

Passivating selective contacts for silicon photovoltaics : solar cells designed by physics

Citation for published version (APA):

Smit, S. (2016). *Passivating selective contacts for silicon photovoltaics : solar cells designed by physics*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Applied Physics and Science Education]. Technische Universiteit Eindhoven.

Document status and date:

Published: 14/06/2016

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Passivating selective contacts for silicon
photovoltaics
Solar cells designed by physics

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een
commissie aangewezen door het College voor
Promoties, in het openbaar te verdedigen
op dinsdag 14 juni 2016 om 16:00 uur

door

Sjoerd Smit

geboren te Waalwijk

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter: prof.dr. H.J.H. Clercx
1^e promotor: prof.dr.ir. W.M.M. Kessels
2^e promotor: prof.dr. M. Zeman (TUD)
leden: prof.dr.ir. P.P.A.M. van der Schoot
prof.dr. J. Gómez Rivas
Prof.Dr. K. Lips (Freie Universität Berlin)
Dr.rer.nat. P.P. Altermatt (Trina Solar Ltd.)
adviseur: dr. P.C.P. Bronsveld (ECN Solar Energy)

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Funding information

This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs (FLASH project 1.4: Advanced Interface Engineering for Si Heterojunction Solar Cells; STW project number 12167).



connecting innovators

Contents

1	Introduction	7
2	The physics of solar cells: semiconductor physics enhanced by thermodynamics	23
2.1	The theory of non-equilibrium thermodynamics	27
2.1.1	Literature discussion	29
2.2	Overview of NET_{PV}	29
2.3	The thermodynamic formulation of solar cell	
	physics step 1: Equilibrium and thermodynamic variables . . .	31
2.3.1	Entropy	40
2.3.2	Generalisation to non-equilibrium	43
2.4	The thermodynamic formulation of solar cell	
	physics step 2: Interaction with the environment	47
2.5	The thermodynamic formulation of solar cell	
	physics step 3: Continuity equations	50
2.5.1	The entropy source term q_s	53
2.5.2	Loss processes in the bulk of a solar cell and the entropy generation rate	54
2.6	The thermodynamic formulation of solar cell	
	physics step 4: Onsager theory	57
2.7	The thermodynamic formulation of solar cell	
	physics step 5: Symmetry, the Onsager reciprocal relations and the Curie principle	64
2.7.1	The Onsager reciprocal relations	64
2.7.2	The Curie principle	65
2.8	Wrap up: what does thermodynamics tell us?	71
3	The physics of solar cells: semiconductor physics enhanced by thermodynamics	77
3.1	The thermodynamics of the selective membrane model	77
3.2	The selective membrane model	78
3.2.1	What is selectivity?	89

3.3	The selectivity of Schottky-type junctions	92
3.4	The selectivity of homojunctions	97
3.5	The selectivity of heterojunctions	99
3.6	Generation and recombination in selective membranes	105
3.7	Conclusions	109
4	Variational method for the minimisation of entropy generation in solar cells	115
4.1	Introduction	115
4.2	Entropy production in solar cells	117
4.3	Variational solar cell optimisation	123
	Appendices	127
4.A	Boundary conditions for the VEGM method	127
5	Metal-oxide-based hole-selective tunnelling contacts for crystalline silicon solar cells	133
5.1	Introduction	133
5.2	Materials and methods	135
5.3	Theory	138
5.3.1	The ZnO/Al ₂ O ₃ / <i>c</i> -Si system as a selective contact . . .	138
5.3.2	Modeling of the tunnelling currents	138
5.3.3	Considerations for improving the contact resistance and surface passivation	140
5.4	Results and discussion	143
5.4.1	Passivation of <i>c</i> -Si by Al ₂ O ₃ /ZnO stacks	143
5.4.2	Charge transport through Al ₂ O ₃ /ZnO stacks	147
5.4.3	Improved <i>c</i> -Si passivation by <i>a</i> -Si:H/Al ₂ O ₃ /ZnO stacks	147
5.5	Conclusions and outlook	149
6	Properties of graphene and its potential applications in silicon solar cells	155
6.1	The electrical and optical properties of graphene	156
6.1.1	The graphene dispersion relation and density of states .	156
6.1.2	Conduction and photon absorption in graphene	162
6.2	Junctions of graphene and silicon	170
6.2.1	Equilibrium conditions	171
6.2.2	Non-equilibrium conditions	173
6.2.3	Graphene as a carrier selective contact	181
6.3	Conclusions	182
	Appendices	185
6.A	Approximate expressions for the graphene carrier densities . . .	185

A Mathematical formulation of the charge transport in solar cells	195
A.1 The drift-diffusion equations	195
A.2 The Poisson-Boltzmann equation	199
Summary	205
List of publications	207
Acknowledgements	208
CV	210

Chapter 1

Introduction

As the title of this thesis makes abundantly clear, this work is about the physics of solar cells. Our choice for this title deserves some extra clarification, since the physics of solar cells is a very broad and developed field and one would probably expect to find such a title on a student handbook instead of a PhD thesis. There already exist sophisticated software packages that have implemented detailed physical and mathematical models and that are capable of predicting the efficiencies of complex cell designs, so the physics of solar cells might appear to be a closed chapter for the most part; a finished work. Then why do we devote a PhD thesis to it? Our short answer to that question is, bluntly, that we still need to understand solar cells better than we currently do. Naturally, when we talk about understanding solar cells, we talk about improving them¹ since photovoltaics (PV) is not a purely academic subject that we study just for the science of it. The question then rises what the role of physics is (or should be) in the design of solar cells. The ideal we normally have of science and physics is that we study a subject to learn about it and then use this knowledge improve what whatever it was that needed improvement. However, this ideal is not as straightforward to realise as we would often like and the step from knowledge to improvement is not always trivial (as is often assumed, unfortunately). We can ask a computer to calculate the efficiency a certain cell design and we would get a very accurate answer, but what we would much rather do is ask the computer to calculate a cell design that achieves, e.g., a 26 % efficiency given certain practical constraints. The physics of solar cells may provide us with very accurate models, but in the end it seems that for solar cell design we still need something else; something nebulous we usually call “ingenuity”, “art” or “deeper understanding” or something like that.

¹The term “improvement” can mean many different things in this context. Indeed, there is a very fruitful discussion to be had about what we mean by “improvement of solar cells”. For the purposes of this work, we will mostly be talking about cell efficiency since that gives us a clear physical and thermodynamical measure of performance. For the argument in this paragraph, though, any other metric of performance can be substituted if necessary.

So let us try to analyse this limitation of current solar cell physics a bit further. The first thing to note is that physics in general is, in its roots, a science based on an alternating loop of observation and prediction. A physicist observes a system and based on his observations he makes predictions. Next, he tests his predictions by further observations. However, if we want to improve a solar cell, we are not primarily interested in continually making predictions and then observing if the cell does indeed follow the mathematical model we made of it. Instead, what we are effectively trying to achieve is to “force” a particular observation we are interested in: the cell that performs best (see Figure 1.1). So while physics is very good at making good models of solar cells, that does not immediately mean that these models are immediately useful for making good cells. One could say that the problem of solar cell design is an inverse problem (i.e, one is trying to improve the observation rather than the model that makes the predictions) compared to ordinary empirical science.²

At the start of this research project in November 2011, we set out to gain a better understanding of the crystalline silicon (*c*-Si) heterojunction solar cell (abbreviated as SHJ cell) in order to come up with other potentially interesting and cost-efficient ways to create a cell out of a wafer. At the time there was still some debate about the success of the heterojunction between hydrogenated amorphous silicon (*a*-Si:H) and *c*-Si (see Figure 1.2), which was –and at the time of writing still is– achieving unusually high open circuit voltages (V_{OC}) for a silicon solar cell. Early 2012, De Wolf *et al.* of EPFL published a review of this type of solar cell [1] and from this paper it becomes clear that the success of the *a*-Si:H/*c*-Si heterojunction was only partially understood. The low defect density of the *a*-Si:H/*c*-Si interface was mentioned as an important contributor to the high V_{OC} , but it was clear that this alone cannot be the whole story. De Wolf *et al.* write:

“Key to the success of SHJ devices is the separation of highly recombination-active (ohmic) contacts from the crystalline surface by insertion of a passivating, semiconducting film with a wider bandgap [2]. For SHJ devices, ideally, charge trickles through this buffer layer sufficiently slowly to build up a high voltage, but fast enough to avoid carriers recombining before being collected. The buffer layer may thus be considered as a semi-permeable membrane for carrier extraction [3].”

It is very significant to note that the authors here make no use of terms like “space charge region” (SCR), “drift current”, and “diffusion current” that

²Without getting to far ahead of ourselves, it is worth mentioning here that there is one branch of physics that specialises in these kinds of optimisation (i.e., engineering) problems. We are talking, of course, about thermodynamics. There is perhaps a certain irony in the fact that the physical theory that deals with optimisation of energy conversion is also the one that does not really care about microscopic details. Our first instinct as scientist to understand something is to take it apart, but are more details always the best way to get to a better device?

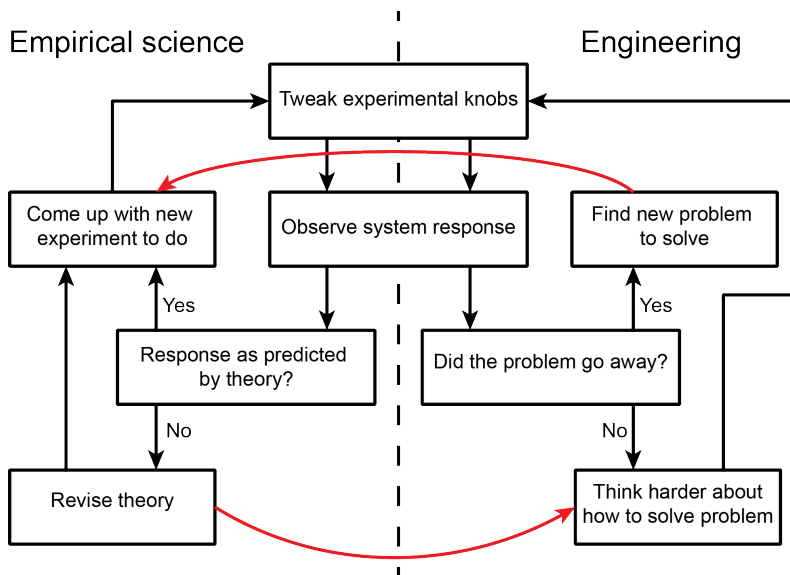


Figure 1.1: The contrast between learning about solar cells and improving them. Both disciplines are rooted in experiment and observation, but for science the ultimate goal is to improve the mathematical models of solar cells while engineering is concerned with improving the cell itself. Of course, the two are not mutually exclusive: engineering can lead to interesting new questions science can investigate (top red arrow) while the scientific models can help to find new solutions to engineering problems (bottom red arrow).

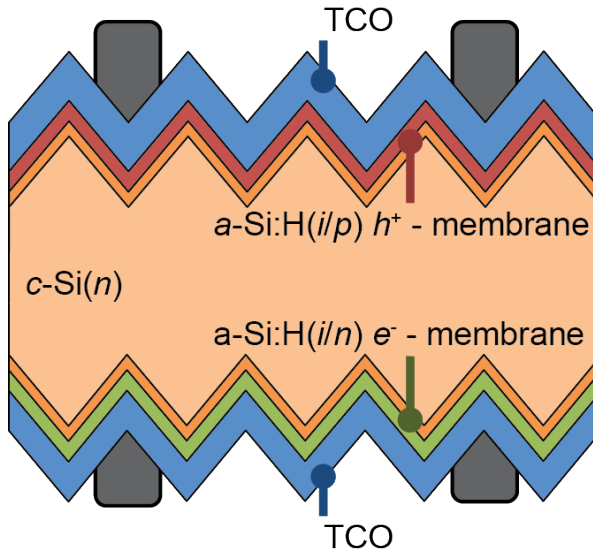


Figure 1.2: Schematic layout of the $a\text{-Si:H}/c\text{-Si}$ heterojunction cell as described by, e.g., De Wolf *et al.* [1]. It is fabricated by depositing films of intrinsic $a\text{-Si:H}$ on both sides of a texturised silicon wafer (usually n -type), followed by films of doped $a\text{-Si:H}$ of opposite polarities on either side. Typically, the $a\text{-Si:H}$ films are have a combined thickness of around 10 nm. These films form the electron and hole selective membranes that allow the different charge carriers to be extracted at opposite sides of the wafer. On top of the $a\text{-Si:H}$ films one then deposits TCO (short for transparent conductive oxide) films that provide low-resistance current pathways to the metal contacts. Commonly used TCOs include tin-doped indium oxide (ITO) and aluminium-doped zinc oxide (ZnO:Al). SHJ cells of this simple planar design have proven to be relatively easy to fabricate in only a few low-temperature ($< 200\text{ }^\circ\text{C}$) processing steps while still achieving very respectable efficiencies; some of them rivalling the world record such as Panasonic’s 750 mV SHJ cell from 2011 [4]. In fact, a modified version (also by Panasonic) of the SHJ cell holds the current world record for single junction $c\text{-Si}$ cells [5]. What makes this record cell different, is that it has both the n -type and p -type doped $a\text{-Si:H}$ films on the rear side of the cell in a configuration often called “interdigitated back contacts” or IBC. The IBC SHJ solar cell design is a solution to the main drawback of the planar SHJ cell, namely its parasitic absorption. As it turns out (and as we will discuss in greater depth later), the photons absorbed in the $a\text{-Si:H}$ films in the SHJ cell contribute relatively little to usable current. Thus, by putting the $a\text{-Si:H}$ films at the back of the cell in the IBC design, the amount of absorption in the amorphous films is minimised so that it can provide more current. The main disadvantage of the IBC SHJ cell is that it is significantly more complex to design and fabricate.

are traditionally used to explain how homojunction solar cells work. Instead, they rely on the relatively new idea of semi-permeable membranes, first coined by Peter Würfel. Reading his book *Physics of Solar Cells* made it clear that Würfel has a very different view of solar cells than the one presented in most text books on semiconductor physics, even though Würfel never really disputes the validity of the equations those text books use. Würfel demonstrates that there is more to the physics of solar cells than just the equations that describe them: there is also a bigger picture –an overall philosophy if you will– consisting of easily identifiable elements and a clear understanding of how they interact.

So while the equations of solar cell physics are mostly clear and cut, the big picture view of solar cells is still developing. But what does that mean, exactly? How can such a framework improve? Needless to say, a general idea like Würfel’s selective membrane model should not clash with the fundamental solar cell equations; otherwise it would be no more than a superstition or misconception. This means that its predictive power (the usual metric by which we judge a physical theory) should not be different than that of the basic solar cell equations. It also means that we cannot choose between two competing pictures purely based on their predictive power: both should be physically correct to begin with. Instead, if we have to choose between two such pictures, we should look at how much they simplify the process of improving solar cells. In other words, we are looking for a language that allows us to have meaningful, clear discussions about solar cells and that makes it easy to find design flaws in cells we want to improve. Thus, a good overall picture of solar cells is very important to solve the inverse problem of solar cell design mentioned earlier.

In the summer of 2012 the author of this thesis was lucky enough to meet Würfel in person and have some discussions with him. It became apparent that, at the time, he was not very optimistic about the general acceptance of his ideas. Three years later, we think that his reservations are, fortunately, unwarranted: during the many discussions with different researchers in (as well as outside of) the field, we noticed that the semi-permeable membrane model has a universal appeal due to its intuitive simplicity and most people we introduced it to were quick to adopt it. In fact, we have heard people admit more than once that they never felt like they really understood much about solar cells before we introduced them to Würfel’s ideas. In our own experience, it is a very powerful way of thinking about solar cells that makes the discussions about solar cell design much more transparent and less naive compared to what we had before (see Figure 1.3). Still, even though the selective membrane model is steadily gaining ground, the old solar cell picture (left in Figure 1.3) is still commonly found in literature, especially in articles from authors that originate from other fields (such as graphene processing, see e.g., [6, 7]) that pay a short visit to the world of PV. Even in the field of silicon PV there are still authors who, e.g., attribute the success of the SHJ cell to –in our opinion quite erroneously– the inversion layer in the *c*-Si [8]. In fact,

the reader will see in Chapter 5 (which is a paper we wrote in 2013) that we had some trouble in the past as well to completely shed our old conceptions of solar cells and adopt the selective membrane model.

Würfel’s ideas raise new and interesting questions that need to be answered. What exactly is a semi-permeable membrane in the context of solar cells and how do you make one? What makes it effective? Apparently the bandgap of such a membrane is important (see the quote of De Wolf *et al.*); why? And is that the only thing that matters? What makes the semi-permeable membrane model a useful way to look at solar cells and how does it follow from the basic underlying physics? Questions like these we have tried to address over the last four years and in this thesis we hope to answer them to the reader. As always turns out to be the case, simple ideas can become pretty complicated when you examine them closely, so it is important to guard against pitfalls and oversimplification.

If rules of thumb and heuristic explanations are not investigated and their limitations explained, the risk exists that they become rigid laws in the minds of researchers that steer them away from potentially interesting directions of thinking. Take, for example, the often used rule that the V_{OC} of a single-junction cell cannot be larger than the bandgap of the semiconductor. This rule is often taken for granted, but what does it mean, exactly? For one, what do we exactly mean by “the bandgap”: the optical or the electrical bandgap? Looking at Figure 1.4 it seems that the electrical bandgap is rather immaterial to the question, even though that is the number most people have in mind. Indeed, it has already been pointed out by Yablonovitch [9] (among others) that a very effective cell should have a dense internal photon field, just in the situation shown in the figure. The next (rather difficult) question is then what the exact limitations are when it comes to confining the internal photons by energy-selective mirrors; a question that would remain unasked if one would not venture to take a close look at the rule that the V_{OC} cannot be larger than the bandgap. Unfortunately, for *c*-Si based cells the possibility of achieving a V_{OC} larger than 1.124 V can be ruled out because Auger recombination will always dominate over radiative recombination in silicon³ [10]. Still, Figure 1.4 demonstrates why it is important to take a hard, critical look at broad ideas about the way solar cells work; the selective membrane model is no exception and in Chapter 3 we will take a closer look at it. In that section we demonstrate that the selective membrane model is indeed a useful way of thinking about solar cells. The main reason for this is that the functions of the selective membranes and the photon absorber in the cell are so dissimilar that they require very different material properties to be effective at their job. Because of this complementarity of functions, it makes sense to mentally

³This is also the reason that the photonic design principles of solar cells will not be treated in great depth in this thesis, since the main focus is on *c*-Si based cells. In Sections 2.4 and 2.5 we will briefly deal with this issue in broad terms.

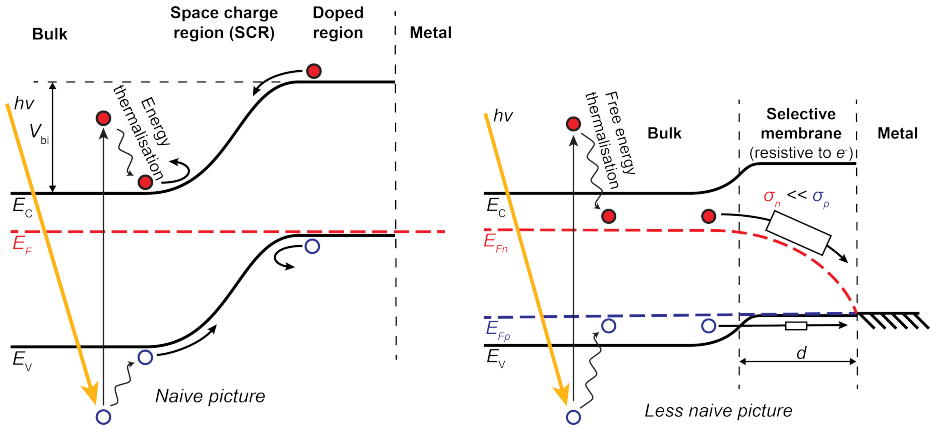


Figure 1.3: Two competing pictures of solar cell junctions. The left picture is a rather naive visualisation of the operation of a solar cell for several reasons. First of all, it puts great emphasis on the built-in voltage V_{bi} and the space charge region (SCR) and the forces the electric field exert on the carriers. However, the electrostatic force is only part of the total (thermodynamic) force that the carriers experience. Drawing an electron that falls of a electrostatic potential hill makes sense if that electron exists in a vacuum between two plates of a capacitor, but not in a solar cell where the motion of a single electron is both unpredictable and unimportant. Moreover, the left picture leaves the incorrect impression that the electric field drives the current, but an electric field cannot perform net work since it is a conservative force. The left picture also depicts the thermalisation of the energy of an electron-hole pair after excitation. This is also slightly naive, since the (total) energy of an electron-hole pair is, thermodynamically speaking, rather inconsequential: the total energy is not necessarily the amount of energy one can convert into work. In fact, the diagram shown on the left is in equilibrium; in this situation *none* of the energy of the excited pair can be extracted as work in the same way one cannot extract thermal energy from a silicon wafer and use that to power a machine. The right picture (Würfels') is a thermodynamically less naive presentation of a solar cell and it puts the emphasis on the quasi Fermi levels $E_{F_{n,p}}$ that represent the *free energy* of the ensembles of electrons and holes. This also means that the red and blue dots do not represent individual electrons and holes: they represent the average of a large number of particles. These averaged particles do not feel electrostatic forces (the gradients of $E_{C,V}$) but electrochemical forces (the gradients of $E_{F_{n,p}}$). Because E_{F_n} and E_{F_p} come together in the metal, there will always be electrons flowing towards the hole-collecting contact and the only way to do something about that is by making sure that the electrons encounter much more resistance (represented by rectangular resistor symbols) than the holes on their way from the bulk to the contact. This asymmetry between electron and hole resistance is what we call selectivity. Note that the SCR in this picture suddenly becomes rather unimportant: as it turns out, the SCR contributes only a little to selectivity of the membrane (see Section 3.3). The free energy also allows us to provide a new, more insightful, view of thermalisation losses associated with photon absorption, because the free energy is the maximum energy one can hope to extract as useful work from the system. For example, by considering *free energy thermalisation* rather than *energy thermalisation* it suddenly becomes obvious why photon absorption in selective membranes is less effective than in the bulk of the cell: in the membranes the Fermi level splitting is less and the free energy thermalisation losses are larger. In practise, these thermalisation losses manifest themselves as *parasitic absorption*; a well-known problem in *a*-Si:H-based SHJ cells [1].

divide the cell into these different parts and understand how they work and interact. Furthermore, this complementarity also immediately makes it clear that heterojunction cells provide the design flexibility we need to push the cell efficiency to its limit.

Halfway through 2012 we started playing around with ideas for making new heterojunction concepts based on Würfel’s selective membrane model. Our research group has had significant success in the past with the material Al_2O_3 , which is a surprisingly good passivation layer for silicon [11]. Passivation (the last word in the title of this thesis that has gone unmentioned so far) is the practise of making the surface of a silicon wafer (where it is not contacted by metal leads) less recombination active. This is important, because ordinarily surfaces are very defective and poorly passivated surfaces will cause significant losses in a cell. The reason that aluminium oxide is a good passivation layer is because of the low defect density of the $\text{Al}_2\text{O}_3/c\text{-Si}$ interface and because of the negative fixed charge in this interface that generates a hole inversion layer at the $c\text{-Si}$ surface [12]. Under the motto “If you have a good hammer, why not look for some more nails to hit?”, we started to think if it would be possible to extend the application of Al_2O_3 to selective membranes. In Chapter 5 the results of this research are described. One of our conclusions at the time was that one on hand we need a thin (~ 1 nm) film of Al_2O_3 to make it possible for the holes to tunnel through the oxide into the ZnO (which we used as a transparent conductor), but on the other hand that very thin oxide films also lead to large recombination losses and low carrier lifetimes. This result should have tipped us off that the contact we were trying to make is actually not selective enough. Indeed, we were trying to make a hole-selective membrane by creating a hole inversion layer in the $c\text{-Si}$ and we now know (see Section 3.3) that such an inversion layer simply cannot provide enough selectivity to make a good cell; an insight that we simply did not have at the time⁴. The Al_2O_3 itself, unfortunately, does not provide hole-selectivity either⁵. At the time we were still a bit stuck in the past in that regard and we focussed a lot on passivation (interface defect reduction) and on the hole inversion layer (i.e., the SCR). Of course, passivation is very important when it comes to making selective heterojunction membranes: precisely because the inversion layer cannot provide selectivity, the defects at the $c\text{-Si}$ interface need to be reduced as much as possible or you lose your carriers before the selective membrane even gets to do something (see also Section 3.6). So in this thesis when we are talking about using heterojunctions to make a selective contacts, we really mean *passivating selective contacts*. However, when it comes to making selective contacts, passivation is only the first step: between the passivated $c\text{-Si}$ interface and the metal there needs

⁴In fact, much of the work in Chapter 3 is motivated precisely by our initial inability to get a better understanding of the results from the $\text{Al}_2\text{O}_3/\text{ZnO}$ experiments.

⁵If anything, quite the opposite: looking at the tunnelling barriers in Figure 5.2, the Al_2O_3 itself should be electron selective

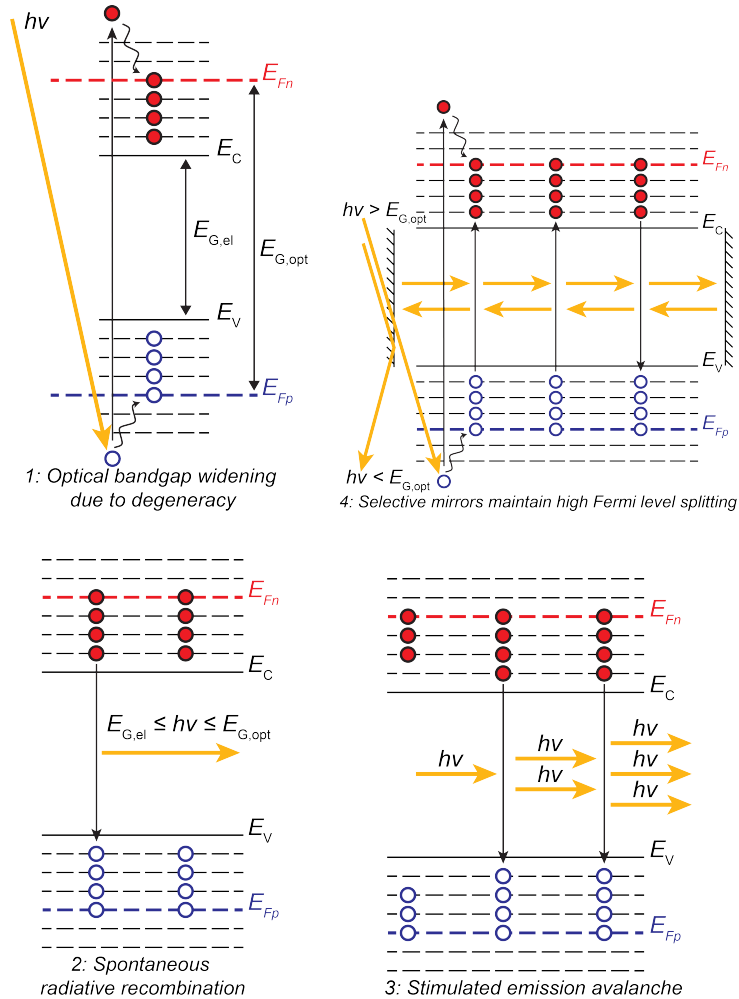


Figure 1.4: Is it possible to make a cell with a higher voltage than its bandgap? (Counterclockwise from the top-left.) 1: If enough carriers are injected into the semiconductor, both the electrons and holes become degenerate. The optical bandgap $E_{G,opt}$ (the minimal photon energy that can be absorbed by the semiconductor) widens and becomes equal to the Fermi level splitting instead of the electrical bandgap $E_{G,el}$. 2: Photons are then emitted by radiative recombination in the energy range between $E_{G,el}$ and $E_{G,opt}$. 3: The emitted photons will de-excite nearby electrons due to stimulated emission, generating even more photons and causing rapid loss of electron-hole pairs. 4: To avoid this avalanche reaction, it is necessary to confine the photons in the $[E_{G,el}, E_{G,opt}]$ range by selective mirrors so that internal equilibrium between the photons and e^-h^+ plasma can form, just like in a semiconductor laser. In this case we clearly have $V_{OC} < E_{G,opt}$, since the V_{OC} is the maximum possible Fermi level splitting. However, it seems likely that $V_{OC} > E_{G,el}$ is possible if the photons inside of the cell can be confined well enough by the mirrors. Interestingly, in this situation the free energy thermalisation losses in the cell are significantly lower than if the electron-hole pairs were non-degenerate.

to be some sort of layer that provides a source of selectivity, otherwise too many minority carriers will end up in the wrong metal contact, which is just as bad as interface recombination. Sometimes the term *passivating contacts* (or *passivated contacts*) can also be heard in the solar cell world, but we think that this term gives the wrong impression that it is somehow possible to passivate a metal; to make it less recombination active. This is simply not the case: all you can really do is make it as difficult as possible for one type of carrier to reach the metal in the first place.

We then tried to salvage the situation by inserting a little bit of *a*-Si:H between the *c*-Si and Al₂O₃, even though this was the material we initially set out to avoid. The unfortunate thing about *a*-Si:H is that it is a strong photon absorber, which is something you do not want out of a selective membrane since it leads to parasitic absorption (see Figure 1.3, Section 3.5, and also De Wolf *et al.* [1]). As expected, the *a*-Si:H improved the measured carrier lifetimes. At the time we attributed this effect mainly to a reduction of the defect density at the *c*-Si/*a*-Si:H interface (compared to the original Al₂O₃/*c*-Si interface), but the interface defects are probably only part of the story. Looking back at the data in Chapter 5 with the insights we have gained since then, we now think it is very likely that the main reason for the improved carrier lifetimes should be attributed to a significant increase in the resistance the electrons experience on their way to the contact (for which we used ZnO). In a sense, the fixed charges in the Al₂O₃ act as a sort of induced *p*-type doping in the *a*-Si:H. So while the induced SCR in the *c*-Si only has a very low selectivity, the SCR induced in the *a*-Si:H actually *is* resistive enough to make a good selective contact. However, we also found that the hole-selective contacts we fabricated with these materials (ZnO, *a*-Si:H and Al₂O₃) were too resistive for holes. Here, we think that the original conclusions from Chapter 5 still stand and that the main problem is the choice of ZnO as a TCO (transparent conductive oxide), which simply does not have a large enough electron affinity to efficiently collect holes (as demonstrated by modeling the contact like an Esaki diode; see Figure 5.2 and Figure 5.3). Ultimately, the idea behind the TCO/*a*-Si:H/Al₂O₃ stacks is still sound enough to warrant further investigation of this concept and those related to it (such as the successful MoO_{*x*}-based hole contacts by Battaglia *et al.* [13, 14]). The attractive feature of these concepts is that they use a thin, transparent metal oxide layer to replace the doped *a*-Si:H in traditional SHJ cells so that only the intrinsic *a*-Si:H is needed. So while the problem of parasitic absorption cannot be completely solved by fixed-charge dielectrics such as Al₂O₃, at least they provide feasible alternatives to doped *a*-Si:H without having to discard the good qualities of intrinsic *a*-Si:H.

Graphene is another material that caught our attention as potentially interesting for the fabrication of selective membranes. Of course, graphene is a very unique material with remarkable properties matched by a remarkable hype to go with it. It should not come as a surprise that many researchers

have already made solar cells by putting graphene on a silicon wafer –just to see what happens– but so far this field has been explored mainly by experts of graphene and not so much by solar cell experts. We therefore thought it was a good idea to review this very active field from our own perspective to see if graphene can indeed live up to the hype when it comes to PV. This is the subject of Chapter 6, where we tried to give an overview of the physics of graphene that are relevant for a solar cell researcher to make the entry into this field a little simpler. Furthermore, these physics are then used to see if the electrical properties of graphene are indeed remarkable enough to make a relevant contribution to the toolbox of the silicon solar cell designer. As it turns out, the answer to this question is a tentative “yes”. For one, the conductance of a single sheet of graphene can, in principle, be high enough to act as a transparent electrode for a *c*-Si solar cell. However, this high conductance is far from straightforward to achieve since the 2D material is rather sensitive to its surroundings and its substrate. As a selective contact it shows more potential since it seems to behave like a metal with a tunable work function and an extremely low (for metal standards) effective Richardson constant. In fact, the Richardson constant of graphene could potentially be so low that the graphene/silicon Schottky junction could be the exception to the old rule of thumb that Schottky junctions cannot make good solar cells. Unfortunately, the literature data so far cannot prove for certain that this is actually possible and more focussed experiments are necessary to confirm this hypothesis. Additionally, graphene is known to be a material that is difficult to dope (and good doping control is definitely necessary for PV applications) and difficult to process, so graphene is certainly not a material that will just solve any problem without creating new ones as the hype might sometimes lead you to believe. Still, it is not unlikely that at some point it will earn its place in the field of PV.

Parallel to the selective membrane model we also developed a thermodynamical view of solar cells in this thesis. Indeed, Würfel’s book is also rather explicit about the importance of thermodynamics when it comes to understanding solar cells and in a sense this is a natural thing to do since thermodynamics is the natural language of energy conversion. The mere idea to view a solar cell as a thermodynamic energy converter that turns heat from the sun into work is already a valuable concept to broaden the discussion about what solar cells are (and are not) capable of. It may seem like a trivial idea, but many fellow physicists we have spoken to (especially those not directly involved with PV) were surprised at first when presented with the idea of applying thermodynamics to solar cells⁶. There seems to be a sort of implicit consensus that electrical devices are not really in the realm of thermodynamics for some reason. Some

⁶Even though it is actually not a new idea at all; it can already be found the famous Shockley-Queisser paper [15] and in earlier works cited by that paper

physicists also seem to have an aversion to the thought that sunlight is a form of heat because it does not come in the form of thermal vibrations⁷.

So if a thermodynamic treatment of solar cells is not new, what is our motivation for presenting one? The thing about thermodynamics is that there are different ways to approach the subject depending on the a priori assumptions you make about your system. Our goal is to make thermodynamics accessible to solar cell researchers and that means we want to stay relatively close to the semiconductor physics they are (hopefully) already familiar with. We do this to demonstrate that the thermodynamics of solar cells is more than just academic discussions about the upper efficiency limit of hypothetical devices: the science of thermodynamics is, from its very roots, an engineering toolbox⁸. Yet in the solar cell community it is not usually recognised as such and we think that this is because the languages of thermodynamics and semiconductor physics have been rather distinct. That is not to say that it is all bad: the efforts of Würfel and other authors have already helped to spread some thermodynamic ideas in the PV community, such as the view that the quasi Fermi levels represent the free energy (i.e., the energy one can hope to extract as useful work) of electrons and holes. However, there are still open ends and many works (including Würfel's) rely on traditional equilibrium thermodynamics and then make a sudden leap to the non-equilibrium problem of describing a working solar cell without really motivating how such a leap can be made.

In our opinion, the missing link between thermodynamics and semiconductor physics is the book *Non-equilibrium Thermodynamics* by De Groot and Mazur [16]. In its own field, it is a well-known work that is known for being thorough yet rather dense. When cast in the right form, the theory of De Groot and Mazur can produce the thermodynamic equivalent of semiconductor physics without much trouble, as we will show in Chapter 2. In the end, this means that one does not have to learn many new things to discuss solar cells thermodynamically: semiconductor physics actually were a form of non-equilibrium thermodynamics all along. By recognising this equivalence, we suddenly gain access to all kinds of thermodynamical tools that were hidden before, such as the entropy generation rate that tells us where in the cell (and

⁷To be fair, there are some peculiarities about photons that cannot be glossed over. What makes photons strange is that they cannot interact with each other: if you throw a bunch of light particles together, they will not equilibrate because they cannot interact. However, they *can* equilibrate with a dense collection of charged particles, such as the plasma of the sun's surface. Thus, in the sun the photons can reach equilibrium with the photosphere and because of the zeroth law of thermodynamics this also means that they are in equilibrium with each other, even though they do not feel each other. Then they are sent off to earth and maintain that equilibrium until they interact with the solar cell (ignoring the atmosphere). Thus, the interaction of sunlight with a solar cell can be viewed as a form of heat transfer from the photosphere to that cell.

⁸After all, Carnot was an engineer who was very much interested in solving practical problems.

how much) potential output power is being wasted.

Then in Chapter 4 we will use the derived expression of the entropy generation rate to set up a mathematical model called VEGM (variational entropy generation minimisation) that can predict efficient cell designs. VEGM is, in principle, the mathematical answer to the inverse problem that was posed at the start of this chapter: how can you make a computer spit out the most efficient cell for a given set of practical constraints? The reader may notice that Chapter 4 does not actually contain a computer calculation that implements the VEGM model. This is not for lack of trying, but it turned out that the differential algebraic equations produced by VEGM are not straightforward to implement into an existing solver (we tried COMSOL Multiphysics[®], but ran into convergence issues) and at the moment we are of the opinion that this implementation requires dedicated effort from an expert on numerical computation.

To conclude this introduction, we would like to present a brief helicopter view of the content in the next chapters in Figure 1.5. The central theme of this dissertation is the optimisation problem of solar cells we introduced in the first two paragraphs of this chapter: how do you “invert” the known solar cell physics so that you can predict which cell is the most efficient? This question is approached from two different angles, namely that of the selective membrane model and that of thermodynamics. The thermodynamics provide the physical foundation necessary to quantitatively understand the loss mechanisms in solar cells while the selective membrane model provides a qualitative language that helps to get a better intuitive grasp of the overall design of a cell as well as the strengths and weaknesses of its components.

We hope that between these two complementary angles the reader finds new insights that will help him or her come up with fruitful new ideas about how to improve the cell he or she is working on. Of course, if you are left with any questions, you are kindly invited to contact the author. You are kindly referred to the contact details in the CV on page 210.

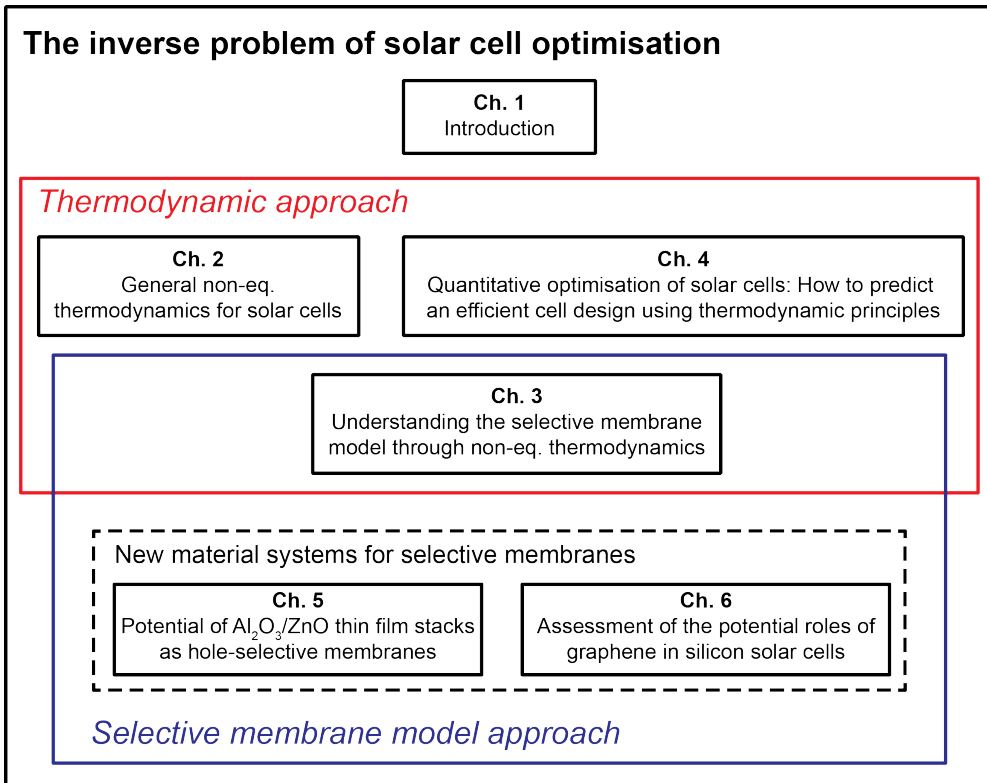


Figure 1.5: Overview of the themes in this thesis and the contents of the chapters.

References Chapter 1

- [1] Stefaan De Wolf, Antoine Descoeurdes, Zachary C. Holman, and Christophe Ballif. High-efficiency Silicon Heterojunction Solar Cells: A Review. *green*, 2(1):7–24, January 2012. ISSN 1869-876X. doi: 10.1515/green-2011-0018.
- [2] E. Yablonovitch, T. Gmitter, R. M. Swanson, and Y. H. Kwark. A 720 mV open circuit voltage SiO_x:c-Si:SiO_x double heterostructure solar cell. *Applied Physics Letters*, 47(11):1211, 1985. ISSN 00036951. doi: 10.1063/1.96331.
- [3] Peter Würfel. *Physics of Solar Cells*. Wiley-VCH Verlag GmbH & Co, Weinheim, 2005. ISBN 3-527-40428-7.
- [4] Martin A Green, Keith Emery, Yoshihiro Hishikawa, Wilhelm Warta, and Ewan D Dunlop. Solar cell efficiency tables (version 43). *Progress in Photovoltaics: Research and Applications*, 22(1):1–9, jan 2014. ISSN 10627995. doi: 10.1002/pip.2452.
- [5] Martin A. Green, Keith Emery, Yoshihiro Hishikawa, Wilhelm Warta, and Ewan D. Dunlop. Solar cell efficiency tables (version 46). *Progress in Photovoltaics: Research and Applications*, 23(7):805–812, jul 2015. ISSN 10627995. doi: 10.1002/pip.2637.
- [6] Yu Ye and Lun Dai. Graphene-based Schottky junction solar cells. *Journal of Materials Chemistry*, 22(46):24224, 2012. ISSN 0959-9428. doi: 10.1039/c2jm33809b.
- [7] Xinming Li, Hongwei Zhu, Kunlin Wang, Anyuan Cao, Jinqun Wei, Chunyan Li, Yi Jia, Zhen Li, Xiao Li, and Dehai Wu. Graphene-on-silicon Schottky junction solar cells. *Advanced materials (Deerfield Beach, Fla.)*, 22(25):2743–8, July 2010. ISSN 1521-4095. doi: 10.1002/adma.200904383.
- [8] J. Poortmans M. Ghannam, G. Shehadah, Y. Abdulraheem. Basic Understanding of the Role of the Interfacial Inversion Layer in the Operation of Silicon Solar Cells with a-Si/c-Si Heterojunction (HIT). In *28th EU*

PVSEC Proceedings, page 822, 2013. doi: 10.4229/28thEUPVSEC2013-2BO.2.6.

