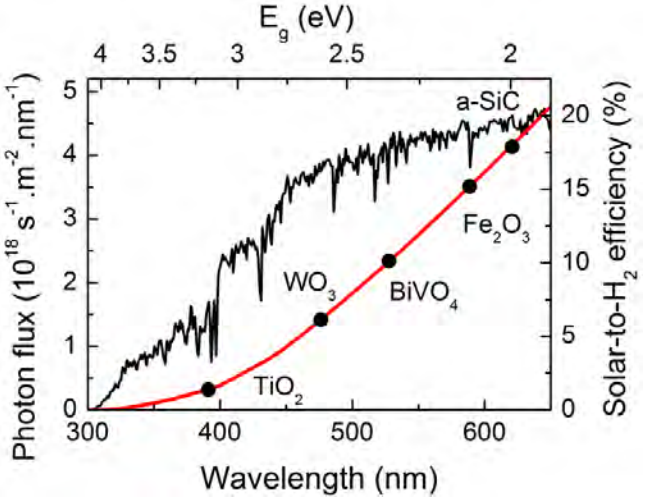

Figure 21.3: The solar-to-hydrogen efficiency and bandgap of different potential photocathode materials. Also the photon flux is shown [127].

The absorption and catalysis problems can be tackled by carefully choosing the semiconductor material and its corresponding catalyst. Several materials can be considered for solar water splitting. Some of the most popular materials studied for this application are titanium dioxide (TiO_2), tungsten oxide (WO_3), bismuth vanadate ($BiVO_4$), hematite (Fe_2O_3) or amorphous silicon carbide (a-SiC). As shown in Fig. 21.3, $BiVO_4$, Fe_2O_3, and a-SiC have the most promising potential solar-to-hydrogen efficiency. , as shown in the graph. Hematite and silicon carbide are the best choice for the optimal absorption of light, since they have band gap energies closest to the optimal bandgap of 2.1 eV. However, if also other factors such as the band position or stability are considered, materials like bismuth vanadate may also be a viable option. Because of their large bandgap energies, the other materials are not considered.

The overall efficiency of water splitting depends on the catalytic efficiency and the separation efficiency of the photoelectrode. The *catalytic efficiency* can be improved by placing a catalyst on the semiconductor surface. For example, for a ceBiVO4 photoanode the inclusion of a cobalt phosphate [$Co_3(PO_4)_2$] catalyst on the surface will ease the oxidation reaction and hence increase the water splitting efficiency. The *separation efficiency* can be improved by introducing an electric field inside the material. One way to do this is to introduce gradient doping, starting from no doping at the surface to 1%

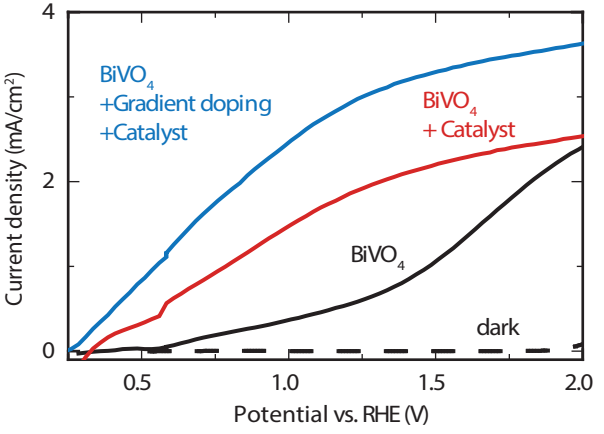

Figure 21.4: The effect on using a catalyst and gradient doping on the performance of a $BiVO_4$ photoanode [127].

doping close to the back of the electrode. With gradient doping a depletion region is created in between the semiconductor and the electrolyte. As a result, electrons will move more easily from the electrolyte into the semiconductor. Combining both effects can highly improve the efficiency of the overall device. Both the effects of the catalyst and the gradient doping are shown in Fig. 21.4, where the current density of the illuminated photoanode is shown in dependence of the applied voltage with respect to the *reversible hydrogen electrode* (RHE). The RHE is a special electrode with the property that the measured potential does not changed when the pH value of the solution is changed.

As we already mentioned above, for the photoelectrochemical (PEC) water splitting process, a potential difference of at least 1.23 V plus the overpotential ΔV must be present. The value of ΔV will depend on the materials and electrolyte used, but it is usually around 0.8V. Both potentials added will result in the total potential difference required to drive the redox reaction. This voltage will be partially covered by the potential difference created within the photoelectrode when light shines on it. However, depending on the material, the PEC only creates 0.6 V of the required voltage. For the extra potential that is required to allow water splitting, the photoelectrode can be combined with solar cells that provide the extra potential needed for the reaction to happen. The combination of a photoelectrode and a solar cell forms a photoelectrochemical device.

In Fig. 21.5 (a) such a PEC device is sketched. The photoelectrode (in this case a photoanode) is connected to a solar cell in series. The photoanode is connected to the positive contact of the solar cell. The negative contact of the solar cell is connected to another electrode via an external circuit. This counter electrode may or may not be photoactive. The electric circuit is closed by the electrolyte. The band diagram of this device is shown in Fig. 21.5 (b).

Since the photoelectrode and the solar cell are connected in series, the same current will go through both the devices, similar as for multi-junction solar cells. In this device, the light is utilised better then it would be in the solar cell alone. In the photoelectrode, the photons with energies exceeding that of the photoelectrode bandgap will be absorbed. The fraction of the light that has not been absorbed or reflected, called the *transmitted spectrum*, reaches the solar cell where it can be absorbed to generate the extra potential difference required for water splitting. The solar cell must be optimised for the transmitted spectrum, which is different from the standard AM1.5 spectrum that is usually used for solar cell optimisation.

As any semiconductor device, the photoelectrode has its own characteristic *J-V* curve. When a solar cell and a photoelectrode are combined, the conditions at which they will work can be estimated by studying their *J-V* curve characteristics. Figure 21.6 shows the *J-V* curves

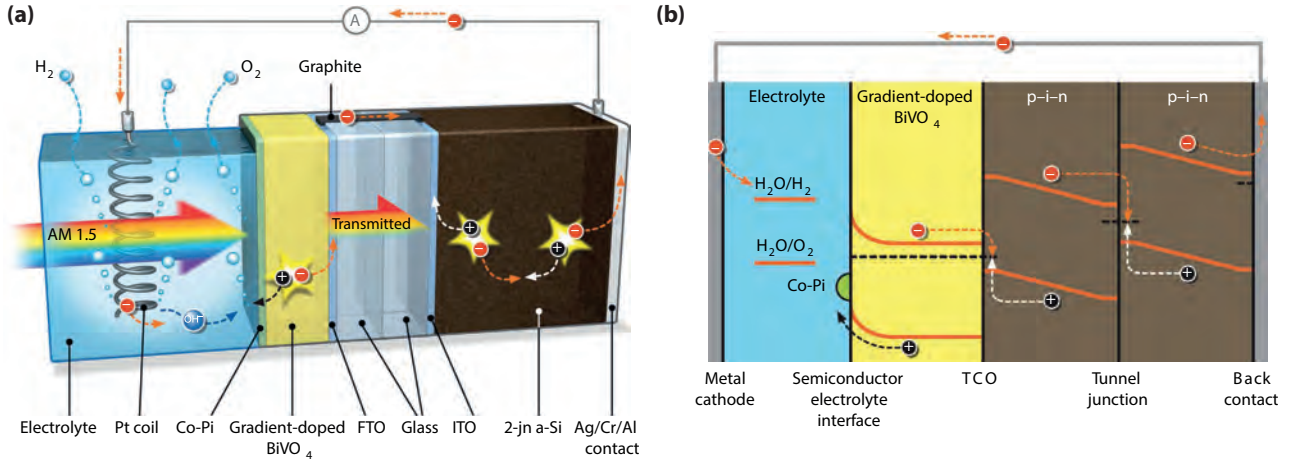

Figure 21.5: (a) Illustrating a photoelectrochemical (PEC) device consisting of a $BiVO_4$ photoanode and a tandem a-Si:H/a-Si:H solar cell. (b) The band diagram of this PEC device [127].

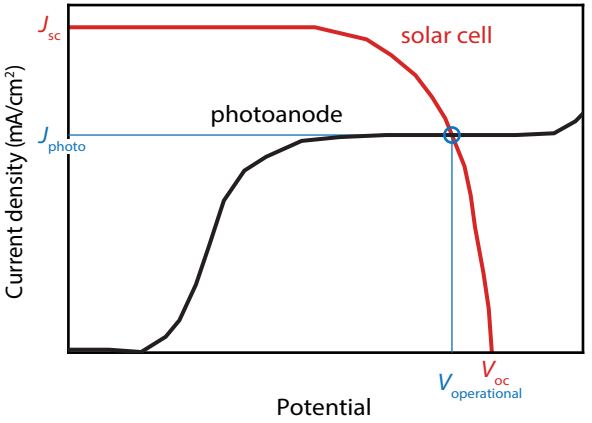

Figure 21.6: The *J*-*V* characteristics of the photoelectrode and the solar cell. The operating point is at the crossing of the two.

of the solar cell and of the photoanode in this case. Since both elements are connected in series, the current of both the solar cell and the photoelectrode must be the same. Hence, the operational point will be where both *J*-*V* curves cross.

Now we will estimate the *solar-to-hydrogen efficiency* η_{STH} that is related to the amount of produced hydrogen. This amount is directly calculated by the current density measured at the operating point, which we can understand by realising that current is the flow of electric charge per time.

When we assume that all generated charges produced are involved in the production of hydrogen, the overall solar-to-hydrogen conversion efficiency, is described by

$$\eta_{\mathrm{STH}} = \frac{J_{\mathrm{ph}} \times 1.23\,\mathrm{V}}{P_{\mathrm{in}}}, \tag{21.11}$$

where J_{ph} is the operational current density, P_{in} is the irradiance arriving at the PEC device, usually it will be 1000 W/m^2 from the AM 1.5 spectrum.

The only free variable in Eq. (21.11) is the current density: the higher the operational current, the more hydrogen is produced and the higher the device efficiency. The major focus of current academic research therefore is to improve the current density. Besides utilising gradient doping water-oxidation catalysts also other improvements can be done to increase the performance of PEC devices. These are optimising the materials and

layer thicknesses in the used solar cell and texturing of the photoanode for better light trapping.

All these concepts were applied at the Delft University of Technology, where a bismuth vanadate photoanode was combined with a double junction amorphous silicon solar cell and a platinum cathode. This device achieved a solar-to-hydrogen efficiency of $\eta_{\mathrm{STH}} = 4.9\%$, which is actually the highest efficiency reported for such devices based on a metal oxide photoanode [127].

Appendix

A

Derivations in Electrodynamics

A.1 The Maxwell equations

The four Maxwell equations couple the electric and magnetic fields to their sources, *i.e.* electric charges and current densities, and to each other. They are given by

$$\nabla \cdot \mathbf{D}(\mathbf{r},t) = \rho_F(\mathbf{r}), \tag{A.1a}$$

$$\nabla \times \mathbf{E}(\mathbf{r},t) = -\frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t}, \tag{A.1b}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r},t) = 0, \tag{A.1c}$$

$$\nabla \times \mathbf{H}(\mathbf{r},t) = +\frac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t} + \mathbf{J}_F(\mathbf{r}), \tag{A.1d}$$

where $\mathbf{r}$ and t denote location and time, respectively. $\mathbf{D}$ is the *electric displacement*, $\mathbf{E}$ the *electric field*, $\mathbf{B}$ is the *magnetic induction* and $\mathbf{H}$ is the *magnetic field*. ρ_F is the *free charge density* and $\mathbf{J}_F$ is the *free current density*.

The electric displacement and field are related to each other via

$$\mathbf{D} = \epsilon\epsilon_0\mathbf{E}, \tag{A.2a}$$

where ϵ is the relative permittivity of the medium in that the fields are observed and $\epsilon_0 = 8.854 \times 10^{-12}$ As/(Vm) is the permittivity *in vacuo*. Similarly, the magnetic field and induction are related to each other via

$$\mathbf{B} = \mu\mu_0\mathbf{H}, \tag{A.2b}$$

where μ is the relative permeability of the medium in

that the fields are observed and $\mu_0 = 4\pi \cdot 10^{-7}$ Vs/(Am) is the permeability of vacuum. Eqs. (A.2) are only valid if the medium is *isotropic*, *i.e.* ϵ and μ are independent of the direction. We may assume all the materials important for solar cells to be *nonmagnetic*, *i.e.* $\mu \equiv 1$.

A.2 Derivation of the electromagnetic wave equation

We now derive de electromagnetic wave equations in source-free space, $\rho_F \equiv 0$ and $\mathbf{j}_F \equiv 0$. For the derivation of the electromagnetic wave equations we start with applying the rotation operator $\nabla\times$ to the second Maxwell equation, Eq. (A.1b),

$$\nabla \times (\nabla \times \mathbf{E}) = -\nabla \times \left(\frac{\partial \mathbf{B}}{\partial t}\right). \tag{A.3}$$

Now we take the fourth Maxwell equation, Eq. A.1d, with $\mathbf{j}_F = 0$ and Eqs. (A.2),

$$\frac{1}{\mu\mu_0}\nabla \times \mathbf{B} = \epsilon\epsilon_0 \frac{\partial \mathbf{E}}{\partial t},$$

and substitute it into Eq. (A.3),

$$\nabla \times (\nabla \times \mathbf{E}) = -\epsilon\epsilon_0\mu\mu_0 \left(\frac{\partial^2 \mathbf{E}}{\partial t^2}\right). \tag{A.4}$$

By using the relation

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \Delta\mathbf{E} = -\Delta\mathbf{E}, \tag{A.5}$$

we find

$$\Delta\mathbf{E} = \epsilon\epsilon_0\mu\mu_0 \left(\frac{\partial^2 \mathbf{E}}{\partial t^2}\right). \tag{A.6}$$

In Eq. (A.5) we used that we are in source free space, *i.e.* $\nabla\mathbf{E} = 0$. Equation (A.6) is the *wave equation* for the *electric field*. Note that the factor

$$\frac{1}{\epsilon\epsilon_0\mu\mu_0}$$

has the unit of $(\mathrm{m/s})^2$, *i.e.* a speed to the square. In easy terms, it is the squared propagation speed of the wave.[1] We now set

$$c_0^2 := \frac{1}{\epsilon_0\mu_0} \tag{A.7}$$

and

$$n^2 = \epsilon, \tag{A.8}$$

where c_0 is the speed of light *vacuo* and n is the *refractive index* of the medium. Since we also assume $\mu \equiv 1$, we finally obtain for the wave equation for the electric field

$$\Delta\mathbf{E} - \frac{n^2}{c_0^2}\left(\frac{\partial^2 \mathbf{E}}{\partial t^2}\right) = 0. \tag{A.9a}$$

[1] In reality, if the medium is absorbing, things are getting much more complex.

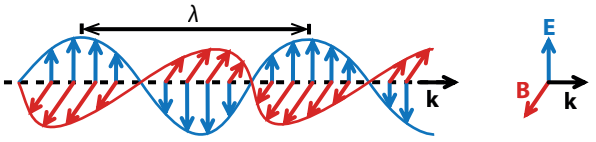

Figure A.1: Illustrating the **E** and **B** field and the **k** vector of a plane electromagnetic wave.

In a similar manner we can derive the wave equation for the *magnetic field*,

$$\Delta \mathbf{H} - \frac{n^2}{c_0^2}\left(\frac{\partial^2 \mathbf{H}}{\partial t^2}\right) = 0. \tag{A.9b}$$

A.3 Properties of electromagnetic waves

In section A.2, we found that plane waves can be described by

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 \cdot \mathrm{e}^{ik_z z - i\omega t}, \tag{A.10a}$$
$$\mathbf{H}(\mathbf{x}, t) = \mathbf{H}_0 \cdot \mathrm{e}^{ik_z z - i\omega t}. \tag{A.10b}$$

In this section we study some general properties of plane electromagnetic waves, illustrated in Fig. A.1.