- [9] Eli Yablonovitch and Owen D. Miller. The Opto-Electronics of Solar Cells. *IEEE Photonics Society News*, 27(1), 2013.
- [10] Armin Richter, Martin Hermle, and Stefan W. Glunz. Reassessment of the Limiting Efficiency for Crystalline Silicon Solar Cells. *IEEE J. Photovolt.*, 3(4):1184–1191, oct 2013. ISSN 2156-3381. doi: 10.1109/JPHOTOV.2013.2270351.
- [11] G. Dingemans and W. M. M. Kessels. Status and prospects of Al₂O₃-based surface passivation schemes for silicon solar cells. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 30(4):040802, 2012. ISSN 07342101. doi: 10.1116/1.4728205.
- [12] Bram Hoex, S. B. S. Heil, E. Langereis, M. C. M. van de Sanden, and W. M. M. Kessels. Ultralow surface recombination of *c*-Si substrates passivated by plasma-assisted atomic layer deposited Al₂O₃. *Applied Physics Letters*, 89(4):042112, 2006. ISSN 00036951. doi: 10.1063/1.2240736.
- [13] Corsin Battaglia, Xingtian Yin, Maxwell Zheng, Ian D Sharp, Teresa Chen, Stephen McDonnell, Angelica Azcatl, Carlo Carraro, Biwu Ma, Roya Maboudian, Robert M Wallace, and Ali Javey. Hole Selective MoO_x Contact for Silicon Solar Cells. *Nano Letters*, 14(2):967–971, feb 2014. ISSN 1530-6984. doi: 10.1021/nl404389u.
- [14] Corsin Battaglia, Silvia Martín de Nicolás, Stefaan De Wolf, Xingtian Yin, Maxwell Zheng, Christophe Ballif, and Ali Javey. Silicon heterojunction solar cell with passivated hole selective MoO_x contact. *Applied Physics Letters*, 104(11):113902, mar 2014. ISSN 0003-6951. doi: 10.1063/1.4868880.
- [15] William Shockley and Hans J. Queisser. Detailed Balance Limit of Efficiency of p-n Junction Solar Cells. *J. Appl. Phys.*, 32(3):510, 1961. ISSN 00218979. doi: 10.1063/1.1736034.
- [16] S.R. de Groot and Peter Mazur. *Non-equilibrium Thermodynamics*. Dover Publications, Mineola, New York, 1984. ISBN 0-486-64741-2.

Chapter 2

The physics of solar cells: semiconductor physics enhanced by thermodynamics

Symbols used in Chapter 2 and Chapter 3

Sym.	Definition	Sym.	Definition
A_i	General thermodynamic variable when U is used as the potential. E.g., S , V . $i = 1, \dots, l$.	G	Generation rate of electron-hole pairs ($\text{cm}^{-3} \text{s}^{-1}$).
\mathbf{B}	Magnetic field (T).	G_{net}	Net generation rate of electron-hole pairs $G - R$ ($\text{cm}^{-3} \text{s}^{-1}$).
B_i	General thermodynamic variable when S is used as the potential. $i = 1, \dots, l$.	\mathbf{J}_i	Thermodynamic current (or flow) of quantity i . $i \in \{U, S, Q, n, p\}$. The particle currents $\mathbf{J}_{n,p}$ are in $\text{cm}^{-2} \text{s}^{-1}$.
$D_{n,p}$	Electron/hole diffusion coefficient ($\text{cm}^2 \text{s}^{-1}$).	$J_{n,p}$	1D electron/hole particle flux ($\text{cm}^{-2} \text{s}^{-1}$).
e	Elementary charge (C).	J_{SC}	Short circuit current density (A cm^{-2}).
E_C	Conduction band energy level $E_C = -\phi - \chi$ (eV).	k_B	Boltzmann constant (J K^{-1}).
E_F	Equilibrium Fermi level (eV).	L_{ij}	Scalar transport coefficient that relate force \mathbf{F}_j to flux \mathbf{J}_i .
\mathbf{E}	Electric field (V cm^{-1}).	L_{ij}	Second order tensor transport coefficient that relate force \mathbf{F}_j to flux \mathbf{J}_i .
$E_{Fn,p}$	Quasi Fermi energies for electrons/holes (eV).	n	Electrons density (cm^{-3}).
E_G	Semiconductor bandgap (eV).	n_i	Intrinsic carrier density: $n_i^2 = N_C N_V \exp(-E_G/k_B T)$ (cm^{-3}).
E_V	Valence band energy level: $E_V = -\phi - \chi - E_G$, (eV).	n_{eh}	Density of electron-hole pairs (cm^{-3}).
f_i	Conjugate variable to A_i when U is the potential: $f_i = \partial U / \partial A_i$, $i = 1, \dots, l$.	$N_{C,V}$	Conduction/valence band effective density of states (cm^{-3}).
\mathbf{F}_i	Thermodynamic force. $i \in \{U, S, Q, n, p\}$.		
$F_{1/2}$	Fermi-Dirac integral.		
$F_{1/2}^{\leftarrow}$	Inverse of the Fermi-Dirac integral.		
FF	Fill factor.		
g_i	Conjugate variable to B_i when S is the potential: $g_i = \partial S / \partial B_i$, $i = 1, \dots, l$.		

Sym.	Definition
$N_{n,p}$	Total number of electrons/holes (dimensionless).
N_{dop}	Semiconductor doping density (cm^{-3}).
N_{eh}	Total number of electron-hole pairs (dimensionless).
p	Hole density (cm^{-3}).
P	Pressure (N cm^{-2}).
q_i	Source term in continuity equation for quantity i . $i \in \{U, S, n, p\}$.
Q	Heat (J).
Q_{np}	Total electron-hole charge divided by e (cm^{-3}).
R	Recombination rate of electron-hole pairs ($\text{cm}^{-3} \text{s}^{-1}$).
R_{Rad}	Radiative recombination rate of electron-hole pairs ($\text{cm}^{-3} \text{s}^{-1}$).
\tilde{R}	Pseudo resistance ($\text{C } \Omega \text{ cm}^2$).
\check{R}	Intrinsic resistance ($\text{C } \Omega \text{ cm}^2$).
s	Volumetric entropy density ($\text{J K}^{-1} \text{ cm}^{-3}$).
S	Entropy (J K^{-1}).
\tilde{S}	Pseudo conductance ($\text{S cm}^{-2} \text{ C}^{-1}$).
\check{S}	Intrinsic conductance ($\text{S cm}^{-2} \text{ C}^{-1}$).
t	Time (s).
T	Temperature (K).
u	Volumetric internal energy density (J cm^{-3}).

Sym.	Definition
U	Internal energy (J).
V	Volume (cm^3).
V_{OC}	Open circuit voltage (V).
V_T	Thermal voltage = $k_{\text{B}}T/e$ (V).
x	Position coordinate (cm); can be 1, 2, or 3 dimensional.
β	Inverse temperature $1/T$ (K^{-1}).
γ	Index that indicates that a symbol pertains to photons.
δ	Variational differential. E.g., when $f(x)$ is a function of position, $\delta f(x)$ represents a small deviation of that function.
$\Delta\eta_{eh}$	Fermi level splitting or electrochemical potential per electron-hole pair (eV). $\Delta\eta_{eh} = \eta_n + \eta_p = E_{Fn} - E_{Fp}$.
Δn	Injection level; excess carrier density (cm^{-3}).
ϵ	Dielectric constant $\epsilon_r \epsilon_0$ (F cm^{-1}).
$\eta_{n,p}$	Electron/hole electrochemical potential (eV). $\eta_n = E_{Fn}$, $\eta_p = -E_{Fp}$.
$\tilde{\eta}_{n,p}$	$\beta\eta_{n,p}$.
$\Delta\tilde{\eta}_{eh}$	$\beta\Delta\eta_{eh}$.
η_{PC}	Photovoltaic conversion efficiency, PCE (dimensionless).

Sym.	Definition
$\mu_{n,p}$	Electron/hole mobility ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$).
$\mu_{n,p}^c$	Electron/hole chemical potential (eV).
$\nu_{n,p}$	Selectivity of an electron/hole selective membrane (dimensionless).
ρ_{net}	electron-hole charge density divided by e (cm^{-3}).
σ_{irr}	Total entropy generation rate per unit volume ($\text{J K}^{-1} \text{s}^{-1} \text{cm}^{-3}$).
Σ_{irr}	Total entropy generation rate ($\text{J K}^{-1} \text{s}^{-1}$).
$\sigma_{n,p}$	Electron/hole conductivity (S cm^{-1}).
Φ_Q	Conjugate variable of Q_{np} when U is the potential: $\Phi_Q = \partial U / \partial Q_{np}$ (J).
$\tilde{\Phi}_Q$	Conjugate variable of Q_{np} when S is the potential: $\tilde{\Phi}_Q = \partial S / \partial Q_{np}$ (J).
ϕ	Electrostatic energy (eV).
χ	Electron affinity (eV).

2.1 The theory of non-equilibrium thermodynamics

The description of solar cells is usually framed in the language of semiconductor physics (SCP) because many photovoltaic devices rely on the unique properties of semiconductors to work. The literature on SCP has accumulated a large wealth of detailed models (such as recombination and mobility models) that can be used to accurately predict the efficiencies of industrial device designs. Paradoxically, this level of detail can also be detrimental to the global understanding of the device because only a computer can work out the complex SCP models so that it quickly becomes unclear which details are the most important and what parts of the device need to be addressed to yield the largest gains in efficiency. Indeed, for a long time there have been misconceptions in the PV field about the general principles of photovoltaic energy conversion such as the idea that the electric field in the space charge region of a junction is responsible for driving the photocurrent [1, 2].

An alternative way to look at solar cells is by considering it as a thermodynamic system. The motivation for this viewpoint is that, unlike most semiconductor devices, a solar cell is first and foremost an energy convertor and the natural language of energy conversion are thermodynamics. Specifically, since a solar cell is always operating under non-equilibrium¹ conditions, it should be described by non-equilibrium thermodynamics (NET) rather than equilibrium thermodynamics (ET). The advantage of thermodynamics is that it provides methods to characterise losses of free energy (i.e., useful output power). In literature many different thermodynamic descriptions of solar cell physics can be found [4, 5, 6, 7, 8, 9, 1, 10, 11] and addition, several works have presented thermodynamic frameworks for SCP in general [12, 13, 14, 15] without specific application to PV. Yet despite these efforts, the SCP view of solar cells is still by far more common than the thermodynamic view, especially in engineering domain of *c*-Si based cells. About the reasons for this difference one can hypothesise:

- The PV community simply does not see the value of an alternative model as long as SCP does a good job of predicting cell efficiencies.
- Thermodynamics is thought of as a subject that is mainly useful for academic discussions about the maximum possible efficiency of highly hypothetical devices under idealised circumstances. Indeed, many existing works on solar cell thermodynamics focus on a very broad description of the device rather than on the detailed engineering that is necessary to improve the efficiency of practical designs.
- NET is a relatively new field that is still developing and thus has not fully

¹There seems to be no general consensus about the spellings “non-equilibrium” vs. “nonequilibrium”. In this work the spelling by De Groot and Mazur [3] will be followed.

crystallised yet. This is in stark contrast to the SCP description of solar cells, which has been solidified into a uniform shape. As was mentioned, different formulations of solar cell thermodynamics exist and potentially this gives the impression that there are no established methods yet one can safely rely on.

- The PV community is simply not aware enough of the progress in the area of thermodynamics to appreciate the value of the many tools that have become available throughout the last few decades.
- A combination of some or all of these factors.

Under scrutiny, these objections do not hold up. Van Vliet [12], Lindefelt [14], and Parrot [15] already demonstrated that SCP can quite readily be incorporated into NET as a specialised area. Furthermore, the use of thermodynamics is very common in other engineering disciplines (chemistry being the prime example), so the question rises why this should be different for solar cells. The main goal of this chapter is to build up a comprehensive treatment of NET applied to solar cell physics that will be referred to as NET_{PV} . The end product of this treatment is a theory that combines the strengths of SCP (i.e., detailed models for the engineering of practical solar cell designs) and of NET (i.e., generality and a clear interpretation of free energy losses).

In NET_{PV} , certain assumptions relevant to PV will be made from the start to allow for simplifications and to avoid unnecessary abstraction. The common grounds between NET_{PV} and SCP will be made as clear as possible by choosing notation familiar from SCP, especially in Chapter 3. Where necessary, it will be stated how NET_{PV} interprets or describes physical phenomena differently compared to SCP. Special attention will be drawn to cases where NET_{PV} potentially extends SCP or even disagrees with it.

To give a firm foundation for NET_{PV} , the book “Non-equilibrium Thermodynamics” by De Groot and Mazur (abbreviated as GM) [3] is used as the starting point. This work, first published in 1962 and republished in 1984, is widely considered to be a standard work in its field and is commonly referenced by more modern works [16, 17, 18]. Their theory for the description of non-equilibrium systems (abbreviated as NET_{GM}) is very well equipped to describe the energy conversion processes in solar cells. One of the main goals of this chapter is to present an accessible and comprehensive application of NET_{GM} to solar cells, since the book by GM is known to be rather dense. It should be noted, though, that the thermodynamic description of solar cells requires some unique considerations related to the interaction between the device and the sunlight. These subtleties are not treated in the sources mentioned above and deserve some extra attention. The main references for the thermodynamics of light are the works by Würfel [8, 1] and Landsberg [19].

It is assumed that the reader of this thesis is familiar with the basics of the SCP description of solar cells, but for clarity a brief summary is given in Appendix A.1.

2.1.1 Literature discussion

Perhaps surprisingly, it seems that most authors in the field of *c*-Si PV (even those that have given thermodynamic descriptions of solar cells) are unaware of the work on non-equilibrium thermodynamics by GM, even though some (but certainly not all) of its methods have found their way into the field of PV indirectly. Unfortunately, the book by GM is not indexed in major publication databases, making it difficult to determine precisely how influential it is in various fields of research or in the field of PV in particular. All that can be said for sure is that none of the work by GM can be found in the 2-generation backward citation maps of the works by Brendel [10], Greulich [11], Harder [9], P. Würfel [8], and U. Würfel [2].

The closest link between the PV literature and the work by GM seems to be the work of J.E. Parrot, who refers to GM in his work “Thermodynamic theory of transport processes in semiconductors” [15]. However, in his earlier works on solar cell thermodynamics [4, 6] it seems that influences from GM are absent. An expression for the irreversible entropy generation rate in a solar cell is obtained in his article “Thermodynamics of solar cell efficiency” [6], but this derivation relies on an earlier paper [13] that uses microscopic theory based on the Boltzmann transport equation rather than the macroscopic theory of GM.

2.2 Overview of NET_{PV}

Here in Section 2.2, a brief overview will be given of NET_{PV} with its methods and goals. This should provide the reader with a map that will hopefully reinforce a sense of direction while also providing some motivation as to why NET_{PV} is worth learning.

These are the steps and ingredients that go into NET_{PV} :

1. Because of the fact that microscopic physics cannot give a definitive foundation for thermodynamics, the first step towards a mathematical model for NET_{PV} is perhaps the most crucial one: the choice of variables. In NET , the choice of variables is motivated by ET and to do so, it is necessary that the system under study (i.e., a solar cell in our case) has an equilibrium state that more or less resembles the non-equilibrium state one wishes to describe. Therefore, the ET description of a solar cell will be described in Section 2.3 and its meaning elaborated on. It is assumed that the reader has a basic understanding of ET, but to make sure that the reader is on the right page, Section 2.3 will also give a brief overview

of the subject in general terms. This section does not aim to give a detailed account of all of the mathematical methods from ET, but focuses more on the underlying assumptions and ideas that are useful for NET_{PV} as well.

It will also be explained how the equilibrium description can be extended to non-equilibrium situations: this is done by dividing the cell into many small subsystems that are each approximately in equilibrium and then assigning thermodynamics variables to every subsystem. In a sense, the NET description of the cell is very similar to the ET description, but the number of thermodynamic variables is significantly greater in the non-equilibrium situation. The potential pitfalls in the generalisation step from equilibrium to non-equilibrium are also discussed in Section 2.3.

2. The description of operating solar cells mostly focuses on non-equilibrium steady state conditions, but any system that is insulated from its environment will return to equilibrium. This means that the operating state of a cell is inherently characterised by its environment. The interaction of a solar cell with its surroundings is not straightforward due to the non-local nature of light absorption and will be considered in Section 2.4.
3. After specifying the relevant thermodynamic variables (step 1) and the interaction of the cell with its environment (step 2), the next step is to obtain the equations that describe the cell. This is done by generalising the first law of thermodynamics to a local form and then writing down continuity equations for the conserved quantities. Section 2.5 will discuss this step.
4. The continuity equations introduce new quantities, namely the fluxes (i.e., currents) of the conserved quantities. To describe these fluxes, the theory of Onsager [20] is used, which provides a general framework for obtaining flux-force relations. To find the thermodynamically correct fluxes and forces in Onsager's theory, the local entropy production rate is of crucial importance. The flux-force relations will be discussed in Section 2.6.
5. Finally, symmetry considerations (such as isotropy of the material and reversibility of microscopic physical laws) can be used to establish some important simplifications. The most important of these are the Onsager reciprocal relations and they will be discussed in Section 2.7. There are also material symmetries relevant to solar cells and these will be discussed by using what is known as the *Curie principle*.

Finally, Section 2.8 will wrap up the discussion about NET_{PV} by giving an overview of the main take-home messages.

2.3 The thermodynamic formulation of solar cell physics step 1: Equilibrium and thermodynamic variables

The goal of Section 2.3 is twofold. First of all, the aim is to get the reader to appreciate what a thermodynamic description of a solar cell actually encompasses. This is done by giving a broad introduction to ET in general terms, especially about its underlying assumptions rather than the mathematical formalism. Whereas the mathematical formalism of ET is well established, the motivation for arriving at that formalism can differ significantly from author to author. This section does not follow one source in particular, though most of the influences come from the texts by Callen [21], Chandler [22] and Zemansky [23]. These books (especially Callen) are recommended for the critical reader who needs more background than the brief discussion below.

The second aim of this section is more practical, namely to explain how ET can be applied to solar cells. This application is intermixed with the general discussion of ET to ground some of the more abstract ideas with practical examples relevant to the reader. However, a solar cell is not an ideal demonstration system for all thermodynamic notions, so other examples will be used as well. The works by Peter Würfel, e.g., his book “Physics of Solar Cells” [1], are important sources for the ET description of solar cells.

Thermodynamics is the study of macroscopic systems based on macroscopic variables and this means that one cannot fully reproduce thermodynamics from microscopic laws alone. This gives thermodynamics a unique position in physics: it never attempts to reduce a system to its elementary parts or to give a first principles explanation of a system (based on, e.g., quantum mechanics). Instead, thermodynamics can be seen as a sort of recipe that prescribes how to build a model for a system. In this recipe there are certain requirements that the model should satisfy (such as energy conservation), but thermodynamics also leave degrees of freedom in the way that the problem can be described rather than prescribing everything a-priori. Because of this freedom a physicist is allowed to insert new ideas into the model if that is deemed necessary, giving thermodynamics a wide range of application.

It would be incorrect to say that thermodynamics is completely disconnected from microscopic phenomena and fundamental laws, since it is well known that statistical mechanics provide a valuable link between microscopic and macroscopic views. However, to establish this link, statistical mechanics rely on its own set of assumptions that (in general) cannot be derived from first principles, such as the assumption of ergodicity and the assumption that all microscopic states with equal energy are equally likely to occur. The thermodynamic and microscopic descriptions of a system therefore have an interesting mutual dependency: the microscopic description can illuminate the underlying

mechanisms behind macroscopic observations, while the macroscopic laws provide a benchmark for the microscopic models and assumptions. In this work, most of the emphasis will be on the macroscopic point of view, though some ideas from and links with microscopic physics will be briefly indicated.

Thermodynamics rely only on a small set of assumptions that will be briefly reviewed here. It is common to take as the first assumption the law of conservation of energy, which has been solidified in the celebrated First Law of Thermodynamics. There is a reason why energy is such a crucial quantity for thermodynamics, even though there is a whole range of other conserved quantities (momentum, angular momentum, charge, etc.): energy is the only conserved quantity that has many different manifestations. Whereas quantities such as momentum only manifest themselves macroscopically in one way (e.g., the centre of mass motion of the system), energy has many forms: mass, kinetic energy, electromagnetic fields, chemical potential and more. Most physicists will agree that the first law is among the most secure in the whole of physics and is unlikely to be overturned any time soon, making the foundation of thermodynamics very firm. In a sense, thermodynamics can be seen as the branch of physics that tries to make the very most out of the law of conservation of energy.

Before continuing, it is interesting to note that it is not strictly necessary to start with the law of conservation of energy. In Callen's view (chapter 21), symmetries of the fundamental laws of physics are the real foundations of thermodynamics. Central to this view is *Noether's theorem*, which is a purely mathematical result that states that “for every continuous symmetry of the dynamical behaviour of a system there exists a conservation law for that system”. As an example, the (continuous) time-translation symmetry of the laws of physics (i.e., the physics of today are the same as those of today + Δt for any given Δt) implies the existence of a conserved quantity, namely energy. By the same argument, space-translation symmetry implies conservation of momentum and rotational symmetry implies conservation of angular momentum.

The law of conservation of energy, at its bare minimum, does not predict much yet. All it says is that associated with any system there is a number called internal energy U and that this number U will not change unless you allow energy to flow into the system or leave the system. The law of conservation of energy does not prescribe how to measure U or how it can be exchanged between systems. To be of any value, additional rules need to be imposed on how U transforms. In thermodynamics, this is done by assuming that the state of the system is determined by U as well as a number of additional extrinsic macroscopic variables, which will be referred to as A_i in general, with the index i running from 1 till l . The A_i are sometimes also called *thermodynamic coordinates*. The A_i are measurable properties of the system itself (as opposed to space and time coordinates, that are independent of the system), such as

volume V and number of particles of species α , N_α . The variables A_i are required to be extensive, meaning that the magnitude of every A_i is linear in the size of the system. In practice, this means that if two identical samples of a certain material (say, two samples of gas) are added together, then the A_i (e.g., the volume) for the composite system is also twice as great². Every system also has an extensive property S called entropy that, because of its ubiquity, will always be associated with $i = 1$ (i.e., $A_1 = S$). For now it suffices just to know that entropy exists; its meaning will be discussed in a separate subsection below.

In general, the A_i can be identified with some special combination of the microscopic coordinates of the system. These combinations are special in that they are essentially constant in time, even though most microscopic coordinates fluctuate on a time scale of 10^{-15} s, making them macroscopically completely inaccessible. To quote Callen:

“Only those few particular combinations of atomic coordinates that are essentially time independent are macroscopically observable.”

In addition, macroscopic coordinates also need to be well defined over a large length scale to be measurable. To give an example of how microscopic freedoms can give rise to macroscopic coordinates, consider the phonon modes of a solid in the long wavelength limit $\lambda \rightarrow \infty$, i.e., macroscopic vibrations. Phonons with large λ are simply sound waves and their frequency ω is given by $\omega = 2\pi v_{\text{sound}}/\lambda$ (with v_{sound} the speed of sound in the solid). As can be seen, ω vanishes in the limit $\lambda \rightarrow \infty$. To realise why this is significant, consider what would happen if this were not the case: if the phonons with macroscopic length scales would have a nonzero frequency, they would be fluctuating and the solid would exhibit random expansions and contractions. This would mean that the volume of the solid would not be constant. It can be concluded that the vanishing frequency of the long-wavelength phonons gives rise to a macroscopically well-defined quantity, namely volume. It is not an accident that the long-wavelength phonons have zero frequency: this is a consequence of a quantum mechanical result known as *Goldstone’s theorem* (see Callen chapter 21). It is beyond the scope of this section to go into details about Goldstone’s theorem, but here it suffices to state that, like Noether’s theorem, it is again a theorem about (broken) symmetry. This emphasises the central role of symmetries in thermodynamics.

The values of the variables A_i together characterise the *thermodynamic state* of a system $\{A_i\} = \{S, A_2, \dots, A_l\}$. It is assumed that the internal energy

²Note that this means that extra care has to be taken when the interface of a substance contains energy. For example, when two identical samples of a liquid with a strong surface tension are added together, the volume of the composite system will be twice large, but not the surface. In this case, it is necessary to treat the surface as a (2 dimensional) thermodynamic system of its own.

is a unique function of the thermodynamic state, so that it is possible to write³:

$$U = U(S, A_2, \dots, A_l). \quad (2.1)$$

The exact functional form of $U(S, A_2, \dots, A_l)$ is known only in the simplest cases (e.g., the ideal gas), but all that really matters is that it exists in principle. This statement is stronger than it might initially sound: by assuming $U = U(S, A_2, \dots, A_l)$, it is immediately asserted that the history of the system is not important: all that matters is the values of the variables A_i right now. The mere fact that U has the functional form postulated here is the basis for many useful thermodynamic results, such as the Maxwell relations. In this light it is also common to refer to U as the fundamental thermodynamic potential.

It is commonly assumed that U is also differentiable with respect to the extensive variables A_i . This makes it possible to relate changes in U to changes in A_i . This is important, because changes in quantities are generally much more experimentally accessible than absolute values of those quantities. The partial differentials $\partial U/\partial A_i$ are measurable thermodynamic quantities and are denoted by f_i :

$$f_i(S, \dots, A_l) = \frac{\partial U(S, \dots, A_l)}{\partial A_i}. \quad (2.2)$$

Because both U and the A_i are extensive, the f_i are independent of the size of the system (i.e., the differential divides out the scale of the system). This makes the f_i intensive quantities. The reader may recall from ET that it is possible to use a Legendre transformation to trade a coordinate A_i with its conjugate f_i as independent variable. This procedure gives rise to different thermodynamic potentials, such as the Helmholtz free energy $F = U - TS$ that depends on temperature T rather than S . However, for NET_{PV} the Legendre transformation is not of great interest so it will not be discussed in detail here.

To give a concrete example of how thermodynamic coordinates are picked, a solar cell under equilibrium is now considered. A solar cell is commonly modelled to a good degree of accuracy as a container with a mixture of two gases: an electron gas and a hole gas. The thermodynamic coordinates that describe such a gas mixture are entropy S , volume V , the number of free electrons N_n , and the number of free holes⁴ N_p . This choice of variables is not unique: depending on the situation, different and/or more variables can be used as well. For example, it is also possible to replace N_n and N_p by two different coordinates that are sometimes more convenient: the number of electron-hole

³By immediately postulating that $U = U(S, A_2, \dots, A_l)$, several of the finer points of thermodynamics (that are outside of the current scope) are glossed over. The critical reader is referred to the textbooks mentioned at the start of this section.

⁴Free electrons and free holes will be indicated in the rest of the text simply by electrons and holes for brevity unless confusion can arise.

pairs N_{eh} and the total electron-hole charge Q_{np} . It should also be clarified that S is not only the entropy of the electron-hole gas but also of the background lattice of the semiconductor. In principle, it is possible to split up S in the entropy of the electrons S_e^- , the entropy of the holes S_h^+ , and the entropy of the lattice S_{lat} . However, it will be assumed throughout (both for equilibrium and non-equilibrium conditions) that all three of these subsystems are in thermal equilibrium so that they have the same temperature. For this reason it is only necessary to consider the sum $S = S_e^- + S_h^+ + S_{\text{lat}}$. Furthermore, at this stage it is important to note that the energy in the field of photons in the solar cell is explicitly not included in the thermodynamic description, even though the photons obviously play a crucial role in the operation of the device. In Section 2.4 it will be explained that (and why) the photons will be considered as a part of the environment rather than the solar cell, though this is a matter of choice and convenience rather than necessity.

The partial differentials corresponding to S , V , N_n , and N_p are temperature T , (negative) pressure $-P$, and the electrochemical potentials $\eta_{n,p}$. The electrochemical potentials deserve some extra explanation. Like their names suggest, $\eta_{n,p}$ are the sum of electrostatic energy ($+\phi$ for holes, $-\phi$ for electrons) and chemical energy $\mu_{n,p}^c$ ⁵. Electrostatic energy, in turn, is related to charge while chemical energy is related to an amount of particles. In a solar cell, however, charge and particle number are both quantised in one and the same entity, namely the electron (and consequently also the hole). Because of this fundamental coupling between charge and particle number it is impossible to remove charge from the system without removing particles and the distinction between electrostatic and chemical energy cannot be made in practise. Thus, only the electrochemical energies are thermodynamically relevant. Electrochemical energies (or potentials, when divided by e) are not mysterious new quantities either: it was shown by Würfel [1] that a regular Volt meter actually measures electrochemical potential rather than electrostatic potential, as is often believed.

The electrochemical potentials are basically the same quantities as the quasi Fermi levels commonly used in semiconductor physics [1]. The only difference is one minus sign for the holes: $E_{Fn} = \eta_n$, $E_{Fp} = -\eta_p$. In a thermodynamic context the minus sign can make equations confusing to read, so the choice was made to use $\eta_{n,p}$ in this chapter. The intensive variables relevant for a solar

⁵Furthermore, there is also a contribution to $\eta_{n,p}$ from the material in which the electrons and holes are located. These contributions are expressed using the electron affinity χ and bandgap E_G , see Eqs. (A.3) in Appendix A.1

cell are defined by:

$$T = \frac{\partial U(S, V, N_n, N_p)}{\partial S} = \left(\frac{\partial U}{\partial S} \right)_{V, N_n, N_p}, \quad (2.3a)$$

$$-P = \frac{\partial U(S, V, N_n, N_p)}{\partial V} = \left(\frac{\partial U}{\partial V} \right)_{S, N_n, N_p}, \quad (2.3b)$$

$$\eta_n = \frac{\partial U(S, V, N_n, N_p)}{\partial N_n} = \left(\frac{\partial U}{\partial N_n} \right)_{S, V, N_p}, \quad (2.3c)$$

$$\eta_p = \frac{\partial U(S, V, N_n, N_p)}{\partial N_p} = \left(\frac{\partial U}{\partial N_p} \right)_{S, V, N_n}. \quad (2.3d)$$

As will be discussed in the next section, T is always the intensive variable corresponding to S . The somewhat peculiar notation on the rightmost sides of Eqs. (2.3) is used in thermodynamics to remind the reader what remaining variables are considered to be the independent freedoms of the system. This also means that for this choice of variables the intensive quantities are functions of S , V , N_n , and N_p as well (e.g., $T = T(S, V, N_n, N_p)$). This is especially important when Legendre transformations are used to trade extensive coordinates for intensive ones.

The idea that U is a differentiable function of several coordinates is commonly stated briefly using the Gibbs form of the first law. The general form is:

$$dU = \sum_{i=1}^l f_i dA_i. \quad (2.4)$$

The differential d is used here to denote what is called an exact differential. This simply means that U can be considered to be a function of the A_i , as is required. If the differentials were inexact (denoted by \bar{d}), it would not be possible to consider U as a function of the A_i . When Eq. (2.4) is applied to solar cells, it becomes:

$$dU = T dS - P dV + \eta_n dN_n + \eta_p dN_p. \quad (2.5)$$

In a sense, Eq. (2.5) can be considered as a condensed version of Eqs. (2.3). Eq. (2.5) can also be given in terms of N_{eh} and Q_{np} by using the following transformations:

$$dN_{\text{eh}} = \frac{1}{2} (dN_n + dN_p), \quad (2.6a)$$

$$dQ_{np} = dN_p - dN_n. \quad (2.6b)$$

Note that Q_{np} is defined here in units of cm^{-3} . With these definitions, the first law can be written as:

$$\begin{aligned} dU &= T dS - P dV + (\eta_n + \eta_p) dN_{\text{eh}} + \frac{1}{2}(\eta_p - \eta_n) dQ_{np}, \\ dU &= T dS - P dV + \Delta\eta_{eh} dN_{\text{eh}} + \Phi_Q dQ_{np}. \end{aligned} \quad (2.7)$$

Here $\Delta\eta_{eh} = (\eta_p + \eta_n)$ is the free energy per electron-hole pair and $\Phi_Q = (\eta_p - \eta_n)/2$ is the conjugate variable for the charge. The latter should not be confused with the electrostatic energy ϕ since it is not possible to extract charge from the cell without extracting electrons or holes. In other words, Φ_Q accounts for both the changes in electrostatic energy as chemical energy when charge is removed/added to the system. If the carrier statistics (either Maxwell-Boltzmann or Fermi-Dirac) expressions for the electrons and holes are used (Eqs. (A.3) in Appendix A.1), then the thermodynamic energy per charge Φ_Q is

$$\Phi_Q = \phi + \chi + \frac{E_G}{2} + \frac{k_B T}{2} \left[F_{1/2}^{\leftarrow} \left(\frac{p}{N_V} \right) - F_{1/2}^{\leftarrow} \left(\frac{n}{N_C} \right) \right] \quad (2.8)$$

and it can be seen that it includes ϕ , the electron affinity χ , the bandgap E_G , as well as terms that take into account that charge has to be changed by adding (or removing) electrons/holes in such a way that the e^-h^+ plasma stays in equilibrium.

The discussion above should make it clear that the first step in making a thermodynamic description of a system is to identify the variables A_i for the system under study and that this first step is crucial to the success of the thermodynamic approach. If not all relevant variables are found, there is a great chance that thermodynamics cannot be applied unless special conditions are met. If one tries to describe the system as $U(S, \dots, A_{l-1})$ (i.e., without taking A_l into account), the thermodynamics will only work as long as A_l stays constant during the experiment (and is therefore essentially still under experimental control) or if U is insensitive to A_l (meaning that A_l is simply not relevant to the physics at hand).

Suppose, for example, one would like to describe a pure gas (such as N_2) in a sealed-off container and one would identify S and V as the relevant coordinates. In this example, the number of molecules N has been deliberately left out, perhaps as a simplification. The proper thermodynamic function for the internal energy is denoted by $U_{\text{gas}}(S, V, N)$. Because the container is sealed-off, $N = N_0$ is constant and for the sake of the experiment at hand, the description in terms of just $\{S, V\}$ will work just as well as that in terms of $\{S, V, N\}$. The internal energy for this experiment is denoted by $\tilde{U}_{\text{gas}}(S, V) = U_{\text{gas}}(S, V, N_0)$ and $\tilde{U}_{\text{gas}}(S, V)$ will be referred to as a *reduced potential*. However, the moment N is no longer constant (think of, e.g. a safety valve that releases gas when

the pressure reaches a critical value), the reduced potential $\tilde{U}_{\text{gas}}(S, V)$ will no longer contain enough information to give a thermodynamic description of the system. Conversely, by using the original description $\{S, V, N\}$ one has already made a lot of implicit assumptions about the physics relevant for the gas. For example, all possible chemical reactions that N_2 can have with itself have been neglected because they are so rare in practice.

Note how the volume of the cell can be omitted in the solar cell description $U(S, V, N_n, N_p)$ proposed earlier in this section. Of course U is not truly independent of V , but this (reduced) representation of U is still useful because the volume of a cell is constant to a very good approximation under most experimental conditions. For this reason, V will be omitted in most discussions of solar cell thermodynamics in this work to simplify the mathematics. However, care should be taken because certain thermodynamic results cannot be used when working with reduced potentials. For example, the Euler equation⁶ states that if:

$$U(\lambda S, \lambda A_2, \dots, \lambda A_l) = \lambda U(S, A_2, \dots, A_l), \quad (2.9)$$

for every scaling factor λ , then:

$$U = \sum_{i=1}^l f_i A_i. \quad (2.10)$$

The Euler equation applies to every system where all the extensive A_i correspond to effects of the same dimensionality (e.g., all A_i are bulk-related variables or they are all surface-related variables). For a solar cell, S , N_n , and N_p are all bulk coordinates, however if one wishes to use the Euler equation, it is necessary to include V as a variable as well. The correct expression for U is:

$$U = TS + \eta_n N_n + \eta_p N_p - PV. \quad (2.11)$$

Here, $P = -(\partial U / \partial V)_{S, N_n, N_p}$ is the pressure of the (ideal) free electron-hole gas. The pressure is the sum of the partial pressures of the electrons and holes:

$$P = P_e^- + P_h^+ = \frac{k_B T}{V} (N_n + N_p). \quad (2.12)$$

Note that a naive application of the Euler equation on the reduced potential $U(S, N_n, N_p)$ would give the incorrect result $U = TS + \eta_n N_n + \eta_p N_p$, which misses the $-PV$ term.

⁶It is somewhat confusing that so many equations and mathematical results carry Euler's name due to his incredible productivity. The Euler equation referred to here is also known as the Euler homogeneous function theorem.

The situation of describing a system while using insufficient thermodynamic variables is somewhat analogous to the phenomenon of degeneracy in quantum mechanics (QM): in QM, degeneracy occurs whenever a certain quantum number no longer uniquely specifies the state of the system and more quantum numbers need to be added to make the specification unique again. In the same way, the energy of the system $\tilde{U}_{\text{gas}}(S, V)$ is degenerate in the numbers S and V because extra information about the particle number N is needed to specify the energy uniquely. The similarity with QM is not just a superficial one: just like in QM, degeneracy of variables in thermodynamics can happen because of symmetries that suddenly become important. A famous example given by Callen is the case of ortho- and para-hydrogen molecules: in an ortho-hydrogen molecule the proton spins of two atoms are in a triplet state, while in para-hydrogen they are in a singlet state. At high temperatures, these two different types of hydrogen gas can be described by a single particle number N_{H} because they behave essentially the same. At very low temperatures, however, the fundamentally different symmetries of ortho- and para-hydrogen necessitate the use of two particle numbers N_{ortho} and N_{para} .

The case of ortho- and para-hydrogen should illustrate that it can be a bit of an art to find coordinates A_i such that the system can be described by a function $U(S, \dots, A_i)$. In other words: thermodynamics provides a prescription for how to build a macroscopic model of a system, but ultimately that model will have to be verified by experiment just like any other physical model. It was demonstrated that too few variables will lead to inconsistencies in the description, so how about the other extreme of the spectrum? Obviously, one can always resort to a fully microscopic description and use all internal quantum mechanical degrees of freedom, which will be denoted by ν_i . So in this case $A_i = \nu_i$. No doubt this will result in a mathematically correct model: no matter what system is under scrutiny or even how close or far away it is from equilibrium, the quantum mechanical equation of motion guarantees that the energy is a unique function of the ν_i . From a practical point of view, though, it is desirable to not invoke more variables than strictly necessary and certainly not as many as the full microscopic description demands. Moreover, as was indicated previously in this section, only very few specific combinations of the ν_i are accessible on macroscopic temporal and spatial scales. Invoking the remaining inaccessible degrees of freedom will only lead to a theory that is completely disconnected from experiment. Still, for the sake of the argument it will sometimes be pretended that the choice $A_i = \nu_i$ is a pseudo-thermodynamic (denoted as: “thermodynamic”) description. As a result, there is now a whole spectrum of thermodynamic descriptions ranging over the number of coordinates used: at one extreme there are the formalisms that use too few variables and at the other end there are those that use too many. A formalism with too few variables is mathematically inconsistent while one with too many is no longer empirical and therefore has no physical content.

The previous discussion shows where the limitation of thermodynamics lies: the moment one loses track of the variables that describe where energy is hiding during the experiment is the moment the thermodynamic description breaks down. In principle, every system, whether it is in thermodynamic equilibrium or not, can be described by adding enough coordinates to the description. After all, the conservation of energy should apply to any system and the energy is, microscopically, only a function of the current state and never of its history or future. Ultimately, one can keep adding variables until every internal freedom has been accounted for, though doing so quickly makes it impossible to exert control over all of these variables.

2.3.1 Entropy

So far the entropy S has only been mentioned in passing as one of the thermodynamic coordinates A_i . The role of S in thermodynamics has a couple of different facets and properties. These will be outlined below, though the following list should be treated as not more than a list of known results: no hard proofs will be presented here.

1. Entropy is an important state variable for the calculation of efficiencies of work-producing apparatus. As an example, consider a heat engine that operates between a hot reservoir at temperature T_h and a cold reservoir at temperature T_c . Then the efficiency is always bounded by $\eta = W/Q_H \leq 1 - T_c/T_h$, with Q_h the amount of heat extracted during one cycle and W the work produced during that cycle. If during the cycle also S amount of entropy is produced, then the work W done by the engine is given by:

$$W = \left(1 - \frac{T_c}{T_h}\right) Q_H - T_c S = \eta_{\max} Q_H - T_c S. \quad (2.13)$$

As can be seen, the production of entropy is in direct relation with the loss of useful work that can be extracted from the engine. A solar cell is no different in that regard: any production of entropy inside of the device diminishes the output power and leads to extra production of low-quality heat.

2. The internal energy $U(S, \dots, A_l)$ is a monotonically increasing function of S and the slope of U vs S is defined to be the temperature T . In short:

$$\left(\frac{\partial U}{\partial S}\right)_{A_2, \dots, A_l} = T > 0. \quad (2.14)$$

Especially the property $\partial U/\partial S > 0$ is very significant: it means that U can be inverted with respect to S so that the entropy can be considered to be a function of U and the remaining A_i :

$$S = S(U, A_2, \dots, A_l). \quad (2.15)$$

By using this inversion it becomes possible to do thermodynamics with S instead of U as the fundamental potential. In fact, it will turn out that for non-equilibrium systems, S is actually a more convenient potential than U because U is a conserved quantity and is therefore more suitable to be used as an independent variable than S . To avoid notational clutter, whenever S is used as a potential, the variables it depends on will be called B_i , with $B_1 = U$ and $B_i = A_i$ for $i \geq 2$. The partial differentials of S are called $g_i = (\partial S / \partial B_i)_{B_j, j \neq i}$ so that the change in entropy of a system can be written as:

$$dS = \sum_{i=1}^l g_i dB_i. \quad (2.16)$$

For a solar cell, the differential form of S can be found by inverting Eq. (2.5):

$$\begin{aligned} dS &= \frac{1}{T} dU - \frac{\eta_n}{T} dN_n - \frac{\eta_p}{T} dN_p \\ &= \beta dU - \tilde{\eta}_n dN_n - \tilde{\eta}_p dN_p \end{aligned} \quad (2.17)$$

The conjugates coordinates β and $\tilde{\eta}_{n,p}$ have been introduced for convenience: the tilde indicates that the quantity has been divided by T . In terms of N_{eh} and Q_{np} , dS is given by:

$$\begin{aligned} dS &= \beta dU - \frac{\Delta\eta_{eh}}{T} dN_{eh} - \frac{\Phi_Q}{T} dQ \\ &= \beta dU - \Delta\tilde{\eta}_{eh} dN_{eh} - \tilde{\Phi}_Q dQ_{np}. \end{aligned} \quad (2.18)$$

3. The Second Law of Thermodynamics states that the entropy of a closed system (meaning: a system that has no exchange of energy of any kind with the outside world) cannot decrease. Closely related to the second law is the maximum entropy principle (explained in detail below), which is a crucial tool for the analysis of the thermodynamic equilibrium of a system. It states that the entropy of a closed equilibrium system takes the highest possible value consistent with the internal constraints imposed on that system.
4. From a microscopic point of view, the entropy $S(U, B_2, \dots, B_l)$ is related to the number of possible microstates of a closed system with macroscopic parameters B_i . This number of microstates will be denoted by $\Omega(U, B_2, \dots, B_l)$ and the famous Boltzmann result then states that:

$$S(U, B_2, \dots, B_l) = k_B \ln \Omega(U, B_2, \dots, B_l). \quad (2.19)$$

Note the dependence of Ω on the variables B_i : not only does Ω depend on the *values* of the B_i , but also on the *choice of the set of variables*. The most extreme example of this second dependency is the “thermodynamic” description based on the full set of microscopic coordinates ν_i : in that case, every “thermodynamic” state is a microstate so that $\Omega(\nu_1, \dots, \nu_l) = 1$ and:

$$S(\nu_1, \dots, \nu_l) = k_B \ln 1 = 0. \quad (2.20)$$

This means that the value of the entropy depends on what is assumed to be known and what not. If this sounds surprising, it is helpful to think of the analogous problem of the interpretation of probabilities. Consider the question “What is the probability that a coin is head side up?”. Obviously, this question cannot be answered in a meaningful way without further details about the coin. Was it flipped just now or did someone put it on its side on the table? Who needs to answer the question and did this person see the coin? Is the coin fair? Questions about probability can only have meaningful answers in a framework of all of the prior knowledge that is involved and assumed. Entropy is no different in that regard (see also, e.g., Öttinger [17]).

The maximum entropy principle deserves some clarification, since it will be used in Section 2.6. A simple example of this principle can be given by considering the equilibrium state of a solar cell. For example, if the cell is insulated from its environment, the system is free to find its own values for N_n and N_p because electrons and holes can be generated and annihilated in pairs. Imagine that the system spontaneously generates a small amount of δN_{eh} electron-hole pairs. Because the cell is insulated, the total energy and charge of the cell cannot change spontaneously, so $\delta U = \delta Q_{np} = 0$. Then from Eq. (2.18) the total first order entropy change can be calculated:

$$\delta S = -\Delta\tilde{\eta}_{\text{eh}}\delta N_{\text{eh}} = -\frac{\Delta\eta_{\text{eh}}}{T}\delta N_{\text{eh}}. \quad (2.21)$$

The maximum entropy principle now states that under equilibrium any such variation δN_{eh} will decrease the entropy so that $\forall \delta N_{\text{eh}} : \delta S \leq 0$. Since δN_{eh} can be both positive and negative, this yields the familiar result that $\Delta\eta_{\text{eh}} = E_{Fn} - E_{Fp} = 0$, i.e., there is no Fermi level splitting in equilibrium and electron-hole pairs are in electrochemical equilibrium and have no free energy. In principle this means that under equilibrium N_{eh} is a thermodynamic variable that can be neglected, but the essential feature of a solar cell is that $\Delta\eta_{\text{eh}} > 0$ so for the generalisation to non-equilibrium N_{eh} is retained.

The imaginary change in N_{eh} is an example of applying an internal constraint. By considering what kind of internal freedoms a closed system has, equilibrium conditions can be obtained. In the example above, only the first

order variation in S was considered, which usually yields a necessary condition for equilibrium, but not a sufficient condition. This is because the first order variations of S usually yields $\delta S = 0$, so if the second order variation $\delta^2 S$ would be positive, the equilibrium would still be unstable. Therefore, to obtain sufficient conditions (often referred to as stability criteria), higher order terms in the variation of S need to be considered as well. The second order variation in entropy for electron-hole variations δN_{eh} is given by:

$$\delta^2 S = \frac{1}{2} \frac{\partial^2 S}{\partial N_{\text{eh}}^2} \delta N_{\text{eh}}^2 = -\frac{1}{2} \left(\frac{\partial \Delta \tilde{\eta}_{\text{eh}}}{\partial N_{\text{eh}}} \right)_{U, Q_{np}} \delta N_{\text{eh}}^2 \leq 0, \text{ so:}$$

$$\left(\frac{\partial \Delta \tilde{\eta}_{\text{eh}}}{\partial N_{\text{eh}}} \right)_{U, Q_{np}} \geq 0. \quad (2.22)$$

This result means that the creation of an electron-hole pair will increase the Fermi level splitting, which is a crucial property that makes it possible to generate useful work from a solar cell.

2.3.2 Generalisation to non-equilibrium

So far only the equilibrium state for a solar cell has been considered, which can be described by the three coordinates S , N_n and N_p . The question that will be addressed now, is how this equilibrium description can be expanded upon to give a description of a solar cell under non-equilibrium conditions. The main idea that will be used here is that a thermodynamic description of a non-equilibrium system is not fundamentally different, except that the choice of variables can be significantly more difficult. In such systems the energy can be distributed over many different phenomena depending on the exact nature of the non-equilibrium situation. For example, in a laminar fluid flow the energy can be traced quite easily and in such a situation NET can be used to obtain transport equations (i.e., the Navier-Stokes equations) for the fluid [3]. However, when the flow transitions from laminar to turbulent it suddenly becomes very difficult to find out how much energy is in the turbulent vortices. Clearly, the energy associated with a vortex is some sort of kinetic energy, but unless the exact flow pattern of the vortex is known, this energy generally cannot be accounted for. It is just as undesirable to calculate the exact flow of a vortex as it is to track heat by considering the kinetic energy of individual particles. Furthermore, in non-equilibrium thermodynamics it can become very difficult to assure that the extensive variables truly scale with the size of the system. If one scales a block of ice to ten times its original size, nothing special happens. However, if a system with a fluid flow is scaled in size, the flow might change completely (going from laminar to turbulent, for example), so in NET it is not always possible to scale extensive variables like in Eq. (2.9).

It suffices to say that there is no single method that can reliably provide a functional set of thermodynamic coordinates that describe any non-equilibrium system, but fortunately there is a method that works well for reasonably well-behaved systems (like solar cells) that do not suffer from complications such as turbulence. This is the method of De Groot and Mazur which basically uses the ET description of a system as a starting point. In NET_{GM} the assumption of local thermodynamic equilibrium (LTE) is used. This means that the system is divided into many small (but still macroscopic) subsystems, each of which is approximately in equilibrium so that they can be described by the same thermodynamic coordinates that were used in the ET description. Each subsystem is labeled with a space coordinate x_j with $j \in \{1, \dots, M\}$ and M the number of subsystems. The entropy S is used as thermodynamic potential and the variables for the system are then denoted by B_{i,x_j} . In general, the coordinates are also allowed to depend on time, though this possibility is not considered here since the main goal is to describe the steady-state operation of a solar cell. The functional dependence of S is now given as:

$$S = S(U_{x_1}, \dots, U_{x_M}, B_{2,x_1}, \dots, B_{2,x_M}, \dots, B_{l,x_1}, \dots, B_{l,x_M}), \quad (2.23)$$

i.e., there are $l \times M$ independent variables. For a solar cell this means that the Gibbs form of the first law in NET_{PV} generalises Eq. (2.17) and becomes:

$$dS = \sum_{j=1}^M \left[\beta_{x_j} dU_{x_j} - \tilde{\eta}_{n,x_j} dN_{n,x_j} - \tilde{\eta}_{p,x_j} dN_{p,x_j} \right]. \quad (2.24)$$

As can be seen, NET_{PV} uses $3M$ coordinates. Obviously, M has to be large enough that there is enough spatial resolution to, e.g., assign a proper temperature T_{x_j} to each subsystem. Yet at the same time $3M$ still has to be much smaller than the number of microscopic coordinates, otherwise NET_{PV} would be overdetermined and would not qualify as a practical macroscopic description. Besides these two requirements, there is no clear criterion for how large M should be. This means that the resulting theory should be independent of M over a wide range of values.

The generalisation of Eq. (2.17) to Eq. (2.24) is mathematically simple, but its physical implications are substantial. It is important to realise that in Eq. (2.17) the three variables U , N_n and N_p (or their conjugates β and $\tilde{\eta}_{n,p}$) are in principle experimentally accessible. However, it cannot be expected that the experimentalist has simultaneous control over all $3M$ variables occurring in Eq. (2.24), so it can be said that a non-equilibrium system is more difficult to control because its description involves more variables. This loss of experimental control can be considered to be the reason that non-equilibrium systems generate entropy.

Because the description of the cell should not depend on the choice of M (as long as M is large enough), it is more convenient to write Eq. (2.24) as

a integral over the cell. The integration domain will be referred to as Ω and the integration densities will be denoted by lower case letters. For example, $dU_{x_j} = du(x) dx^3 = du dx^3$. The dependence of u on x will be suppressed to save space unless extra attention needs to be called to it. Eq. (2.24) then becomes:

$$dS = \iiint_{x \in \Omega} dx^3 (\beta du - \tilde{\eta}_n dn - \tilde{\eta}_p dp) \quad (2.25)$$

$$= \iiint_{x \in \Omega} dx^3 ds. \quad (2.26)$$

Here, s has been introduced as the local entropy density. It is worth noting that in the book by GM slightly different definitions are used for the densities s , u , n , and p since there the densities are normalised to the mass density. In a solar cell with electrons and holes, however, the concept of mass is not very useful and it is more convenient to work with volumetric densities instead.

The notation in Eq. (2.25) is mathematically sloppy since du refers to a change in a thermodynamic variable while dx^3 is simply a volume element for the integration. Nevertheless, for convenience this is the form that will be used and the expression $x \in \Omega$ is placed under the integral to remind the reader that the integration is over x . However, despite its appearance, Eq. (2.25) should still be treated formally as a finite sum: the use of an integral is nothing more than mathematical and notational convenience.

If the maximum entropy principle is applied to Eq. (2.25), it can be demonstrated that under equilibrium all of the intensive variables are constant throughout the system. As before, the e^-h^+ density $n_{eh}(x)$ and charge density $\rho(x)$ are the most convenient coordinates to work with. The first order variation of S is given by:

$$\delta S = \iiint_{x \in \Omega} dx^3 \left(\beta \delta u - \Delta \tilde{\eta}_{eh} \delta n_{eh} - \tilde{\Phi}_Q \delta \rho \right) = 0. \quad (2.27)$$

The $\delta u(x)$, $\delta n_{eh}(x)$, and $\delta \rho(x)$ can be considered to be unknown functions that represent how the system can potentially rearrange its internal configuration and the equality should hold for every conceivable combination of such functions. Since the variation δn_{eh} is unconstrained, it can be concluded again (like in the equilibrium case) that its prefactor has to be zero: $\Delta \tilde{\eta}_{eh} = 0$. Note that this is why $\delta S = 0$ has been imposed, rather than $\delta S \leq 0$. Since the system is closed, the total variations in U and Q_{np} have to be zero:

$$\delta U = \iiint_{x \in \Omega} dx^3 \delta u = 0, \quad (2.28)$$

$$\delta Q_{np} = \iiint_{x \in \Omega} dx^3 \delta \rho = 0. \quad (2.29)$$

These constraints can be imposed by using Lagrange multipliers, but here a different method is introduced based on *displacement vectors* $\delta \mathbf{D}_{u,\rho}(x)$. These displacement vectors describe how the internal energy and charge can be re-distributed through the solar cell. In the continuum limit, the total change in energy or charge in a subvolume at position x is then:

$$\delta u = -\nabla \cdot \delta \mathbf{D}_u, \quad (2.30)$$

$$\delta \rho = -\nabla \cdot \delta \mathbf{D}_\rho. \quad (2.31)$$

The internal constraint that the system is insulated completely can then be imposed by demanding that $\delta \mathbf{D}_{u,\rho} \cdot \hat{\mathbf{n}} = 0$ on the boundary $\partial\Omega$ (with $\hat{\mathbf{n}}$ the unit normal vector). This means that no displacements will take energy or charge in or out of the system. Eq. (2.27) then becomes (recall that $\Delta\eta_{eh} = 0$):

$$\begin{aligned} \delta S &= \iiint_{x \in \Omega} dx^3 \left(-\beta \nabla \cdot \delta \mathbf{D}_u + \tilde{\Phi}_Q \nabla \cdot \delta \mathbf{D}_\rho \right) \\ &= \iint_{x \in \partial\Omega} dx^2 \hat{\mathbf{n}} \cdot \left(-\beta \delta \mathbf{D}_u + \tilde{\Phi}_Q \delta \mathbf{D}_\rho \right) \\ &\quad + \iiint_{x \in \Omega} dx^3 \left(\delta \mathbf{D}_u \cdot \nabla \beta - \delta \mathbf{D}_\rho \cdot \nabla \tilde{\Phi}_Q \right) = 0. \end{aligned} \quad (2.32)$$

The surface integral ($x \in \partial\Omega$) is zero because $\delta \mathbf{D}_{u,\rho} \cdot \hat{\mathbf{n}} = 0$. From the volume integral it can be concluded that $\nabla \beta = \nabla \tilde{\Phi}_Q = 0$, since that is the only way the integral can be zero for all vector fields $\delta \mathbf{D}_{u,\rho}$. Thus, β , $\Delta\tilde{\eta}_{eh} = 0$, and $\tilde{\Phi}_Q$ are all constant as claimed. This justifies why in ET, e.g., only a single temperature is needed to describe the cell: it is possible to use more than one non-local temperature, but the maximum entropy principle can be used to immediately prove that all temperatures are equal anyway. Note how the problem of the number of variables has also sorted itself out in this situation: by pretending that dS can be written in the integral form of Eq. (2.25), one formally assumes an infinite number of coordinates, which is certainly too many from a practical point of view. However, for the equilibrium situation this infinitude of variables can be quickly reduced to three, so this gives a vote of confidence that (as long as the mathematics are applied correctly) the integral form of dS does not lead to big inconsistencies or overdetermination of the problem.

Finally, note that the idea of rearranging internal variables in a closed system only makes sense when extensive coordinates are used as the primary independent variables for the system. Constraints such as Eqs. (2.28) and (2.29) cannot be applied to quantities like (local) temperature and neither is it possible to talk about displacements or flows of temperature. This is one of the reasons why in NET_{GM} extensive coordinates are often preferred over intensive ones.

2.4 The thermodynamic formulation of solar cell physics step 2: Interaction with the environment

So far the discussion of the solar cell thermodynamics has been almost exclusively about closed systems. In fact, the maximum entropy principle was used to demonstrate that in a closed system it does not matter if one divides the cell into subsystems with different temperatures etc.: in the end, the cell will only have just one well-defined temperature. To prove this, the assumption that the cell is a closed system was crucial, otherwise it would not be possible to conclude that the boundary term in Eq. (2.32) is zero.

Interaction with the environment is essential for understanding non-equilibrium situations, especially those that show steady-state behaviour, like a solar cell does. Of course, it would be impossible for a solar cell to generate useful work if it were completely closed off from the rest of the universe and it would be equally impossible to extract that work. However, for a solar cell the interaction with the environment is more complex than for most systems. To see why, it is instructive to first understand how non-equilibrium steady-state situations can occur for simple systems.

In most systems (the solar cell being an important exception, as will be explained) the only interaction with the environment happens at a clearly defined interface that separates the inside world from the outside. For most systems it is possible to draw an imaginary box around the system and to account for the influence of the environment by specifying boundary conditions at the surface of that box. For example, when a metal rod is placed such that one end is in contact with a cold reservoir ($T = T_c$) and the other end is in contact with a hot reservoir ($T = T_h > T_c$), then it can be said that the environment forces two different temperatures at either end of the rod. As was shown in Section 2.3, thermodynamics tries to achieve $\nabla T = 0$ inside of the rod, but because of the BCs imposed by the environment this is simply not possible (unless discontinuities in T are allowed, but those would not make physical sense). After a while, a steady-state non-equilibrium situation is achieved where the temperature in the rod assumes a non-uniform distribution over its length that gradually goes from hot to cold and stays constant in time. Of course, in the long run the combined system of “rod+environment” will reach an equilibrium nonetheless: the rod will conduct heat from hot to cold and eventually the temperature difference will be equalised. However, the key aspect here is that this global equilibration time is much slower than the time it takes the rod to arrive at an internal steady-state non-equilibrium situation. Note, however, that it is not always the case that a system reaches a steady state just because the BCs provided by the environment change much slower than the internal dynamics of the system. As an example, a steady stream of

fluid that is guided through a partially obstructed canal can develop turbulence and the flow will not reach a steady state even though the BCs have no time dependence. In this case, the erratic behaviour is intrinsic to the system itself.

In non-equilibrium situations there is always generation of entropy and since this entropy cannot accumulate inside of the system (because of the steady-state assumption), entropy has to be ejected from the system continuously. For example, consider the metal rod again over a short period of time dt . During this time, the hot reservoir injects a certain amount of heat $dQ_h > 0$ into the rod. It is assumed that the hot reservoir and the local temperature of the rod are equal, so dQ_h is reversible heat and therefore the hot reservoir also injects an amount of entropy $dS_h = dQ_h/T_h$ into the rod. Because of the steady-state condition, the same amount of heat $dQ_c = -dQ_h$ is ejected from the rod at the cold reservoir, along with an amount of entropy $dS_c = dQ_c/T_c$. Note that $dS_h + dS_c < 0$, yet the steady-state assumption asserts that the entropy of the rod cannot have changed, so clearly extra entropy dS_{irr} has been created inside of the rod such that $dS_h + dS_c + dS_{\text{irr}} = 0$. This is the entropy generation associated with the irreversible process of heat conduction inside of the rod. The main conclusion to draw from this thought experiment is that irreversible entropy production is related to gradients in intensive variables ($\nabla T \neq 0$) and flows of extensive variables (in this example the flow of energy that heats up the cold reservoir). In Section 2.5 this relation will be discussed more thoroughly.

In the case of a solar cell, the interaction with the environment is more complex than for the heat conducting metal rod. This is because the environment is providing energy to the cell in the form of a constant stream of photons that do not interact with the cell at a well-defined interface. In fact, some of the low energy photons can travel all the way through the cell and reflect on internal interfaces several times before either being absorbed or exiting the cell. This means that if one chooses to do the energy bookkeeping by putting an imaginary box around the cell and then imposing BCs on the edges of that box, it is important to keep track not only of the cell itself but also of the photons inside in order to make the sure the balance adds up. As a consequence, it is then required to include thermodynamic coordinates for the energy in the photon field and to model transport of energy due to radiative recombination (RR) and reabsorption of photons, which is a formidable task. A model whereby the photons are considered as part of the thermodynamic system will be referred to as a bilateral model. This terminology refers to the fact that the photons can affect the electrons and vice versa. The complexity of such a description invites the use of additional simplifications. Such a simplification can be to consider the photons as part of the environment, meaning that the photons affect the electrons in the cell, but not the other way around. This is called a unilateral description and in this case there are no thermodynamic variables that describe the photons inside the cell. To see how the thermodynamic description of a solar cell (which is bilateral in principle) can be simplified to

a unilateral one, it is instructive to first consider how the problem is usually handled in semiconductor physics.

In solar cell simulations (especially for *c*-Si devices), it is common practise to separate the optical and electrical description of the device. First, an optical simulation is done using, e.g., a Lambert-Beer model or ray tracing. The optical simulation calculates how the light propagates through the cell and how much light is absorbed in the semiconductor at every position and then the result of the optical model is used as an input for the electrical model of the device that calculates the electron and hole densities from transport models. In this procedure, the absorption profile is calculated from known material properties (i.e., the complex dielectric function or refractive index) and does not depend on the non-equilibrium situation of the cell (e.g., the electron and hole densities, which are not even known yet at this first stage). The reason this unilateral coupling of optical and electrical models generally works well, is that there is a great asymmetry in the way photons interact with the electrons/holes in the cell. Whereas the main source of photons is the external influx from the sun, the main source of electrons and holes is the generation due to photons. The exceptions to this rule are high-efficiency devices based on direct-bandgap semiconductors: in those, the radiative recombination (and possibly even stimulated emission) provides a significant contribution to the photon field [24], making it absolutely necessary to couple the optical and electrical models bilaterally rather than unilaterally. Thus, if one wishes to push the efficiency of solar cells to the absolute limit, it becomes essential to explicitly include the thermodynamic coordinates of the photons into the model.

The question of whether to use a unilateral or bilateral coupling of photons and electrons can be rather subtle. It is discussed briefly in the paper by Lindefelt [14], where he remarks that for the accuracy of a calculation it is best to either use a full bilateral coupling between electrical and optical model or to use a fully unilateral model. The in-between option, whereby RR is included as a loss mechanism for the electron-hole pairs but not as a source term for the photon field, is less accurate according to Lindefelt and he bases that claim on a paper [25] that numerically compares these three methods for a GaAs device. The idea is that in RR the photons that are generated are often recycled since the strengths of emission and absorption are proportional to each other. Thus, if in a unilateral model the RR is included, reabsorption is not considered and the recombination losses are overestimated. In the modelling of *c*-Si solar cells, however, it is not uncommon to use a unilateral model together with RR. The situation for *c*-Si (which has an indirect bandgap) is different than for GaAs (a direct bandgap semiconductor), since in *c*-Si the RR is generally much smaller than the Auger recombination and because the reabsorption of photons is very unlikely due to the very low absorption coefficient for photons with energies close to the *c*-Si bandgap.

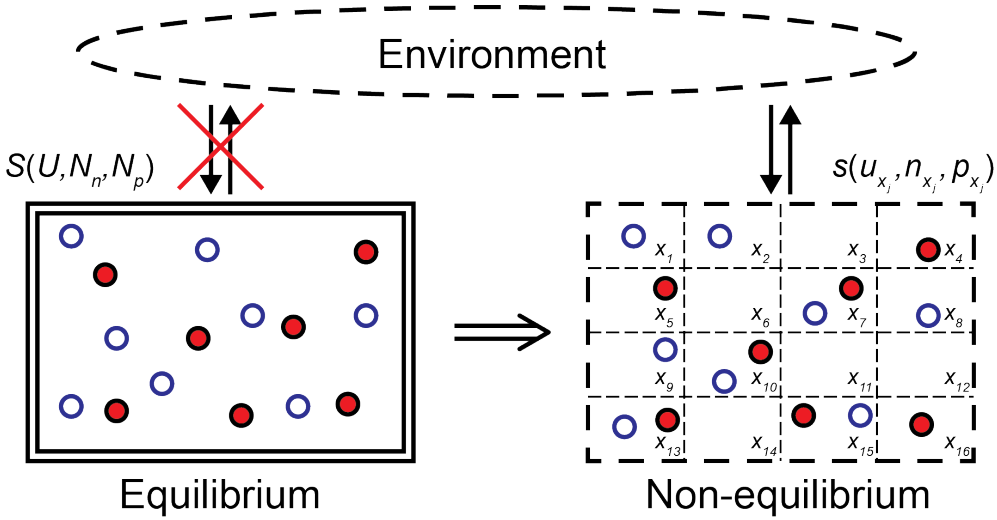


Figure 2.1: Schematic representation of the extension of equilibrium to non-equilibrium thermodynamics for a solar cell: the equilibrium variables become localised and the system is no longer insulated from its environment.

The preceding discussion shows that it is not always entirely clear how to simplify a bilateral description to a unilateral one, but the main idea is clear: first the absorption profile of the photons is calculated from predetermined material parameters and then this generation profile is used as a source term for the thermodynamic processes in the solar cell. These source terms will be addressed in Section 2.5. So to conclude, the environment couples to a solar cell in two distinct ways: at the cell boundaries there is the thermal and electrical contact with the environment which is described by boundary conditions. Inside of the cell there is the interaction between the light field and the electron-hole gas which is (approximately) taken into account through source terms.

2.5 The thermodynamic formulation of solar cell physics step 3: Continuity equations

Figure 2.1 illustrates the main ideas that were discussed so far in Sections 2.3 and 2.4. At this point it is time to develop the mathematics that describe the solar cell. It was explained how the equilibrium solar cell description was generalised to non-equilibrium by dividing the cell in M small subsystems with coordinates x_j and assigning thermodynamic coordinates to each subsystem. This made it possible to write dS as a sum of the local entropy ds_{x_j} over the

subsystems, with ds given by:

$$ds = \beta du - \tilde{\eta}_n dn - \tilde{\eta}_p dp. \quad (2.33)$$

Eq. (2.33) can be considered to be the local form of the first law. Recall that, for mathematical convenience, the sum over x_i was treated as a volume integral and that the densities u , n , and p were used as the independent thermodynamic coordinates for the solar cell.

The description of a working solar cell does not only deal with densities but also with flows. Flows (denoted by \mathbf{J}) are introduced into the model by postulating continuity equations for the densities of extensive variables, just like in SCP. However, a significant difference between NET_{GM} and SCP is that in the former the entropy continuity equation is of central interest while in SCP the main focus tends to be on the conservation laws for electrons and holes. The continuity equations for s , u , n , and p are given by:

$$\frac{\partial s}{\partial t} + \nabla \cdot \mathbf{J}_s = q_s, \quad (2.34a)$$

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{J}_u = q_u, \quad (2.34b)$$

$$\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{J}_n = q_n, \quad (2.34c)$$

$$\frac{\partial p}{\partial t} + \nabla \cdot \mathbf{J}_p = q_p. \quad (2.34d)$$

The symbol q is used to represent source terms in the continuity equations and the $\partial/\partial t$ terms have been included for completeness because they will be necessary for deriving the expression for q_s in the next subsection. For the remainder of this chapter, however, the time derivatives will be set to zero because only steady-state solutions are of interest here. Eqs. (2.34) put constraints on what flows can be considered, but they do not give a unique definition since \mathbf{J} only appears inside of a divergence in each equation. The unique definitions of the flows will be given in Section 2.6, but first the continuity equations, and the entropy source term q_s in particular, will be considered on their own merits.

The source terms $q_{u,n,p}$ in Eqs. (2.34) are, in principle, known quantities and will be discussed first because they are important for deriving q_s . They are:

$$q_u = \epsilon_\gamma(x)G(x) - \epsilon_{\gamma,\text{Rad}} R_{\text{Rad}}, \quad (2.35a)$$

$$q_n = G(x) - R(n, p, \dots), \quad (2.35b)$$

$$q_p = G(x) - R(n, p, \dots). \quad (2.35c)$$

Here, G is the generation rate of electron-hole pairs obtained from an optical model, ϵ_γ the average absorbed photon energy, $\epsilon_{\gamma,\text{Rad}}$ the average energy of

photons emitted by radiative recombination⁷, and R (R_{Rad}) the (radiative) electron-hole recombination rate. At this point it is important to make a clear distinction between two different type of sources: those that are due to interaction with the environment and those that are due to internal irreversible processes in the system. This will be especially important when considering the entropy source term, since there is a fundamental difference between entropy that is generated due to irreversible dynamics and entropy that is transferred between the system and its surroundings.

Section 2.4 discussed that in a solar cell the photons interact with the cell all the way throughout the bulk. If the photons were represented by a set of thermodynamic coordinates, then such an interaction would not change the energy density and $q_u = 0$. However, the choice for a unilateral coupling between photons and electrons necessitates the use of a nonzero source term $q_u > 0$ which represents the interaction between the environment and the cell. Note that the source term for u depends on whether or not the background semiconductor lattice of the cell is considered as part of the thermodynamic system or not⁸: if only the electron-hole gas would considered (assuming it is non-degenerate), then q_u would instead be given by:

$$q_{u,\text{electron-hole gas}} = (E_G + 3k_B T)G - \epsilon_{\gamma,\text{Rad}} R_{\text{Rad}} + H, \quad (2.36)$$

with E_G the bandgap of the semiconductor. The $3k_B T$ term is the average kinetic energy of an electron-hole pair and H is the energy transferred between the lattice and the e^-h^+ gas⁹. To put it succinctly: q_u in Eq. (2.35a) represents all the energy transferred to and from the cell by the photons and includes the heating of the semiconductor lattice, whereas $q_{u,\text{electron-hole gas}}$ in Eq. (2.36) only represents the energy transferred to and from the electron-hole pairs.

Looking at the source terms $q_{n,p}$, it can be seen that these have two different contributions: the generation G that represents an interaction with the environment and the recombination R that is mostly due to the internal dynamics of the cell. The exception to the latter is radiative recombination R_{Rad} , which also represents a coupling to the environment. For now, $R_{\text{Rad}} = 0$ will be assumed for the reasons given by Lindelfelt (see Section 2.4) and this choice also makes the distinction between internal dynamics (R) and interaction with

⁷Which is equal to the semiconductor bandgap if broadening effects are ignored

⁸recall from Section 2.3 that the choice was made to include the lattice in the thermodynamic description presented here.

⁹Usually in the form of heat, hence the symbol H . Note that the energy thermalisation loss upon the excitation (the average difference between the absorbed photon energy and the energy of an electron hole pair) of an e^-h^+ pair is not included in H because only thermalised electrons and holes are considered to be part of the system. This means that the thermalisation energy transfer is directly from the photons to the lattice, both of which are part of the environment here.

the environment (G) more clear. In situations where RR is of interest, it will be accounted for explicitly by writing $R + R_{\text{Rad}}$.

2.5.1 The entropy source term q_s

Now that the role of the other source terms is clear, it is time to obtain an expressions for q_s and for \mathbf{J}_s and to interpret their meaning. Deriving them is done by combining the local form of the first law Eq. (2.33) with the continuity equations (2.34). The former is used to obtain $\partial s/\partial t$:

$$\frac{\partial s}{\partial t} = \frac{\partial s}{\partial u} \frac{\partial u}{\partial t} + \frac{\partial s}{\partial n} \frac{\partial n}{\partial t} + \frac{\partial s}{\partial p} \frac{\partial p}{\partial t} \quad (2.37)$$

$$= \beta \frac{\partial u}{\partial t} - \tilde{\eta}_n \frac{\partial n}{\partial t} - \tilde{\eta}_p \frac{\partial p}{\partial t}. \quad (2.38)$$

Next, the continuity equations are used to eliminate all $\partial/\partial t$:

$$q_s - \nabla \cdot \mathbf{J}_s = \beta (q_u - \nabla \cdot \mathbf{J}_u) - \tilde{\eta}_n (q_n - \nabla \cdot \mathbf{J}_n) - \tilde{\eta}_p (q_p - \nabla \cdot \mathbf{J}_p) \quad (2.39)$$

The next step is to push β and $\tilde{\eta}_{n,p}$ inside of the divergences by using the product rule $f \nabla \cdot \mathbf{w} = \nabla \cdot (f \mathbf{w}) - \mathbf{w} \cdot \nabla f$. This gives:

$$\begin{aligned} q_s - \nabla \cdot \mathbf{J}_s &= \beta q_u - \tilde{\eta}_n q_n - \tilde{\eta}_p q_p + \mathbf{J}_u \cdot \nabla \beta - \mathbf{J}_n \cdot \nabla \tilde{\eta}_n - \mathbf{J}_p \cdot \nabla \tilde{\eta}_p \\ &\quad - \nabla \cdot (\beta \mathbf{J}_u - \tilde{\eta}_n \mathbf{J}_n - \tilde{\eta}_p \mathbf{J}_p). \end{aligned} \quad (2.40)$$

At this point there are two methods that are commonly used in literature to obtain expressions for q_s and \mathbf{J}_s . Lindefelt [14], for example, proceeds by postulating that:

$$T \mathbf{J}_s = \mathbf{J}_Q = \mathbf{J}_u - \eta_n \mathbf{J}_n - \eta_p \mathbf{J}_p, \quad (2.41)$$

with \mathbf{J}_Q the heat flow. This definition of \mathbf{J}_s appears to be motivated by Eq. (2.33) and seems reasonable, but there is no clear argument why Eq. (2.41) necessarily follows from Eq. (2.33) since flows are not a type of differential. The advantage of postulating (2.41) is that q_s then follows immediately from Eq. (2.40):

$$q_s = \beta q_u - \tilde{\eta}_n q_n - \tilde{\eta}_p q_p + \mathbf{J}_u \cdot \nabla \beta - \mathbf{J}_n \cdot \nabla \tilde{\eta}_n - \mathbf{J}_p \cdot \nabla \tilde{\eta}_p. \quad (2.42)$$

Another convenient form for the entropy source term is obtained by replacing \mathbf{J}_u with \mathbf{J}_Q using Eq. (2.41):

$$T q_s = q_u - \eta_n q_n - \eta_p q_p - \mathbf{J}_Q \cdot \frac{\nabla T}{T} - \mathbf{J}_n \cdot \nabla \eta_n - \mathbf{J}_p \cdot \nabla \eta_p. \quad (2.43)$$

The second method to obtain Eqs. (2.41) and (2.42) is found in, e.g., the book by De Groot and Mazur [3]. Here it is asserted that in Eq. (2.40) the parts

inside of the divergence have to be equal (yielding Eq. (2.41)) as well as the parts outside of the divergence (yielding Eq. (2.42)). The authors assert that this split of Eq. (2.40) is uniquely determined by two additional requirements, namely that of Galilei invariance and the requirement that q_s becomes zero when global thermodynamic equilibrium is reached. This assertion is made without proof, however, so it remains unclear which method is the best to justify Eqs. (2.41) and (2.42). Fortunately, this does not mean that the accuracy of these equations is in serious doubt, since there are also works that derive very similar results from microscopic principles (see, e.g., Parrot [13] and Greulich [11]).

2.5.2 Loss processes in the bulk of a solar cell and the entropy generation rate

Having derived q_s in Eq. (2.42), it is time to discuss its physical meaning. As was stated earlier, the entropy source term can be split up into two different contributions: entropy that is transferred from the environment (i.e., the sunlight) to the cell $q_{s,\text{env}\rightarrow\text{cell}}$ and entropy that is generated due to irreversible processes $q_{s,\text{irr}}$. The irreversible contribution can again be split up into entropy generation due to (non-radiative) recombination and due to transport phenomena. According to the second law of thermodynamics, irreversible processes always generate net entropy, even locally. Thus, the entropy source term for the solar cell q_s can be written as:

$$q_s = q_{s,\text{env}\rightarrow\text{cell}} + q_{s,\text{irr}}, \quad (2.44a)$$

$$q_{s,\text{env}\rightarrow\text{cell}} = \beta q_u - (\tilde{\eta}_n + \tilde{\eta}_p)G, \quad (2.44b)$$

$$q_{s,\text{irr}} = q_{s,\text{rec}} + q_{s,\text{trans}} \geq 0, \quad (2.44c)$$

$$q_{s,\text{rec}} = (\tilde{\eta}_n + \tilde{\eta}_p)R \geq 0, \quad (2.44d)$$

$$q_{s,\text{trans}} = \mathbf{J}_u \cdot \nabla \beta - \mathbf{J}_n \cdot \nabla \tilde{\eta}_n - \mathbf{J}_p \cdot \nabla \tilde{\eta}_p \geq 0. \quad (2.44e)$$

It was discussed in Section 2.3 that the creation of entropy goes at the cost of useful work that can be extracted from the cell, making the entropy source term a useful tool for understanding loss processes. However, it is important to keep in mind that neither of the terms q_s and $q_{s,\text{irr}}$ are equal to the total entropy production rate in the complete photovoltaic process, since there is also an entropy source term associated with the photon field, which will be referred to as $q_{s,\gamma}$ (with γ referring to photons). Due to the unilateral treatment of the interaction of photons with the cell, this term remained outside of the scope of the discussion, but of course it does matter in the greater scheme. The exact form of $q_{s,\gamma}$ can be deduced from the entropy formula for radiation presented by Landsberg (Eq. (5.6) in his article [19]) and depends on the spectral shape of the radiation field (which varies throughout the cell) and on the angular distribution of the photon propagation directions. So, for example,

a collimated laser beam has a very low entropy because it has a very low spread in both energy and angular distribution. Sunlight is also well-collimated but has a large spread in its photon energy distribution and therefore also a larger entropy. There is generally no entropy generation associated with the (reversible) propagation of light through a refractive medium, so if all light propagation in the cell is reversible then all contributions to $q_{s,\gamma}$ are due to absorption and emission of photons. Irreversible (i.e., non-refractive) scattering of light, such as Rayleigh scattering and scattering with free electrons, can also increase the entropy of the photons by increasing the angular spread of the radiation [1]. These scattering contributions can be included in $q_{s,\gamma}$ as well, if necessary.

Usually $q_{s,\gamma}$ is negative because the light field is transferring energy to the solar cell, though it can be positive if there is locally more radiative recombination than absorption or if the light is being irreversibly scattered. Furthermore, the loss of entropy from the radiation field due to loss of energy $q_{s,\gamma}$ does not necessarily cancel against the gain in entropy in the cell due to absorption (or emission) of energy βq_u . Only if the transfer of energy between the cell and the radiation field is reversible (which is usually not the case) are these terms equal and opposite. In this work, the symbol σ_{irr} will be used to indicate the total entropy production rate in the combined system “solar cell + photons”. The second law states that the total entropy production rate is always larger than zero:

$$\begin{aligned}\sigma_{\text{irr}} &= q_s + q_{s,\gamma} + R_{\text{Rad}}(\tilde{\eta}_n + \tilde{\eta}_p) \\ &= q_{s,\gamma} + \beta q_u + \beta \Delta \eta_{eh}(R + R_{\text{Rad}} - G) \\ &\quad + \mathbf{J}_Q \cdot \nabla \beta - \beta \mathbf{J}_n \cdot \nabla \eta_n - \beta \mathbf{J}_p \cdot \nabla \eta_p \geq 0.\end{aligned}\tag{2.45}$$

Recall the convention to mention R_{Rad} explicitly when necessary. The terms in Eq. (2.45) capture all local loss processes that can occur in the bulk of a solar cell. No exact form for $q_{s,\gamma}$ will be given or used here, but some general considerations about the role of $q_{s,\gamma}$ in solar cell design are in order and these are presented below. In particular, it will be motivated why $q_{s,\gamma}$ is ultimately not of great importance for the optimisation of many types of solar cells, especially those based on *c*-Si. To see why this is the case, it is important to understand the physical meaning of the entropy generation processes and what possible practical methods are available to reduce the corresponding losses. To put it in different words: for the purpose of solar cell optimisation it is unnecessary to consider entropy generation terms that belong to loss processes one is not willing to (or incapable to) address anyway.

The bulk losses can be categorised and interpreted as follows:

- Generation related losses. These are losses associated with the absorption of photons and the production of e^-h^+ pairs. The most important loss

in a *c*-Si cell is of course the thermalisation of an e^-h^+ pair directly after the absorption of a photon with energy $\geq E_G$. The entropy generation associated with thermalisation loss is $\beta(q_u - \Delta\eta_{eh}G)$. In practice, the most feasible method to reduce thermalisation losses is to increase the product $\Delta\eta_{eh}G$ as much as possible. This can be achieved by letting generation take place in parts of the cell with a large Fermi level splitting and by building multi-junction cells with different bandgaps for each junction. Note that the benefit of increasing E_G is that this also increases $\Delta\eta_{eh}$. Often, great emphasis is put on the value of E_G , but ultimately it is the free energy per e^-h^+ pair $\Delta\eta_{eh}$ that determines how much useful energy can be extracted per absorbed photon. For illustration, recall that for an electron-hole gas satisfying Maxwell-Boltzmann statistics $\Delta\eta_{eh}$ is given by:

$$\Delta\eta_{eh} = E_G + k_B T \ln \left(\frac{np}{N_C N_V} \right), \quad (2.46)$$

which not only shows that $\Delta\eta_{eh}$ increases with E_G , but also that other quantities matter just as well. Indeed, it is even possible to increase $\Delta\eta_{eh}$ by using materials with low N_C and N_V as was also noted by Osterloh [26]. Since $N_{C,V} \sim (m_{n,p}^*)^{3/2}$, this means that it is beneficial to use materials with a low effective mass.