Substituting Eq. (A.10a) into the first Maxwell equation, Eq. (A.1a), yields

$$\begin{aligned}
\nabla \mathbf{D} = \epsilon\epsilon_0 \nabla \mathbf{E} &= 0, \\
\nabla \mathbf{E} &= 0, \\
\nabla \left(\mathbf{E}_0 \mathrm{e}^{ik_z z - i\omega t}\right) &= 0, \\
E_{x,0} \underbrace{\frac{\partial}{\partial x}\mathrm{e}^{ik_z z - i\omega t}}_{=0} + E_{y,0} \underbrace{\frac{\partial}{\partial y}\mathrm{e}^{ik_z z - i\omega t}}_{=0} + & \\
+ E_{z,0} \frac{\partial}{\partial z}\mathrm{e}^{ik_z z - i\omega t} &= 0, \\
i E_{z,0} k_z \mathrm{e}^{ik_z z - i\omega t} &= 0, \\
E_{z,0} &= 0,
\end{aligned} \tag{A.11a}$$

where we used the notation $\mathbf{E}_0 = \left(E_{x,0}, E_{y,0}, E_{z,0}\right)$. In a similar manner, by substituting Eq. (A.10b) into the first Maxwell equation, Eq. (A.1c), we obtain

$$H_{z,0} = 0. \tag{A.11b}$$

Thus, neither the electric nor the magnetic fields have components in the propagation direction but only components perpendicular to the propagation direction (the x- and y-directions in our case).

Without loss of generality we now assume that the electric field only has an x-component, $\mathbf{E}_0 = (E_{x,0}, 0\,0)$. Substituting this electric field into the left hand side of the second Maxwell equation, Eq. (A.1b), yields

$$\nabla \times \mathbf{E} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial x} \\ \frac{\partial}{\partial x} \end{pmatrix} \times \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial y}E_z - \frac{\partial}{\partial z}E_y \\ \frac{\partial}{\partial z}E_x - \frac{\partial}{\partial x}E_z \\ \frac{\partial}{\partial x}E_y - \frac{\partial}{\partial y}E_x \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{\partial}{\partial z}E_x \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ E_{x,0}\frac{\partial}{\partial z}e^{ik_z z - i\omega t} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ iE_{x,0}k_z e^{ik_z z - i\omega t} \\ 0 \end{pmatrix}. \tag{A.12}$$

The right hand side of Eq. (A.1b) yields

$$-\frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t} = -\mu_0 \frac{\partial \mathbf{H}(\mathbf{r},t)}{\partial t} = -\mu_0 \begin{pmatrix} \frac{\partial}{\partial t}H_x \\ \frac{\partial}{\partial t}H_y \\ \frac{\partial}{\partial t}H_z \end{pmatrix} = -\mu_0 \begin{pmatrix} H_{x,0}\frac{\partial}{\partial t}e^{ik_z z - i\omega t} \\ H_{y,0}\frac{\partial}{\partial t}e^{ik_z z - i\omega t} \\ 0 \end{pmatrix} = -\mu_0 \begin{pmatrix} -i\omega H_{x,0}e^{ik_z z - i\omega t} \\ -i\omega H_{y,0}e^{ik_z z - i\omega t} \\ 0 \end{pmatrix}. \tag{A.13}$$

Thus, we obtain for Eq. (A.1b)

$$\begin{pmatrix} 0 \\ iE_{x,0}k_z e^{ik_z z - i\omega t} \\ 0 \end{pmatrix} = -\mu_0 \begin{pmatrix} -i\omega H_{x,0}e^{ik_z z - i\omega t} \\ -i\omega H_{y,0}e^{ik_z z - i\omega t} \\ 0 \end{pmatrix}, \tag{A.14}$$

i.e. only the y-component of the magnetic field survives. The electric and magnetic components are related to each other via

$$E_{x,0}k_z = \mu_0 \omega H_{y,0} \tag{A.15}$$

By substituting Eq. (4.4) into Eq. (A.15), we find

$$H_{y,0} = \frac{n}{c\mu_0}E_{x,0} = \frac{n}{Z_0}E_{x,0}, \tag{A.16}$$

where

$$Z_0 = c\mu_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 376.7\,\Omega \tag{A.17}$$

is the *impedance of free space*.

In summary, we found the following properties of the electromagnetic field.

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

$$\mathbf{k}\cdot\mathbf{E}_0 = \mathbf{k}\cdot\mathbf{H}_0 = \mathbf{H}_0\cdot\mathbf{E}_0 = 0. \tag{A.18}$$

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

along the y-direction, this ratio is given by

$$H_{y,0} = \frac{n}{c\mu_0} E_{x,0} = \frac{n}{Z_0} E_{x,0}, \tag{A.19}$$

where

$$Z_0 = c\mu_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 376.7\,\Omega \tag{A.20}$$

is the *impedance of free space*.

The derivations in this section were done for plane waves. However, it can be shown that the properties of electromagnetic waves summarised in the itemisation above are valid for electromagnetic waves in general.

B

Derivation of homojunction *J-V* curves

In this appendix we present the derivations for the *J-V* curves in dark and under illumination for *p-n* homojunctions, as they were discussed in Section

B.1 Derivation of the *J-V* characteristic in dark

When an external voltage, V_a, is applied to a *p-n* junction the potential difference between the *n*-type and *p*-type regions will change and the electrostatic potential across the space-charge region will become $(\psi_0 - V_a)$. Under the forward-bias condition an applied external voltage decreases the electrostatic-potential difference across the *p-n* junction. The concentration of the minority carriers at the edge of the space-charge region increases exponentially with the applied forward-bias voltage but it is still much lower that the concentration of the majority carriers (low-injection conditions). The concentration of the majority carriers in the quasi-neutral regions do not change significantly under forward bias. The concentration of charge carriers in a *p-n* junction under forward bias is schematically presented in Fig. B.1

The concentrations of the minority carriers at the edges of the space-charge region, electrons in the *p*-type semiconductor and holes in the *n*-type semiconductor after

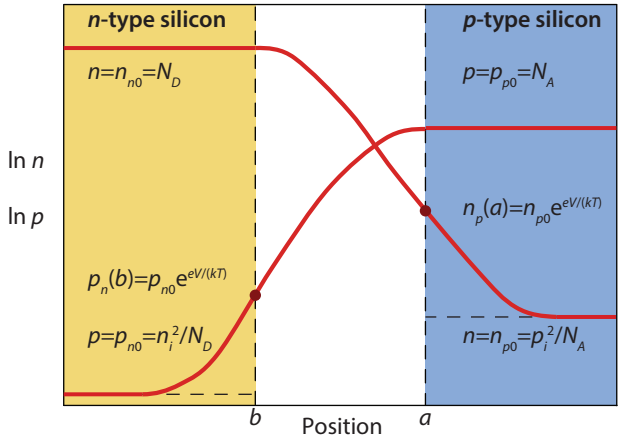

Figure B.1: Concentration profiles of mobile charge carriers in a p-n junction under forward bias (red lines). Concentration profiles of carriers under thermal equilibrium are shown for comparison (black lines).

applying forward-bias voltage are described by

$$n_p(a) = n_{p0}\,\mathrm{e}^{\frac{eV_a}{kT}} = \frac{n_i^2}{N_A}\,\mathrm{e}^{\frac{eV_a}{kT}}, \tag{B.1a}$$

$$p_n(b) = p_{n0}\,\mathrm{e}^{\frac{eV_a}{kT}} = \frac{n_i^2}{N_D}\,\mathrm{e}^{\frac{eV_a}{kT}}. \tag{B.1b}$$