The term $q_{s,\gamma}$ contains generation related contributions that depend on the angular distribution of the photon propagation directions and the spectral distribution of the photon energies. Those contributions basically characterise how well the photon field can be trapped inside of the cell to absorb as much of the incident light as possible. In the design of a solar cell this problem is usually optimised beforehand and separately from the electrical design. Thus, considering $q_{s,\gamma}$ becomes important once the solar cell has been improved so extensively that disjointed optimisation of optical and electrical design can no longer yield any gains. Some of the optimisations that are possible by considering $q_{s,\gamma}$ are highly academical, though. For example, it is possible to redirect photons emitted from the cell (through thermal radiation or radiative recombination) back towards the sun as much as possible. This way, the lifetime of the sun is extended and more solar energy can be extracted in the long run (see also Würfel [1] section 2.5). Of course such an optimisation scheme is not very interesting in practice. This should help to illustrate that for solar cell engineering it is not necessary to consider every entropy generation contribution imaginable and that one should always consider the physical content of the loss terms one is trying to reduce. An interesting discussion of optical design principles is presented by Polman [27].

- Recombination related losses. Non-radiative recombination is accounted

for by the product $\Delta\eta_{eh}R$ and since it is undesirable to reduce $\Delta\eta_{eh}$ this just means that the total recombination over the whole cell

$$R_{\text{tot}} = \iiint_{x \in \Omega} R \, dx^3, \quad (2.47)$$

needs to be kept as low as possible to reduce the total entropy generation associated with recombination. This is achieved by reducing the number of recombination-active defects (i.e. making the integrand of R_{tot} smaller) and by keeping the cell as thin as possible (i.e., making the integration domain Ω smaller). The entropy generation due to RR requires some extra consideration since this process makes positive contributions to the terms $q_{s,\gamma}$ and $\beta\Delta\eta_{eh}R_{\text{Rad}}$ while giving a negative contribution to βq_u (since energy is ejected from the cell). Because of this, the entropy generation due to RR is usually lower than due to other recombination processes. The physical reason for this is that the photons emitted due to RR have a well-defined energy which makes it possible to use special optical tricks (energy-selective mirrors etc.) to keep them inside of the cell. However, it is still a spontaneous process, so the total contribution to all three terms together has to be ≥ 0 . This example further illustrates how the practical limitations can influence the interpretation of loss processes: if one is not considering to use special optical techniques to capture photons from RR, then for the optimisation of the cell it is unnecessary to consider the entropy generation terms corresponding to RR.

The remaining loss terms $\mathbf{J}_Q \cdot \nabla\beta$, $-\beta\mathbf{J}_n \cdot \nabla\eta_n$, and $-\beta\mathbf{J}_p \cdot \nabla\eta_p$ are the transport losses and they will be the subject of Section 2.6 below.

2.6 The thermodynamic formulation of solar cell physics step 4: Onsager theory

So far no specific mathematical form has been given for the currents \mathbf{J}_s , \mathbf{J}_u , \mathbf{J}_n , and \mathbf{J}_p . As was briefly alluded to in Section 2.4, the NET_{GM} view of flows is that they always arise from gradients in local intensive variables: temperature gradients drive heat flows; pressure gradients drive mass flows; gradients in chemical potential cause diffusion etc.¹⁰ Interestingly, in SCP a common viewpoint is that diffusive flows are driven by gradients in density (which is, from a thermodynamic point of view, a local *extensive* variable) rather than

¹⁰In fact, the idea that gradients in thermodynamic quantities drive flows is a useful starting point to separate the interesting transport phenomena from trivial ones such as uniform motion of the whole system.

chemical potential. The diffusive flow $\mathbf{J}_n^{\text{diff}}$ of electrons (with density n) is given in Fick's law:

$$\mathbf{J}_n^{\text{diff}} = -D_n \nabla n, \quad (2.48)$$

with D_n the diffusion constant. If the current is considered to be driven by the chemical potential μ_n^c , however, the expression should be written instead as:

$$\mathbf{J}_n^{\text{diff}} = -\frac{\sigma_n}{e^2} \nabla \mu_n^c. \quad (2.49)$$

If the electrons are considered to be an ideal gas, then $\mu_n^c = k_B T \ln(n/N_C)$ (see also Eqs. (A.3) in Appendix A.1) and:

$$\mathbf{J}_n^{\text{diff}} = -\frac{\mu_n n}{e} \nabla \left[k_B T \ln \left(\frac{n}{N_C} \right) \right] = -\frac{k_B T}{e} \mu_n \nabla n. \quad (2.50)$$

The motivation for writing the proportionality constant as a conductivity $\sigma_n = e\mu_n n$ is that $\nabla \mu_n^c$ is a force (unit: N) just like $\nabla \phi$ and it is common practice to write currents driven by forces as “mobility \times density \times force”. Note that the two rightmost expressions in Eq. (2.49) are only valid when the electrons can be described by Boltzmann statistics and when T and N_C are constant. Indeed, these are crucial assumptions for the validity of Fick's law as well. By comparing Eqs. (2.48) and (2.49) the well-known Einstein relation $D_n = \mu_n k_B T / e$ is obtained, demonstrating that μ_n in Eq. (2.49) is the familiar carrier mobility from SCP. Usually the Einstein relation is derived by comparing diffusion to another flow-driving force such as gravity, but this is not strictly necessary: all that is needed is to identify the chemical potential as the driving force of diffusion.

In NET_{GM} the gradients of intensive variables that drive flows are called “thermodynamic forces” or “generalised forces”. In some works (such as Le Bellac [28]) those two terms are already reserved for other quantities and then new jargon is introduced by the author (“affinities”, for example), but in this work the terminology based on forces is used throughout because this expresses the central idea most intuitively. When the context is clear the thermodynamic forces will simply be referred to as forces for brevity.

The identification of the right thermodynamic forces and the flows is not trivial if one has to do this based on first principles, but fortunately NET_{GM} has a standard procedure for obtaining the correct flux-force pairs (and the transport relations between them) from the entropy generation rate. The method is based on the work by Lars Onsager [20] and on Einstein fluctuation theory [29] but the details of those works will not be discussed here. The main result can be summarised as follows as follows:

1. First an expression for the irreversible entropy generation rate in the semiconductor ($q_{s,\text{irr}}$) is written out. As was shown in Section 2.5 there

are various possible forms for $q_{s,\text{irr}}$ that can be obtained by substitutions, such as:

$$q_{s,\text{irr}} = (\tilde{\eta}_n + \tilde{\eta}_p)R + \mathbf{J}_u \cdot \nabla\beta - \mathbf{J}_n \cdot \nabla\tilde{\eta}_n - \mathbf{J}_p \cdot \nabla\tilde{\eta}_p, \quad (2.51)$$

or:

$$q_{s,\text{irr}} = \beta\Delta\eta_{eh}R - \mathbf{J}_Q \cdot \frac{\nabla T}{T^2} - \mathbf{J}_n \cdot \beta\nabla\eta_n - \mathbf{J}_p \cdot \beta\nabla\eta_p. \quad (2.52)$$

Both of these expressions can be used, resulting in different flux-force pairs. Of course, the different choices for $q_{s,\text{irr}}$ lead to equivalent thermodynamic descriptions of the fluxes in the solar cell, though there often is a practical preference for a particular expression for $q_{s,\text{irr}}$. In this case it will turn out that Eq. (2.52) is more convenient to work with than Eq. (2.51).

2. From each term in $q_{s,\text{irr}}$ a flux-force pair of the form $\mathbf{J}_i \cdot \mathbf{F}_i$ is identified, so the terms in Eq. (2.51) yield the pairs:

$$\mathbf{J}_u \quad \leftrightarrow \quad \mathbf{F}_u = -\frac{\nabla T}{T^2} = \nabla\beta, \quad (2.53a)$$

$$\mathbf{J}_n \quad \leftrightarrow \quad \mathbf{F}_n = -\nabla(\beta\eta_n), \quad (2.53b)$$

$$\mathbf{J}_p \quad \leftrightarrow \quad \mathbf{F}_p = -\nabla(\beta\eta_p). \quad (2.53c)$$

Alternatively, one can identify from Eq. (2.52) the pairs:

$$\mathbf{J}_Q \quad \leftrightarrow \quad \mathbf{F}_Q = -\frac{\nabla T}{T^2} = \nabla\beta, \quad (2.54a)$$

$$\mathbf{J}_n \quad \leftrightarrow \quad \mathbf{F}_n = -\beta\nabla\eta_n, \quad (2.54b)$$

$$\mathbf{J}_p \quad \leftrightarrow \quad \mathbf{F}_p = -\beta\nabla\eta_p. \quad (2.54c)$$

As can be seen, care should be taken with the symbols $\mathbf{F}_{n,p}$: depending on the form of $q_{s,\text{irr}}$ that is used the forces take different forms. For the rest of Section 2.6 the forces $\mathbf{F}_{Q,n,p}$ from Eqs. (2.54) will be used because they have the advantage that the inverse temperature β does not appear under the gradient in $\mathbf{F}_{n,p}$.

The perceptive reader may have observed that the term $\beta\Delta\eta_{eh}R$ has not been discussed here. Indeed, associated with this term are the *scalar flux* R and the *scalar force* $\beta\Delta\eta_{eh}$, but these can be ignored for now and will be discussed in Section 2.7.

3. Finally, Onsager's theory asserts that the fluxes (measured in the reference frame of the cell) can be written as a general linear combination of

the forces. The coefficients in this linear combination are denoted by the symbols L_{ij} with i and j labelling the fluxes and forces:

$$\mathbf{J}_Q = L_{QQ}\mathbf{F}_Q + L_{Qn}\mathbf{F}_n + L_{Qp}\mathbf{F}_p, \quad (2.55a)$$

$$\mathbf{J}_n = L_{nQ}\mathbf{F}_Q + L_{nn}\mathbf{F}_n + L_{np}\mathbf{F}_p, \quad (2.55b)$$

$$\mathbf{J}_p = L_{pQ}\mathbf{F}_Q + L_{pn}\mathbf{F}_n + L_{pp}\mathbf{F}_p. \quad (2.55c)$$

These equations can be stated more compactly as:

$$\mathbf{J}_i = \sum_j L_{ij}\mathbf{F}_j \quad (2.56)$$

It should be noted that these equations are valid for transport in an isotropic system. In Section 2.7 the consequences and limitations of this assumption will be analysed, but for now the assumption of isotropy will be used to simplify the discussion.

In Eqs. (2.55), the diagonal terms $L_{QQ}\mathbf{F}_Q$, $L_{nn}\mathbf{F}_n$, and $L_{pp}\mathbf{F}_p$ account for well known transport phenomena such as heat diffusion due to a temperature gradient (i.e., Fourier's law) and charge transport due to gradients in the electrochemical potentials. The off-diagonal elements like $L_{Qn}\mathbf{F}_n$ account for effects such as the Peltier effect whereby heat is transported along with the flow of electrons. The cross-terms $L_{np}\mathbf{F}_p$ and $L_{pn}\mathbf{F}_n$ represent momentum exchange between electrons and holes due to microscopic collisions [14] arising from short-range interactions between electrons and holes that are quantum mechanical in nature. For example, at close proximity an electron-hole pair can temporarily form an exciton, which is a quantum mechanical state similar to that of the hydrogen atom (i.e., a bound state between two oppositely charged particles). During this binding the two particles can exchange momentum and in this manner the electrons can exert a short-range force on the holes and vice versa. Macroscopically this force then gives rise to the off-diagonal transport terms $L_{np}\mathbf{F}_p$ and $L_{pn}\mathbf{F}_n$. It is expected that this effect becomes stronger in materials with a strong exciton binding energy, which are usually materials with a low dielectric constant (see, e.g., Ibach and Lüth [30]) such as polymers. The effect of these short-range interactions should be distinguished from the long range Coulombic interactions between electrons and holes in a solar cell that are accounted for by the electrostatic potential ϕ (which modifies $\eta_{n,p}$ and $\mathbf{F}_{n,p}$) and the Poisson equation (see Eq. (A.1) in Appendix A.1).

Due to the linear form of Eqs. (2.55) it is sometimes said that NET_{GM} is a linear theory, but care should be taken when interpreting this statement because NET_{GM} only asserts that the fluxes are always linear in the forces, (meaning that the L_{ij} cannot depend on the $\mathbf{F}_{Q,n,p}$). However, the coefficients

L_{ij} are not necessarily constant: they can depend on other thermodynamic quantities such as T , $\eta_{n,p}$, n , p and on x explicitly. For example, the electron electrical conductivity and L_{nn} can be related as $L_{nn} = T\sigma_n/e^2 = T\mu_{nn}/e$, so L_{nn} is not constant. When Eqs. (2.55) are inserted into the continuity equations Eqs. (2.34), the resulting transport equations are therefore generally non-linear¹¹.

The flux-force equations (2.55) can be used to rewrite the transport dissipation term $q_{s,\text{trans}}$ from Eqs. (2.44):

$$\begin{aligned} q_{s,\text{trans}} &= \mathbf{J}_Q \cdot \mathbf{F}_Q + \mathbf{J}_n \cdot \mathbf{F}_n + \mathbf{J}_p \cdot \mathbf{F}_p \quad (\text{by definition of the forces}) \\ &= \sum_j \left(L_{Qj} \mathbf{F}_j \cdot \mathbf{F}_Q + L_{nj} \mathbf{F}_j \cdot \mathbf{F}_n + L_{pj} \mathbf{F}_j \cdot \mathbf{F}_p \right) \\ &= \sum_{i,j} L_{ij} \mathbf{F}_i \cdot \mathbf{F}_j \geq 0. \end{aligned} \quad (2.57)$$

The inequality $q_{s,\text{trans}} \geq 0$ immediately imposes a constraint on the coefficients L_{ij} : since the L_{ij} are independent of the forces \mathbf{F}_j , the inequality $q_{s,\text{trans}} \geq 0$ has to hold for every possible combination of forces \mathbf{F}_Q , \mathbf{F}_n , and \mathbf{F}_p . This is only possible if the elements L_{ij} form a positive-definite matrix, meaning that all of its eigenvalues are positive. Consequently, the diagonal elements L_{QQ} , L_{nn} and L_{pp} are all positive as well. The off-diagonal elements can be negative, but their absolute values are bounded. For example, a simple bound for the elements L_{np} and L_{pn} is given by:

$$4L_{nn}L_{pp} \geq (L_{np} + L_{pn})^2. \quad (2.58)$$

In Section 2.7, the Onsager relations are discussed and these require that $L_{np} = L_{pn}$, so that $L_{nn}L_{pp} \geq L_{np}^2$. The off-diagonal coefficients have a significant influence on the transport losses that occur in the cell. To illustrate this, consider for the moment only the dissipation due to the electron and hole forces in a 1D solar cell. For simplicity it is assumed that $F_Q = 0$. There is a total charge current $J_{\text{charge}} = e(J_p - J_n)$ running through the cell and then the problem is to find the ideal values of $F_{n,p}$ for a given set of transport coefficients. This optimum is found by solving a constrained minimisation problem for $q_{s,\text{trans}}$. Using the method of Lagrange multipliers, the following quantity Λ is minimised with respect to F_n , F_p , and the multiplier λ :

$$\begin{aligned} \Lambda &= q_{s,\text{trans}} + \lambda \left[\frac{J_{\text{charge}}}{e} - (J_p - J_n) \right] = L_{nn}F_n^2 + L_{pp}F_p^2 + 2L_{np}F_nF_p \\ &\quad + \lambda \left[\frac{J_{\text{charge}}}{e} - (L_{pp}F_p + L_{np}F_n - L_{nn}F_n - L_{np}F_p) \right]. \end{aligned} \quad (2.59)$$

¹¹see also the remarks from the authors in preface to the Dover edition of their work [3].

By setting $\partial\Lambda/\partial F_n = \partial\Lambda/\partial F_p = \partial\Lambda/\partial\lambda = 0$ and solving, the following solution is obtained:

$$F_p = -F_n = \frac{J_{\text{charge}}/e}{L_{nn} + L_{pp} - 2L_{np}}, \quad (2.60)$$

$$q_{s,\text{trans}} = \frac{(J_{\text{charge}}/e)^2}{L_{nn} + L_{pp} - 2L_{np}}. \quad (2.61)$$

This shows that the ideal way to drive a current J_{charge} is by making sure that the electron and hole forces are equal and opposite. Recall that $L_{nn} \geq 0$ and $L_p \geq 0$, so due to the constraint $L_{nn}L_{pp} \geq L_{np}^2$ the entropy production is always positive. As can be seen, $q_{s,\text{trans}}$ is smaller for materials with greater L_{nn} and L_{pp} because those materials that are more conductive. Furthermore, the entropy production is lower for materials where $L_{np} < 0$ than for those with $L_{np} \geq 0$. The physical reason is that whenever $L_{np} < 0$, the electron force \mathbf{F}_n pushes the holes in the opposite direction and vice versa, thereby facilitating charge separation. To put it differently: the electron force is helping the holes and the other way around. Since the entropy generation terms are $\mathbf{J}_n \cdot \mathbf{F}_n$ and $\mathbf{J}_p \cdot \mathbf{F}_p$, it is beneficial to drive the electron current \mathbf{J}_n towards its collecting contact by forces other than \mathbf{F}_n (and vice versa). Conversely, when $L_{np} > 0$ the electrons and holes tend to be transported together in the same direction, making the photovoltaic conversion less efficient. Indeed, it is known that this co-directional transport of electrons and holes is one of the limiting factors in organic solar cells. For a material in which $L_{nn} = L_{pp}$ and that achieves the thermodynamically allowed limit $L_{np} = -\sqrt{L_{nn}L_{pp}}$, the reduction of charge transport losses can be up to 50% compared to $L_{np} = 0$. This demonstrates that both the diagonal and the off-diagonal transport coefficients can have a significant influence on the rate of entropy production.

Traditionally, the SCP description of solar cells does not consider the off-diagonal transport coefficients L_{np} and L_{pn} . Presumably, the error that occurs from this (implicit) assumption is not severe given the success with which SCP can model photovoltaic devices. However, this does not necessarily mean that $L_{np} = L_{pn} = 0$, since it is always possible to transform Eqs. (2.55) to a diagonal form (i.e., through diagonalisation of the matrix L_{ij}). By doing so, the effects of L_{np} and L_{pn} are incorporated into L_{nn} and L_{pp} (i.e., the electron and hole conductivities). However, the diagonalisation of L_{ij} also muddles the distinction between electrons and holes and this has repercussions for other empirical models such as the expression for the recombination rate and the mobility. It is therefore worth exploring whether greater accuracy and mathematical and/or physical simplicity can be achieved for these empirical models when the off-diagonal transport coefficients L_{np} and L_{pn} are explicitly incorporated in the transport equations.

Conversely, off-diagonal transport coefficients can also arise when a change in variables is made to a transport problem that was originally diagonal. Such

a transformation of fluxes and forces can sometimes provide advantages that weigh up against the added complexity of having a more complicated matrix of transport coefficients. As an example, it is possible to transform the electron and hole currents $\mathbf{J}_{n,p}$ to the current of electron-hole pair $\mathbf{J}_{\text{eh}} = 1/2(\mathbf{J}_n + \mathbf{J}_p)$ and the charge current $\mathbf{J}_c = \mathbf{J}_p - \mathbf{J}_n$. This is exactly analogous to the situation in Section 2.3 (Eqs. (2.5) and (2.7)) where one can choose between using the numbers of electrons and hole (N_n and N_p) as thermodynamic variables or, instead, the number of electron-hole pairs (N_{eh}) and the electron-hole charge (Q_{np}). The advantage of this transformation is, of course, that the charge current has no source term (i.e., $\nabla \cdot \mathbf{J}_c = 0$) since charge is always conserved. The accompanying forces for \mathbf{J}_{eh} and \mathbf{J}_c are:

$$\mathbf{J}_{\text{eh}} \quad \leftrightarrow \quad \mathbf{F}_{\text{eh}} = -\beta \nabla \Delta \eta_{\text{eh}} = -\beta \nabla (\eta_n + \eta_p), \quad (2.62a)$$

$$\mathbf{J}_c \quad \leftrightarrow \quad \mathbf{F}_c = -\beta \nabla \Phi_Q = \frac{\beta}{2} \nabla (\eta_n - \eta_p). \quad (2.62b)$$

It can be verified that these are indeed correct flux-force pairs since:

$$\mathbf{J}_{\text{eh}} \cdot \mathbf{F}_{\text{eh}} + \mathbf{J}_c \cdot \mathbf{F}_c = \mathbf{J}_n \cdot \mathbf{F}_n + \mathbf{J}_p \cdot \mathbf{F}_p, \quad (2.63)$$

which demonstrates that the change in variable from \mathbf{J}_n and \mathbf{J}_p to \mathbf{J}_{eh} and \mathbf{J}_c (along with the correct change in forces) leaves the irreversible entropy generation rate (Eq. (2.52)) invariant. Now suppose that the original transport matrix for \mathbf{J}_n and \mathbf{J}_p is diagonal so that $\mathbf{J}_n = L_{nn} \mathbf{F}_n$ and $\mathbf{J}_p = L_{pp} \mathbf{F}_p$. It can then be readily verified that the transport matrix for \mathbf{J}_{eh} and \mathbf{J}_c is not, since some algebra demonstrates that the Onsager relations for the new fluxes are:

$$\mathbf{J}_{\text{eh}} = \frac{L_{nn} + L_{pp}}{4} \mathbf{F}_{\text{eh}} + \frac{L_{pp} - L_{nn}}{2} \mathbf{F}_c, \quad (2.64a)$$

$$\mathbf{J}_c = \frac{L_{pp} - L_{nn}}{2} \mathbf{F}_{\text{eh}} + (L_{nn} + L_{pp}) \mathbf{F}_c. \quad (2.64b)$$

Thus, the convenience of having a flux with zero divergence should be contrasted against the inconvenience of have more complicated flux-force relations.

Finally, it should be noted that in various literature sources a theorem called the *principle of minimum entropy generation* is frequently mentioned. It will not be discussed in detail here, but it is worth mentioning that this theorem does not apply to solar cells. This is because a critical condition for the validity of the minimum entropy generation principle is that the transport coefficients L_{ij} are true constants (i.e., independent of thermodynamic variables like n and p as well as x) [3]. This condition is not fulfilled in the case of solar cells so the theorem can be disregarded.

2.7 The thermodynamic formulation of solar cell physics step 5: Symmetry, the Onsager reciprocal relations and the Curie principle

The transport coefficients L_{ij} are purely empirical from a macroscopic point of view, so to specify them further one either has to measure them; calculate them from microscopic theory; or make ad-hoc assumptions about them. Typical theoretical tools that can be used to link the L_{ij} to microscopic phenomena like collisions include the Boltzmann transport equation, Kubo formalism, and fluctuation theory. The books by De Groot [3] and Le Bellac [28] are among the sources that can be referred to for more details on the microscopic description of the transport coefficients.

However, even without microscopic models for the L_{ij} it is possible to obtain some important universal results for the transport coefficients. For example, in Section 2.6 it was already mentioned that the L_{ij} have to form a positive definite matrix because entropy cannot decrease. Recall from Section 2.3 that thermodynamics is a theory about symmetries of the underlying microscopic physics of a system and here in Section 2.7 two important symmetries along with their consequences for the transport coefficients will be discussed, namely time-reversal symmetry and material isotropy.

2.7.1 The Onsager reciprocal relations

Time reversal symmetry is a fundamental property of all microscopic laws that ultimately govern the behaviour of everyday macroscopic objects. Newton's laws of motion, Maxwell's laws of electromagnetic fields and quantum mechanics are all symmetric under the reversal of the flow of time. In some cases one has to transform certain quantities when time is reversed (e.g., magnetic fields get a minus sign when time is reversed) but ultimately it is not possible to distinguish the forward and reverse directions of time flow based on microscopic physics: only on the macroscopic scale is there a clear breaking of symmetry that manifests itself as the second law of thermodynamics¹². The microscopic time reversal symmetry has interesting macroscopic consequences that are known as the Onsager reciprocal relations. Applied to a solar cell, these relations state that the matrix of the transport coefficients is symmetric,

¹²Even though much work has been devoted to explaining how microscopically reversible laws can give rise to macroscopically irreversible phenomena, this question is still an active topic of debate and will not be discussed here since it is far beyond the scope of the current text.

i.e.:

$$L_{Qn} = L_{nQ}, \quad (2.65a)$$

$$L_{Qp} = L_{pQ}, \quad (2.65b)$$

$$L_{np} = L_{pn}, \quad (2.65c)$$

or simply $L_{ij} = L_{ji}$. This result may seem surprising at first. The sceptical reader is advised to verify that, e.g., the coefficients L_{Qn} and L_{nQ} have the same units before accepting that they are, in fact, equal. The Onsager relations present an important link between phenomena that superficially appear unrelated such as the Peltier effect and the Seebeck effect. The derivation of the reciprocal relations will not be given here; it can be found in various literature sources [20, 3]. However, it is possible to gain some heuristic understanding of the reciprocal relations by considering that time reversal also reverses cause and effect. If one would assert that a flow of electrons \mathbf{J}_n causes a flow of heat \mathbf{J}_Q along with the electrons (Peltier effect), then the principle of time-reversal symmetry states that it is equally valid to assert that a flow of heat \mathbf{J}_Q causes a flow of electrons \mathbf{J}_n (Seebeck effect). In other words, the reciprocal relations express the idea that, macroscopically, it is impossible to distinguish which of the two phenomena is the cause and which is the effect. It should be emphasised that the principle of time reversal symmetry does *not* imply that the flows \mathbf{J}_Q , \mathbf{J}_n , and \mathbf{J}_p are reversible: macroscopic flows generate entropy and are therefore always irreversible.

2.7.2 The Curie principle

The *Curie principle* (named after Pierre Curie) is the general idea that a symmetric system cannot give rise to phenomena that break the symmetries of the original system. For example, a gas in a container is completely isotropic in every direction and therefore it will conduct heat in the same way regardless of the direction of the heat current. In other words, if the heat conduction is described by Fourier's law $\mathbf{J}_Q = -(L_{QQ}/T^2)\nabla T$ from Eqs. (2.55), then L_{QQ} has to be a scalar constant for the gas rather than a tensor. The Curie principle is a useful concept, but it has limitations. A good example of a violation of the Curie principle is (ironically) the cooling of a ferromagnet from above its Curie temperature T_c to below. At $T > T_c$, the magnet is completely isotropic (or at least has a highly symmetrical crystal structure). Yet once it is cooled to $T < T_c$, the symmetry of the system spontaneously breaks and the material develops a net magnetisation \mathbf{M} . In more general it can be said that phase transitions can give rise to broken symmetries in otherwise completely symmetric systems, so care should be taken when applying the Curie principle when phase transitions can occur. Fortunately, this is not a very important limitation in solar cell physics.

Most parts of a solar cell (such as the bulk of a *c*-Si wafer) are to good approximation isotropic and it is therefore instructive to review what consequences this isotropy has on the macroscopic physics of the device. This way, one can also appreciate what kind of special phenomena can be expected in parts of the device where the isotropic symmetry is broken. While formulating the flux-force equations Eqs. (2.55) it was already assumed that the system is isotropic to simplify the discussion, because for a non-isotropic system significantly more terms are needed in Eqs. (2.55). To understand why, recall that the flux-force pair in Eqs. (2.54) were obtained from the terms that occur in the irreversible entropy generation rate $q_{s,\text{irr}}$. In Section 2.6, $q_{s,\text{irr}}$ was suggestively written out in vector form, but is of course also possible to expand the inner products in Cartesian components as follows:

$$\begin{aligned}
q_{s,\text{irr}} &= \beta \Delta \eta_{eh} R - \mathbf{J}_Q \cdot \nabla \beta - \mathbf{J}_n \cdot \beta \nabla \eta_n - \mathbf{J}_p \cdot \beta \nabla \eta_p \\
&= \beta \Delta \eta_{eh} R - J_{Q,x} \frac{\partial \beta}{\partial x} - J_{Q,y} \frac{\partial \beta}{\partial y} - J_{Q,z} \frac{\partial \beta}{\partial z} \\
&\quad - J_{n,x} \beta \frac{\partial \eta_n}{\partial x} - J_{n,y} \beta \frac{\partial \eta_n}{\partial y} - J_{n,z} \beta \frac{\partial \eta_n}{\partial z} \\
&\quad - J_{p,x} \beta \frac{\partial \eta_p}{\partial x} - J_{p,y} \beta \frac{\partial \eta_p}{\partial y} - J_{p,z} \beta \frac{\partial \eta_p}{\partial z}.
\end{aligned} \tag{2.66}$$

This form of $q_{s,\text{irr}}$ suggests that there actually are nine flux-force pairs related to heat and charge transport instead of three. Consequently, there are also nine flux-force relations with nine terms in each equation instead of the three terms shown in Eqs. (2.55). This means that strictly speaking, each of the coefficients L_{ij} in Eqs. (2.55) is actually a second order tensor rather than a scalar. For clarity, tensors will be denoted in roman symbols as:

$$\mathbf{L}_{ij} = \begin{pmatrix} L_{ij}^{xx} & L_{ij}^{xy} & L_{ij}^{xz} \\ L_{ij}^{yx} & L_{ij}^{yy} & L_{ij}^{yz} \\ L_{ij}^{zx} & L_{ij}^{zy} & L_{ij}^{zz} \end{pmatrix}. \tag{2.67}$$

The Onsager reciprocal relations for these tensors are:

$$\mathbf{L}_{ij} = \mathbf{L}_{ji}^T, \quad i, j \in \{Q, n, p\}, \tag{2.68}$$

where T denotes the transpose operation. Note that Eq. (2.68) also implies that \mathbf{L}_{QQ} , \mathbf{L}_{nn} and \mathbf{L}_{pp} are always symmetric.

The reason it is possible to reduce the tensors \mathbf{L}_{ij} to scalars L_{ij} is the Curie principle: if the system does not exhibit broken symmetry between the three Cartesian directions, then neither should the transport equations. In fact, it is not even completely necessary for the system to be completely isotropic: even in a system that only has cubic symmetry (which is a common symmetry type for many crystals, such a *c*-Si) the tensorial transport coefficients still

reduce to scalars and for the purposes of this text cubic symmetry can be considered equivalent to complete isotropy. In the works of GM and Kuiken, general mathematical methods are given that can reduce the complexity of the transport coefficients for any given type of symmetry of the system [3, 16].

However, the Curie principle does more than just assert that the flux-force relations are isotropic. At this point the reader is asked to recall from Section 2.6 that $q_{s,\text{irr}}$ in Eq. (2.52) also has the term $\beta\Delta\eta_{eh}R$ that yields the scalar flux-force pair $R \leftrightarrow \beta\Delta\eta_{eh}$. Since $\beta\Delta\eta_{eh}$ is a thermodynamic force, it should also be able to drive the currents $\mathbf{J}_{Q,n,p}$. In other words, in Eqs. (2.55) the following terms are also expected:

$$\mathbf{J}_Q = \dots + \mathbf{L}_{QR} \beta\Delta\eta_{eh}, \quad (2.69a)$$

$$\mathbf{J}_n = \dots + \mathbf{L}_{nR} \beta\Delta\eta_{eh}, \quad (2.69b)$$

$$\mathbf{J}_p = \dots + \mathbf{L}_{pR} \beta\Delta\eta_{eh}. \quad (2.69c)$$

As can be seen, the transport coefficients \mathbf{L}_{QR} , \mathbf{L}_{nR} , and \mathbf{L}_{pR} are necessarily vectors since they relate the scalar force $\beta\Delta\eta_{eh}$ to the vectorial fluxes $\mathbf{J}_{Q,n,p}$. Conversely, there should also be a flux-force relation for the scalar flux R :

$$R = \mathbf{L}_{RQ} \cdot \mathbf{F}_Q + \mathbf{L}_{Rn} \cdot \mathbf{F}_n + \mathbf{L}_{Rp} \cdot \mathbf{F}_p + L_{RR}\beta\Delta\eta_{eh}. \quad (2.70)$$

Again, the coefficients \mathbf{L}_{RQ} , \mathbf{L}_{Rn} , and \mathbf{L}_{Rp} have to be vectors, since the inner product is the only way to make a general linear transformation from the vectorial forces $\mathbf{F}_{Q,n,p}$ to the scalar flux R . Of course, there are also Onsager relations for the vectorial transport coefficients \mathbf{L}_{iR} and \mathbf{L}_{Ri} that have been introduced just now:

$$\mathbf{L}_{QR} = \mathbf{L}_{RQ}, \quad (2.71a)$$

$$\mathbf{L}_{nR} = \mathbf{L}_{Rn}, \quad (2.71b)$$

$$\mathbf{L}_{pR} = \mathbf{L}_{Rp}. \quad (2.71c)$$

The cross effects between recombination and transport in Eqs. (2.69) and (2.70) were not mentioned in Section 2.6 because these cross phenomena do not occur in an isotropic medium. This is again a consequence of the Curie principle: in an isotropic system no preferential direction can exist, so the vectors like \mathbf{L}_{QR} and \mathbf{L}_{RQ} have to be zero since they are material properties. In other words it can also be said that in an isotropic medium the scalar force $\beta\Delta\eta_{eh}$ cannot drive a vectorial current since there is no direction the current can be forced towards.

In conclusion, for a fully non-isotropic system, it is necessary to generalise

Eqs. (2.55) to:

$$\mathbf{J}_Q = L_{QQ}\mathbf{F}_Q + L_{Qn}\mathbf{F}_n + L_{Qp}\mathbf{F}_p + \mathbf{L}_{QR}\beta\Delta\eta_{eh}, \quad (2.72a)$$

$$\mathbf{J}_n = L_{nQ}\mathbf{F}_Q + L_{nn}\mathbf{F}_n + L_{np}\mathbf{F}_p + \mathbf{L}_{nR}\beta\Delta\eta_{eh}, \quad (2.72b)$$

$$\mathbf{J}_p = L_{pQ}\mathbf{F}_Q + L_{pn}\mathbf{F}_n + L_{pp}\mathbf{F}_p + \mathbf{L}_{pR}\beta\Delta\eta_{eh}, \quad (2.72c)$$

$$\mathbf{R} = \mathbf{L}_{RQ} \cdot \mathbf{F}_Q + \mathbf{L}_{Rn} \cdot \mathbf{F}_n + \mathbf{L}_{Rp} \cdot \mathbf{F}_p + L_{RR}\beta\Delta\eta_{eh}. \quad (2.72d)$$

Here it is understood that terms such as $L_{QQ}\mathbf{F}_Q$ represent matrix multiplications. Having compared the isotropic flux-force relations with the fully non-isotropic form in Eqs. (2.72), it is time to consider what kind of departures from isotropy can be expected in a solar cell. The first possibility is that the bulk material of the cell itself has an inherent anisotropy due to its crystal structure. However, many materials relevant for solar cells are either amorphous (*a*-Si:H, polymers) or have cubic crystal symmetry and both of these types of materials can be considered isotropic to a good degree of accuracy. Material isotropy will therefore not be considered in detail. Secondly, and more importantly, the isotropy of the system can be reduced by the presence of electric or magnetic fields (see also Kuiken [16] for a detailed mathematical treatment of symmetry breaking due to electromagnetic fields). It is well known that electric fields are commonly encountered in solar cells and in, e.g., *c*-Si-based cells they can be as strong as 10^6 V m^{-1} , which is enough to alter the electrical properties of the crystal. To discuss the macroscopic consequences of such an electric field, it is assumed that the field \mathbf{E} is in the z direction (so that $\mathbf{E} = E\hat{z}$) in an otherwise isotropic medium. The tensorial transport coefficients in the general flux-force relations in Eqs. (2.72) can then be shown to reduce to the form:

$$L_{ij} = \begin{pmatrix} L_{ij}^\perp & 0 & 0 \\ 0 & L_{ij}^\perp & 0 \\ 0 & 0 & L_{ij}^\parallel \end{pmatrix}, \quad i, j \in \{Q, n, p\}. \quad (2.73)$$

As can be seen, transport in the plane perpendicular to \mathbf{E} is still isotropic, but a different coefficient appears for the transport along \mathbf{E} . Furthermore, the vectorial transport coefficients all become scalar multiples of \mathbf{E} :

$$\mathbf{L}_{QR} = \mathbf{L}_{RQ} = L_{QR}\mathbf{E}, \quad (2.74a)$$

$$\mathbf{L}_{nR} = \mathbf{L}_{Rn} = L_{nR}\mathbf{E}, \quad (2.74b)$$

$$\mathbf{L}_{pR} = \mathbf{L}_{Rp} = L_{pR}\mathbf{E}. \quad (2.74c)$$

Heuristically, it can be said that broken symmetry makes it possible to couple scalar and vector processes. Because the electric field is the source of this broken symmetry, the transport coefficients \mathbf{L}_{QR} , \mathbf{L}_{nR} , and \mathbf{L}_{pR} that couple the scalar and vectorial processes have to be co-directional to \mathbf{E} . In the first order, the prefactors L_{QR} , L_{nR} , and L_{pR} are independent of the strength of \mathbf{E} ,

but for strong electric fields they can become dependent on the field strength $\mathbf{E} \cdot \mathbf{E}$.

Thus, the broken symmetry due to the electric field can influence the currents of the electrons and holes by the terms:

$$\mathbf{J}_n = \dots + L_{nR} \mathbf{E} \beta \Delta \eta_{eh}, \quad (2.75a)$$

$$\mathbf{J}_p = \dots + L_{pR} \mathbf{E} \beta \Delta \eta_{eh}. \quad (2.75b)$$

It should be emphasised again that this contribution is not directly due to the electrostatic force of the electric field on the charge carriers, since this force is already accounted for in \mathbf{F}_n and \mathbf{F}_p in the flux-force equations. In fact, the electric field does not act like a thermodynamic force at all in this case but only a factor in a transport coefficient: $\beta \Delta \eta_{eh}$ is the thermodynamic force that drives the current. From a microscopic point of view, the electric field creates an asymmetry between recombination of electrons/holes with different momenta by distorting the recombination potential, as shown in Figure 2.2. Thus, if electrons (holes) with a certain momentum are more likely to recombine than those with the opposite momentum, this will macroscopically influence the current \mathbf{J}_n (\mathbf{J}_p). The idea that electric fields affect carrier recombination is not new and has been described by microscopic physics (see, e.g., Schenk [31]). On the other hand, the conclusion that the recombination therefore also acts as a force for $\mathbf{J}_{n,p}$ is not immediately obvious from microscopic physics (and has not been pointed out before, to the author's knowledge) while this consequence follows in a straightforward manner from a thermodynamic description. This illustrates how macroscopic physics can help to illuminate the consequences of microscopic processes and correlate seemingly unrelated phenomena. From a pure modeling point of view it is probably more natural to write the recombination contributions to $\mathbf{J}_{n,p}$ as:

$$L_{nR} \beta \Delta \eta_{eh} \mathbf{E} = C_{nR} R \mathbf{E}, \quad (2.76)$$

$$L_{pR} \beta \Delta \eta_{eh} \mathbf{E} = C_{pR} R \mathbf{E}, \quad (2.77)$$

where the coefficients C_{nR} and C_{pR} measure the asymmetry in recombination probability between electrons/holes moving with or against the electric field \mathbf{E} .

Having demonstrated that in the space-charge region of a solar cell it is possible for cross-phenomena to occur between recombination and charge transport, the question arises how this affects the generation of entropy. As was illustrated in Section 2.6 for the off-diagonal transport coefficients $L_{np} = L_{pn}$, the understanding of cross-phenomena can help to make the device more efficient. When considering the coupling coefficients L_{nR} and L_{pR} , it can be seen that it would be beneficial to look for materials where $L_{nR} \leq 0$ and $L_{pR} \geq 0$ since in that case the scalar force $\beta \Delta \eta_{eh}$ will provide extra driving terms for $\mathbf{J}_{n,p}$ that guide the electrons and holes to their respective contacts.

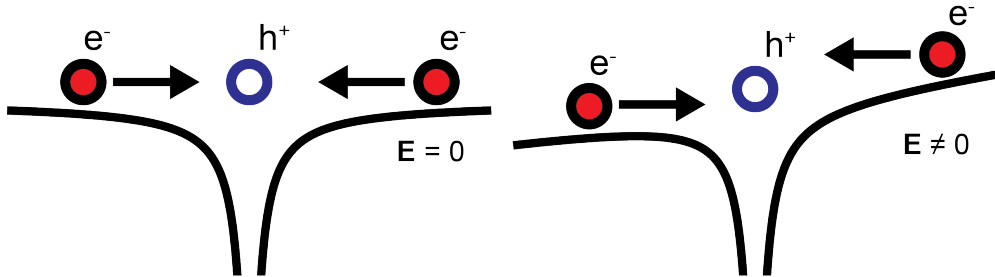


Figure 2.2: Comparison of the Coulomb potentials of electrons near a hole. In the presence of an external electric field \mathbf{E} the symmetry between electrons moving left and those moving right is broken and one of the two can become more likely to recombine. In the recombination mechanism described by Schenk [31] (Figure 1 in his article), the electron moving from left to right in the second image would be the one more likely to recombine.