Since we assume that there is no electric field in the quasi-neutral region the current-density equations of carriers reduce to only diffusion terms and are not coupled by the electric field. The current is based on the diffusive flows of carriers in the quasi-neutral regions and is determined by the diffusion of the minority carriers. The minority-carriers concentration can be calculated separately for both quasi-neutral regions. The electron-current density in the quasi-neutral region of the p-type semiconductor and the hole-current density in the quasi-neutral region of the n-type semiconductor are described by

$$J_n = \quad eD_N\frac{\mathrm{d}n}{\mathrm{d}x}, \tag{B.2a}$$

$$J_p = -eD_P\frac{\mathrm{d}p}{\mathrm{d}x}. \tag{B.2b}$$

The continuity equations (4.34) for electrons and holes in steady-state ($\partial n/\partial t = 0$ and $\partial p/\partial t = 0$) can be writ-

ten as

$$\frac{1}{e}\frac{\mathrm{d}J_N}{\mathrm{d}x} - R_N + G_N = 0, \tag{B.3a}$$

$$-\frac{1}{e}\frac{\mathrm{d}J_P}{\mathrm{d}x} - R_P + G_P = 0. \tag{B.3b}$$

Under *low-injection conditions*, a change in the concentration of the majority carriers due to generation and recombination can be neglected. However, the recombination-generation rate of minority carriers depends strongly on the injection and is proportional to the excess of minority carriers at the edges of the depletion region. The recombination-generation rate of electrons, R_n, in the p-type semiconductor and holes, R_p, in the n-type semiconductor is described by Eqs. (),

$$R_n = \frac{\Delta n}{\tau_n}, \tag{B.4a}$$

$$R_p = \frac{\Delta p}{\tau_p}, \tag{B.4b}$$

where Δn is the excess concentration of electrons in the p-type semiconductor with respect to the equilibrium concentration n_{p0}, and τ_n is the electrons (minority carriers) lifetime. Δp is the excess concentration of holes in the n-type semiconductor with respect to the equilibrium concentration p_{n0} and τ_p is the holes (minority carriers) lifetime. Δn and Δp are given by

$$\Delta n = n_p(x) - n_{p0}, \tag{B.5a}$$

$$\Delta p = p_n(x) - p_{n0}. \tag{B.5b}$$

By combining Eqs. (B.2a), (B.3a) and (B.4a) we obtain

$$D_N \frac{\mathrm{d}^2 n_p(x)}{\mathrm{d}x^2} = \frac{\Delta n}{\tau_n} - G_N. \tag{B.6a}$$

This equation describes the diffusion of electrons in the p-type semiconductor. Similarly, by combining Eqs. (B.2b), (B.3b) and (B.4b) we obtain

$$D_P \frac{\mathrm{d}^2 p_n(x)}{\mathrm{d}x^2} = \frac{\Delta p}{\tau_p} - G_P, \tag{B.6b}$$

that describes the diffusion of holes in the n-type semiconductor.

Now we substitute $n_p(x)$ from Eq. (B.5a) and $p_n(x)$ from Eq. (B.5b) into Eqs. (B.6a) and (B.6b), respectively. By using that $\mathrm{d}^2 n_{p0}/\mathrm{d}x^2 = d^2 p_{n0}/\mathrm{d}x^2 = 0$, and in dark $G_N = G_P = 0$ we obtain

$$\frac{\mathrm{d}^2 \Delta n}{\mathrm{d}x^2} = \frac{\Delta n}{D_N \tau_n}, \tag{B.7a}$$

$$\frac{\mathrm{d}^2 \Delta p}{\mathrm{d}x^2} = \frac{\Delta p}{D_P \tau_p}. \tag{B.7b}$$

The electron-concentration profile in the quasi-neutral region of the p-type semiconductor is given by the general solution to Eq. (B.7a),

$$\Delta n(x) = A \exp\left(\frac{x}{L_N}\right) + B \exp\left(-\frac{x}{L_N}\right),$$

where $L_N = \sqrt{D_N \tau_n}$ [see Eq. (7.28a)] is the electron minority-carrier diffusion length. The origin of the x axis is set to the edge of the depletion region in the p-type semiconductor and denoted as a in Fig. B.1. For the p-type semiconductor infinite thickness is assumed (*approximation of the infinite thickness*). The constants A and B can be determined from the boundary conditions

1. At $x = 0$, $n_p(a) = n_{p0} \exp(qV_a/kT)$,
2. n_p is finite at $x \to \infty$, therefore $A = 0$.

With these boundary conditions the solution for the concentration profile of electrons in the p-type quasi-neutral region is found to be

$$n_p(x) = n_{p0} + n_{p0}\left(e^{\frac{eV_a}{kT}} - 1\right) e^{-\frac{x}{L_N}}. \tag{B.8a}$$

The hole concentration profile in the quasi-neutral region of the n-type semiconductor is given by the general solution to Eq. (B.7a),

$$\Delta p(x') = A' \exp\left(\frac{x'}{L_P}\right) + B' \exp\left(-\frac{x'}{L_P}\right),$$

where $L_P = \sqrt{D_P \tau_p}$ [see Eq. (7.28b)] is the hole minority-carrier diffusion length. The origin of the x'-axis ($x' := -x$) is set at the edge of the depletion region in the n-type semiconductor and denoted as b in Fig. B.1. Here, we use the approximation of infinite thickness for the n-type semiconductor. The constants A' and B' can be determined from the boundary conditions

1. At $x' = 0$, $p_n(b) = p_{n0} \exp(qV_a/kT)$,
2. $p - n$ is finite at $x' \to \infty$, therefore $A' = 0$.

We thus find the concentration profile of holes in the quasi-neutral region of the n-type semiconductor to be

$$p_n(x') = p_{n0} + p_{n0}\left(e^{\frac{eV_a}{kT}} - 1\right) e^{-\frac{x}{L_P}}. \tag{B.8b}$$

When substituting the corresponding concentration profiles of minority carriers (Eqs.B.8) into Eqs. (B.2) we obtain for the current densities

$$J_N(x) = \frac{e D_N n_{p0}}{L_N}\left(e^{\frac{eV_a}{kT}} - 1\right) e^{-\frac{x}{L_N}}, \tag{B.9a}$$

$$J_P(x') = \frac{e D_P p_{n0}}{L_P}\left(e^{\frac{eV_a}{kT}} - 1\right) e^{-\frac{x}{L_P}}. \tag{B.9b}$$

Under the assumptions that the effect of recombination and thermal generation of carriers in the depletion region can be neglected, which means that the electron and hole current densities are essentially constant across the depletion region, one can write for the current densities at the edges of the depletion region

$$J_N|_{x=0} = J_N|_{x'=0} = \frac{e D_N n_{p0}}{L_N}\left(e^{\frac{eV_a}{kT}} - 1\right), \tag{B.10a}$$

$$J_P|_{x'=0} = J_P|_{x=0} = \frac{e D_P p_{n0}}{L_P}\left(e^{\frac{eV_a}{kT}} - 1\right). \tag{B.10b}$$

The total current density flowing through the *p-n* junction at the steady state is constant across the device therefore we can determine the total current density as the sum of the electron and hole current densities at the edges of the depletion region,

$$J(V_a) = J_N|_{x=0} + J_P|_{x=0} \qquad \text{(B.11)}$$
$$= \left(\frac{e D_N n_{p0}}{L_N} + \frac{e D_P p_{n0}}{L_P}\right)\left(e^{\frac{eV_a}{kT}} - 1\right).$$

Using Eqs. (8.1b) and (8.2b), Eq. (B.11) can be rewritten as

$$J(V_a) = J_0\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right], \qquad \text{(B.12)}$$

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

$$J_0 = \left(\frac{e D_N n_i^2}{L_N N_A} + \frac{e D_P n_i^2}{L_P N_D}\right). \qquad \text{(B.13)}$$

Equation (B.12) is known as the *Shockley equation* that describes the current-voltage behaviour of an ideal *p-n* diode. It is a fundamental equation for microelectronics device physics.