It is worth mentioning that there is a significant difference between the description of (non-radiative) recombination between SCP and NET. In SCP, bulk recombination rates are usually parameterised as:

$$R = \alpha(np - n_i^2), \quad (2.78)$$

with n_i the intrinsic carrier density and α a parameter (not necessarily a constant) that is commonly a polynomial in powers of n and p . The factor $(np - n_i^2)$ (which is essentially the exponent of $\Delta\eta_{eh}$) occurs naturally to make sure that $R = 0$ under equilibrium conditions. With Eq. (2.78) it is possible to conveniently model collisions between two, three or more carriers as terms with varying powers of n and p . On the other hand, NET prescribes the bulk recombination as a scalar flux-force relation:

$$R = L_{RR}\beta\Delta\eta_{eh} = L_{RR}k_B \ln\left(\frac{np}{n_i^2}\right), \quad (2.79)$$

since the remaining terms from Eq. (2.72d) do not occur in the bulk. There is no true disagreement between Eqs. (2.78) and (2.79), since the parameter L_{RR} is not necessarily constant and because Eq. (2.79) also yields $R = 0$ under equilibrium. However, the strong non-linearity of L_{RR} makes Eq. (2.79) less practical for the modeling of recombination than Eq. (2.78). In their own work, GM already remark that the scalar flux-force relations for chemical reactions are often not very useful because of the strong non-linearity of the coefficients (L_{RR} in this case) that occur in real systems. This suggests that strong deviations from chemical equilibrium often happen in practice. Indeed, the strong non-equilibrium of electrons and holes is precisely what characterises a solar cell and what allows it to generate power.

2.8 Wrap up: what does thermodynamics tell us?

The mathematical formalism of NET_{GM} has strong parallels with that of SCP. Both rely on the continuity equations and use empirical laws for the fluxes. The main difference is that SCP also introduces the Poisson equation for the electrostatic potential ϕ and the Boltzmann (or Fermi-Dirac) expressions for the carrier densities:

$$n = N_{\text{C}} \exp\left(\frac{E_{Fn} + \phi + \chi}{k_{\text{B}}T}\right), \quad (2.80a)$$

$$p = N_{\text{V}} \exp\left(-\frac{E_{Fp} + \phi + \chi + E_{\text{G}}}{k_{\text{B}}T}\right). \quad (2.80b)$$

Furthermore, SCP has many detailed models for various transport and recombination phenomena.

There are, however, no limitations to incorporate these elements into NET. The Poisson equation and carrier statistics can be included as auxiliary equations while the transport and recombination models simply prescribe the values of the empirical transport coefficients L_{ij} . This means that any discussion about solar cells framed in terms of SCP can be translated to NET and vice versa, so that both disciplines can benefit from the strengths of the other. For SCP, the main enrichments coming from NET are the addition of entropy as a tool for discussing losses of useful output power and the systematic analysis of symmetries (and their consequences) in the system. It would not be an exaggeration to say that SCP actually is a form of non-equilibrium thermodynamics that is often just not recognised as such. By making this link explicit, an invaluable extra level of depth becomes available for the interpretation of results and calculations from SCP. Due to its generality, NET can also help to ground the mathematical formalism of SCP on fundamental principles and to clarify assumptions that are used in SCP.

Moreover, NET can help to point out phenomena predicted by the empirical transport relations (Eq. (2.55) and (2.72)) that are possibly relevant to investigate. Examples are the cross-effects between electron and hole transport (characterised by the coefficient L_{np} , see Section 2.6) or those between charge transport and recombination (characterised by the coefficients \mathbf{L}_{nR} and \mathbf{L}_{pR} , see Section 2.7). An important lesson that is learnt from NET, is that the discovery of new phenomena (e.g., a new recombination process) also has consequences for other physical processes in the cell as predicted by the Onsager reciprocal relations. Finally, NET also demonstrates that the ultimate solar cell design with the highest possible efficiency can only be achieved if the entropy generation rates in both the cell and the photon field are considered simultaneously since both are an essential in the photovoltaic conversion process (see Section 2.4 and the discussion following Eq. (2.45)). This means that, at some point, the unilateral treatment of the light (which is still prevalent

for most solar cell modeling tools) will reach its limits and no further design improvement can be achieved without switching to a bilateral description of the coupling between cell and light.

It can be concluded that there are no arguments left against the use of thermodynamics in solar cell research: The apparent objections raised at the start of Section 2.1 can be countered by modern NET and in a sense SCP can be viewed as a particular application of the work by GM. Non-equilibrium thermodynamics offer tools that are not only useful for academic research about the ultimate efficiency potential, but they can also be combined with SCP to assist the down-to-earth engineering of current cell designs.

References Chapter 2

- [1] Peter Würfel. *Physics of Solar Cells*. Wiley-VCH Verlag GmbH & Co, Weinheim, 2005. ISBN 3-527-40428-7.
- [2] Uli Würfel, Andres Cuevas, and Peter Würfel. Charge Carrier Separation in Solar Cells. *IEEE Journal of Photovoltaics*, 5(1):461–469, 2015. ISSN 2156-3381. doi: 10.1109/JPHOTOV.2014.2363550.
- [3] S.R. de Groot and Peter Mazur. *Non-equilibrium Thermodynamics*. Dover Publications, Mineola, New York, 1984. ISBN 0-486-64741-2.
- [4] J.E. Parrott. Theoretical upper limit to the conversion efficiency of solar energy. *Solar Energy*, 21(3):227–229, 1978. ISSN 0038092X. doi: 10.1016/0038-092X(78)90025-7.
- [5] A. De Vos and H. Pauwels. On the thermodynamic limit of photovoltaic energy conversion. *Appl. Phys.*, 25(2):119–125, June 1981. ISSN 0340-3793. doi: 10.1007/BF00901283.
- [6] J.E. Parrott. Thermodynamics of solar cell efficiency. *Solar Energy Materials and Solar Cells*, 25(1-2):73–85, 1992. ISSN 09270248. doi: 10.1016/0927-0248(92)90017-J.
- [7] C. H. Henry. Limiting efficiencies of ideal single and multiple energy gap terrestrial solar cells. *J. Appl. Phys.*, 51(8):4494, 1980. ISSN 00218979. doi: 10.1063/1.328272.
- [8] Peter Würfel. Thermodynamic limitations to solar energy conversion. In *Physica E: Low-Dimensional Systems and Nanostructures*, volume 14, pages 18–26, 2002. ISBN 1386-9477.
- [9] Nils-Peter Harder and Peter Würfel. Theoretical limits of thermophotovoltaic solar energy conversion. *Semiconductor Science and Technology*, 18(5):S151–S157, 2003. ISSN 0268-1242. doi: 10.1088/0268-1242/18/5/303.
- [10] R. Brendel, S. Dreissigacker, N.-P. Harder, and P. P. Altermatt. Theory of analyzing free energy losses in solar cells. *Applied Physics Letters*, 93(17):173503, 2008. ISSN 00036951. doi: 10.1063/1.3006053.

- [11] Johannes Greulich, Hannes Höffler, Uli Würfel, and Stefan Rein. Numerical power balance and free energy loss analysis for solar cells including optical, thermodynamic, and electrical aspects. *Journal of Applied Physics*, 114(20):204504, 2013. ISSN 00218979. doi: 10.1063/1.4832777.
- [12] K.M. Van Vliet. Entropy production and choice of heat flux for degenerate semiconductors and metals. *physica status solidi (b)*, 78:667, 1976. ISSN 03701972. doi: 10.1002/pssb.2220780227.
- [13] J.E. Parrott. Transport theory of semiconductor energy conversion. *Journal of Applied Physics*, 53(12):9105–9111, 1982. ISSN 00218979. doi: 10.1063/1.330422.
- [14] U. Lindefelt. Heat generation in semiconductor devices. *Journal of Applied Physics*, 75:942–957, 1994. ISSN 00218979. doi: 10.1063/1.356450.
- [15] J.E. Parrott. Thermodynamic theory of transport processes in semiconductors. *IEEE Transactions on Electron Devices*, 43, 1996. ISSN 0018-9383. doi: 10.1109/16.491259.
- [16] Gerard D.C. Kuiken. *Thermodynamics of irreversible processes*. John Wiley & Sons, Ltd., Chichester, England, 1994. ISBN 0471948446.
- [17] Hans Christian Öttinger. *Beyond Equilibrium Thermodynamics*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.
- [18] Signe Kjelstrup and Dick Bedeaux. *Non-Equilibrium Thermodynamics of Heterogeneous Systems*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2008.
- [19] P. T. Landsberg and G. Tonge. Thermodynamic energy conversion efficiencies. *Journal of Applied Physics*, 51(7):R1, July 1980. ISSN 00218979. doi: 10.1063/1.328187.
- [20] Lars Onsager. Reciprocal Relations in Irreversible Processes. I. *Physical Review*, 37(4):405–426, February 1931. ISSN 0031-899X. doi: 10.1103/PhysRev.37.405.
- [21] Herbert B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. John Wiley & Sons, Inc., New York, 2 edition, 1985.
- [22] David Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, New York, 1987. ISBN 0-19-504276-X.
- [23] Mark Waldo Zemansky and Richard Dittman. *Heat and Thermodynamics : An Intermediate Textbook*. McGraw-Hill International Book Company, London, 7 edition, 1997.

- [24] Eli Yablonovitch and Owen D. Miller. The Opto-Electronics of Solar Cells. *IEEE Photonics Society News*, 27(1), 2013.
- [25] E. Velme. Numerical modelling of semiconductor structures including electron-hole scattering and recombination radiation recycling effect. *Period. Polytech. Elec. Eng.*, 33(3):141–150, 1989.
- [26] Frank E Osterloh. Maximum Theoretical Efficiency Limit of Photovoltaic Devices: Effect of Band Structure on Excited State Entropy. *The Journal of Physical Chemistry Letters*, 5(19):3354–3359, October 2014. ISSN 1948-7185. doi: 10.1021/jz501740n.
- [27] Albert Polman and Harry A Atwater. Photonic design principles for ultrahigh-efficiency photovoltaics. *Nature Materials*, 11(3):174–7, January 2012. ISSN 1476-1122. doi: 10.1038/nmat3263.
- [28] Michel Le Bellac, Fabrice Mortessagne, and G. George Batrouni. *Equilibrium and Non-Equilibrium Statistical Thermodynamics*. Cambridge University Press, Cambridge, 2004. ISBN 0 521 82143 6.
- [29] A. Einstein. Zur allgemeinen molekularen Theorie der Wärme [AdP 14, 354 (1904)]. *Annalen der Physik*, 14(S1):154–163, February 2005. ISSN 00033804. doi: 10.1002/andp.200590003.
- [30] Harald Ibach and Hans Lüth. *Solid-State Physics*. Springer-Verlag Berlin Heidelberg New York, Berlin, Heidelberg, 3 edition, 2003. doi: Phys-RevB.83.165113.
- [31] A. Schenk. An improved approach to the ShockleyReadHall recombination in inhomogeneous fields of space-charge regions. *Journal of Applied Physics*, 71(7):3339, 1992. ISSN 00218979. doi: 10.1063/1.350929.

Chapter 3

The physics of solar cells: semiconductor physics enhanced by thermodynamics

3.1 The thermodynamics of the selective membrane model

Chapter 2 and Appendix A.1 review the most important physical equations that are needed to describe a solar cell and to predict its efficiency. The equations, however, are only part of the complete understanding of how a solar cell works and, more importantly, how it can be improved. The set of equations that describes the charge transport in a cell is highly nonlinear and even solving numerically on a computer is not entirely straightforward. In a sense, the transport equations only replace the problem of understanding the solar cell with another (not necessarily easier) problem: that of understanding the mathematics underlying the equations. It is no surprise then, that different concepts and simplifications have been developed to understand solar cells on a more heuristic level, such as the selective-membrane concept by P. Würfel [1] and U. Würfel [2]. Indeed, such simplifications are of crucial importance to the field of PV (and science in general) because they make the exchange of ideas between people more efficient.

The goal of this section is therefore to present and justify a heuristic description of solar cells that uses NET_{PV} from Chapter 2 as the main physical language. First, the selective-membrane concept as developed by P. Würfel will be presented briefly and will then be used as the main motivation for the thermodynamic discussion that follows.

3.2 The selective membrane model

A typical single-junction solar cell consists of three parts that are essential for the operation of the device: a *photon absorber* (or just absorber for short) which forms the centre of the cell and two *selective membranes* (one for electrons and one for holes) that are located between the absorber and the metal contacts. The function of the absorber is to convert the incoming light into useful work in the form electrochemical free energy, meaning that the photon absorber has to be good at converting light into free charge carriers and also at sustaining the resulting electron-hole plasma. The function of the selective membranes is to ensure that the electrochemical energy generated in the absorber is transported out of the cell as efficiently as possible in the form of electrical current. The selective membranes are highly conductive to one type of carrier and highly resistive to the other type. A membrane that allows electrons (holes) to pass while blocking holes (electrons) is called an electron (hole) membrane or an n -membrane (p -membrane). The combination of a selective membrane plus metal contact is referred to as a *selective contact*. Due to the large asymmetry in the electron and hole conductivities between the two different selective membranes, electrons and holes are much more likely to exit the cell through their respective contacts (see Figure 3.1). It is important to note that the selective contacts are ideally completely transparent to light so that there is no charge carrier generation inside of them. Heuristically, the reason for this is that holes generated in the middle of the n -membrane in Figure 3.1 encounter approximately the same resistance in the path towards the electron collecting contact as in the path towards the bulk of the cell (and ultimately the p -contact). Thus, holes generated in the selective contact are likely to be collected by the wrong electrode, leading to current loss. This phenomenon is known as parasitic absorption and it will be discussed in more detail in Section 3.6.

As can be seen, the concepts of conductivity and resistivity are key to the understanding of selective membranes and one of the goals of this chapter is to give these heuristic concepts a quantitative basis that is straightforward to interpret. To do so, it is useful to first consider the most common type of solar cell, namely the doped homojunction where the absorber and the selective membranes consist of the same material. To create the selective membranes, a high doping concentration is used to drastically increase the conductivity of one type of charge carrier. In Figure 3.2 the band diagram of a p -contact and part of the photon absorber of the cell is shown. For simplicity, the cell under discussion is assumed to be 1D. The band diagram (as well as all the other diagrams in this chapter) is shown for operating conditions, since under these conditions it is the most sensible to analyse the influence of the cell design on the entropy generation in the cell.

In the bulk of the cell the quasi Fermi levels $E_{F_{n,p}}$ are almost completely

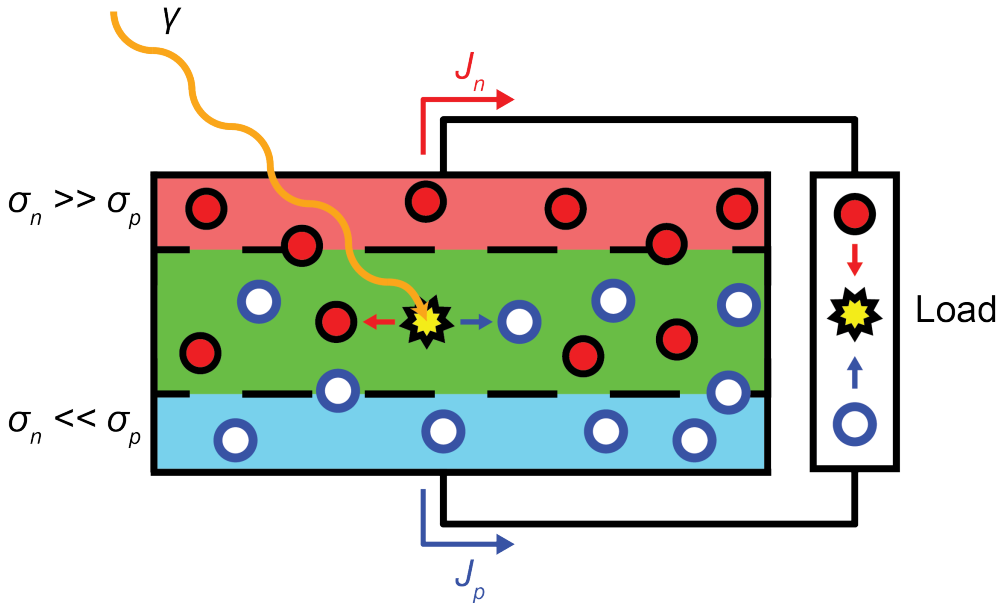


Figure 3.1: Schematic overview of a working solar cell. The green part of the cell is the photon (γ) absorber and the red and blue parts are the electron and hole selective membranes. The asymmetry between electron conductivity (σ_n) and hole conductivity (σ_p) drives a current through the load. In practice it is not possible to differentiate between electron and hole current inside of a metal, but the underlying idea is that the selective contacts ensure that the easiest (i.e., least resistive) recombination pathway for the carriers is through the metal contacts and the external load.

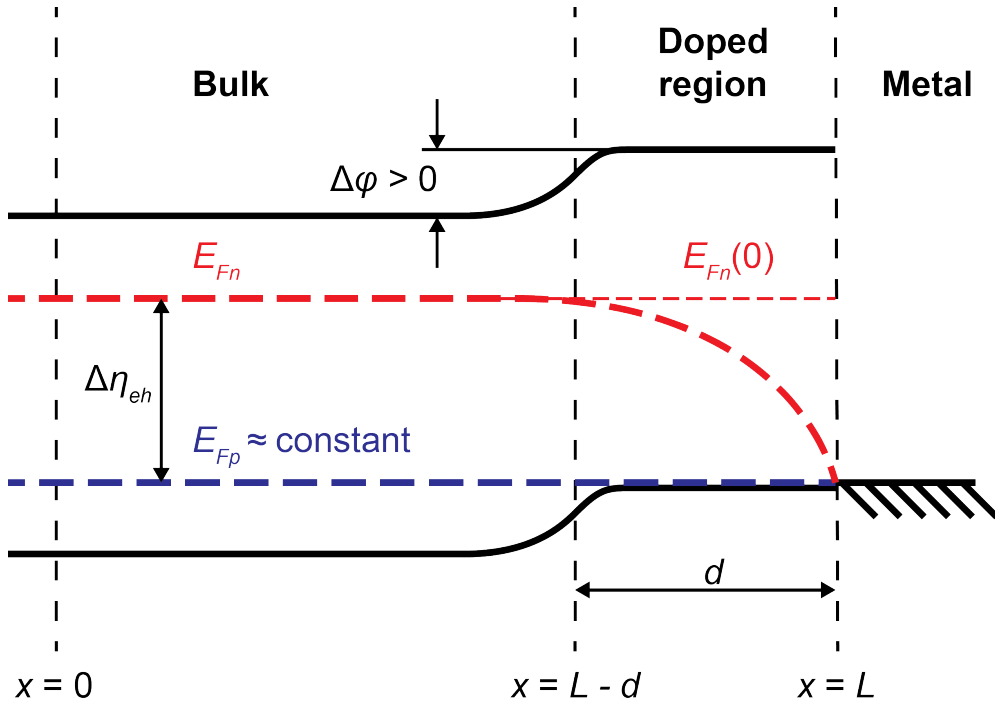


Figure 3.2: Sketch of the band diagram of a p -contact under operating conditions. The contact consists of a highly p -doped region between the absorber and the metal. The quasi Fermi levels are assumed to be continuous, but in practice it is possible to have a discontinuity in E_{Fn} across semiconductor-metal interface. As will be demonstrated in the discussion of the Schottky contact in Section 3.3, this discontinuity does not complicate the discussion of the selective contacts significantly.

flat¹. Of course, there have to small gradients in E_{Fn} and E_{Fp} to drive the carriers to their collecting contacts, but for a material with a high enough mobility (such as *c*-Si) these gradients are typically so small that E_{Fn} and E_{Fp} can be considered constant throughout the bulk of the cell. As a rule of thumb it should be remembered that the quasi Fermi levels E_{Fn} and E_{Fp} are almost completely straight in those parts of the cell where there are many electrons or holes respectively². If this were not the case, the gradient in E_{Fn} (E_{Fp}) would drive a large amount of electrons (holes) in such a way that the gradient is nullified.

Bulk transport losses are generally small in the 1D cell for materials with a high mobility like *c*-Si and they will therefore not be discussed in detail. However, it is interesting quickly compare the charge transport in the bulk with the ideal transport requirement from Section 2.6. This requirement stated that, for given transport coefficients L_{nn} , L_{pp} , the optimal way to drive a given total charge current J_{charge} is by making the electron and hole thermodynamic forces equal and opposite (see Eq. (2.60)). This means that it is desirable to have $E'_{Fn} = E'_{Fp}$ in the bulk of the cell. Fortunately, it is not necessary to worry about this requirement, since the bulk of the cell is charge-neutral as long as the background doping is uniform and from charge-neutrality it can be derived that:

$$\begin{aligned}
0 &= \frac{d}{dx} [p - n + N_{\text{dop}}] \\
&= \frac{d}{dx} \left[N_V \exp\left(-\frac{E_{Fp} - E_V}{k_B T}\right) - N_C \exp\left(-\frac{E_C - E_{Fn}}{k_B T}\right) + N_{\text{dop}} \right] \\
&= \frac{1}{k_B T} \left[N_V (E'_V - E'_{Fp}) \exp\left(-\frac{E_{Fp} - E_V}{k_B T}\right) - \right. \\
&\quad \left. N_C (E'_{Fn} - E'_C) \exp\left(-\frac{E_C - E_{Fn}}{k_B T}\right) \right] = 0.
\end{aligned} \tag{3.1}$$

Since $E'_C = E'_V$, it follows that $E'_{Fn} = E'_C = E'_{Fp}$ and that charge neutrality ensures that the charge transport is optimal³.

Having determined that the transport losses in the bulk are not of great concern, it remains to analyse the scalar entropy generation processes in the photon absorber: the absorption of photons and generation/recombination of electron-hole pairs. The entropy source term for the scalar processes in the bulk is denoted by $q_{s,\text{bulk}}$. Assuming a constant Fermi level splitting $E_{Fn} - E_{Fp} =$

¹From this point onward it is more convenient to work with the quasi Fermi levels $E_{Fn,p}$ rather than of $\eta_{n,p}$, since $E_{Fn,p}$ are easier to draw in a band diagram. Recall that $\eta_n = E_{Fn}$ and $\eta_p = -E_{Fp}$.

²Or, to be more precise: in those parts of the cell where the electron or hole conductivity is large

³Actually, Eq. (3.1) can be zero while $E'_{Fn} \neq E'_{Fp}$, but this typically only happens around very strong local sinks or sources of electron-hole pairs.

$\Delta\eta_{eh}$, the total entropy generation (per unit area of the cell) in the bulk is approximately (see also Sections 2.5 and 2.6):

$$\begin{aligned} \iiint_{\text{bulk}} Tq_{s,\text{bulk}} dx &= \iiint_{\text{bulk}} (q_u + \Delta\eta_{eh}[R(\Delta\eta_{eh}) - G]) dx \\ &= I_u - \Delta\eta_{eh}I_{\text{gen}} + \Delta\eta_{eh}d_{\text{bulk}}R(\Delta\eta_{eh}). \end{aligned} \quad (3.2)$$

Here, I_u is the total absorbed energy flux (per unit area); I_{gen} is the total generation rate of electron hole pairs in the bulk; and d_{bulk} is the thickness of the bulk. Recall that R is generally an increasing function of $\Delta\eta_{eh}$. It is important to direct attention to the fact that the thermodynamic thermalisation losses associated with photon absorption are given by $I_u - \Delta\eta_{eh}I_{\text{gen}}$ and not, as commonly pictured, by $I_u - E_G I_{\text{gen}}$ (see Figure 3.3). In other words, from a thermodynamic perspective it is important to consider the free energy $\Delta\eta_{eh}$ of the generated pairs and not the total energy. This is because part of the total energy of the electron-hole plasma simply cannot be extracted as useful work in the same way that it is not possible to cool the cell down and extract the thermal energy as work.

In Eq. (3.2, the fluxes I_u and I_{gen} are more or less fixed by the properties of the incident light and the bandgap of the semiconductor, leaving $\Delta\eta_{eh}$ and d_{bulk} as the main optimisation parameters to reduce the entropy generation in the bulk. In practice, this means that d_{bulk} should be as low as possible without impairing the ability of the cell to absorb all of the incident photons. Indeed, it is well-known that thin cells tend to have a higher performance (mainly due to a higher V_{OC}) than thick cells (see, e.g., De Wolf [3]). The Fermi level splitting $\Delta\eta_{eh}$ is controlled by maintaining a potential across the cell and the optimal potential is commonly known as the max-power voltage V_{mpp} . Thus, if the cell is completely limited by its bulk, the optimal potential is found making the best trade-off between the terms $-\Delta\eta_{eh}I_{\text{gen}}$ (i.e., the total production of free energy) and $\Delta\eta_{eh}dR(\Delta\eta_{eh})$ (i.e., the loss of free energy).

It can be seen that the Fermi level splitting in the bulk of the cell is crucial to the operation of the device, so it is useful to consider for a moment what makes Fermi level splitting possible in the first place. As was demonstrated in Section 2.3.1, Fermi level splitting is only possible in a non-equilibrium state of the cell since $\Delta\eta_{eh} = 0$ in equilibrium. Put differently, the Fermi level splitting represents the failing of electrons and holes to reach electrochemical equilibrium fast enough to counteract the external influences (the incoming light and the voltage that is maintained over the cell) that are continuously disturbing the equilibrium. In the semiconductor, an electron is forced to make one big change in energy to recombine with a hole due to the bandgap of the material which does not allow the electron to occupy a state with an intermediate energy. Since such a big change in energy is relatively improbable, the recombination rate is much lower than if the electron has states available with intermediate

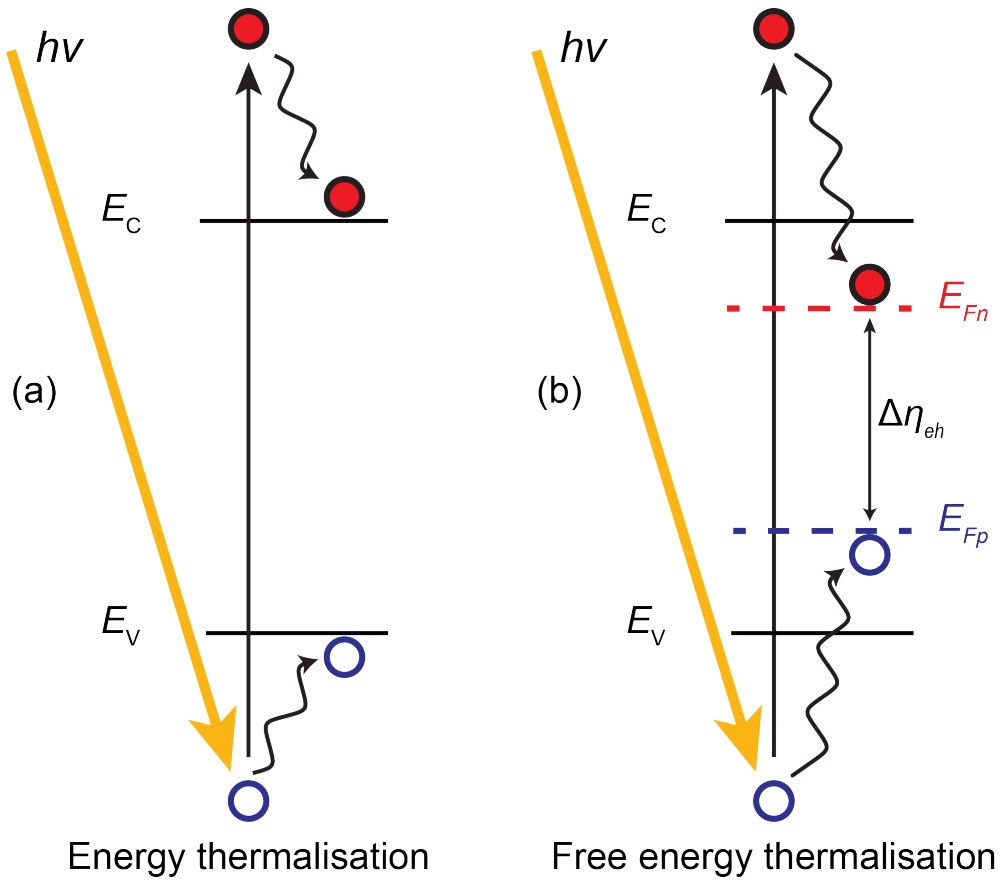


Figure 3.3: Two different visualisation of thermalisation losses after photon absorption. Figure (a) show the thermalisation losses of the *energy* of an e^-h^+ pair, while (b) shows the thermalisation losses of the *free energy* of the same pair. Of course there are no energy states in the band gap, so (b) should not be interpreted as if the electron and hole are actually occupying forbidden states.

energies that would allow the electron to lose its energy gradually step-by-step (such as when the semiconductor material is impure and has defect states in the forbidden gap). This is why Fermi level splitting can be achieved in semiconductors but not in metals, since in a metal there are (by the defining property of a metal) many intermediate states that make it very easy for the electrons to reach chemical equilibrium.

Next the selective contacts will be considered. To simplify the discussion, a p -contact will be taken as a specific example, but what follows applies to electron contacts with equal validity. Thus, holes are the majority carriers and the electrons are the minority carriers for now⁴. Consider a short interval $x \in [0, L]$ close to the a hole-collecting contact, with $x = 0$ a point in the bulk and $x = L$ the position of the metal. The first thing to note about the band diagram of the p -contact in Figure 3.2 is that E_{Fp} is almost completely straight because of the very high hole density. Due to the small gradient, the transport of holes across the p -membrane is almost lossless, as can be seen when writing out the transport losses $q_{s,\text{trans}}$ (see Eq. (2.57)) for the problem at hand. There are three equivalent forms for $q_{s,\text{trans}}$ that can be obtained by substitution of the flux-force relations (specifically: Eqs. (A.4) in Appendix A.1):

$$Tq_{s,\text{trans}} = -E'_{Fn}J_n + E'_{Fp}J_p \quad (3.3a)$$

$$= \frac{\sigma_n}{e^2}(E'_{Fn})^2 + \frac{\sigma_p}{e^2}(E'_{Fp})^2 \quad (3.3b)$$

$$= \frac{e^2 J_n^2}{\sigma_n} + \frac{e^2 J_p^2}{\sigma_p}. \quad (3.3c)$$

Recall that here the chosen unit of $E_{Fn,p}$ is eV and that of $J_{n,p}$ is $\text{cm}^{-2} \text{s}^{-1}$. Eq. (3.3c) shows that for a given current J_p (i.e., the photocurrent), $q_{s,\text{trans}}$ is small since σ_p is very large. Furthermore, Eq. (3.3a) demonstrates that if for some reason (e.g., Fermi level pinning at the interface) the hole quasi Fermi level bends upward, a lot of entropy will be generated in the p -contact. A straight E_{Fp} is therefore an absolutely necessary condition for a good p -contact.

It can also be seen in Figure 3.2 that E_{Fn} and E_{Fp} meet at the metal contact. As mentioned, the reason for this is that inside of the metal there can be no Fermi level splitting since metals do not have a bandgap. Note in particular that E_{Fn} bends towards E_{Fp} rather than the other way around. This is because the electron concentration is much lower in the doped region than in the rest of the cell and this low concentration makes it possible to have large

⁴When selective contacts are discussed in this chapter, the terms “minority carriers” and “majority carriers” will always refer to the carriers that are least/most abundant in the selective membrane. This is different from the common terminology where minority/majority carriers are designated with respect to the doping of the bulk of the cell. Thus, in a p -membrane, the majority carriers are holes; in an n -membrane the majority carriers are electrons.

gradients in E_{Fn} . The bending of E_{Fn} is also the reason it is undesirable to have generation inside of the doped region, since the $\Delta\eta_{eh}G$ product is necessarily lower in the p -contact than in the bulk, meaning that the generation of e^-h^+ pairs in the doped region produces less free energy than in the bulk, leading to parasitic absorption. It should be emphasised that the high selectivity of the p -membrane and the decreasing Fermi level splitting inside of it, are strictly related to one another. Simply put, it is not possible to have a selective contact that is efficient at absorbing photons. The reason it makes sense to discuss a solar cells in terms of absorbers and selective membranes can be traced back to this elementary duality between functions.

The gradient of E_{Fn} across the selective contact drives an electron recombination current J_n towards the surface. This gradient and the recombination current cannot be avoided since a large Fermi level splitting in the bulk $\Delta\eta_{eh} = E_{Fn}(0) - E_{Fp}(0) > 0$ is necessary to generate useful work. In other words: $\Delta\eta_{eh}$ is simultaneously the voltage over the cell (this is because the quasi Fermi levels are nearly constant in the bulk) as well as the driving potential difference of the electrons towards the metal. Thus, there will always be contact recombination whenever the cell is producing power. Because the gradient in E_{Fn} is unavoidable, Eq. (3.3b) is the most convenient form of $q_{s,\text{trans}}$ to assess the losses due to the minority carrier transport, which shows that the only way to reduce the entropy generation from the minority carriers is to make the σ_n as low as possible.

To gain insight into the relation between J_n , $\Delta\eta_{eh}$ and the design parameters of the solar cell, an approximate solution to the drift-diffusion equation will given here in a fashion similar to Del Alamo [4]. This approximate solution will then be generalised and applied to other types of selective contacts to assess their potential for application in high-efficiency solar cells. The most important assumption that will be used, is that on the short length scales of the selective contacts, the source terms G and R have negligible influence on the shape of E_{Fn} and E_{Fp} . In a c -Si solar cell, this assumption holds to a good degree of accuracy for the junction on the rear side of the cell (known as the back surface field or BSF), so for the moment a BSF will be discussed. At the front junction of a c -Si cell, though, it is usually not possible to neglect G since the generation will affect the minority carrier quasi Fermi level significantly. This issue will be discussed briefly in Section 3.6 and as it will turn out, the issue of generation in the front membrane can be considered separately from the selectivity of that membrane, except for the case when the generation also significantly affects the majority carrier quasi Fermi level (which is uncommon).

In the BSF, the conductivity can be taken as an explicit function of x , so

that the electron transport equation is:

$$-\frac{\sigma_n(x)}{e^2} E'_{Fn}(x) = J_n = \text{constant}, \quad (3.4)$$

$$E_{Fn}(0) - E_{Fn}(L) = \Delta\eta_{eh}. \quad (3.5)$$

This equation is straightforward to integrate, giving:

$$J_n = \frac{\Delta\eta_{eh}}{\int_0^L \frac{e^2}{\sigma_n(x)} dx}. \quad (3.6)$$

The integral of $1/\sigma_n$ can be interpreted as a resistance (normalised to area), so Eq. (3.6) expresses (as expected) that the resistance of the region $0 \leq x \leq L$ determines the recombination current for a given potential difference $\Delta\eta_{eh}$ over the region. Eq. (3.6) is therefore simply Ohm's law, but unfortunately it is only useful qualitatively since $\sigma_n = en\mu_n$ is a function of E_{Fn} and therefore the resistance of the region $0 \leq x \leq L$ cannot be calculated directly. This makes the ordinary resistance (or conductance) relatively poor tools for understanding the working of a solar cell, so the next paragraphs will be used to develop two alternative concepts of resistance that are easier to compute. To obtain these alternative measures of resistance, first the electron drift-diffusion equation is written in full by expanding σ_n using the Boltzmann exponent for n :

$$\sigma_n(x) = e\mu_n(x)N_C \exp\left(-\frac{E_C - E_{Fn}}{k_B T}\right). \quad (3.7)$$

Thus, the electron transport equations is then:

$$-\frac{\mu_n(x)N_C}{e} \exp\left(-\frac{E_C(x) - E_{Fn}(x)}{k_B T}\right) E'_{Fn}(x) = J_n. \quad (3.8)$$

To solve this equation, it is assumed that $E_C(x)$ is already known. Of course, it is not possible to find the exact shape of the conduction band without solving all three drift-diffusion equations simultaneously, but in most cases it is possible to obtain a good approximation to $E_C(x)$ without much difficulty using, e.g., the Poisson Boltzmann equation under the assumption that E_{Fn} is straight (see Appendix A.2 and the red dashed line in Figure 3.2).

With $E_C(x)$ known, Eq. (3.7) can be solved for E_{Fn} by noting that:

$$\exp\left(\frac{E_{Fn}}{k_B T}\right) E'_{Fn} = \frac{d}{dx} \left[k_B T \exp\left(\frac{E_{Fn}}{k_B T}\right) \right] \quad (3.9)$$

The solution is therefore:

$$\begin{aligned} E_{Fn}(0) - E_{Fn}(L) &= \Delta\eta_{eh} \\ &= -k_B T \ln \left(1 - \frac{eJ_n}{k_B T} \left[\int_0^L \frac{1}{\mu_n(x)N_C(x) \exp\left(-\frac{E_C(x) - E_{Fn}(0)}{k_B T}\right)} dx \right] \right). \end{aligned} \quad (3.10)$$

Rewriting for J_n gives:

$$J_n = V_T \tilde{S}_n \left[1 - \exp \left(-\frac{\Delta\eta_{eh}}{k_B T} \right) \right], \quad \text{with:} \quad (3.11)$$

$$\begin{aligned} \tilde{S}_n &= \left[\int_0^L \frac{1}{\mu_n(x) N_C(x) \exp \left(-\frac{E_C(x) - E_{Fn}(0)}{k_B T} \right)} dx \right]^{-1} \\ &= \left[\int_0^L \frac{1}{\mu_n(x) \tilde{n}(x)} dx \right]^{-1}. \end{aligned} \quad (3.12)$$

Here the symbol \tilde{S}_n is used to denote a quantity that will be called the electron *pseudo conductance* of the doped region. It is the conductance that is calculated by assuming that the quasi Fermi level for the electrons is straight (i.e., $E_{Fn}(x) = E_{Fn}(0)$ for $0 \leq x \leq L$ indicated by the red dashed line in Figure 3.2). The associated electron density $\tilde{n}(x)$ will be referred to as the pseudo electron density and $\tilde{R}_n = \tilde{S}_n^{-1}$ is the electron pseudo resistance⁵. In Eq. (3.11), the last factor is approximately 1 since $\Delta\eta_{eh} \gg k_B T$ under operating conditions. Thus, in a sense, Eqs. (3.6) and (3.11) can be interpreted as if one can either use $\Delta\eta_{eh}$ or $k_B T$ as the driving force for J_n . When $\Delta\eta_{eh}$ is used as the driving force, then the real conductance should be used to calculate the recombination current; when $k_B T$ is used as the driving force, then the pseudo conductance should be used instead.

The pseudo conductance \tilde{S}_n depends exponentially on $\Delta\eta_{eh}$ due to the injection of electrons from the bulk into the doped region, so it is preferable to split off this trivial exponential dependence in the hope that whatever remains is approximately independent of $\Delta\eta_{eh}$. Thus, J_n is written as:

$$J_n = V_T \check{S}_n \left[\exp \left(\frac{\Delta\eta_{eh}}{k_B T} \right) - 1 \right], \quad (3.13)$$

$$= J_{0,n} \left[\exp \left(\frac{\Delta\eta_{eh}}{k_B T} \right) - 1 \right], \quad (3.14)$$

The resulting parameter $\check{S}_n = \tilde{S}_n \exp(-\Delta\eta_{eh}/k_B T)$ will be called the *intrinsic electron conductance* of the selective membrane because it usually does not depend on the injection level in the bulk of the cell. It is related to the well-known recombination parameter $J_{0,n}$ by $J_{0,n} = V_T \check{S}_n$. Some manipulation

⁵Because the default unit of current is $\text{cm}^{-2} \text{s}^{-1}$ in this work, $e\tilde{S}$ yields the conventional measure of area-normalised conductance. Thus $[e\tilde{S}] = \text{S cm}^{-2}$ and $[\tilde{R}/e] = \Omega \text{cm}^2$

yields:

$$\check{S}_n = \left[\int_0^L \frac{1}{\mu_n(x) N_C \exp\left(-\frac{E_C(x) - E_{Fp}(0)}{k_B T}\right)} dx \right]^{-1} \quad (3.15)$$

$$= \left[\int_0^L \frac{1}{\mu_n(x) \check{n}(x)} dx \right]^{-1}. \quad (3.16)$$

In words: \check{S}_n is the electron conductance calculated by assuming that $E_{Fn}(x) = E_{Fp}(0)$. The electron density calculated from this assumption is denoted by \check{n} and $\check{R}_n = \check{S}_n^{-1}$ is the associated resistance. Note that \check{S}_n is almost equal to the electron conductance calculated under equilibrium conditions where $E_{Fn} = E_{Fp} = E_F$ and $\check{n} = n_{\text{eq}}$. The reason it is not identically equal to the equilibrium conductance is because in the bulk of the cell $E_{Fp} \neq E_F$ due to injection of holes and because $E_C(x)$ depends on the operating conditions as well. However, the integral in Eq. (3.15) takes only a small contribution from the bulk, so this effect only has a small influence on \check{S}_n . In the p -membrane, which contributes the most to \check{S}_n , the doping is typically high enough to pin E_{Fp} and E_C to their equilibrium positions regardless of injection level. Thus, even though equilibrium band diagrams cannot provide a complete picture of the operation of solar cells, they can still be useful for making estimates of the intrinsic minority carrier conductivities of selective membranes.

To conclude, there are now three different measures of electron conductance: real, pseudo and intrinsic:

- The real conductance is calculated from the actual electron density n and the corresponding current driving force is the gradient of E_{Fn} .
- The pseudo conductance \check{S}_n is obtained by pretending that E_{Fn} is straight and then calculating the pseudo electron density \check{n} from that assumption. The current driving potential corresponding to the pseudo conductance is $V_T(1 - \exp(-\Delta\eta_{eh}/k_B T)) \approx V_T$. Thus, the pseudo conductance substantiates the intuition that the minority carrier recombination current is thermally driven towards the metal contact.
- The intrinsic conductance \check{S}_n is obtained by pretending that E_{Fp} is also the electron quasi Fermi level and calculating the density \check{n} from that assumption and its current driving potential is $V_T(\exp(\Delta\eta_{eh}/k_B T) - 1)$. The main advantage of \check{S}_n is that it does not depend strongly on the operating conditions, because the exponential dependence of the current on $\Delta\eta_{eh}$ has been absorbed into the driving force. This makes \check{S}_n approximately a device parameter, though care should be taken to not mistake it for a true constant: depending on the exact nature of the device \check{S}_n can still depend on $\Delta\eta_{eh}$ and if it does, it should be evaluated at max-power conditions to be of use as a device parameter.

3.2.1 What is selectivity?

At this point it is a good moment to reflect on the meaning of the word “selectivity” that has been used a couple of times so far. Indeed, this word is mentioned more and more frequently in the field of PV, yet no quantitative measure of selectivity has been proposed (to the author’s knowledge), though the recombination parameter J_0 is commonly mentioned as an important quantity connected to selectivity. However, J_0 is not a true measure of selectivity, since it does not measure the asymmetry between electron and hole conductance: a contact could have a low J_0 by simply being poorly conductive to both electrons and holes without actually preferring one over the other. To address this caveat, a selective contact can then be defined by having not only a low J_0 but also a low majority carrier resistivity (as has been assumed throughout this chapter).

Clearly, these two requirements are a good characterisation of a selective contact, yet they fail to precisely define what selectivity is by itself and a mathematical definition of selectivity is still absent. Given the discussion in the previous paragraphs, the first idea that comes to mind for defining selectivity, is to introduce a dimensionless number that is obtained by taking the quotient of some measure of minority carrier resistance and some measure of majority carrier resistance. Since it is desirable to have a figure of merit that is more or less independent of the voltage over the device, the intrinsic resistance \check{R}_n is the most suitable candidate for the electron resistance of the p -membrane. For the holes it can be said that the real, pseudo and intrinsic resistances are all equivalent and will simply be denoted by \check{R}_p . Thus, the selectivity of a p -membrane would then be \check{R}_n/\check{R}_p . In terms of the usual definitions of J_0 (in A cm^{-2}) and contact resistance $\rho_c = \check{R}_p/e$ (in Ωcm^2) the selectivity can be expressed as $\nu_p = V_T/(J_0\rho_c)$, confirming that high selectivity requires both a low J_0 and a low ρ_c .

However, this measure of selectivity is somewhat unsatisfactory. To see why, it is necessary to answer the question why exactly selectivity is important for a solar cell. Intuition would probably say that a high selectivity is necessary to ensure that the losses due to minority carrier recombination in the contacts are kept minimal, but this idea is actually not necessarily true and certainly not the complete story. As an example, consider a simplified model of a solar cell that is only limited by contact recombination (and perhaps other recombination mechanisms that can be described by an ideal diode equation). This means that E_{Fn} and E_{Fp} are approximately constant in the bulk⁶. In addition, it is assumed that the intrinsic conductivities of the n -contact and the p -contact are both constant, so that the losses in the cell can be described by a single recombination parameter $J_0 = k_B T(\check{S}_n + \check{S}_p)$. Since transport losses

⁶It is possible to incorporate bulk resistance in the following argument, but it makes the formulae more complex yet it does not affect the result in Eq. (3.20) significantly.

are neglected, the total charge current through the cell $J_{\text{charge}} = J_p - J_n$ is given by an ideal illuminated diode equation:

$$J_{\text{charge}} = J_{\text{gen}} - J_0 \left(\exp \left(\frac{\Delta\eta_{eh}}{k_B T} \right) - 1 \right), \quad (3.17)$$

with J_{gen} the photo generation current density. To extract as much free energy from the cell as possible, it is necessary to maintain an optimal potential $\Delta\eta_{eh} = \Delta\eta_{\text{mpp}}$ (i.e., the max-power potential) across the cell that maximises the product $\Delta\eta_{eh} J_{\text{charge}}$. This maximisation of an ideal solar cell is a well-known exercise and leads to the following transcendental equation for η_{mpp} :

$$J_{\text{charge}}(\Delta\eta_{\text{mpp}}) - \frac{\Delta\eta_{\text{mpp}}}{k_B T} J_0 \exp \left(\frac{\Delta\eta_{\text{mpp}}}{k_B T} \right) = 0. \quad (3.18)$$

The solution to this equation is not of real interest here: what should be noted is that the second term in Eq. (3.18) is almost equal to the power loss due to contact recombination of minority carriers P_{loss} . After all, the recombination current is given by Eq. (3.14) (in which the -1 term can be safely omitted) and the dissipation is obtained by integrating $-E'_{Fn} J_n$ over the p -membrane and $E'_{Fp} J_p$ over the n -membrane. Since $J_{n,p}$ are constant, these integrations results in the loss terms $\Delta\eta_{\text{mpp}} |J_n|$ and $\Delta\eta_{\text{mpp}} |J_p|$. Thus the total power loss due to recombination at the max power point is then:

$$P_{\text{loss}} = \Delta\eta_{\text{mpp}} J_0 \left(\exp \left(\frac{\Delta\eta_{\text{mpp}}}{k_B T} \right) - 1 \right) \approx \Delta\eta_{\text{mpp}} J_0 \exp \left(\frac{\Delta\eta_{\text{mpp}}}{k_B T} \right). \quad (3.19)$$

According to Eq. (3.18) the power loss due to recombination is:

$$P_{\text{loss}} = k_B T J_{\text{charge}}(\Delta\eta_{\text{mpp}}). \quad (3.20)$$

What makes this result noteworthy and perhaps seemingly paradoxical, is that P_{loss} is nearly independent of J_0 , since it is well-known that for an ideal solar cell (i.e., a cell described by a diode equation) the max-power current is close to J_{gen} , regardless of the recombination parameter. So if the power loss is approximately constant, what is the use of improving J_0 ? The answer is that a lower J_0 leads to a higher $\Delta\eta_{\text{mpp}}$, meaning that the majority carriers obtain more free energy (or, conversely, that there is less entropy generation associated with photon absorption (see Eq. (3.2) and Figure 3.3). Thus, even though all of the recombination occurs in the contacts, the most important thermodynamic losses occur in the bulk of the cell. The optimisation of J_0 (and therefore the selectivity of the contacts) is therefore less about minimising losses in the contacts and more about maximising gains in the bulk! This example also serves to illustrate that the optimisation of solar cells is a strongly non-local problem, meaning that a design flaw in one part of the cell might lead to power losses in a completely different part.

Thus, the problem with defining selectivity as \check{R}_n/\check{R}_p is that these two resistances affect the cell in completely different ways: \check{R}_p (linearly) affects the ohmic losses in the p -contact, while \check{R}_n (logarithmically) affects the generation losses in the bulk. A better way to define selectivity is obtained by considering the electrochemical potential. It is the job of a selective contact to enable a large $\Delta\eta_{eh}$ in the bulk of the cell while also losing as little free energy as possible when the majority carriers are conducted out of the cell. From Eq. (3.3a) it was derived that the gradient in E_{Fp} is a good measure of the extraction efficiency of holes, since J_p is fixed and equal to the photo current. Integrating E'_{Fp} over the selective contact yields (analogously to Eq. (3.6)):

$$\Delta E_{Fp} = E_{Fp}(L) - E_{Fp}(0) = eJ_p\check{R}_p = eJ_{\text{gen}}\check{R}_p. \quad (3.21)$$

This potential drop should be compared to the max-power potential $\Delta\eta_{\text{mpp}}$ that the contact can sustain in the bulk of the cell. However, since $\Delta\eta_{\text{mpp}}$ is slightly difficult to compute, it is more convenient to look at the open circuit potential $\Delta\eta_{\text{OC}}$ that the contact can sustain (i.e., the potential at which $J_n = J_{\text{gen}}$). It is given by:

$$\Delta\eta_{\text{OC}} = k_{\text{B}}T \ln \left(\frac{J_{\text{gen}}}{J_0} + 1 \right). \quad (3.22)$$

Thus, a selective contact is expected to have a high $\Delta\eta_{\text{OC}}$ and a small ΔE_{Fp} . The selectivity of a p -membrane ν_p will now be defined as:

$$\nu_p = \frac{\Delta\eta_{\text{OC}}}{E_{Fp}(L) - E_{Fp}(0)} = \frac{V_T \ln \left(\frac{J_{\text{gen}}}{J_0} + 1 \right)}{J_{\text{gen}}\check{R}_p}. \quad (3.23)$$

The physical interpretation of ν_p becomes clear once its inverse is considered: ν_p^{-1} is the fractional voltage loss due to ohmic dissipation in the p -contact if the cell would be completely limited by just that contact.

Having defined the selectivity, it should be noted that selectivity by itself is not enough to evaluate the performance of a selective contact. For example, the contact could be selective because its majority carrier resistance is extremely low rather than having a low J_0 . Such a contact would not be very useful for a solar cell. Indeed, no single parameter is capable of characterising a selective contact completely by itself since any assessment necessarily requires consideration of both electrons and holes and at the very least two figures of merit are necessary for this task. In a sense, J_0 and ρ_c already provide this complete characterisation, so the added value of discussing selectivity is not that it provides a new figure of merit that yields more information. Instead, the added value provided by the terminology of selectivity is the fact that the word *selectivity* draws attention to concepts of resistance for the understanding of the operation of solar cells.

After this general discussion of selective contacts, the equations for the pseudo and intrinsic conductances/resistances will be used in the next three subsections to assess the selectivity of three important kinds of selective contact: the homojunction, the heterojunction, and the Schottky-junction. To keep the formulae simple, most of the attention will be directed towards \check{S}_n and \check{R}_n (rather than ν_p), since these two parameters are the most crucial limiting factors for the contact performance in the examples that will be given.

3.3 The selectivity of Schottky-type junctions

In a Schottky-type junction illustrated in Figure 3.4a, the most important feature is a space charge region (SCR) in the semiconductor directly adjacent to the metal contact. In the ideal case, the source of this SCR is the mismatch between the metal workfunction Φ_M and the intrinsic Fermi level of the semiconductor (see, e.g., Sze [5] or Ibach & Lüth [6]). For the moment this ideal model of Schottky junctions will be considered; non-ideal factors will be discussed afterwards.

In the Schottky SCR, one type of carrier (in this example the electrons) is almost completely depleted, making the SCR more resistive to that carrier type. Schottky-type junctions are still occasionally proposed as selective contacts that are potentially relevant for solar cells, so it is a useful exercise to analyse their potential. For example, several authors have proposed devices based on graphene/*c*-Si junctions [7, 8, 9]. To determine the selectivity potential of a Schottky junction, the intrinsic electron conductance \check{S}_n of the SCR will be estimated. Let the position $x = 0$ be the edge of the SCR in the semiconductor and $x = L$ the position of the metal contact, noting that the width of the SCR depends on the Fermi level splitting $\Delta\eta_{eh}$ in the bulk and on the total bend banding across the SCR $\Delta\phi$. Next, the expression for \check{S}_n can be transformed by noting that $\check{n}p$ is approximately the equilibrium $np = n_i^2$ product since \check{n} is the electron density computed from E_{Fp} (recall that $E_{Fp} \neq E_F$, though). Furthermore, the mobility will also be assumed constant over the SCR. Using these approximations, the intrinsic electron resistance \check{R}_n becomes:

$$\check{R}_n \approx \frac{1}{\mu_n n_i^2} \int_0^L p \, dx \approx \frac{1}{\mu_n n_i^2} (Q_{\text{SCR}} - LN_{\text{dop}}). \quad (3.24)$$

Here, Q_{SCR} is the total amount of space charge (in units of cm^{-2}) and N_{dop} is the background charge of the doping in the semiconductor (in units of cm^{-3}). The reason the integral of p can be rewritten like this, is because the contribution of the electron density to the total space charge is negligible.

Eq. (3.24) is useful to make an estimation of how much space charge is needed to obtain a contact that is selective enough for a high efficiency cell. For example, if the aim is to make a *c*-Si solar cell with a V_{OC} of 750 mV,

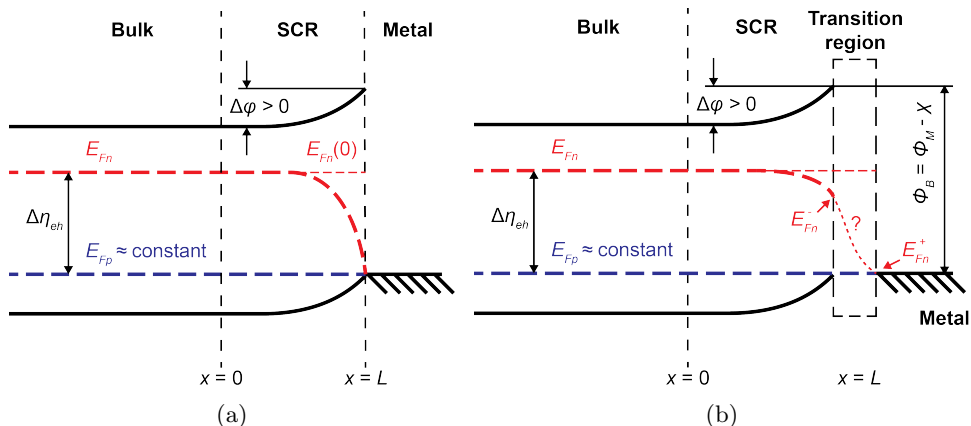


Figure 3.4: Sketch of the band diagram of a hole-selective Schottky contact under working conditions. In (a) the interface is assumed to be non-resistive, resulting in a continuous electron quasi Fermi level. In (b) the interface is modelled as a transition region that links the bulk descriptions of the two materials. The exact nature of this transition region is not known, but the current that crosses it can, for example, be modelled by a thermionic emission model. The resistance of the region creates a discontinuity $E_{Fn}^- - E_{Fn}^+ > 0$ across the interface.

then the recombination current should not exceed the short circuit current $J_{SC} \approx 43 \text{ mA cm}^{-2}$ in Eq. (3.13) (with $\Delta\eta_{eh} = 750 \text{ meV}$). This means that the intrinsic resistance \bar{R}_n/e should be more than $2.6 \cdot 10^{12} \Omega \text{ cm}^2$ and $J_{0,n} < 10 \text{ fA cm}^{-2}$. To achieve this, a total space charge $Q_{SCR} > 4 \cdot 10^{16} \text{ cm}^{-2}$ is necessary in the Schottky contact according to Eq. (3.24) (the doping density has no influence on this result for realistic values of N_{dop}). To put this number into context, values of Q_{SCR} between 10^{12} and 10^{13} cm^{-2} are commonly encountered in *c*-Si inversion/accumulation layers⁷.

The Poisson Boltzmann (PB) equation in Appendix A.2 can be used to find Q_{SCR} as a function of the total band bending and a straightforward calculation shows that a value of $Q_{SCR} = 4 \cdot 10^{16} \text{ cm}^{-2}$ would require more than 80 eV of band bending in the *c*-Si: an amount that is completely unfeasible in practice. For a *c*-Si Schottky junction, the bandgap of silicon (1.124 eV) is a reasonable upper limit to the amount of (equilibrium) band bending that can be achieved. This upper limit corresponds to an amount of space charge of $|Q_{SCR}| < 1 \cdot 10^{14} \text{ cm}^{-2}$ (regardless of the voltage over the cell). According to Eq. (3.24), the lower limit for the recombination parameter is then $J_0 > 4 \cdot 10^3 \text{ fA cm}^{-2}$, giving an upper limit of $V_{OC} < 600 \text{ mV}$ for space-charge limited Schottky cells.

⁷This range of charge is, not coincidentally, also the typical amount of fixed charge commonly encountered in heavily charged oxide passivation films on *c*-Si such as Al_2O_3 [10].

From this estimation can be concluded that the SCR in a *c*-Si Schottky junction can only achieve a limited amount of selectivity. In fact, in the field of device modeling, the SCR in *c*-Si cells can frequently be neglected without introducing significant error in the simulation (see, e.g., the COBO model by Brendel [11] and the Quokka model by Fell [12]) precisely because the SCR possesses so little selectivity. Clearly, if a SCR-based cell is to achieve a high efficiency, then it is necessary to introduce some extra source of selectivity between the *c*-Si and the metal. At this point it is worth considering the nature of the *c*-Si/metal interface in more detail, since it has been significantly simplified in the discussions until now. The assumption that E_{Fn} is continuous across the interface essentially assumes that this interface is truly two-dimensional and therefore does not present any resistance to the carriers that cross it. In reality this is not true and it is more accurate to regard the interface as a transition region of small but finite thickness where the bulk descriptions of the two joining materials fail because of their close proximity (see Figure 3.4b). The exact nature of this transition region is generally unknown, making it difficult to formulate a closed transport equation for this region. The most convenient way to deal with this uncertainty is to use an empirical model that describes the current through the transition region by imposing a modified boundary condition on J_n . In a sense, this means that the transition region is integrated out of the transport equation and by doing so, new empirical parameters are introduced that need to be correlated with experiment. The process of integrating out the transition region also causes a discontinuity in E_{Fn} ; the values of E_{Fn} on either side of the region are denoted by E_{Fn}^- and E_{Fn}^+ as indicated in Figure 3.4b.

A commonly used approximation for the transport of electrons between the semiconductor and the metal is the thermionic emission (TE) model. This model essentially asserts that the electrons on both sides of the interface behave like Fermi gases that are separated by an energy barrier of height $\Phi_B = \Phi_M - \chi$, with Φ_M the metal work function and χ the electron affinity of the semiconductor. Only an exponentially small fraction of the electrons in the metal have enough kinetic energy to cross the energy barrier Φ_B , while electrons from the semiconductor are free to move down the energy barrier into the bulk. Thus, the current $J_{M,S}$ from the metal into the semiconductor is:

$$eJ_{M,S} = A^*T^2 \exp\left(-\frac{\Phi_B}{k_B T}\right), \quad (3.25)$$

with A^* an empirical parameter called the effective Richardson constant, which characterises the microscopic details of the interface. Typically, A^* is on the same order of magnitude as the universal Richardson constant A_0 . At $T = 300$ K it is given by:

$$A_0 T^2 = \frac{4\pi m_e k_B^2 T^2 e}{h^3} = 1.08 \cdot 10^7 \text{ A cm}^{-2}. \quad (3.26)$$

Often, A^* is slightly smaller than A_0 due to the quantum mechanical probability that electrons reflect from the energy barrier even if (classically) they have enough energy to cross the interface. Furthermore, A_0 only takes the form in Eq. (3.26) if the bands in the metal and semiconductor are parabolic and if the charge carriers are not confined to less than 3 dimensions.

The current $J_{S,M}$ from the semiconductor into the metal depends on the surface electron density $n(L)$ and the conduction band density of states N_C :

$$eJ_{S,M} = A^*T^2 \frac{n(L)}{N_C}. \quad (3.27)$$

There is a special motivation for using this form of $J_{S,M}$ that becomes apparent when $J_{TE} = J_{S,M} - J_{M,S}$ is written out and manipulated into the following form:

$$\begin{aligned} J_{TE} &= \frac{A^*T^2}{e} \left[\frac{n(L)}{N_C} - \exp\left(-\frac{\Phi_B}{k_B T}\right) \right], \\ &= \frac{A^*T^2}{e} \frac{n(L)}{N_C} \left[1 - \exp\left(-\frac{E_{Fn}^- - E_{Fn}^+}{k_B T}\right) \right]. \end{aligned} \quad (3.28)$$

This form suggests that J_{TE} is driven by the electrochemical potential difference $E_{Fn}^- - E_{Fn}^+$ across the transition region, just like J_n is driven by the $\Delta\eta_{eh}$ in Eq. (3.11). Note especially that J_{TE} becomes zero whenever $E_{Fn}^- = E_{Fn}^+$, showing once more that the gradient in Fermi level across the interface is the driving force behind the thermionic emission, with electrons flowing from high to low E_{Fn} just like in the rest of the cell. Indeed, it would be a violation of the second law of thermodynamics if electrons would flow from low to high E_{Fn} across the interface since they would be gaining (rather than losing) free energy from being transported.

This suggests that the TE interface can be seen as a (zero-dimensional and non-linear) electron resistance not fundamentally different from the rest of the cell. Indeed, if the same analysis that led to Eq. (3.11) is carried out for the Schottky contact with TE transport at the interface, the pseudo and intrinsic resistances are given by:

$$\tilde{R}_n = \frac{k_B T N_C}{A^* T^2 \tilde{n}(L)} + \int_0^L \frac{1}{\mu_n(x) \tilde{n}(x)} dx, \quad (3.29a)$$

$$\check{R}_n = \frac{k_B T N_C}{A^* T^2 \tilde{n}(L)} + \int_0^L \frac{1}{\mu_n(x) \tilde{n}(x)} dx, \quad (3.29b)$$

$$= \frac{k_B T}{A^* T^2} \exp\left(\frac{\Phi_B}{k_B T}\right) + \int_0^L \frac{1}{\mu_n(x) \tilde{n}(x)} dx. \quad (3.29c)$$

Thus, the total resistance the electrons encounter is simply a sum of the resistance of the SCR (given by the integral) and the interface. Eqs. (3.29a) and (3.29b) show that interface resistance is found by calculating $V_T/J_{S,M}$ from \tilde{n} or

\tilde{n} . Eq. (3.29c) (which is obtained by assuming that the majority carrier Fermi level E_{Fp} is constant throughout the transition region) show that the intrinsic resistance of the transition region is a constant that only depends on the temperature, the barrier height, and the effective Richardson constant. Note that, perhaps surprisingly, the intrinsic resistance of the TE interface increases exponentially with Φ_B , even though the electrons cross the semiconductor/metal barrier in the energetically downward direction. This demonstrates once more that, when discussing transport of charge carriers, the total energy of the carriers is relatively unimportant since only the free energy drives the current. Thus, the formula $\check{R}_n \sim \exp(\Phi_B/k_B T)$ should not be interpreted as if the electrons have to cross an energy barrier, but rather that the barrier depletes the electron density near the interface.

It is an interesting exercise to calculate how large the energy barrier should be to salvage the $V_{OC} = 750$ mV goal for *c*-Si Schottky junctions that was proposed earlier. If the Richardson constant is taken as $A^* = A_0$, then $\Phi_B > 1.251$ eV will yield a high enough interface resistance to achieve this goal without taking into account the SCR. This is a rather significant number and it is not straightforward to achieve such an energy barrier in practice. However, the value of A^* can be significantly lower than A_0 for some junctions. As an example, graphene/*c*-Si junctions, which will be discussed in Chapter 6, have an effective Richardson constant⁸ of more than three orders of magnitude lower than A_0 [13], which brings down the requirement on the energy barrier to $\Phi_B > 1.026$ eV; a value that is physically feasible for a graphene/*c*-Si junction.

As can be seen, if a *c*-Si Schottky junction is to be successful as a selective contact, the metal-semiconductor interface is of great importance since that is the part where most of its selectivity will need to be derived from. Primarily, this means that metal-semiconductor interface needs to be of high quality: if the interface is defective, then the electrons will recombine before reaching the TE interface and the resistance of the interface becomes irrelevant⁹. Unfortunately, interface defects are common in Schottky-type junctions since the close proximity of the metal makes impurities in semiconductor very likely. Moreover, even if one succeeds in making an atomically sharp interface, there will be states in the metal that decay into the semiconductor where they generate a high density of midgap states called MIGS (Metal Induced Gap States [6, 14]). These MIGS will often influence the barrier height Φ_B significantly and may even cause E_{Fp} to bend due to Fermi level pinning. It is therefore very difficult to fabricate Schottky junctions that can serve as a good selective contact as long as the interface consists of nothing more than a metal on a (lowly doped)

⁸as will be explained in Chapter 6, these junctions actually do not have Richardson constant of the form in Eq. (3.26), but this detail is unimportant for now.

⁹In fact, as Section 3.6 will demonstrate, the recombination that happens before the carriers have crossed the interface can simply be regarded as a parallel resistance that shunts the interface resistance, thus limiting its selectivity.

semiconductor. An appealing solution is to insert a thin film of oxide between the semiconductor and the metal that can prevent impurities from reaching the semiconductor while simultaneously subduing the MIGS. For example, in graphene/*c*-Si based junctions, the inclusion of an SiO₂ film has already been shown to lead to a better V_{OC} [9]. However, even with an oxide buffer film, the Schottky junction still needs to have a high intrinsic Φ_B and a low A^* to act as a selective contact.

3.4 The selectivity of homojunctions

The next selective membrane that will be considered is the homojunction shown in Figure 3.2, which is currently the most common type used in industry. To assess the selectivity potential of the doped *p*-membrane, a block-shaped doping profile of width d and uniform doping N_{dop} is assumed. For such a profile, \tilde{S}_n and \check{S}_n are straightforward to calculate since the resistance integrals (Eqs. (3.12) and (3.15)) are completely dominated by the interval $L - d \leq x \leq L$ and because the valence and conduction bands are flat in the selective membrane. Note that the SCR in the bulk induced by the homojunction is ignored here: just like in the Schottky cell, the SCR in a homojunction can only achieve a very limited amount of selectivity and its contribution to \check{S}_n is negligible.

If $\Delta\phi$ is used to denote the band bending across the junction (see Figure 3.2), then the pseudo electron conductivity is approximately:

$$\tilde{S}_n \approx \frac{n(0)\mu_{n,BSF}}{d} \exp\left(-\frac{\Delta\phi}{k_B T}\right). \quad (3.30)$$

Here, $\mu_{n,BSF}$ is the electron mobility in the doped region (referred to as back surface field for convenience). In Eq. (3.30) the semiconductor-metal resistance of the interface at $x = L$ has been neglected (hence the continuous E_{Fn} in Figure 3.2), but it can be added in the same fashion as was done for Schottky junctions in Eq. (3.29a).

It is appealing to think of $\Delta\phi$ as an energy barrier that blocks electrons from diffusing into the metal. This view is perhaps not even completely wrong, but it is important to remember that $\Delta\phi$ is not constant but depends on the carrier densities in the cell. This expression for \tilde{S}_n can therefore be deceptive. The intrinsic conductance does not have this problem and some manipulation yields \check{S}_n for a block-shaped doping profile:

$$\check{S}_n = \frac{\mu_{n,BSF} n_{i,BSF}^2}{d N_{dop,BSF}} = \frac{\mu_{n,BSF} n_{eq}}{d}, \quad (3.31)$$

with $n_{i,BSF}^2$ the equilibrium *np* product in the doped region, $N_{dop,BSF}$ the *p*-type doping density in the BSF, and n_{eq} the equilibrium electron density in the

BSF. As was alluded to in Section 3.2, the high doping concentration in the BSF ensures that \tilde{S}_n is a true constant in the homojunction cell. Eq. (3.31) shows the three main parameters that influence the selectivity of a doped membrane: the minority carrier mobility, the equilibrium electron density and the depth of the doping profile. In practice, however, $\mu_{n,\text{BSF}}$ and n_{eq} both depend on the doping density $N_{\text{dop,BSF}}$, since the mobility decreases with increasing $N_{\text{dop,BSF}}$. This is because doping creates scattering centres that impede the transport of electrons. The dependence of n_{eq} on the doping level is more complicated, since n_{eq} depends in the intrinsic carrier density $n_{i,\text{BSF}}$, which in turn also depends on the doping density. There are two different phenomena that influence $n_{i,\text{BSF}}$, namely bandgap narrowing (BGN) and the transition from Boltzmann to Fermi-Dirac (FD) statistics for the holes. BGN tends to increase the intrinsic carrier density because there is more thermal excitation of electron hole pairs when the bandgap shrinks. FD statistics, on the other hand, tend to decrease $n_{i,\text{BSF}}$. To see why, first consider the intrinsic carrier density in the bulk n_i where Boltzmann statistics apply. Under equilibrium there is a single Fermi level $E_{Fn} = E_{Fp} = E_F$ and thus:

$$\begin{aligned} n_i^2 &= n_0 p_0 = N_C \exp\left(\frac{E_F - E_C}{k_B T}\right) N_V \exp\left(-\frac{E_F - E_V}{k_B T}\right) \\ &= N_C N_V \exp\left(\frac{-E_G}{k_B T}\right). \end{aligned} \quad (3.32)$$

The exponential dependence on E_G is the reason why n_i increases with decreasing bandgap. However, in a highly doped BSF where the holes become degenerate, the intrinsic carrier density changes to:

$$n_{i,\text{BSF}}^2 = n_0 p_0 = N_C \exp\left(\frac{E_F - E_C}{k_B T}\right) N_V F_{1/2}\left(\frac{E_V - E_F}{k_B T}\right). \quad (3.33)$$

For high doping densities, $E_F < E_V$ and the argument of the Fermi Dirac integral becomes positive. For large arguments, the Fermi function can be approximated as a step function so that:

$$F_{1/2}\left(\frac{E_V - E_F}{k_B T}\right) \sim \left(\frac{E_V - E_F}{k_B T}\right)^{\frac{3}{2}}. \quad (3.34)$$

Thus, by increasing the p -type doping, E_F is shifted downward so that the electrons are depleted exponentially in E_F while the hole density only increases proportionally to $(E_V - E_F)^{3/2}$. Overall, this leads to a decrease in $n_{i,\text{BSF}}$ since the decreasing exponent will win over the increasing power law.

The bandgap narrowing with doping has been measured and modelled extensively, such as in the work by Yan and Cuevas [15, 16] where an easy-to-use model for BGN in c -Si is presented. From experiments it is known that for c -Si,

the equilibrium minority carrier density decreases with increasing N_{dop} so that highly doped contacts are always better at reducing the contact recombination. This is particularly true for the degenerately doped regime where the minority carrier density is more strongly depleted due to FD statistics (see Figure 3.5). Furthermore, at high doping densities the mobility μ_n is also reduced, which impedes electron transport towards the contact even more. However, there is a limit to the solubility of dopants in semiconductors (see, e.g., Jaeger [17], Kooi [18], and Vick *et al.* [19] for the solubility of dopants in silicon) and once this limit has been achieved, the only design parameter that can increase the selectivity of the homojunction is the depth of the doping profile d . A quick back-of-the-envelope calculation shows that achieving a J_0 of less than 10 fA cm^{-2} with a block-shaped doping profile requires that $d > 10 \mu\text{m}$. This should be compared to typical *c*-Si junction depths that are frequently no more than $1 \mu\text{m}$.

Potentially, the selectivity of the doped membrane can be made arbitrarily large by making it deep enough, but in practice the intrinsic recombination inside of the membrane (e.g., Auger recombination) will impose a lower limit on the recombination current into the membrane. This is because the resistance encountered by a minority carrier to the point where it recombines is significantly lower if that point is somewhere in the middle of the doped region than if the electron needs to travel all the way to the metal contact. Thus, at some critical diffusion depth, recombination inside of the doped region becomes dominant over recombination at the contact and at that point it will be no use to make the diffusion even deeper. Furthermore, it is not desirable to make the junction arbitrarily deep for another reason, namely the fact that the doped region has the same absorption coefficient as the bulk of the cell. Deeper doping profiles will therefore result in more generation inside of the selective membranes, which should be avoided. Of course, for the selective contact at the rear of the cell, one could also choose to make the absorber very thick so that no light reaches the selective membrane, but this choice is at odds with the ideal situation where the cell should be as thin as possible (see the discussion following Eq. (3.2)).

3.5 The selectivity of heterojunctions

From the discussion of the homojunction cell and selective contacts so far, it can be concluded that selective membranes ideally have significantly different material properties than the photon absorber:

1. The bulk of the cell needs to have a high absorption coefficient while the selective membrane should be completely transparent if possible. Alternatively, if the selective membranes are absorptive, then it is necessary

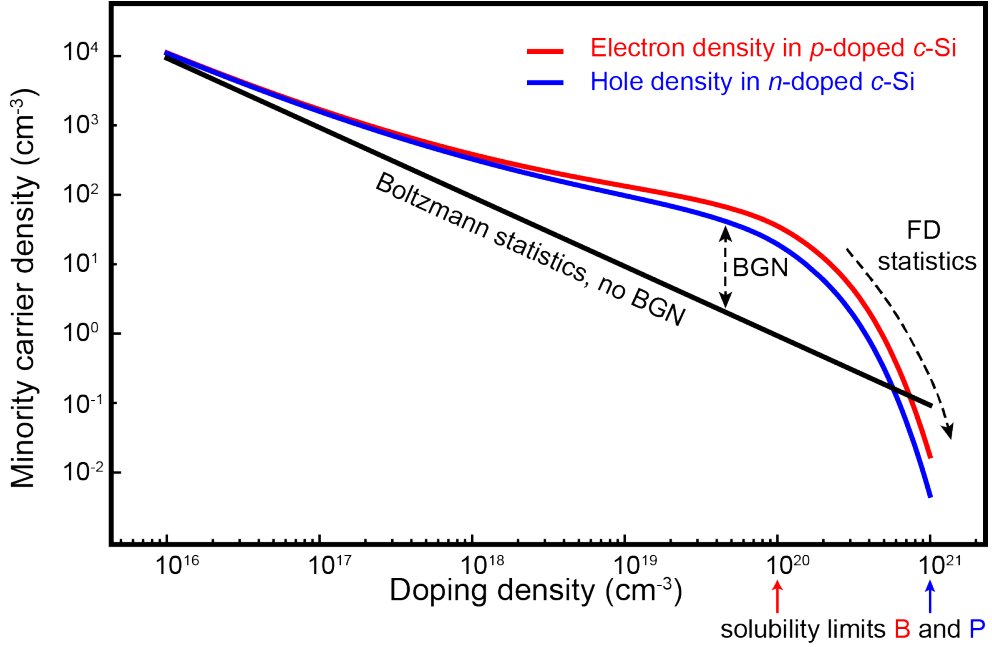


Figure 3.5: Equilibrium minority carrier densities in doped c -Si as a function of the doping density. The black line shows the minority carrier density if n_i is assumed constant (at $9.65 \cdot 10^9 \text{ cm}^{-3}$) (i.e., if the carriers are assumed to obey Boltzmann statistics and if the bandgap is assumed constant). The deviations of the red (electron density) and blue (hole density) lines from the black line are due to BGN and Fermi-Dirac statistics (calculated with the model by Yan and Cuevas [16]). Initially, the BGN opposes the decrease of the minority carrier density with doping density. However, at the point where FD statistics start to become influential, the minority carrier density decreases at an accelerated rate with doping. As shown by Eq. (3.31), it is favourable to have an equilibrium minority carrier density that is as low as possible. The plot shows that degenerately doped c -Si is therefore the most suitable for making selective contacts. The arrows show the solubility limits of boron [19] (red) and phosphorus [18] (blue) in c -Si, demonstrating that n -type doped c -Si can achieve a much lower minority carrier density.

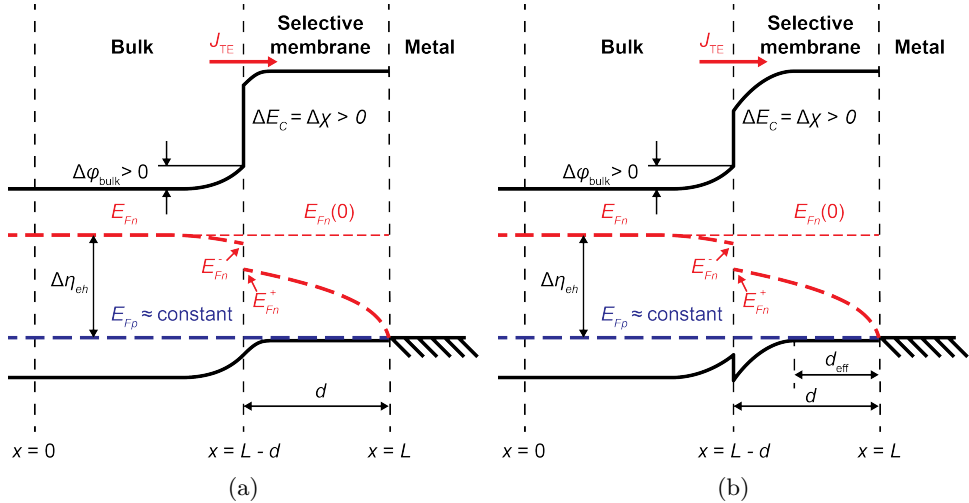


Figure 3.6: Sketch of the band diagram of a selective heterojunction hole contact. Interface defects are assumed zero here. In (a) the ideal situation is depicted where the valence band is continuous while the conduction band has an offset. In (b) there is also an offset in the valence band: this leads to a larger SCR in the selective membrane, decreasing its effective thickness d_{eff} and making it less resistive to electrons. Depending on the exact nature of the heterojunction, the SCR caused by the valence band offset might also pose a significant resistance to the hole current which would result in ohmic losses (i.e., FF loss) in the contact.

to direct the light inside of the cell in such a way that the selective membranes do not receive photons with energies above their bandgap.

2. The mobility of the bulk should be high for both electrons and holes, while a hole (electron) selective membrane ideally has a high hole (electron) mobility and a very low electron (hole) mobility. In fact, it is possible to make a working cell by having selective membranes that only differ in their electron and hole mobility [2]. Typically, though, it is most important that the minority carrier mobility is low since the majority carrier conductivity can be increased by doping if necessary.
3. The photon absorber is preferably a (defect-free) direct bandgap semiconductor since that promotes radiative recombination over other intrinsic recombination mechanisms, making it possible to recycle the photons emitted by recombining charge carriers. The selective membranes on the other hand, should simply have as little recombination as possible, making indirect bandgap materials preferable.

These very different requirements suggest that for homojunction solar cells it will always be difficult to find one material that performs well as both an

absorber and as a selective membrane. Heterojunctions do not have this intrinsic limitation of conflicting material properties, giving them significantly more design freedom to achieve higher efficiencies. A frequently proposed ideal heterostructure is one where the bands of the majority carriers are continuous across the interface whereas the bands of the minority carriers have a large discontinuity. This is illustrated in Figure 3.6a for a hole-collecting contact. The heterojunction p -membrane can be analysed in the same fashion as the doped p -membrane in Figure 3.2, namely by solving the electron transport equation under the assumption that the source term G_{net} is negligible and that E_{Fp} is flat. In fact, the former approximation is often much more accurate for a selective membrane with a high bandgap than for a homojunction. The assumption $G_{\text{net}} = 0$ also means that interface defects are neglected. Of course, for a good selective contact it is vital to have as little recombination-active interface defects as possible: the point of a selective membrane is to place a buffer between the electron-hole pairs in the bulk and the recombination-active metal, so if there are many interface defects at the absorber/selective membrane interface the purpose of the heterojunction is completely defeated. Achieving a low interface defect density is, however, a significant practical challenge in heterojunction solar cells.

To calculate \tilde{S}_n and \check{S}_n for the heterojunction in Figure 3.6, the interface with the metal at $x = L$ is assumed free of resistance, though its resistance can be added without difficulty as before. The heterojunction interface at $x = L - d$, on the other hand, is assumed to be resistive (as evident from the discontinuity in E_{Fn}) and will be modelled using a TE model. For simplicity, the Anderson model [20] is used to describe the conduction band offset, so $\Delta E_C = \Delta\chi$ with $\Delta\chi$ (defined positive in Figure 3.6) the difference between the two electron affinities of the bulk and the p -membrane. Thus, J_{TE} is given by:

$$\begin{aligned}
 J_{\text{TE}} &= \frac{A^*T^2}{e} \left[\frac{n^-}{N_{C^-}} \exp\left(-\frac{\Delta\chi}{k_{\text{B}}T}\right) - \frac{n^+}{N_{C^+}} \right] \\
 &= \frac{A^*T^2}{e} \frac{n^-}{N_{C^-}} \exp\left(-\frac{\Delta\chi}{k_{\text{B}}T}\right) \left[1 - \exp\left(\frac{E_{Fn^+} - E_{Fn^-}}{k_{\text{B}}T}\right) \right]. \quad (3.35)
 \end{aligned}$$

Here, quantities to the left (right) of the discontinuity at $x = L - d$ are indicated with a superscript minus (plus) sign. Notice how, once more, the difference in E_{Fn} across the interface is the driver of the TE current.

With the thermionic emission given, \tilde{R}_n and \check{R}_n can be calculated:

$$\tilde{R}_n = k_B T \frac{N_{C,\text{bulk}}}{A^* T^2 \tilde{n}^- (L-d)} \exp\left(\frac{\Delta\chi}{k_B T}\right) + \int_0^L \frac{1}{\mu_n(x) \tilde{n}(x)} dx, \quad (3.36)$$

$$\check{R}_n = k_B T \frac{N_{C,\text{bulk}}}{A^* T^2 \check{n}^- (L-d)} \exp\left(\frac{\Delta\chi}{k_B T}\right) + \int_0^L \frac{1}{\mu_n(x) \check{n}(x)} dx, \quad (3.37)$$

$$\Delta\chi = \chi_{\text{bulk}} - \chi_{\text{sel}} \geq 0,$$

$$\Delta\phi_{\text{bulk}} = \phi(0) - \phi(L-d) \geq 0,$$

$$\tilde{n}(x) = N_C(x) e^{-\frac{E_C(x) - E_{Fn}(0)}{k_B T}} = N_C(x) e^{\frac{E_{Fn}(0) + \phi(x) + \chi(x)}{k_B T}},$$

$$\check{n}(x) = N_C(x) e^{-\frac{E_C(x) - E_{Fp}(0)}{k_B T}} = N_C(x) e^{\frac{E_{Fp}(0) + \phi(x) + \chi(x)}{k_B T}}.$$

Here, the subscript ‘‘bulk’’ indicates material properties of the photon absorber and the subscript ‘‘sel’’ is for material properties of the selective membrane. Thus, \tilde{S} , and \check{S} are the sum of bulk and interface resistances, indicating that the heterojunction interface indeed possesses a point-like resistance. The pseudo or intrinsic interface resistance is calculated by assuming a straight E_{Fn} ; then finding \tilde{n} or \check{n} in the absorber at the interface; then using that density to calculate the TE current going into the p -membrane (ignoring the reverse current from the p -membrane into the semiconductor); and then finally dividing that TE current by the thermal voltage.