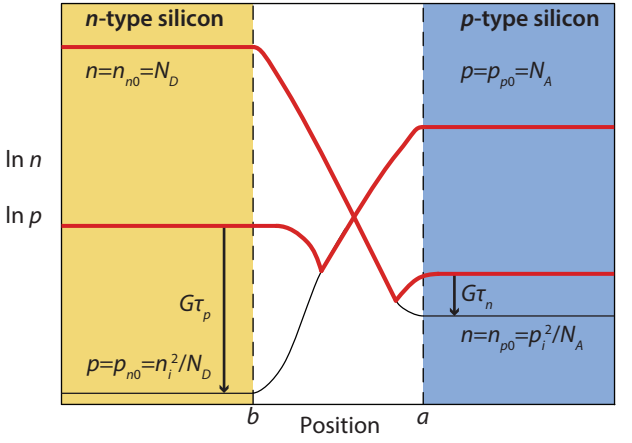

Figure B.2: Concentration profiles of mobile charge carriers in an illuminated *p-n* junction with uniform generation rate G (red line). Concentration profiles of charge carriers under equilibrium conditions are shown for comparison (black line).

B.2 Derivation of the *J-V* characteristic under illumination

When a *p-n* junction is illuminated the additional electron-hole pairs are generated through the junction. In case of moderate illumination the concentration of majority carriers does not change significantly while the concentration of minority carriers (electrons in the *p*-type region and holes in the *n*-type region) will strongly increase. In the following section it is assumed that the photo-generation rate, G, is uniform throughout the *p-n* junction, this is called the *uniform generation-rate approximation*. This assumption reflects the situation where the device is illuminated with a long-wavelength light, which is weakly absorbed by the semiconductor. The concentration of charge carriers in a *p-n* junction with uniform photo-generation rate is schematically presented in Fig. B.2.

Eqs. (B.6a) described have seen the steady-state situation for minority carriers. In section B.1 we then proceeded with assuming that the generation rate is zero. Now we analyse the case where the junction is illuminated, *i.e.* the generation rate is not zero. The equations thus can be rewritten to

$$\frac{\mathrm{d}^2\Delta n}{\mathrm{d}x^2} = \frac{\Delta n}{D_N \tau_n} - \frac{G}{D_N}, \tag{B.14a}$$

$$\frac{\mathrm{d}^2\Delta p}{\mathrm{d}x^2} = \frac{\Delta p}{D_P \tau_p} - \frac{G}{D_P}. \tag{B.14b}$$

Under the assumption that G/D_N and G/D_P are constant, the general solution to Eq. (B.14) is

$$\Delta n\,(x) = G\tau_n + C\,\mathrm{e}^{\frac{x}{L_N}} + D\,\mathrm{e}^{-\frac{x}{L_N}}, \tag{B.15a}$$

$$\Delta p\,(x') = G\tau_p + C'\,\mathrm{e}^{\frac{x'}{L_P}} + D'\,\mathrm{e}^{-\frac{x'}{L_P}}. \tag{B.15b}$$

The constants in Eqs. (B.15) can be determined from the same boundary conditions as were used in the analysis of the *p-n* junction in dark. The particular solution for the concentration profile of electrons in the quasi-neutral region of the *p*-type semiconductor and holes in the quasi-neutral region of the *n*-type semiconductor is described by

$$n_p\,(x) = n_{p0} + G\tau_n + \left[n_{p0}\left(\mathrm{e}^{\frac{eV}{kT}} - 1\right) - G\tau_n\right]\mathrm{e}^{-\frac{x}{L_N}}, \tag{B.16a}$$

$$p_n\,(x') = p_{n0} + G\tau_p + \left[p_{n0}\left(\mathrm{e}^{\frac{eV}{kT}} - 1\right) - G\tau_p\right]\mathrm{e}^{-\frac{x'}{L_P}}. \tag{B.16b}$$

By substituting these concentration profiles of minority carriers into Eqs. (B.2), we obtain for the current densit-

ies

$$J_N(x) = \frac{eD_N n_{p0}}{L_N}\left(e^{\frac{qV}{kT}} - 1\right) e^{-\frac{x}{L_N}} - eGL_N e^{-\frac{x}{L_N}}, \tag{B.17a}$$

$$J_P(x') = \frac{eD_P p_{n0}}{L_P}\left(e^{\frac{qV}{kT}} - 1\right) e^{-\frac{x'}{L_P}} - eGL_P e^{-\frac{x'}{L_P}}. \tag{B.17b}$$

In case of an ideal *p-n* junction the effect of recombination in the depletion region was neglected. However, the contribution of photo-generated charge carriers to the current in the depletion region has to be taken into account. The contribution of optical generation from the depletion region to the current density is given by

$$J_N|_{x=0} = e\int_{-W}^{0} -G\,dx = -eGW, \tag{B.18a}$$

$$J_P|_{x'=0} = e\int_{-W}^{0} -G\,dx' = -eGW. \tag{B.18b}$$

The total current density flowing through the *p-n* junction in the steady state is constant across the junction. Under the *superposition approximation* we thus can determine the total current density as the sum of the electron and hole current densities at the edges of the depletion region,

$$\begin{aligned} J(V_a) =& J_N|_{x=0} + J_P|_{x=0} \\ =& \left(\frac{eD_N n_{p0}}{L_N} + \frac{eD_P p_{n0}}{L_P}\right)\left(e^{\frac{eV_a}{kT}} - 1\right) \\ & - eG(L_N + L_P + W) \end{aligned} \tag{B.19}$$

Eq. (B.19) can be rewritten as

$$J(V_a) = J_0\left[\exp\left(\frac{qV_a}{kT}\right) - 1\right] - J_{\text{ph}}, \tag{B.20}$$

where J_{ph} is the photo-current given by

$$J_{\text{ph}} = eG(L_N + L_P + W). \tag{B.21}$$

A number of approximations have been made in order to derive the analytical expressions for the current-voltage characteristics of an ideal *p-n* junction in dark and under illumination. These approximations are summarised below:

- The depletion-region approximation
- The Boltzmann approximation
- Low-injection conditions
- The superposition principle
- Infinite thickness of doped regions
- Uniform generation rate

The derived expressions describe the behaviour of an ideal *p-n* junction and help us understand the basic processes behind the operation of the *p-n* junction. However, they do not fully and correctly describe real *p-n* junctions. For example, the thickness of a real *p-n* junction is limited, which means that the recombination at the surface of the doped regions has to be taken into account. The thinner the *p-n* junction, the more important surface recombination becomes. The surface recombination modifies the value of the saturation-current density. Further it was assumed that there are no recombination-generation processes in the depletion region. However, in real *p-n* junctions, the recombination in the depletion region represents a substantial loss mechanism. These and other losses in a solar cell are discussed in Chapter 10.

C

Atomic force microscopy and statistical parameters

C.1 Atomic force microscopy (AFM)

AFM[1] is a powerful technique with which the morphology of nano-textured surfaces can be determined [128]. During an AFM measurement a tiny probe with a tip radius of several nanometers is brought so close to the surface that atomic forces between the surface and the probe become important. Due to these forces the surface morphology can be determined. Under *amplitude modulation* the amplitude and the phase of the tip will change during the scan depending on the Van der Waals forces, *i.e.* the distance between the tip and the sample. Several modes can be used. We used the so called *tapping mode* in which the probe is oscillating at its resonance frequency. From the obtained amplitude and phase pictures the height profile of the investigated sample can be obtained.

The height profile is not the *real* height profile of the sample but a convolution of the height profile and the shape of the tip. A blunt tip therefore will lead to a blurry picture. If the tip is broken and therefore consists of two or more peaks, the features on the scan will be doubled. Further, scanning the surface too fast will lead to contact loss between the tip and the sample at steep features. Therefore the AFM measurements have to be performed very carefully.

Figure C.1 shows a picture of a typical tip of AFM

[1] This appendix is taken from the PhD thesis of Klaus Jäger [53].

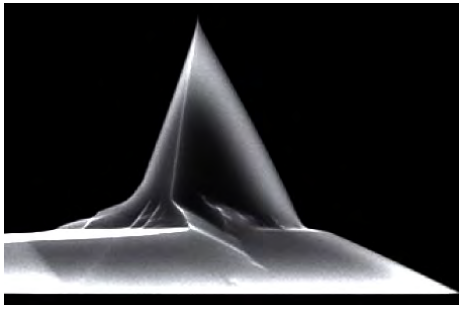

Figure C.1: The NSG 10 gold-coated single-crystal silicon tip used throughout this project [129].

probes. This so-called NSG 10 tip is made from antimony doped single-crystal silicon and coated with gold. Its tip radius is typically 6 nm, with a guaranteed radius of maximal 10 nm [129].

C.2 Statistical parameters

The AFM scan reveals three dimensional data of the nano-textured surface via the height function $z(x, y)$, which can be used to extract statistical properties of the samples. To compare the different samples we use two different parameters, the *root-mean-squared (rms) roughness* σ_r and the *correlation length* ℓ_c.

The rms roughness in principle is the standard deviation of the height profile. It is defined as

$$\sigma_r = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(z_i - \bar{z})^2}, \tag{C.1}$$

where N is the number of data points, z_i is the height of the i^{th} datapoint and $\bar{z}$ is the average height. It becomes clear from the definition that σ_r is insensitive to the lateral feature sizes. Samples with very small or large lateral features both can have the same rms roughness.

The correlation length gives an indication of the lateral feature sizes. Its derivation is less straightforward then that of σ_r: it has to be extracted from autocorrelation function (ACF) and/or the height-height correlation function (HHCF) [130]. For a discrete set of data, the two-dimensional ACF is given by

$$\mathrm{ACF}(\tau_x, \tau_y) = \frac{1}{(N-n)(M-m)} \cdot \sum_{l=1}^{N-n}\sum_{k=1}^{M-m} z(k\delta + \tau_x,\, l\delta + \tau_y)\, z(k\delta,\, l\delta), \tag{C.2}$$

where δ is the distance between two data points, $m = \tau_x/\delta$ and $n = \tau_y/\delta$. For AFM scans usually the one-dimensional ACF along the fast scanning axis (x) is

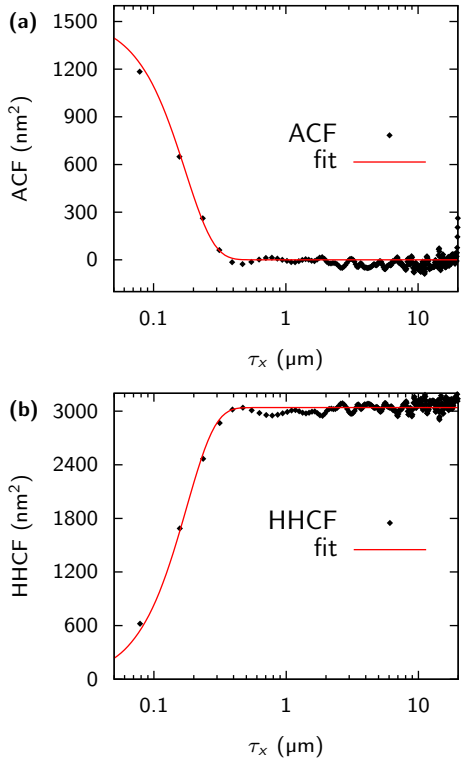

Figure C.2: The ACF, HHCF and fitted Gaussian functions for SnO_2:F.

used:

$$\mathrm{ACF}_x(\tau_x) = \mathrm{ACF}(\tau_x, 0) = \frac{1}{N(M-m)} \sum_{l=1}^{N} \sum_{k=1}^{M-m} z(k\delta + \tau_x,\, l\delta)\, z(k\delta,\, l\delta). \tag{C.3}$$

The one-dimensional HHCF is given by

$$\mathrm{HHCF}_x(\tau_x) = \frac{1}{N(M-m)} \sum_{l=1}^{N} \sum_{k=1}^{M-m} \left[z(k\delta + \tau_x,\, l\delta) - z(k\delta, l\delta)\right]^2. \tag{C.4}$$

To determine ℓ_c, Gaussian functions can be fitted to the ACF and the HHCF. They are given by

$$\mathrm{ACF}_x^{\mathrm{fit}}(\tau_x) = \sigma_r^2 \exp\left(-\frac{\tau_x^2}{\ell_c^2}\right), \tag{C.5}$$

$$\mathrm{HHCF}_x^{\mathrm{fit}}(\tau_x) = 2\sigma_r^2 \left[1 - \exp\left(-\frac{\tau_x^2}{\ell_c^2}\right)\right], \tag{C.6}$$

respectively. The correlation length then is the length at which the (fitted) ACF has decayed to $1/e$ of its highest value. Instead of applying Eq. (C.1) directly, the rms roughness also can be obtained by fitting Gaussian functions to the ACF and/or the HHCF.

Figure C.2 shows the ACF and the HHCF and fitted Gaussian functions for the SnO_2:F sample depicted in Fig. 13.3 (a). Instead of fitting to Gaussian functions, some authors also use exponential functions for fitting. For this sample, Gaussians lead to better fits.

Bibliography

[1] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Vol. 1 (Addison-Wesley, Menlo Park, California, 1963) Chap. 4.

[2] OECD/IEA, *2013 Key World Energy Statistics* (Paris, France, 2013).

[3] TomTheHand, (2013), commons.wikimedia.org/wiki/File:Brent_Spot_monthly.svg and commons.wikimedia.org/wiki/File:Oil_Prices_1861_2007.svg.

[4] OECD/IEA, *World Energy Outlook* (Paris, France, 2013).

[5] OECD/IEA, *Energy Balances of non-OECD Countries 2010* (Paris, France, 2010).

[6] R. A. Rhode, (2006), www.globalwarmingart.com/wiki/File:Carbon_Dioxide_400kyr_Rev_png.

[7] T. F. Stocker, D. Qin, G.-K. Plattner, M. M. B. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, eds., *Climate Change 2013 - The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, United Kingdom, 2014).

[8] L. Freris and D. Infield, *Renewable Energy in Power Systems* (John Wiley & Sons Inc, Chichester, United Kingdom, 2008).

[9] M. Beerepoot and A. Marmion, *Policies for renewable heat* (OECD/IEA, Paris, France, 2012).

[10] The Worldbank, (2014), wdi.worldbank.org/table/3.7#.

[11] Centraal Bureau voor de Statistiek, (2014), statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=80030NED&D1=0,3&D2=0&D3=a&D4=14-15&HD=140818-2139&HDR=T,G1,G3&STB=G2.

[12] Ministry of Energy and Mines, (2007).

[13] G. Masson, M. Latour, M. Rekinger, I.-T. Theologitis, and M. Papoutsi, "Global Market Outlook for Photovoltaics 2013-2017," European Photovoltaic Industry Association, Brussels, Belgium (2013).

[14] G. Masson, S. Orlandi, and M. Rekinger, "Global Market Outlook for Photovoltaics 2014-2018," European Photovoltaic Industry Association, Brussels, Belgium (2014).

[15] S. Chowdhury, U. Sumita, A. Islam, and I. Bedja, Energy Policy **68**, 285 (2014).

[16] K. Ardani and R. Margolis, *2010 Solar Technologies Market Report*, Tech. Rep. DOE/GO-102011-3318 (U.S. Department of Energy, Energy Efficiency & Renewable Energy, Washington, DC, 2011).

[17] R. P. Raffaele, *Current Trends in Photovoltaics*, Santa Barbara Summit on Energy Efficiency (The Institute for Energy Efficiency, Santa Barbara, California, 2011).

[18] V. Shah, J. Booream-Phelps, and S. Mie, "2014 Outlook: Let the Second Gold Rush Begin," Deutsche Bank, New York, New York (2014).

[19] A. Einstein, Ann. Phys. (4) **322**, 132 (1905).

[20] D. Neamen, *Semiconductor Device Physics: Basic Principles*, 4th ed. (McGraw-Hill, New York, NY, USA, 2012).

[21] P. Würfel, *Physics of Solar Cells* (WILEY-VCH Verlag, Weinheim, Germany, 2005).

[22] G. Conibeer, Mater. Today **10**, 42 (2007).

[23] W. Shockley and H. J. Queisser, J. Appl. Phys. **32**, 510 (1961).

[24] NASA/Goddard, (2012), www.nasa.gov/mission_pages/sunearth/multimedia/Sunlayers.html.