The procedure for calculating the resistance of a selective membrane outlined in this section generalises to arbitrary stacks of materials in a straightforward fashion, provided that the underlying assumptions (1D charge transport; low generation and recombination rates in selective regions; TE transport at the interfaces; and E_{Fp} constant) remain valid. The transport resistance from the bulk of the layers is taken into account by the integral in Eq. (3.36), while each interface adds an extra resistance that depends on its band offset $\Delta\chi$; effective Richardson constant A^* ; and the pseudo density at the interface \tilde{n} .

It is possible to estimate the intrinsic resistance \check{R}_n of the selective contact in Figure 3.6 by making the assumption that the p -membrane has a uniform and high doping density so that almost all of the band bending occurs in the bulk of the cell. The resistance of the membrane is then the following sum of the interface and bulk contributions:

$$\check{R}_n = \check{R}_{n,\text{TE}} + \check{R}_{n,\text{sel}}, \quad (3.38a)$$

$$\check{R}_{n,\text{TE}} = \frac{k_B T}{A^* T^2} \frac{N_{\text{dop,sel}}}{N_{V,\text{sel}}} \exp\left(\frac{E_{G,\text{sel}}}{k_B T}\right), \quad (3.38b)$$

$$\check{R}_{n,\text{sel}} = \frac{d N_{\text{dop,sel}}}{\mu_{n,\text{sel}} n_{i,\text{sel}}^2} \quad (3.38c)$$

Eqs. (3.38) demonstrate that the most important tuning parameter to suppress J_0 is the bandgap of the selective membrane $E_{G,\text{sel}}$, since $n_{i,\text{sel}}$ decreases

exponentially with $E_{G,\text{sel}}$ and also because $E_{G,\text{sel}}$ increases \tilde{R}_n in junctions where TE is the limiting transport process. As a rule of thumb it can be said that for solar cells limited by the selectivity of their contacts, the V_{OC} increases linearly with the bandgap of the selective contact. However, it should not be overlooked that the electron mobility in the p -membrane as well as its thickness have significant influence on the performance of the contact as well. To put it a bit more strongly: if the membrane is not thick enough, it is unable to function as a sufficient resistance to minority carriers. The conduction band offset in Figure 3.6 should therefore not be merely considered as a barrier to electrons, because this view leads to the incorrect conclusion that it is only the interface that provides the selectivity towards electrons. While it is true that the interface also contributes to the electron resistance, Eqs. (3.38) show that the interface is only part of the story and that the bulk of the selective membrane is equally or more important.

The previous discussion does not change much if there is also a valence band offset such as shown in Figure 3.6b). In that case, the thickness of the membrane d should be replaced with the effective thickness d_{eff} (as shown in Figure 3.6b) in Eqs. (3.38) and the interface resistance $\tilde{R}_{n,\text{TE}}$ needs to be revised because the band offset influences the electron density at the interface. Depending on the exact nature of the heterojunction, a large offset in the valence band can also potentially give rise to a SCR (either in the bulk or in the selective membrane) that is resistive to the holes. Even though space charge regions usually do not have a significant resistance when it comes to reducing minority carrier contact recombination, in SCRs where the majority carriers are depleted there can be significant resistive losses because the majority carrier current needs to be high under operating conditions. Fortunately, a heterojunction contact generally gives the cell a high V_{OC} , meaning that the injection level in the bulk of the cell is also high. This high injection level, in turn, also leads to extra hole injection into the resistive SCR, increasing its conductance. As can be seen, valence band offsets can be –but do not need to be– detrimental to the function of a hole-selective contact and in the end it depends on the details of the cell whether or not they limit its efficiency. Unfortunately, band offsets for the majority carriers are often practical hindrances that are difficult to control, meaning that one has to engineer around them.

It can be concluded that heterojunctions have significant advantages as selective membranes over homojunctions. First of all, heterojunctions have more degrees of freedom than homojunctions to tune the selectivity of the membrane for one type of carrier, which makes it possible to keep the selective membranes very thin. Second of all, heterojunctions offer the possibility to decouple the optical and electrical optimisation of the selective membrane. By using a non-absorbing material for the selective contacts, the problem of parasitic absorption can, in principle, be solved completely. It is an extra

plus that the demands of transparency and selectivity both demand a material with a high bandgap. It should be emphasised, though, that absorption in a wide-bandgap selective membrane is always detrimental despite the fact that the electron-hole pairs generated in such a wide-bandgap material have more energy than those that are generated in the bulk of the cell. The crucial point is that the *free* energy of the electron-hole pairs $\Delta\eta_{eh}$ is what really matters and $\Delta\eta_{eh}$ is the largest in the bulk and decreases towards the metal contact. As was noted in the beginning of section this chapter, this is because the minority carrier quasi Fermi level bends most in the part of the cell where the minority carrier conductivity is lowest, i.e., in the selective membranes. Heterojunctions are no exception to this.

The *a*-Si:H/*c*-Si heterojunction cell from Panasonic (which holds the current world record for *c*-Si based technology [21]) is a good example to illustrate the advantages of heterojunctions. Most notably, this cell can achieve a good V_{OC} of 740 mV due to the higher selectivity of the *a*-Si:H/*c*-Si junction compared to an ordinary diffused junction. The wide bandgap and the low mobility of the *a*-Si:H are the crucial to the success of the *a*-Si:H/*c*-Si heterojunction cell. Unfortunately, though, this cell is non-ideal in one very important aspect, namely the optical properties of the materials: *a*-Si:H is a direct bandgap semiconductor with a high absorption coefficient while *c*-Si has an indirect bandgap and a low absorption coefficient. Even though the bandgap of *a*-Si:H is larger than that of *c*-Si, it is not large enough to prevent significant parasitic absorption in the selective membranes [3]. The most successful method so far to deal with this problem is to fabricate both the *n*-contact and the *p*-contact on the rear of the cell, because only low-energy photons (that are not absorbed by the *a*-Si:H) can penetrate all the way through the *c*-Si bulk.

3.6 Generation and recombination in selective membranes

At the beginning of Section 3.2 the assumption was made that there is no generation or recombination of carriers in the selective membrane, but this assumption cannot always be made in practise. To see how generation and recombination affect the discussion of selective contacts, a small modification of the original problem is considered. Just like before, E_{Fp} is assumed to be straight in the *p*-membrane and the electron transport equation is solved on the interval $0 \geq x \geq L$, with $x = 0$ a point in the bulk and $x = L$ the metal contact. The metal contact is assumed resistanceless, so that E_{Fn} is continuous at $x = L$.

This time, though, there is a point-like source/sink of electrons somewhere in this interval at $x = a$ that generates/recombines an amount J_{source} of electrons per unit of time. It is straightforward to solve the electron transport

equation on the subintervals $[0, a]$ and $[a, L]$ and then to patch those two solutions together. For notational convenience, the intrinsic resistance¹⁰ of a subinterval $[x_1, x_2] \subseteq [0, L]$ will be denoted as:

$$\check{R}_n(x_1, x_2) = \int_{x_1}^{x_2} \frac{1}{\mu_n \check{n}} dx. \quad (3.39)$$

The electron current flowing from the bulk into the selective membrane is then:

$$J_n(0) = \frac{V_T}{\check{R}_n(0, L)} \left[\exp\left(\frac{\Delta\eta_{eh}}{k_B T}\right) - 1 \right] - \frac{J_{\text{source}}}{1 + \frac{\check{R}_n(0, a)}{\check{R}_n(a, L)}}. \quad (3.40)$$

In Eq. (3.40), the first term is simply the metal recombination current that would flow in the absence of a source/sink, while the second term shows how J_{source} affects the total current. Heuristically, Eq. (3.40) can be interpreted as an ordinary electrical circuit consisting of two series resistances $\check{R}(0, a)$ and $\check{R}(a, L)$ with a current source between them (see Figure 3.7a). This demonstrates that the intrinsic resistance behaves just like one would expect from a real resistance.

For a thermodynamic discussion of generation and recombination, it is also useful to calculate the Fermi level splitting at the source/sink:

$$E_{F_n}(a) - E_{F_p} = \Delta\eta_{eh} + k_B T \ln \left[\frac{\frac{\check{R}_n(a, L)}{\check{R}_n(0, a)} + e^{\frac{-\Delta\eta_{eh}}{k_B T}} \left(1 + \frac{J_{\text{source}} \check{R}_n(a, L)}{V_T} \right)}{1 + \frac{\check{R}_n(a, L)}{\check{R}_n(0, a)}} \right]. \quad (3.41)$$

As an example of how to interpret Eq. (3.40) and (3.41), consider the case where point a is deep within the selective membrane so that $\check{R}_n(0, a) \gg \check{R}_n(a, L)$. In this situation the source is essentially shielded from the bulk and only a small fraction of the source influences the current $J_n(0)$ in the bulk. If $J_{\text{source}} = -J_{\text{rec}} < 0$ acts as a recombination centre, this means that the recombination is not very detrimental to the operation of the cell, since the carriers would have recombined in the metal anyway even if J_{rec} were zero. Thermodynamically speaking, the Fermi level splitting at $x = a$ given by Eq. (3.41) is significantly less than $\Delta\eta_{eh}$ in this case so that the recombination does not generate much entropy. Conversely, if the recombination centre is close to the bulk so that $\check{R}_n(0, a) \ll \check{R}_n(a, L)$, then J_{rec} adds directly to the total recombination current, giving it a significant impact on the cell performance. In that case, the recombination happens at a point with high Fermi level splitting.

As an alternative view, it is also possible to model the recombination as a parallel resistor as shown in Figure 3.7b. The recombination resistance $\check{R}_{n, \text{rec}}$

¹⁰For this discussion the intrinsic resistance will be used, though it is equally possible to use the pseudo resistance.

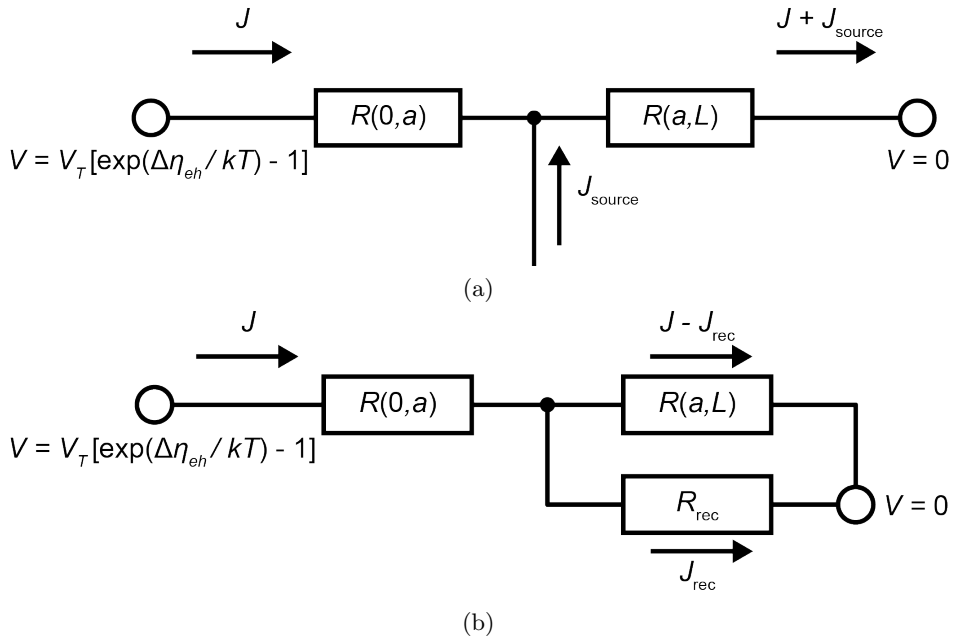


Figure 3.7: The equivalent circuit of the metal electron recombination current through a selective membrane with (a) a current source in the middle or (b) a parallel resistor that models bulk recombination. If the recombination rate is proportional to the np product, then the parallel resistor in (b) is constant; otherwise it is non-ohmic. Recall that the driving force V that is shown here for the equivalent circuit corresponds to the intrinsic resistance and should therefore not be confused with E_{Fn} . For extra clarity it should be noted that these circuits only correspond to the minority electron current; the majority hole current is not considered here.

is not constant for general recombination processes, but there is an important class of loss mechanisms that can be modelled by an ohmic resistance, namely those where the recombination rate is proportional to the np product. To find $\check{R}_{n,\text{rec}}$, the recombination current is first stated in its general form using its recombination parameter $J_{0,\text{rec}}$ and is then rewritten as follows:

$$\begin{aligned} J_{\text{rec}} &= J_{0,\text{rec}} \left[\frac{np}{n_i^2} - 1 \right] \\ &= \frac{V_T}{\check{R}_{n,\text{rec}}} \left[\exp \left(\frac{E_{Fn}(a) - E_{Fp}}{k_B T} \right) - 1 \right]. \end{aligned} \quad (3.42)$$

Thus, $\check{R}_{n,\text{rec}} = V_T/J_{0,\text{rec}}$. The fact that $\check{R}_{n,\text{rec}}$ is constant, is due to the exponential dependence on $E_{Fn}(a) - E_{Fp}$.

For generation in a p -membrane the discussion is similar as for recombination. If the generation takes place deep in the p -membrane (i.e., $J_{\text{source}} > 0$ and $\check{R}_n(0, a) \gg \check{R}_n(a, L)$), then Eq. (3.40) shows that most of the electrons generated at $x = a$ are not transported into the bulk but are lost in the hole-collecting metal contact. In this case, the generation of carriers does not generate free energy due to the low Fermi level splitting at the source, leading to parasitic absorption. If $\check{R}_n(0, a) \ll \check{R}_n(a, L)$, then the generated electrons are transported into the bulk and contribute to the power output of the cell. Perhaps surprisingly, the generation of carriers in the p -membrane has no influence on its selectivity towards carriers coming from the bulk, since these carriers still experience the same resistance regardless of the value of J_{source} . This means that the selectivity of the contact can be discussed independently from its parasitic absorption. The exception to this rule is when J_{source} is so large that not only E_{Fn} but also E_{Fp} (and thereby \check{R}_n) gets affected by the generation. The high density of holes in the p -membrane makes that unlikely, though.

The main goal of Section 3.6 is to substantiate the intuition that a selective membrane can be considered as a series of resistances with sources/sinks between them and Eq. (3.40) expresses this idea quantitatively and succinctly for one source. The method outlined here can be extended in a straightforward fashion to include more than one sinks/sources separated by resistances. When this procedure is taken to the continuum limit, one essentially obtains the Green's function formalism for solving the electron transport equation and the mathematically inclined reader will perhaps recognise in Eq. (3.40) a disguised form of a Green's function for the electron current. However, the Green's function formalism will not be further developed here, since it will not help to clarify the message of this section. Moreover, the Green's function formalism is complicated further by the fact that it can only provide an implicit solution for electron transport if the recombination is a function of E_{Fn} , which is usually

the case¹¹.

3.7 Conclusions

The most important conclusion from this chapter is that a cell can be divided into two parts with complementary functions: the photon absorber and the selective membranes. The absorber converts photons into free electrochemical energy and the selective membranes make sure that the electrons and holes exit the cell through different leads so that a net electric current can flow. These two functions cannot be mixed: a photon absorber is necessarily not selective while a selective membrane is necessarily less efficient at converting photons into electrochemical energy than the absorber. Because of this, the selective membrane model presents a meaningful way to divide a cell into functional parts and to discuss the thermodynamic losses in each part, making it a useful language for discussions about optimisation. As it turns out, not many new tools are necessary to do so: the familiar recombination parameter J_0 can be interpreted as a type of minority carrier resistance that is related to parameters of the selective membrane such as its thickness, mobility and bandgap and the Richardson constants of interfaces. In many cases, this resistance can be estimated with relatively simple formulae. Furthermore, the thermodynamic view of solar cells puts a great importance on the quasi Fermi levels $E_{F_{n,p}}$ that represent the free energy of the electrons and holes. The most important realisation here is that free energy is simply more thermodynamically relevant than total energy, since one can in general not expect to extract all energy from the cell: only free energy is available to generate work. It is therefore recommended to start viewing thermalisation as free energy thermalisation as shown in Figure 3.3b.

Finally, because the photon absorber and the membranes have complementary functions it becomes natural to look for different materials to fulfil these functions optimally. In short, the heterojunction cell is the way forward if we really want to push the efficiency limit of solar cells.

¹¹This situation is similar to the one where Green's function formalism is used for solving Schrödinger's equation in quantum mechanics. See, for example, Chapter 11.4 on the Born approximation in Griffiths [22]

References Chapter 3

- [1] Peter Würfel. *Physics of Solar Cells*. Wiley-VCH Verlag GmbH & Co, Weinheim, 2005. ISBN 3-527-40428-7.
- [2] Uli Würfel, Andres Cuevas, and Peter Würfel. Charge Carrier Separation in Solar Cells. *IEEE Journal of Photovoltaics*, 5(1):461–469, 2015. ISSN 2156-3381. doi: 10.1109/JPHOTOV.2014.2363550.
- [3] Stefaan De Wolf, Antoine Descoeurdes, Zachary C. Holman, and Christophe Ballif. High-efficiency Silicon Heterojunction Solar Cells: A Review. *green*, 2(1):7–24, January 2012. ISSN 1869-876X. doi: 10.1515/green-2011-0018.
- [4] J.A. del Alamo and R.M. Swanson. The physics and modeling of heavily doped emitters. *IEEE Transactions on Electron Devices*, 31(12):1878–1888, December 1984. ISSN 0018-9383. doi: 10.1109/T-ED.1984.21805.
- [5] S.M. Sze and K. Ng Kwok. *Physics of Semiconductor Devices*. John Wiley & Sons, Inc., 2007. ISBN 978-0-471014323-9.
- [6] Harald Ibach and Hans Lüth. *Solid-State Physics*. Springer-Verlag Berlin Heidelberg New York, Berlin, Heidelberg, 3 edition, 2003. doi: Phys-RevB.83.165113.
- [7] Xinming Li, Hongwei Zhu, Kunlin Wang, Anyuan Cao, Jinqian Wei, Chunyan Li, Yi Jia, Zhen Li, Xiao Li, and Dehai Wu. Graphene-on-silicon Schottky junction solar cells. *Advanced materials (Deerfield Beach, Fla.)*, 22(25):2743–8, July 2010. ISSN 1521-4095. doi: 10.1002/adma.200904383.
- [8] Yu Ye and Lun Dai. Graphene-based Schottky junction solar cells. *Journal of Materials Chemistry*, 22(46):24224, 2012. ISSN 0959-9428. doi: 10.1039/c2jm33809b.
- [9] Yi Song, Xinming Li, Charles Mackin, Xu Zhang, Wenjing Fang, Tomás Palacios, Hongwei Zhu, and Jing Kong. Role of Interfacial Oxide in High-Efficiency GrapheneSilicon Schottky Barrier Solar Cells. *Nano Letters*, 15(3):150225082743008, 2015. ISSN 1530-6984. doi: 10.1021/nl505011f.

- [10] G. Dingemans and W. M. M. Kessels. Status and prospects of Al₂O₃-based surface passivation schemes for silicon solar cells. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 30(4):040802, 2012. ISSN 07342101. doi: 10.1116/1.4728205.
- [11] Rolf Brendel. Modeling solar cells with the dopant-diffused layers treated as conductive boundaries. *Progress in Photovoltaics: Research and Applications*, 20(1):31–43, jan 2012. ISSN 10627995. doi: 10.1002/pip.954.
- [12] Andreas Fell. A Free and Fast Three-Dimensional/Two-Dimensional Solar Cell Simulator Featuring Conductive Boundary and Quasi-Neutrality Approximations. *IEEE Transactions on Electron Devices*, 60(2):733–738, feb 2013. ISSN 0018-9383. doi: 10.1109/TED.2012.2231415.
- [13] Dhiraj Sinha and Ji Ung Lee. Ideal Graphene/Silicon Schottky Junction Diodes. *Nano Letters*, 14(8):4660–4664, aug 2014. ISSN 1530-6984. doi: 10.1021/nl501735k.
- [14] J Tersoff. Schottky Barrier Heights and the Continuum of Gap States. *Physical Review Letters*, 52(6):465–468, feb 1984. ISSN 0031-9007. doi: 10.1103/PhysRevLett.52.465.
- [15] Di Yan and Andres Cuevas. Empirical determination of the energy band gap narrowing in highly doped n+ silicon. *Journal of Applied Physics*, 114(4):044508, 2013. ISSN 00218979. doi: 10.1063/1.4816694.
- [16] Di Yan and Andres Cuevas. Empirical determination of the energy band gap narrowing in p+ silicon heavily doped with boron. *Journal of Applied Physics*, 116(19):194505, November 2014. ISSN 0021-8979. doi: 10.1063/1.4902066.
- [17] Richard C. Jaeger. *Introduction to Microelectronic Fabrication: Modular Series on Solid State Devices Volume V*. Prentice Hall, 2 edition, 2001. ISBN 978-0201444940.
- [18] E. Kooi. Formation and Composition of Surface Layers and Solubility Limits of Phosphorus During Diffusion in Silicon. *Journal of The Electrochemical Society*, 111(12):1383, 1964. ISSN 00134651. doi: 10.1149/1.2426010.
- [19] G. L. Vick and K. M. Whittle. Solid Solubility and Diffusion Coefficients of Boron in Silicon. *Journal of The Electrochemical Society*, 116(8):1142, 1969. ISSN 00134651. doi: 10.1149/1.2412239.
- [20] R.L. Anderson. Experiments on Ge-GaAs heterojunctions. *Solid-State Electronics*, 5(5):341–351, September 1962. ISSN 00381101. doi: 10.1016/0038-1101(62)90115-6.

- [21] Martin A. Green, Keith Emery, Yoshihiro Hishikawa, Wilhelm Warta, and Ewan D. Dunlop. Solar cell efficiency tables (version 46). *Progress in Photovoltaics: Research and Applications*, 23(7):805–812, jul 2015. ISSN 10627995. doi: 10.1002/pip.2637.
- [22] David J Griffiths. *Introduction to Quantum Mechanics*. Pearson Education, Inc., 2nd edition, 2005. ISBN 0-13-191175-9.

Chapter 4

Variational method for the minimisation of entropy generation in solar cells

Abstract

In this work a method is presented to extend traditional solar cell simulation tools to make it possible to calculate the most efficient design of practical solar cells. The method is based on the theory of non-equilibrium thermodynamics, which is used to derive an expression for the local entropy generation rate in the solar cell, making it possible to quantify all free energy losses on the same scale. The framework of non-equilibrium thermodynamics can therefore be combined with the calculus of variations and existing solar cell models to minimise the total entropy generation rate in the cell to find the most optimal design. The variational method is illustrated by applying it to a homojunction solar cell. The optimisation results in a set of differential algebraic equations, which determine the optimal shape of the doping profile for given recombination and transport models.

4.1 Introduction

Following the analysis of the efficiency potential of single-junction solar cells by Shockley and Queisser [1], the upper limit on the efficiency of solar cells has been determined in literature under various idealised assumptions (see, e.g., [2, 3, 4]). For many practical types of solar cells, however, it is still unclear how to realise the highest possible efficiency because of the presence of specific (non-ideal) loss mechanisms that limit the efficiency potential of the cell. For example, in solar cells based on crystalline silicon (*c*-Si), Auger recombination and surface recombination at the metal contacts limit the attainable efficiency

considerably, whereas in an ideal solar cell only radiative recombination would occur. Furthermore, these non-ideal loss mechanisms often have complex interactions in the cell design, making it difficult to find the optimal trade-off between them. In homojunction cells, for example, the high doping densities in the emitter and back surface field (BSF) function to limit the recombination at the metal contact, but simultaneously give rise to increased Auger recombination.

To aid in the practical design of these kinds of non-ideal solar cells, detailed models have been developed to predict their efficiency. Programs like PC1D [5], AFORS-HET [6], and ASA [7] are all *efficiency-predictive* in this sense. Efficiency-predictive simulations are commonly used in parametric optimisation studies where the simulation results of the solar cell design are used to improve the design iteratively and provide a cost efficient alternative for experimental trials. However, the parameter space for such optimisation problems can become very large if many different parameters are optimised simultaneously. For example, to optimise the doping profile $\rho(x)$ of a homojunction cell in complete generality, it is in principle possible to parametrise the function $\rho(x)$ using a spline function with a high density of control nodes, but this will lead to an unfeasibly large parameter space.

The goal of this work is to introduce the *design-predictive* method of Variational Entropy Generation Minimisation (VEGM), which aims to circumvent iterative optimisation altogether. The VEGM method extends the scope of current efficiency-predictive methods by making it possible to directly predict the optimal solar cell design for a given set of physical constraints and non-ideal loss mechanisms that depend on the materials used. The main difference with efficiency-predictive methods is in the figure of merit used to optimise the solar cell. Iterative methods based on efficiency-predictive models optimise the solar cell based on figures of merit such as the output power P_{Out} , open circuit voltage V_{OC} , and short circuit current J_{SC} . These figures of merit are determined on the boundary of the solar cell and it is therefore often non-trivial to understand how changes to the design in the *bulk* of the cell will affect P_{Out} , V_{OC} and J_{SC} on the *boundary*. The key point of the new approach is to de-emphasise the importance the usual boundary figures of merit and instead focus on the total rate of entropy production Q_S as a new bulk figure of merit. Unlike P_{Out} , V_{OC} and J_{SC} , the total rate of entropy production Q_S is an integral over the bulk of the cell and thus combines local information from the whole cell. Therefore, Q_S can be optimised using variational methods to immediately predict the most efficient cell given a set of geometrical and physical constraints. In this work the Euler-Lagrange (EL) method will be used to minimise Q_S , but in principle any variational optimisation method can be used.

The type of solar cell considered here is a device that converts blackbody radiation from the sun into electrochemical energy (in the form of electron-

hole pairs) and then into work (in the form of an electric current) [8]. Like any other device that converts heat into work, the solar cell produces entropy under operation and the more entropy it produces, the more energy is lost in the form of low-grade heat. Therefore, the problem of maximising P_{Out} is equivalent to that of minimizing Q_S , which is an integral of the local entropy production rate q_s over the domain of the cell Ω :

$$Q_S = \int_{\Omega} q_s \, dV. \quad (4.1)$$

It is important to note that Q_S should be minimised under the constraints that the solar cell is operating at maximum power conditions under illumination, since the absolute minimum entropy generation rate is trivially $Q_S = 0$ in the dark at thermodynamic equilibrium.

Eq. (4.1) shows that q_s is the central quantity that is required for the variational optimisation of a solar cell. The expression for q_s will be derived in the following paragraphs using non-equilibrium thermodynamics, which is a field that has not found widespread use in the solar cell community. For this reason the method of De Groot and Mazur [9] is used, which is conceptually simple and can be related to traditional semiconductor physics (as commonly practised in the field of photovoltaics) with relatively little effort. A general background of this method for non-equilibrium thermodynamics can also be found in the books of, e.g., Kuiken [10], Le Bellac [11], or Kjelstrup [12]. General thermodynamic treatments of semiconductors physics can be found in the works of, e.g., Lindefelt [13] and Parrott [14].

4.2 Entropy production in solar cells

To derive q_s , three components from the theory of non-equilibrium thermodynamics are required: the first law of thermodynamics; the assumption of local thermodynamic equilibrium (LTE); and the continuity equations for the extensive variables. The derivation for q_s presented below is based on the aforementioned literature sources and has been reduced to the essentials as much as possible for the sake of accessibility.

For a solar cell in global equilibrium, the relevant thermodynamic quantities are: entropy S ; internal energy U ; temperature T ; volume V ; the electron/hole electrochemical potentials $\eta_{n,p}$; and finally the total number of electrons and holes $N_{n,p}$. The electrochemical potentials are defined as:

$$\eta_n = \mu_n^c - \phi, \quad (4.2a)$$

$$\eta_p = \mu_p^c + \phi, \quad (4.2b)$$

with $\mu_{n,p}^c$ the chemical potentials for the electrons and holes, and ϕ the electrostatic energy. Because of the fundamental coupling between charge and

particle, η_n and η_p are the relevant thermodynamic potentials for the electrons and holes and there is no need to consider ϕ and $\mu_{n,p}^c$ separately [8]. Related to $\eta_{n,p}$ are the more commonly used quasi-Fermi energies $E_{Fn,p}$, given by $\eta_n = E_{Fn}$ and $\eta_p = -E_{Fp}$. It is assumed that pressure is not an important thermodynamic quantity for a solar cell because the kinetic energy and momentum of the electrons/holes do not play a significant role in the operation of the device.

It should be noted that there is also entropy associated with the angular distribution of the radiation field and the consideration of this distribution can therefore also help to improve solar cells [15] and in principle it is possible to minimise the entropy generation rate of the total system consisting of the solar cell and the radiation field taken together. For this work, however, only the entropy balance of the solar cell itself will be considered and the radiation field will be regarded as a source term for the processes in the semiconductor. This simplifying approximation is well-suited for material systems where radiative recombination is not the main loss mechanism, such as *c*-Si [13].

With the quantities identified in the previous paragraph, the first law of thermodynamics for a solar cell in global equilibrium can be stated in terms of S :

$$dS = \frac{1}{T} dU + \frac{-E_{Fn}}{T} dN_n + \frac{E_{Fp}}{T} dN_p. \quad (4.3)$$

Observe that in this formulation, U , N_n , and N_p are considered to be the independent thermodynamic variables and S the fundamental potential, so $S = S(U, N_n, N_p)$. To generalise Eq. (4.3) to (non-equilibrium) operating conditions, LTE is assumed. Under this assumption, the system is divided into finite volumes that are small enough to reach equilibrium with their surroundings on a macroscopic timescale, yet microscopically large so that they can be accurately described by thermodynamic variables. It is furthermore assumed that the functional form of Eq. (4.3) remains valid locally. The local form of Eq. (4.3) is stated using volumetric densities of the extensive variables, which are indicated by a lower case letter: entropy density s , internal energy density u , and the electron/hole densities n , p . These quantities are all dependent on time t and position x , but these dependencies will not be stated explicitly in equations unless confusion can arise. The local form of Eq. (4.3) is therefore:

$$ds = \frac{1}{T} du + \frac{-E_{Fn}}{T} dn + \frac{E_{Fp}}{T} dp. \quad (4.4)$$

The densities s , u , n and p satisfy local continuity equations. Let $q_{s,u,n,p}$ and $\mathbf{J}_{s,u,n,p}$ denote the local sources and currents of s , u , n , and p respectively.

Then:

$$\frac{\partial s}{\partial t} + \nabla \cdot \mathbf{J}_s = q_s, \quad (4.5a)$$

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{J}_u = q_u, \quad (4.5b)$$

$$\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{J}_n = q_n = G - R, \quad (4.5c)$$

$$\frac{\partial p}{\partial t} + \nabla \cdot \mathbf{J}_p = q_p = G - R. \quad (4.5d)$$

Here, G is the generation rate and R the recombination rate of electron-hole pairs. Note that $\mathbf{J}_{n,p}$ are defined as particle currents (unit: $\text{cm}^{-2}\text{s}^{-1}$) and that the currents $\mathbf{J}_{s,u,n,p}$ are considered in the barycentric frame of the cell (so $\mathbf{J}_{s,u,n,p} = 0$ under global equilibrium conditions). To find q_s , first the t derivative of Eq. (4.4) is taken:

$$\frac{\partial s}{\partial t} = \frac{1}{T} \frac{\partial u}{\partial t} + \frac{-E_{Fn}}{T} \frac{\partial n}{\partial t} + \frac{E_{Fp}}{T} \frac{\partial p}{\partial t} \quad (4.6)$$

From Eqs. (4.5) and (4.6) the time derivatives can be eliminated. The entropy current \mathbf{J}_s can then be eliminated by using the following relation for the currents:¹

$$\mathbf{J}_s = \frac{1}{T} \mathbf{J}_u + \frac{-E_{Fn}}{T} \mathbf{J}_n + \frac{E_{Fp}}{T} \mathbf{J}_p. \quad (4.7)$$

After rearranging, q_s is then obtained:

$$q_s = \frac{1}{T} q_u + \frac{E_{Fn} - E_{Fp}}{T} (R - G) + \mathbf{F}_u \cdot \mathbf{J}_u + \mathbf{F}_n \cdot \mathbf{J}_n + \mathbf{F}_p \cdot \mathbf{J}_p, \quad (4.8)$$

In Eq. (4.8), the forces $\mathbf{F}_{u,n,p}$ are defined by $\mathbf{F}_u = \nabla(1/T)$, $\mathbf{F}_n = \nabla(-E_{Fn}/T)$, and $\mathbf{F}_p = \nabla(E_{Fp}/T)$. The expression for q_s derived here agrees with expressions for power loss in solar cells derived in literature previously [16, 17], but also includes thermal effects through the term $\mathbf{F}_u \cdot \mathbf{J}_u$, which includes entropy production due to transport of heat. The heat flux \mathbf{J}_Q is given by

$$\mathbf{J}_u = \mathbf{J}_Q + E_{Fn} \mathbf{J}_n - E_{Fp} \mathbf{J}_p, \quad (4.9)$$

and can be used to rewrite Eq. (4.8) as:

$$Tq_s = q_u + (E_{Fn} - E_{Fp})(R - G) - \frac{\nabla T}{T} \cdot \mathbf{J}_Q - \nabla E_{Fn} \cdot \mathbf{J}_n + \nabla E_{Fp} \cdot \mathbf{J}_p. \quad (4.10)$$

¹It should be noted that different arguments to obtain this expression for \mathbf{J}_s can be found in literature. We adopt the view that the First Law in the form of Eq. (4.4) determines the relationship between the currents given in Eq. (4.7).

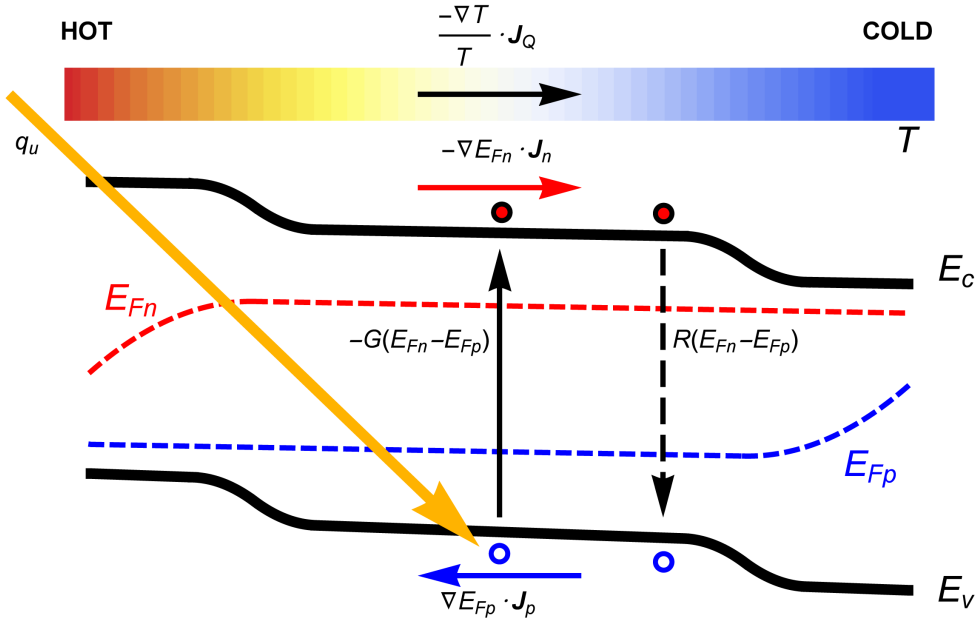


Figure 4.1: Schematic band diagram of an illuminated solar cell that illustrates the physical processes corresponding to the terms in the entropy generation rate as given by Eq. (4.10). It is assumed that the temperature in the cell is higher on the side facing the light, as indicated by the color gradient.

Eq. (4.10) gives the source term for the entropy balance in the solar cell. The first two terms correspond to the transfer of entropy from the radiation field to the solar cell; electron-hole recombination; and the creation of entropy due to the non-radiative relaxation of electron-hole pairs immediately after the absorption of a photon. The last three terms in Eq. (4.10) are entropy production terms that correspond to irreversible transport phenomena. An overview of the physical meaning of the terms in Eq. (4.10) is given in Figure 4.1.

The $\mathbf{F}_{u,n,p}$ introduced in Eq. (4.8) are referred to as forces because they drive the currents $\mathbf{J}_{u,n,p}$. According to Onsager theory [18] the most general relation (for an isotropic medium and linear transport phenomena) between the currents and forces takes the form:

$$\begin{pmatrix} \mathbf{J}_u \\ \mathbf{J}_n \\ \mathbf{J}_p \end{pmatrix} = \begin{pmatrix} L_{uu} & L_{un} & L_{up} \\ L_{nu} & L_{nn} & L_{np} \\ L_{pu} & L_{pn} & L_{pp} \end{pmatrix} \begin{pmatrix} \mathbf{F}_u \\ \mathbf{F}_n \\ \mathbf{F}_p \end{pmatrix}. \quad (4.11)$$

The L_{ij} in Eq. (4.11) are the transport coefficients for the solar cell. Onsager theory states that the matrix L is symmetric (due to microscopic reversibility) and positive-definite (due to the fact entropy cannot spontaneously decrease).

In traditional solar cell modeling, only the transport coefficients L_{nn} and L_{pp} (which are proportional to the electron and hole conductivities) are commonly used, while the remaining transport coefficients are implicitly assumed to be negligible.

In most solar cells, recombination of electron-hole pairs at the boundary $\partial\Omega$ of the cell will also contribute to the total rate of entropy generation. The surface entropy production rate $Q_{S,\text{surf}}$ should be added to the bulk contribution and can be calculated by:

$$Q_{S,\text{surf}} = \int_{\partial\Omega} \frac{E_{Fn} - E_{Fp}}{T} R_{\text{surf}} \, dA. \quad (4.12)$$

In Eq. (4.12), R_{surf} is the surface recombination rate (in $\text{cm}^{-2} \text{s}^{-1}$).

Eqs. (4.8) and (4.10) can be used directly to gain additional insight into simulation results from common solar cell simulation tools such as AFORS-HET. In order to plot power losses due to recombination and ohmic dissipation in a solar cell, it is useful to define the dissipative part of q_s , which will be denoted by q_{diss} :

$$Tq_{\text{diss}} = (E_{Fn} - E_{Fp})R - \frac{\nabla T}{T} \cdot \mathbf{J}_Q - \nabla E_{Fn} \cdot \mathbf{J}_n + \nabla E_{Fp} \cdot \mathbf{J}_p. \quad (4.13)$$

Because q_{diss} is always positive, it can be conveniently plotted as a color map on top of a band diagram calculation to show where the greatest losses of free energy are occurring. This is illustrated in Figure 4.2. However, it is important to remark that in order to optimise the solar cell completely, it is necessary to not only minimise the total dissipation but to also make sure that the generation of electron-hole pairs takes place in regions of the cell where the Fermi level splitting is greatest, as can be seen from the term $-(E_{Fn} - E_{Fp})G$ in Eq. (4.10).