[25] B. Mills, (2007), commons.wikimedia.org/wiki/File:Silicon-unit-cell-3D-balls.png.

[26] S. M. Sze, *Semiconductor Devices Physics and Technology*, 2nd ed. (John Wiley & Sons Inc, New York, NY, 2002).

[27] W. Shockley and W. T. Read, Phys. Rev. **87**, 835 (1952).

[28] R. N. Hall, Phys. Rev. **87**, 387 (1952).

[29] A. Richter, S. W. Glunz, F. Werner, J. Schmidt, and A. Cuevas, Phys. Rev. B **86**, 165202 (2012).

[30] T. Markvart, Appl. Phys. Lett. **91**, 064102 (2007).

[31] T. Markvart, Phys. Status Solidi A **205**, 2752 (2008).

[32] T. Markvart and G. H. Bauer, Appl. Phys. Lett. **101**, 193901 (2012).

[33] R. J. van Overstraeten and R. P. Mertens, *Physics, technology and use of photovoltaics* (A. Hilger, Bristol, United Kingdom, 1986).

[34] Chetvorno, (2012), commons.wikimedia.org/wiki/File:Optical_flat_interference.svg.

[35] D. M. Chapin, C. S. Fuller, and G. L. Pearson, J. Appl. Phys. **25**, 676 (1954).

[36] D. C. Reynolds, G. Leies, L. L. Antes, and R. E. Marburger, Phys. Rev. **96**, 533 (1954).

[37] R. Bergmann and J. Werner, Thin Solid Films **403 - 404**, 162 (2002), proceedings of Symposium P on Thin Film Materials for Photovoltaics.

[38] M. A. Green, K. Emery, Y. Hishikawa, W. Warta, and E. D. Dunlop, Prog. Photovolt: Res. Appl. **22**, 1 (2014).

[39] United States Geological Survey, (2003), http://www.usgs.gov.

[40] K. L. Chopra, P. D. Paulson, and V. Dutta, Prog. Photovolt: Res. Appl. **12**, 69 (2004).

[41] "PHOTON module price index," (2012), 25 May, www.photon-international.com/newsletter/document/65647.pdf.

[42] J. Poortmans and V. Arkhipov, eds., *Thin Film Solar Cells: Fabrication, Characterization and Applications* (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2006).

[43] K. Bädeker, Ann. Phys. (4) **327**, 749 (1907).

[44] T. Koida, H. Fujiwara, and M. Kondo, Jpn. J. Appl. Phys. **46**, L685 (2007).

[45] O. Kluth, *Texturierte Zinkoxidschichten für Silizium-Dünnschichtsolarzellen*, Ph.D. thesis, RWTH Aachen, Forschungszentrum Jülich, Germany (2001).

[46] P. Drude, Ann. Phys. (4) **306**, 566 (1900).

[47] N. W. Ashcroft and N. D. Mermin, *Solid State Physics, HRW International Editions* (Saunders College, Philadelphia, PA, USA, 1976).

[48] D. Mergel and Z. Qiao, J. Phys. D Appl. Phys. **35**, 794 (2002).

[49] F. Ruske, A. Pflug, V. Sittinger, B. Szyszka, D. Greiner, and B. Rech, Thin Solid Films **518**, 1289 (2009).

[50] J. Sap, O. Isabella, K. Jäger, and M. Zeman, Thin Solid Films **520**, 1096 (2011).

[51] E. Fortunato, D. Ginley, H. Hosono, and D. C. Paine, MRS Bull. **32**, 242 (2007).

[52] K. H. Wedepohl, Geochim. Cosmochim. Ac. **59**, 1217 (1995).

[53] K. Jäger, *On the Scalar Scattering Theory for Thin-Film Solar Cells*, Ph.D. thesis, Delft University of Technology, Delft, the Netherlands (2012).

[54] K. Sato, Y. Gotoh, Y. Wakayama, Y. Hayashi, K. Adachi, and N. Nishimura, Rep. Res. Lab., Asahi Glass Co. Ltd. **42**, 129 (1992).

[55] D. Dominé, P. Buehlmann, J. Bailat, A. Billet, A. Feltrin, and C. Ballif, Phys. Status Solidi-R **2**, 163 (2008).

[56] O. Kluth, B. Rech, L. Houben, S. Wieder, G. Schöpe, C. Beneking, H. Wagner, A. Löffl, and H. Schock, Thin Solid Films **351**, 247 (1999).

[57] M. Berginski, J. Hüpkes, W. Reetz, B. Rech, and M. Wuttig, Thin Solid Films **516**, 5836 (2007).

[58] B. Mills, (2007), commons.wikimedia.org/wiki/File:Gallium-arsenide-unit-cell-3D-balls.png.

[59] H. Binder, *Lexikon der chemischen Elemente* (Hirzel Verlag, Stuttgart, Germany, 1999).

[60] A. Tanaka, Toxicol. Appl. Pharm. **198**, 405 (2004).

[61] R. A. Street, *Hydrogenated amorphous silicon* (Cambridge University Press, Cambridge, UK, 1991).

[62] M. A. Wank, *Manipulating the Hydrogenated Amorphous Silicon Growing Surface*, Ph.D. thesis, Delft University of Technology, Delft, the Netherlands (2011).

[63] O. Vetterl, F. Finger, R. Carius, P. Hapke, L. Houben, O. Kluth, A. Lambertz, A. Mück, B. Rech, and H. Wagner, Solar Energy Materials and Solar Cells **62**, 97 (2000).

[64] D. E. Carlson and C. R. Wronski, Appl. Phys. Lett. **28**, 671 (1976).

[65] D. L. Staebler and C. R. Wronski, Appl. Phys. Lett. **31**, 292 (1977).

[66] F. Dross, K. Baert, T. Bearda, J. Deckers, V. Depauw, O. El Daif, I. Gordon, A. Gougam, J. Govaerts, S. Granata, R. Labie, X. Loozen, R. Martini, A. Masolin, B. O'Sullivan, Y. Qiu, J. Vaes, D. Van Gestel, J. Van Hoeymissen, A. Vanleenhove, K. Van Nieuwenhuysen, S. Venkatachalam, M. Meuris, and J. Poortmans, Prog. Photovolt: Res. Appl. **20**, 770 (2012).

[67] B. Mills, (2007), commons.wikimedia.org/wiki/File:Chalcopyrite-unit-cell-3D-balls.png.

[68] M. Contreras, L. Mansfield, B. Egaas, J. Li, M. Romero, R. Noufi, E. Rudiger-Voigt, and W. Mannstadt, in *37th IEEE Photovoltaic Specialists Conference (PVSC)* (2011) pp. 000026–000031.

[69] H. Zhu, J. Hüpkes, E. Bunte, and S. Huang, Appl. Surf. Sci. **256**, 4601 (2010).

[70] A. Chirilă, P. Reinhard, F. Pianezzi, P. Bloesch, A. R. Uhl, C. Fella, L. Kranz, D. Keller, C. Gretener, H. Hagendorfer, D. Jaeger, R. Erni, S. Nishiwaki, S. Buecheler, and A. N. Tiwari, Nat. Mater. **12**, 1107 (2013).

[71] J. Fritsche, D. Kraft, A. Thißen, T. Mayer, A. Klein, and W. Jaegermann, Thin Solid Films **403 - 404**, 252 (2002), proceedings of Symposium P on Thin Film Materials for Photovoltaics.

[72] The Economist, "Solar power: A painful eclipse," (2011), 15 October.

[73] R. U. Ayres and L. Ayres, *A Handbook of Industrial Ecology* (Edward Elgar Publishing, Cheltenham, UK, 2002).

[74] N. Gupta, G. F. Alapatt, R. Podila, R. Singh, and K. F. Poole, Int. J. Photoenergy **2009**, 154059 (2009).

[75] W. van Sark, J. de Wild, J. Rath, A. Meijerink, and R. E. Schropp, Nanoscale Res. Lett. **8**, 81 (2013).

[76] W. G. J. H. M. van Sark, A. Meijerink, and R. E. I. Schropp, "Third generation photovoltaics," (InTech Europe, Rijeka, Croatia, 2012) Chap. 1, p. 1.

[77] D. Jurbergs, E. Rogojina, L. Mangolini, and U. Kortshagen, Appl. Phys. Lett. **88**, 233116 (2006).

[78] O. E. Semonin, J. M. Luther, S. Choi, H.-Y. Chen, J. Gao, A. J. Nozik, and M. C. Beard, Science **334**, 1530 (2011).

[79] K. Tanabe, Electron. Lett. **43**, 998 (2007).

[80] A. Luque, A. Marti, and C. Stanley, Nat Photon **6**, 146 (2012).

[81] D. König, K. Casalenuovo, Y. Takeda, G. Conibeer, J. Guillemoles, R. Patterson, L. Huang, and M. Green, Physica E **42**, 2862 (2010), 14th International Conference on Modulated Semiconductor Structures.

[82] Gehrlicher Solar, (2010), commons.wikimedia.org/wiki/File:Energiepark_Lauingen_Gehrlicher_Solar_AG.JPG.

[83] I. Reda and A. Andreas, *Solar Position Algorithm for Solar Radiation Applications*, Tech. Rep. NREL/TP-560-34302 (National Renewable Energy Laboratory, 2008).

[84] Univ. of Oregon Solar Radiation Monitoring Laboratory, (2008), solardat.uoregon.edu/PolarSunChartProgram.php.

[85] F. Kasten and A. T. Young, Appl. Opt. **28**, 4735 (1989).

[86] E. Laue, Solar Energy **13**, 43 (1970).

[87] C. Honsberg and S. Bowden, (2014), pveducation.org/pvcdrom/properties-of-sunlight/air-mass.

[88] C. Rigollier, O. Bauer, and L. Wald, Solar Energy **68**, 33 (2000).

[89] I. Tobías, C. del Cañizo, and J. Alonso, *Handbook of Photovoltaic Science and Engineering*, edited by A. Luque and S. Hegedus (John Wiley & Sons Ltd, Chichester, England, 2003) Chap. 7, pp. 255–306.

[90] École polytechnique fédérale de Lausanne, PV-Lab, (2012), pvlab.epfl.ch/pv_module_design/module_reliability/efficient_encapsulation.

[91] J. Wohlgemut, *IEC 61215: What it is and isn't*, Tech. Rep. NREL/PR-5200-54714 (National Renewable Energy Laboratory, 2012).

[92] K. Jäger, J. Lenssen, P. Veltman, and E. Hamers, in *28^{th} European Photovoltaic Solar Energy Conference and Exhibition* (2013) p. 2164.

[93] A. Falk, M. Meinhardt, and V. Wachenfeld, in *Proceedings of the European Conference on Power Conversion and Intelligent Motion (PCIM), Nuremberg, Germany* (2009) pp. 14–20.

[94] S. Araujo, P. Zacharias, and R. Mallwitz, IEEE T. Ind. Electron. **57**, 3118 (2010).

[95] J. M. A. Myrzik and M. Calais, in *Power Tech Conference Proceedings, 2003 IEEE Bologna*, Vol. 2 (2003) pp. 8 pp. Vol.2–.

[96] C. Rodriguez and J. Bishop, IEEE T. Ind. Electron. **56**, 4332 (2009).

[97] N. Mohan, T. M. Undeland, and W. P. Robbins, *Power Electronics: Converters, Applications, and Design* (John Wiley & Sons Inc, Hoboken, NJ, 2003).

[98] K. Mertens, *Photovoltaics: Fundamentals, Technology and Practice* (John Wiley & Sons Ltd, Chichester, United Kingdom, 2014) texbook-pv.org.

[99] State of California, (2014), www.gosolarcalifornia.ca.gov/equipment/inverter_tests/summaries/.

[100] C. Honsberg and S. Bowden, (2014), pvcdrom.pveducation.org/BATTERY/charlead.htm.

[101] GeoModel Solar s.r.o., (2014), solargis.info/doc/71.

[102] E. A. de la Breteque, Solar Energy **83**, 1425 (2009).

[103] P. Trinuruk, C. Sorapipatana, and D. Chenvidhya, Renewable Energy **34**, 2515 (2009).

[104] M. K. Fuentes, *A Simplified Thermal Model for Flat-Plate Photovoltaic Arrays*, Tech. Rep. (Sandia National Laboratories, 1987).

[105] A. Jones and C. Underwood, Solar Energy **70**, 349 (2001).

[106] H. E. A. van den Akker and R. F. Mudde, *Fysische transportverschijnselen*, 3rd ed. (VSSD, 2008) p. 299.

[107] M. Muller, B. Marion, and J. Rodriguez, in *Photovoltaic Specialists Conference (PVSC), 2012 38th IEEE* (2012) pp. 000697–000702.

[108] J. Randall and J. Jacot, Renewable Energy **28**, 1851 (2003).

[109] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, Prog. Photovolt: Res. Appl. **19**, 757 (2011).

[110] W. Bower, M. Behnke, W. Erdman, and C. Whitaker, "Performance test protocol for evaluating inverters used in grid-connected photovoltaic systems," (2004), www.gosolarcalifornia.org/equipment/documents/2004-11-22_Test_Protocol.pdf.

[111] D. L. King, S. Gonzalez, G. M. Galbraith, and W. E. Boyson, *Performance Model for Grid-Connected Photovoltaic Inverters*, Tech. Rep. SAND2007-5036 (Sandia National Laboratories, 2007).

[112] Sandia National Laboratories, "Inverter library," (2014), sam.nrel.gov/sites/sam.nrel.gov/files/sam-library-sandia-inverters-2014-1-14.csv.

[113] A. Driesse, P. Jain, and S. Harrison, in *33rd IEEE Photovoltaic Specialists Conference (PVSC)* (2008) pp. 1–6.

[114] A. McEvoy, T. Markvart, and L. Castañer, eds., *Practical Handbook of Photovoltaics*, 2nd ed. (Academic Press, Boston. Massachusetts, 2012).

[115] Deutsche Gesellschaft für Sonnenenergie, *Planning and Installing Photovoltaic Systems: A Guide for Installers, Architects and Engineers*, Planning and Installing Series (Earthscan, London, United Kingdom, 2008).

[116] H. Vogel, *Gerthsen Physik*, 19th ed. (Springer-Verlag, Berlin, Germany, 1997).

[117] "Residential Energy Consumption Survey (RECS) 2009," (2009).

[118] K. Jäger, (2014), personal communication.

[119] V. Ramanathan and Y. Feng, Atmospheric Environment **43**, 37 (2009), atmospheric Environment - Fifty Years of Endeavour.

[120] A. Einstein and L. Szilard, "Refrigeration," (1930), US Patent 1,781,541.

[121] C. Andraka, Energy Procedia **49**, 684 (2014).

[122] KJkolb, (2005), commons.wikimedia.org/wiki/File:Parabolic_trough_solar_thermal_electric_power_plant_1.jpg.

[123] R. Montoya, (2014), share.sandia.gov/news/resources/news_releases/images/2014/areva.jpg.

[124] B. Appel, (2005), commons.wikimedia.org/wiki/File:Dish-stirling-at-odeillo.jpg.

[125] afloresm, (2007), commons.wikimedia.org/wiki/File:PS10_solar_power_tower_2.jpg.

[126] H. Zhang, J. Baeyens, J. Degrève, and G. Cacères, Renew. Sust. Energ. Rev. **22**, 466 (2013).

[127] F. F. Abdi, L. Han, A. H. M. Smets, M. Zeman, B. Dam, and R. van de Krol, Nat. Commun. **4**, 2195 (2013).

[128] F. J. Giessibl, Rev. Mod. Phys. **75**, 949 (2003).

[129] *NT-MDT Catalogue on SPM Accessories*, Tech. Rep. (NT-MDT, Moscow,Russia, 2012) http://www.ntmdt-tips.com/data/media/nt-mdt_afm_probes_29_03_12.pdf.

[130] P. Klapetek, *Characterization of randomly rough surfaces in nanometric scale using methods of modern metrology*, Ph.D. thesis, Masaryk University, Brno, Czech Republic (2003).