It is interesting to note that for a 1 dimensional solar cell at constant temperature and under steady-state operation, the relation between the total entropy generation rate Q_S and output power P_{Out} can be obtained by direct integration. It is assumed that the electrons are extracted at $x = 0$ and the holes at $x = d$. Let $e\Phi = (E_{Fn}(0) - E_{Fp}(d))$ (with e the elementary charge) be the voltage over the cell and $J_0 = e(J_p - J_n)$ be the total charge current (which is constant throughout the cell). Then $P_{\text{Out}} = \Phi J_0$. The total solar energy absorption in the solar cell will be denoted by $Q_U = \int_0^d q_u \, dx$. The recombination rate at the electron contact equals $R_{\text{surf}}(0) = -J_p(0)$ and that at the hole contact $R_{\text{surf}}(d) = J_n(d)$. Using Eqs. (4.5), (4.8), and (4.12), Q_S

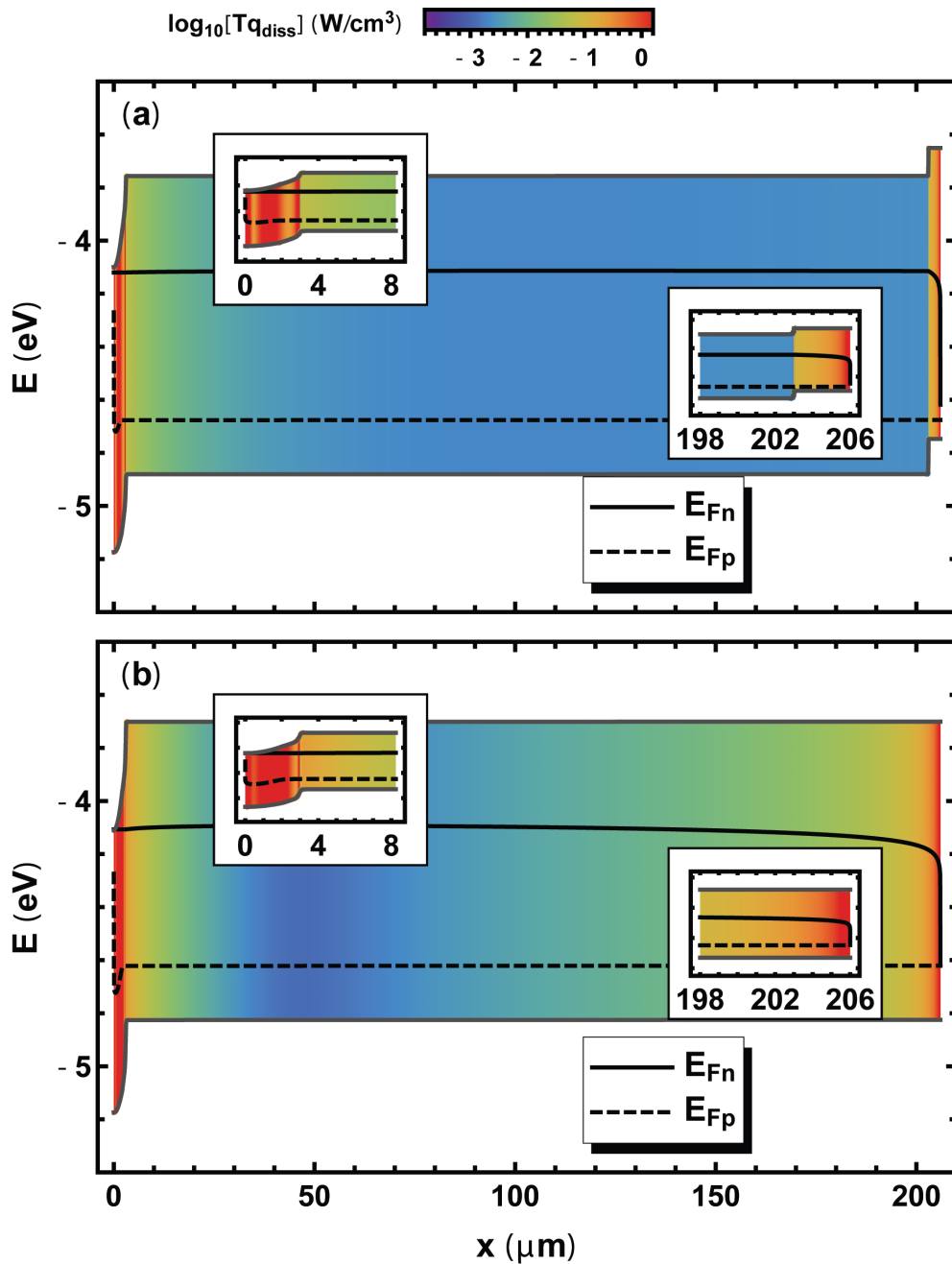


Figure 4.2: Band diagram calculations (performed in AFORS-HET) of a p -type c -Si homojunction solar cell (a) with backside diffusion (BSF) and (b) without BSF. The entropy production rate is shown using a color map. The insets show magnifications near the surface of the cell. The BC $E_{Fn} = E_{Fp}$ was realised by choosing a sufficiently high surface recombination velocity.

is then given by:

$$\begin{aligned}
TQ_S &= \int_0^d Tq_s \, dx + TQ_{S,\text{surf}} \\
&= \int_0^d (q_u - E_{Fn}J'_n + E_{Fp}J'_p - E'_{Fn}J_n + E'_{Fp}J_p) \, dx \\
&\quad + \sum_{x=\{0,d\}} [E_{Fn}(x) - E_{Fp}(x)]R_{\text{surf}}(x) \\
&= Q_U - \Phi J_0 = Q_U - P_{\text{Out}}.
\end{aligned} \tag{4.14}$$

Eq. (4.14) demonstrates that the minimisation of Q_S indeed leads to the maximisation of P_{Out} , provided that Q_U has been given (e.g., from optical modeling) or fixed in some other fashion. This again demonstrates that some care should be taken when minimizing entropy production, since a cell in equilibrium in the dark trivially produces no entropy.

4.3 Variational solar cell optimisation

With Eqs. (4.8) and (4.11), the total entropy generation rate Q_S can now be considered as a functional that depends on all relevant solar cell variables such as n , p , ϕ etc: $Q_S = Q_S[n(x), p(x), \phi(x), \dots]$. This makes it possible to minimise Q_S using variational methods such as the EL method. There are, however, physical laws that need to be satisfied by the solution, so these laws need to be added as constraints to the optimisation procedure. In addition, other practical constraints can be imposed as well. Hence, the VEGM method consists of four steps:

1. The relevant unknowns are identified as well as the physical equations that need to be satisfied.
2. The Lagrange multiplier method is used to construct a functional Λ which, when minimised, will lead to a solution that simultaneously minimises Q_S and satisfies all physical equations and additional constraints.
3. The EL method is used to find the set of differential equations for the stationarity of Λ .
4. The obtained set of equations is solved using a numerical method.

In the next paragraphs, the first three steps of the VEGM method will be illustrated by applying it to a homojunction solar cell to optimise the doping profile $\rho(x)$ (expressed in cm^{-3}). The solar cell will be modelled under the assumptions commonly used in literature. This means that the device is considered at constant temperature T (so $\mathbf{F}_u = 0$) and under steady-state operation. It

is assumed that the boundary of the cell $\partial\Omega$ consists of perfectly passivated surface (meaning that $R_{\text{surf}} = 0$) and metal contacts. At the metal contacts the boundary condition (BC) $E_{Fn} = E_{Fp}$ is assumed since Fermi level splitting is not possible in a metal. With these BCs, $Q_{S,\text{surf}} = 0$ in Eq. (4.12) and only bulk contributions to Q_S need to be considered. This does, however, not mean that recombination at the metal contacts does not generate entropy, but only that the entropy is generated in the bulk rather than at the contact (as can be seen in Figure 4.2).

To obtain a transport model equivalent to ordinary semiconductor physics, the off-diagonal transport coefficients in Eq. (4.11) $L_{un} = L_{nu} = 0$, $L_{up} = L_{pu} = 0$ and $L_{np} = L_{pn} = 0$ are assumed negligible. Taken together, these assumptions amount to neglecting the Seebeck and Peltier effects as well as off-diagonal transport effects between electrons and holes. From a microscopic point of view, the assumption $L_{np} = 0$ means that electron-hole collisions are neglected [13]. The diagonal coefficients for charge transport are $L_{nn} = Tn\mu_n/e$ and $L_{pp} = Tp\mu_p/e$ (with $\mu_{n,p}$ the electron/hole mobilities). The mobilities $\mu_{n,p} = \mu_{n,p}[\rho(x)]$ are assumed to depend on the doping density ρ . The local absorption of solar heat q_u is assumed known from optical modeling, as well as the generation rate $G = G(x)$. For recombination model R the general form

$$R = R(x, n, p, \rho), \quad (4.15)$$

is used, indicating that R depends on n , p , and ρ as well as on x explicitly (due to, e.g., a position-dependent defect density). Furthermore, Boltzmann statistics will be assumed, so $E_{Fn,p}$ are related to n , p , and ϕ by:

$$n = N_C \exp\left(\frac{E_{Fn} + \chi + \phi}{kT}\right), \quad (4.16)$$

$$p = N_V \exp\left(-\frac{E_{Fp} + \chi + E_G + \phi}{kT}\right), \quad (4.17)$$

with $N_{C,V}$ the effective densities of states for the conduction and valence bands, χ and E_G the electron affinity and band gap (all assumed to be constant), and k the Boltzmann constant.

Next, the set of physical functions Y and transport equations Ξ that describe the solar cell are specified. The set of functions Y that will be used is given by: $Y = \{n, p, \phi, \rho\}$. If the equations are stated as $\xi_i = 0$, the set of equations $\Xi = \{\xi_i\}$ is:

$$\begin{aligned} \xi_1 &= \nabla \cdot \left(-\frac{\mu_n n}{e} \nabla E_{Fn} \right) - (G - R), \\ \xi_2 &= \nabla \cdot \left(\frac{\mu_p p}{e} \nabla E_{Fp} \right) - (G - R), \\ \xi_3 &= \nabla^2 \phi + \frac{e^2}{\epsilon} (p - n + \rho). \end{aligned}$$

Here, $\xi_{1,2}$ are the drift-diffusion equations and ξ_3 is the Poisson equation. Because ρ is considered to be an unknown, the set Ξ consists of three equations in the four unknowns Y and is therefore underdetermined. However, the requirement that the solution minimises Q_S is used to make the problem well-posed.

To minimise Q_S subject to the constraints set by Eqs. Ξ , the following functional Λ is minimised:

$$\begin{aligned}\Lambda[Y, \{\lambda_i\}] &= \int \left(Tq_s + \sum_{i=1}^3 \lambda_i(x) \xi_i \right) dV \\ &= \int \left(q_u - (E_{Fn} - E_{Fp})(G - R) - \nabla E_{Fn} \cdot \mathbf{J}_n \right. \\ &\quad \left. + \nabla E_{Fp} \cdot \mathbf{J}_p + \sum_{i=1}^3 \lambda_i(x) \xi_i \right) dV\end{aligned}\quad (4.18)$$

The functional Λ depends on the four functions Y as well as the three scalar Lagrange multiplier functions $\lambda_i(x)$, which ensure that Eqs. Ξ are satisfied for all $x \in \Omega$.

Because Λ depends on 7 scalar functions, there are also 7 EL equations that correspond to the stationarity requirement $\delta\Lambda = 0$. The 3 EL equations for the λ_i reproduce the 3 Eqs. Ξ , while the EL equations for the 4 functions Y generate 4 new equations. The latter set of 4 equations can be simplified by making the substitutions $\lambda_1 = \hat{\lambda}_1 - E_{Fn}$ and $\lambda_2 = \hat{\lambda}_2 + E_{Fp}$. The resulting 4 differential algebraic equations (DAEs) are:

$$\nabla \cdot \left(\frac{\mu_n n}{e} \nabla \hat{\lambda}_1 \right) = n \frac{\partial R}{\partial n} \frac{\hat{\lambda}_1 + \hat{\lambda}_2}{kT} - \frac{\mathbf{J}_n \cdot \nabla \hat{\lambda}_1}{kT} - \frac{e^2 n \lambda_3}{\epsilon kT}, \quad (4.19a)$$

$$\nabla \cdot \left(\frac{\mu_p p}{e} \nabla \hat{\lambda}_2 \right) = p \frac{\partial R}{\partial p} \frac{\hat{\lambda}_1 + \hat{\lambda}_2}{kT} - \frac{\mathbf{J}_p \cdot \nabla \hat{\lambda}_2}{kT} + \frac{e^2 p \lambda_3}{\epsilon kT}, \quad (4.19b)$$

$$\nabla \cdot (\mu_n n \nabla \hat{\lambda}_1 - \mu_p p \nabla \hat{\lambda}_2 + e \nabla \lambda_3) = 0, \quad (4.19c)$$

$$\frac{e^2}{\epsilon} \lambda_3 + \frac{\partial R}{\partial \rho} (\hat{\lambda}_1 + \hat{\lambda}_2) = \frac{\mathbf{J}_n \cdot \nabla \hat{\lambda}_1}{\mu_n} \frac{\partial \mu_n}{\partial \rho} + \frac{\mathbf{J}_p \cdot \nabla \hat{\lambda}_2}{\mu_p} \frac{\partial \mu_p}{\partial \rho}. \quad (4.19d)$$

The BCs for the $\lambda_{1,2,3}$ are obtained by considering the natural BCs that make Λ stationary (see, e.g., Wan [19] and Appendix 4.A at the end of this chapter). The natural BCs depend on the BCs that have been imposed on the variables Y as well as on $Q_{S,\text{surf}}$. For example, if n and p are fixed at the boundaries (i.e., Dirichlet BCs) and $Q_{S,\text{surf}} = 0$ (as has been assumed here), the natural BCs require that $\lambda_1 = \lambda_2 = 0$ on $\partial\Omega$ (note that these BCs are stated in terms of $\lambda_{1,2}$ rather than $\hat{\lambda}_{1,2}$). The BCs for λ_3 depend on those of ϕ : on boundaries with a Dirichlet BC for ϕ the BC $\lambda_3 = 0$ applies, while on boundaries with a Neumann BC on ϕ the BC $\hat{\mathbf{n}} \cdot \nabla \lambda_3 = \hat{\mathbf{n}} \cdot [2(\mathbf{J}_n - \mathbf{J}_p) + (\mu_p p/e) \nabla \lambda_2 - (\mu_n n/e) \nabla \lambda_1]$ applies (with $\hat{\mathbf{n}}$ the outward unit normal vector on $\partial\Omega$).

There are 7 EL equations in 7 the functions Y and $\lambda_{1,2,3}$, so the system of DAEs can be solved by a numerical method to obtain the optimal shape for ρ . This illustrates the trade-off between parametric optimisation and the variational method: in the VEGM method the set of differential equations is larger and has to be derived for each particular optimisation problem, but the numerical solution has to be computed only once and the shape functions such as ρ are optimised in complete generality rather than parametrically.

In conclusion, the framework of non-equilibrium thermodynamics offers a useful perspective for the analysis of local power losses in photovoltaic devices. By considering the entropy generation rate it becomes possible to take into account electrical, optical and thermal losses and compare them on a uniform scale. The total entropy generation rate can be regarded as a functional, making it possible to optimise solar cells by taking advantage of the power of the calculus of variations. The design-predictive method of Variational Entropy Generation Minimisation can take advantage of detailed models for, e.g., charge carrier recombination and mobility reduction due to semiconductor doping and use these models to predict, e.g., the optimal doping profile for a homojunction solar cell. In short, the VEGM method should be considered as an extension of the field of numerical solar cell modeling that makes it possible to find the optimal trade-off between the many complex loss mechanisms that occur in real solar cells.

Acknowledgements

This work was financially supported by the Dutch Technology Foundation STW (Flash Perspectief Programma). The research of W.M.M. Kessels has been made possible by the Dutch Technology Foundation STW and the Netherlands Organisation for scientific Research (NWO, VICI programma). Furthermore, we would like to thank Dr. Stefan Bordihn for his constructive comments on this work.

Appendix

4.A Boundary conditions for the VEGM method

Consider the homojunction solar cell in 1 dimension (1D) at constant T for simplicity. The domain of the cell is the interval $\Omega = [0, d]$ and the boundary of the cell consists of the two points $\partial\Omega = \{0, d\}$. The cell is described by the set of physical functions $Y = \{n, p, \phi, \rho\}$ and transport equations $\Xi = \{\xi_i\}$. This time the effect of the surface entropy generation will be accounted for in generality, so $Q_{S,\text{surf}} \neq 0$. The functional Λ then becomes:

$$\begin{aligned}
 \Lambda[Y, \{\lambda_i\}] &= \int_0^d \left(Tq_s + \sum_{i=1}^3 \lambda_i(x)\xi_i \right) dx + TQ_{S,\text{surf}} \\
 &= \int_0^d \left(q_u - (E_{Fn} - E_{Fp})(G - R) \right. \\
 &\quad \left. - E'_{Fn}J_n + E'_{Fp}J_p + \sum_{i=1}^3 \lambda_i(x)\xi_i \right) dx \\
 &\quad + \sum_{x \in \partial\Omega} [E_{Fn}(x) - E_{Fp}(x)]R_{\text{surf}}(x) \\
 &= \int_0^d L(x) dx + \sum_{x \in \partial\Omega} L_B(x). \tag{4.20}
 \end{aligned}$$

In Eq. (4.20) the symbols L and L_B have been introduced. The part inside of the integral L is often referred to as the Lagrangian of the problem. The boundary term L_B will be called the boundary Lagrangian here, but sometimes it is also referred to as *salvage value* or *terminal payoff* (these terms are more common in variational problems in economics).

To obtain the proper BCs for the variational problem $\delta\Lambda = 0$, it is necessary to consider both the first order variation of L as well as that of L_B . Calculating

this variation gives:

$$\begin{aligned} \delta\Lambda &= \int_0^d \left(\sum_y \frac{\partial L}{\partial y} \delta y + \frac{\partial L}{\partial y'} \delta y' + \frac{\partial L}{\partial y''} \delta y'' + \dots \right) dx \\ &+ \sum_{x \in \partial\Omega} \sum_y \left(\frac{\partial L_B}{\partial y} \delta y + \frac{\partial L_B}{\partial y'} \delta y' + \frac{\partial L_B}{\partial y''} \delta y'' + \dots \right). \end{aligned} \quad (4.21)$$

The sum over y runs over all functions Y as well as the λ_i . Note that $\delta(y') = (\delta y)' = \delta y'$ since the functional differential δ commutes with ordinary derivatives such as d/dx . The EL equations are obtained by repeated application of partial integration on the integral in Eq. (4.21) until no derivatives of the first order variations (i.e., $\delta y'$, $\delta y''$ etc.) appear under the integral anymore:

$$\begin{aligned} \delta\Lambda &= \int_0^d \sum_y \left(\frac{\partial L}{\partial y} - \left(\frac{\partial L}{\partial y'} \right)' + \left(\frac{\partial L}{\partial y''} \right)'' - \dots \right) \delta y \, dx \\ &+ \sum_{x \in \partial\Omega} \sum_y \left\{ \left[\hat{\mathbf{n}} \left(\frac{\partial L}{\partial y'} - \left(\frac{\partial L}{\partial y''} \right)' \right) + \frac{\partial L_B}{\partial y} \right] \delta y \right. \\ &\left. + \left[\hat{\mathbf{n}} \frac{\partial L}{\partial y''} + \frac{\partial L_B}{\partial y'} \right] \delta y' + \dots \right\} \end{aligned} \quad (4.22)$$

Here, $\hat{\mathbf{n}}$ is the 1D unit outward normal vector defined by $\hat{\mathbf{n}}(0) = -1$, $\hat{\mathbf{n}}(d) = 1$. Note that no assumptions have been made on the boundary values of δy and its derivatives yet. The requirement $\delta\Lambda = 0$ has consequence for both the part under the integral and for the boundary terms. Under the integral, the first order variation δy can be any infinitesimal function, so in order for the integral to be zero, the prefactor of δy has to be zero for every point x . Assuming that L does not depend on higher derivatives than y'' , the EL equations are:

$$\frac{d^2}{dx^2} \left(\frac{\partial L}{\partial y''} \right) - \frac{d}{dx} \left(\frac{\partial L}{\partial y'} \right) + \frac{\partial L}{\partial y} = 0. \quad (4.23)$$

The EL Eq. (4.23) is a necessary condition for the stationarity of Λ , but it is not sufficient since the boundary terms in Eq. (4.22) must be zero as well. On $\partial\Omega$, all variations δy and $\delta y'$ are independent of each other. This means that for every term one has to choose between setting the variation equal to zero (i.e. picking a BC for y or y') or setting the prefactor of that variation to zero. The latter choice results in a so-called natural BC.

To illustrate this, consider the boundary terms involving δn and $\delta n'$ on the boundary $x = d$. It is assumed that this is the hole-collecting boundary and therefore $R_{\text{surf}}(d) = \hat{\mathbf{n}} J_n(d)$. Because the variations in n and n' are

independent, each of the terms involving $\delta n(d)$ and $\delta n'(d)$ have to be zero, so:

$$\left(E'_{Fn} + \lambda'_1 + \frac{(\lambda_1 + E_{Fn} - E_{Fp})\phi'}{kT} \right) \delta n(d) = 0, \quad (4.24a)$$

$$(\lambda_1 + E_{Fn} - E_{Fp}) \delta n'(d) = 0. \quad (4.24b)$$

Both Eqs. (4.24a) and (4.24b) have to be satisfied, but $\delta n(d)$ and $\delta n'(d)$ cannot be zero simultaneously since the transport equation for n is second order. This means that a choice has to be made between imposing either a Dirichlet BC for n (so $\delta n(d) = 0$) or a Neumann BC ($\delta n'(d) = 0$). A Dirichlet BC for n leads to a natural BC for λ_1 specified by Eq. (4.24b) while a Neumann BC leads to a natural BC for λ_1 specified by Eq. (4.24a). The BCs for λ_3 are similarly obtained by considering the boundary terms involving $\delta\phi$ and $\delta\phi'$

In solar cell modeling it is often desirable to consider more general types of BC than just Dirichlet and Neumann BCs. Such general BCs can be employed in the VEGM method by the use of Lagrange multipliers. For example, the recombination at the metal contact can be described by an effective surface recombination velocity S_{eff} : $R_{surf} = S_{eff}\Delta n$. Here, Δn is the injection level, which is approximately n at the hole-collecting contact. The desired BC for electron transport is therefore: $J_n(d) = \hat{n}S_{eff}n(d)$. This mixed BC cannot be used directly in the variational approach since it makes the variations in $\delta n(d)$ and $\delta n'(d)$ interdependent. To remedy this, the BC is enforced with a (scalar) Lagrange multiplier λ_n by adding the term $\lambda_n[J_n(d) - \hat{n}S_{eff}n(d)]$ to the boundary Lagrangian L_B . After that, Λ is minimised with respect to λ_n in addition to all the other variables, thus ensuring that the BC $J_n(d) = \hat{n}S_{eff}n(d)$ is satisfied by the solution. The boundary terms for $\delta n(d)$, $\delta n'(d)$ and $\delta\lambda_n$ are then:

$$\left(\left[\frac{\phi'}{kT} - \hat{n} \frac{eS_{eff}}{kT\mu_n} \right] \lambda_n + E'_{Fn} + \lambda'_1 + \frac{(\lambda_1 + E_{Fn} - E_{Fp})\phi'}{kT} \right) \delta n(d) = 0 \quad (4.25)$$

$$(\lambda_1 + \lambda_n + E_{Fn} - E_{Fp}) \delta n'(d) = 0, \quad (4.26)$$

$$(J_n - \hat{n}S_{eff}n) \delta\lambda_n = 0. \quad (4.27)$$

In this case natural BCs have to be used to satisfy Eqs. (4.25-4.27), since $\delta n(d)$, $\delta n'(d)$ and $\delta\lambda_n$ are all nonzero. From Eqs. (4.25) and (4.26) the multiplier λ_n can be eliminated to obtain the (mixed) BC for λ_1 .

References Chapter 4

- [1] William Shockley and Hans J. Queisser. Detailed Balance Limit of Efficiency of p-n Junction Solar Cells. *J. Appl. Phys.*, 32(3):510, 1961. ISSN 00218979. doi: 10.1063/1.1736034.
- [2] C. H. Henry. Limiting efficiencies of ideal single and multiple energy gap terrestrial solar cells. *J. Appl. Phys.*, 51(8):4494, 1980. ISSN 00218979. doi: 10.1063/1.328272.
- [3] A. De Vos and H. Pauwels. On the thermodynamic limit of photovoltaic energy conversion. *Appl. Phys.*, 25(2):119–125, June 1981. ISSN 0340-3793. doi: 10.1007/BF00901283.
- [4] Armin Richter, Martin Hermle, and Stefan W. Glunz. Reassessment of the Limiting Efficiency for Crystalline Silicon Solar Cells. *IEEE J. Photovolt.*, 3(4):1184–1191, oct 2013. ISSN 2156-3381. doi: 10.1109/JPHOTOV.2013.2270351.
- [5] D.A. Clugston and P.A. Basore. PC1D version 5: 32-bit solar cell modeling on personal computers. In *Conference Record of the Twenty Sixth IEEE Photovoltaic Specialists Conference - 1997*, pages 207–210. IEEE, sep 1997. ISBN 0-7803-3767-0. doi: 10.1109/PVSC.1997.654065.
- [6] R. Stangl, C. Leendertz, and J. Haschke. Numerical Simulation of Solar Cells and Solar Cell Characterization Methods: the Open-Source on Demand Program AFORS-HET. In Radu D Rugescu, editor, *Solar Energy*, pages 319–352. InTech, 2010. ISBN 978-953-307-052-0.
- [7] B. Pieters, J. Krc, and M. Zeman. Advanced Numerical Simulation Tool for Solar Cells - ASA5. In *2006 IEEE 4th World Conference on Photovoltaic Energy Conference*, pages 1513–1516. IEEE, 2006. ISBN 1-4244-0016-3. doi: 10.1109/WCPEC.2006.279758.
- [8] Peter Würfel. *Physics of Solar Cells*. Wiley-VCH Verlag GmbH & Co, Weinheim, 2005. ISBN 3-527-40428-7.

- [9] S.R. de Groot and Peter Mazur. *Non-equilibrium Thermodynamics*. Dover Publications, Mineola, New York, 1984. ISBN 0-486-64741-2.
- [10] Gerard D.C. Kuiken. *Thermodynamics of irreversible processes*. John Wiley & Sons, Ltd., Chichester, England, 1994. ISBN 0471948446.
- [11] Michel Le Bellac, Fabrice Mortessagne, and G. George Batrouni. *Equilibrium and Non-Equilibrium Statistical Thermodynamics*. Cambridge University Press, Cambridge, 2004. ISBN 0 521 82143 6.
- [12] Signe Kjelstrup and Dick Bedeaux. *Non-Equilibrium Thermodynamics of Heterogeneous Systems*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2008.
- [13] U. Lindefelt. Heat generation in semiconductor devices. *Journal of Applied Physics*, 75:942–957, 1994. ISSN 00218979. doi: 10.1063/1.356450.
- [14] J.E. Parrott. Thermodynamic theory of transport processes in semiconductors. *IEEE Transactions on Electron Devices*, 43, 1996. ISSN 0018-9383. doi: 10.1109/16.491259.
- [15] Albert Polman and Harry A Atwater. Photonic design principles for ultrahigh-efficiency photovoltaics. *Nature Materials*, 11(3):174–7, January 2012. ISSN 1476-1122. doi: 10.1038/nmat3263.
- [16] R. Brendel, S. Dreissigacker, N.-P. Harder, and P. P. Altermatt. Theory of analyzing free energy losses in solar cells. *Applied Physics Letters*, 93(17):173503, 2008. ISSN 00036951. doi: 10.1063/1.3006053.
- [17] Johannes Greulich, Hannes Höffler, Uli Würfel, and Stefan Rein. Numerical power balance and free energy loss analysis for solar cells including optical, thermodynamic, and electrical aspects. *Journal of Applied Physics*, 114(20):204504, 2013. ISSN 00218979. doi: 10.1063/1.4832777.
- [18] Lars Onsager. Reciprocal Relations in Irreversible Processes. I. *Physical Review*, 37(4):405–426, February 1931. ISSN 0031-899X. doi: 10.1103/PhysRev.37.405.
- [19] Frederick Y.M. Wan. *Introduction to the calculus of variations and its applications*. Chapman and Hall, 1995. ISBN 0-412-05141-9.

Chapter 5

Metal-oxide-based hole-selective tunnelling contacts for crystalline silicon solar cells

Abstract

The goal of this work is to investigate selective hole contacts for crystalline silicon solar cells that are highly transparent, passivate the silicon surface and have low contact resistance. Stacks of Al_2O_3 and ZnO films are suggested for this purpose. The charge transport mechanism through these stacks is tunnelling recombination and it is shown that such stacks can achieve a contact resistance of $\sim 1.5 \Omega \text{ cm}^2$ for an Al_2O_3 thickness of 1 nm. Furthermore, it is demonstrated that the surface passivation of such stacks can be greatly improved by the insertion of a 3 nm film of hydrogenated amorphous silicon ($a\text{-Si:H}$) between the Al_2O_3 and the crystalline silicon, achieving an effective surface recombination velocity of $\sim 20 \text{ cm s}^{-1}$. The stacks with an $a\text{-Si:H}$ layer achieve a contact resistance of $\sim 5 \Omega \text{ cm}^2$. Furthermore, from applying the theory of tunnel diodes to the charge transport through the contact, three important elements have been identified for the reduction of the contact resistance: the negative fixed charge density in the Al_2O_3 ; the doping concentration in the ZnO ; and the dielectric properties of the Al_2O_3 .

5.1 Introduction

In all solar cells that rely on the principle of generation of free electrons and holes, two key components can be identified: the photon absorber and the

selective contacts [1]. In the photon absorber, the energy of the light is converted into chemical energy in the form of electronhole pairs. This conversion is subject to the thermodynamic limit on efficiency, similarly to any other conversion of heat to a form of higher-grade energy. Next, the selective contacts in the cell force the electrons to exit the absorber along a different path than the holes, converting the chemical energy into work by creating an electric current. In principle, the efficiency of this second conversion is not subjected to thermodynamic limitations because no heat is involved. It is the conversion from chemical energy to work that is the focus of the research described in this paper.

The selective contacts in a traditional crystalline silicon (*c*-Si) solar cell are the diffused emitter and the back surface field (BSF). High *n*- and *p*-doping densities of these contacts result in an increase of the conductivity for one carrier type while the conductivity for the other carrier type is simultaneously reduced.

In silicon heterojunction (Si-HJT) solar cells, it is not just the doping in the deposited hydrogenated amorphous silicon (*a*-Si:H) layers that makes the emitter and rear of the cell selective to one type of carrier, but the band alignment at the *a*-Si:H/*c*-Si interface plays a significant role as well [2]. In fact, higher selectivity can be achieved than with diffused junctions by tuning the valence and conduction band offsets at the *a*-Si:H/*c*-Si interface, which leads to the well-known high V_{OC} values of Si-HJT cells [2]. One disadvantage of the *a*-Si:H layer is that they absorb a portion of the light without converting it into usable current, a phenomenon known as parasitic absorption [2].

The goal of our research is to investigate the possibility to create selective contacts by depositing stacks of selected materials on *c*-Si. The general boundary conditions under consideration in this work (other than high selectivity to either electrons or holes) are:

- (a) The interfaces between the layers and the *c*-Si should be well passivated to avoid surface recombination. An effective surface recombination velocity of 100 cm s^{-1} is considered the maximum feasible limit.
- (b) The contacts should be highly transparent. The light absorption of the contact should be less than the parasitic absorption in Si-HJT solar cells to be considered feasible.
- (c) The contact resistance should be low enough to compete with current technologies. Ultimately, the goal is to obtain a contact resistance on the order of $50 \text{ m}\Omega \text{ cm}^2$.

Other considerations are that the deposition methods used should be industrially scalable and the materials used should be abundant and preferably environmentally friendly.

With these requirements in mind, stacks of Al_2O_3 and ZnO films are suggested to ultimately achieve these goals. The choice for Al_2O_3 is motivated by the advances in recent years in silicon surface passivation by Al_2O_3 films, prepared by atomic layer deposition (ALD) or otherwise [3, 4], making Al_2O_3 a very promising material to fulfil requirement (a). The choice for ZnO is based on the combination of its high bandgap (and thus high transparency), low resistivity and natural abundance.

In this work, the focus is on three points:

- (i) To assess the potential of stacks of Al_2O_3 and ZnO films as passivating, transparent and selective contacts.
- (ii) To gain a theoretical understanding of charge transport through these stacks.
- (iii) To find ways to improve their performance.

Point (i) is addressed in Sections 5.3.1 and 5.4.1. Section 5.3.1 will demonstrate how the negative fixed charge density at the $\text{Al}_2\text{O}_3/c\text{-Si}$ interface [4] can be used to create a homojunction that separates electrons from holes. Section 5.4.1 will show the experimental results on the passivation of $c\text{-Si}$ by the $\text{Al}_2\text{O}_3/\text{ZnO}$ stacks.

Point (ii) is addressed in Sections 5.3.1, 5.3.2 and 5.4.2. In Section 5.3.1 the charge transport mechanism through the $\text{Al}_2\text{O}_3/\text{ZnO}$ stacks will be identified as tunnelling recombination and will be discussed in more depth in Section 5.3.2. Section 5.4.2 will show that the measured currents corroborate the theory presented in Section 5.3.2.

Point (iii) is addressed in Sections 5.3.3 and 5.4.3. Section 5.3.3 will identify from a theoretical perspective three key elements in the $\text{Al}_2\text{O}_3/\text{ZnO}$ stacks that influence the current through them and that can be optimised to reduce the tunnelling resistance. It will also discuss the potential of amorphous silicon interlayers to reduce tunnelling resistance and increase passivation. Section 5.4.3 demonstrates that the passivation of the stacks can be significantly improved by the insertion of an $a\text{-Si:H}$ interlayer between the $c\text{-Si}$ and Al_2O_3 . The $a\text{-Si:H}$ films that have been used are significantly thinner than those used in typical Si-HJT cells.

5.2 Materials and methods

Both the surface passivation and the tunnelling current investigations were performed on two types of stacks: $\text{Al}_2\text{O}_3/\text{ZnO:Al}$ (type A) and $a\text{-Si:H}/\text{Al}_2\text{O}_3/\text{ZnO:Al}$ (type B). These stacks are shown in Figure 5.1a.

For the passivation study, low resistivity ($\sim 3\text{--}4\ \Omega\text{ cm}$) n -type and p -type $c\text{-Si}$ wafers ($(281.0 \pm 1.3)\ \mu\text{m}$ thick) were used. The wafers were treated with

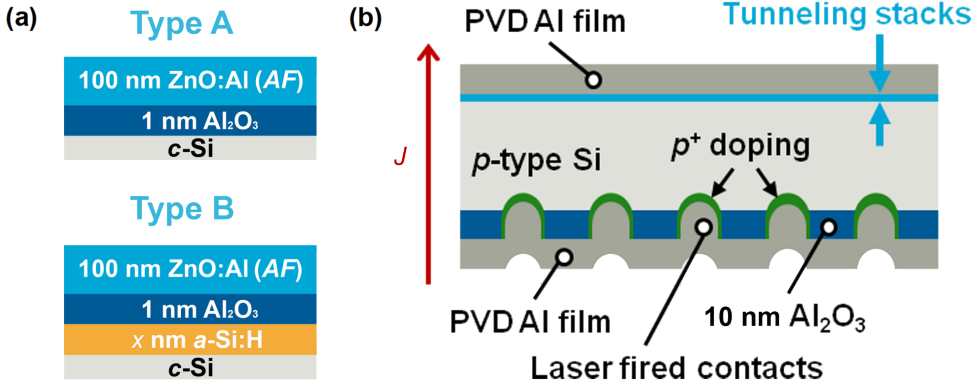


Figure 5.1: (a) The stack configurations of type A and type B. For the ZnO:Al films in stacks of type A, various aluminium fractions (AF) are considered. For stacks of type B, $AF = 0.10$. (b) Schematic cross sectional view of the samples used for J - V measurements.

a diluted HF (1% in DI- H_2O) solution right before carrying out the first film deposition. The surface passivation was always investigated on samples with the thin films deposited on both sides symmetrically.

The 1-nm-thick Al_2O_3 and 100-nm-thick ZnO:Al films were deposited using an ALD open-load reactor (OpAL, Oxford Instruments). A plasma ALD process was used for the deposition of the Al_2O_3 films ($Al(CH_3)_3, O_2$ plasma) at $200^\circ C$ and a thermal ALD process was used to deposit the ZnO films ($Zn(C_2H_5)_2, H_2O$ vapour) at $180^\circ C$. The ZnO films were doped with different Al concentrations by varying the ALD cycle ratio x (1 Al_2O_3 doping ALD cycle after every x ZnO ALD cycles). For stacks of type A, the Al doping concentration (represented as Aluminium fraction (AF)) was varied from $AF = 0$ to 0.31 ; AF being defined as $AF = [Al]/([Al] + [Zn])$, with $[Al]$ and $[Zn]$ are the atomic percentages of Al and Zn as measured by X-ray photoelectron spectroscopy (XPS). For stacks of type B, $AF = 0.10$.

The a -Si:H layers of different thicknesses (1.5–6.0 nm) were deposited at $50^\circ C$ by inductively-coupled plasma chemical vapour deposition (ICP-CVD) (PlasmalabSystem100 ICP 180, Oxford Instruments) using SiH_4 gas.

The stacks of type A were annealed at $400^\circ C$ for 5 min in a controlled N_2 atmosphere before the ZnO:Al film deposition. For stacks of type B, the annealing was carried out at a lower temperature of $300^\circ C$ for 1 min in a controlled N_2 atmosphere to prevent loss of hydrogen in the a -Si:H. The annealing was performed after the complete stack was deposited. The passivation performance of the stacks was evaluated from the effective lifetime, τ_{eff} , of the minority carriers on the silicon wafers. τ_{eff} was determined with photoconductance decay in the transient mode and quasi steady-state-mode (for $\tau_{eff} \leq 100 \mu s$) using a Sinton lifetime tester (WCT 100). The upper level for the effective surface

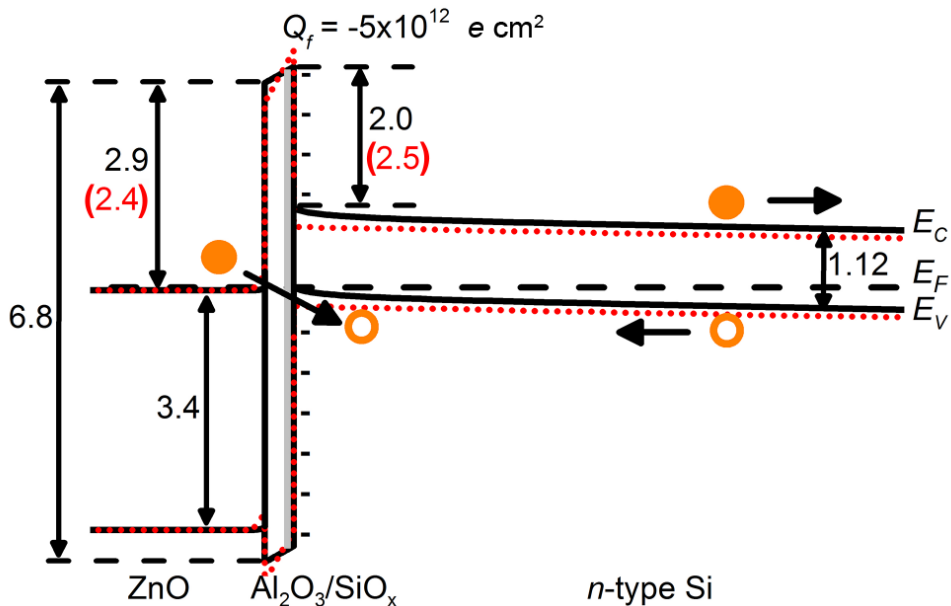


Figure 5.2: Equilibrium band diagram of an Al₂O₃/ZnO stack deposited on 3 Ω cm *n*-type Si. For the Si only the space charge region is shown. The red dotted band diagram shows the sensitivity of the simulation to a variation of the band offsets. It has been simulated using different band offsets (red numbers between brackets) for the ZnO/Al₂O₃ interface and Al₂O₃/c-Si interface.

recombination velocity (SRV), $S_{\text{eff,max}}$, was extracted at an injection level of $1 \cdot 10^{15} \text{ cm}^{-3}$ by the expression $S_{\text{eff,max}} = \text{thickness Si-wafer} / 2\tau_{\text{eff}}$. Spectroscopic Ellipsometry (SE) was employed to measure the thickness of each layer of the stacks. SE was also used to measure the optical bandgap ($E_{G,\text{opt}}$) and the carrier concentration (n_e) of the ZnO:Al films using a Drude model [5].

Current density–voltage (J – V) analysis was carried out to test the tunnelling behavior of stacks type A and type B. The front side of the tunnelling test structure consisted of the stack under investigation and the rear-side of a 10-nm-thick Al₂O₃ single film. On both sides 2-μm-thick Al metal contacts with an area of 1.96 cm² were deposited by physical vapour deposition (PVD). Subsequently, high quality p^+ -doped contacts with a pitch of 200 μm were produced by laser firing [6] on the rear side. Figure 5.1b shows a schematic cross section of the tunnelling samples and the stack configurations used.

5.3 Theory

5.3.1 The ZnO/Al₂O₃/*c*-Si system as a selective contact

To demonstrate how a Al₂O₃ stack can function as a selective hole contact, Figure 5.2 shows the calculated band diagram of the ZnO/Al₂O₃/*c*-Si system, which is based on AFORS-HET simulations [7]. In the simulation, a negative fixed charge density (Q_f) of $5 \cdot 10^{12} e \text{ cm}^{-1}$ [4] (with e the elementary charge) has been assumed at the interface between Al₂O₃ and Si. The values for band gaps and band offsets (eV) have been taken from the literature [4, 8, 9]. It is known that there is also an SiO_x layer between the Al₂O₃ and *c*-Si at the interface [4]. This oxide is not included in the model explicitly, but its presence was taken into account by using an effective dielectric constant $k = 6$. This is valid because only the capacitance of the oxide interlayer has an influence on the equilibrium band diagram calculation. The SiO_x/Al₂O₃ thickness in the model is 5 nm. It should also be noted that the distance of the Fermi level in the ZnO to the conduction band is larger in reality than this simulated band diagram shows. This is due to the fact that AFORS-HET is limited to MaxwellBoltzmann statistics, while FermiDirac statistics are needed to correctly determine the Fermi level in a degenerate semiconductor.

The simulation demonstrates how the Al₂O₃/ZnO stack serves as a hole-selective contact on the *c*-Si. The high density of negative fixed charges at the interface leads to accumulation of holes at the Al₂O₃/*c*-Si interface. Similarly to a diffused junction, in the inversion layer the conductivity for holes is much higher than for electrons, so the inversion layer serves as a homojunction. However, the doping is achieved without adding impurities to the silicon, thus keeping the bulk recombination low.

If the film is thin enough, a tunnelling current through the Al₂O₃ is possible by recombination of electrons from the ZnO and holes in the Si without a significant voltage loss due to the alignment of the ZnO conduction band with the *c*-Si valence band. This is analogous to the case of *n*-type Si heterojunction solar cells where the *p*-type *a*-Si:H is contacted with an *n*-type transparent conductive oxide (TCO) and the current is carried by tunnelling recombination [2, 10].

In a Si-HJT cell, the hole contact consists of a stack of intrinsic and doped *a*-Si:H with a combined thickness of ~ 10 nm [2] with on top a TCO of ~ 75 nm. Therefore, compared to the Si-HJT cell, the Al₂O₃/ZnO stack avoids the use of *a*-Si:H, leading to a higher transparency.

5.3.2 Modeling of the tunnelling currents

From a theoretical perspective, it is expected that the tunnelling recombination junction behaves like an Esaki (or tunnel) diode (see Figure 5.3a) [11], rather

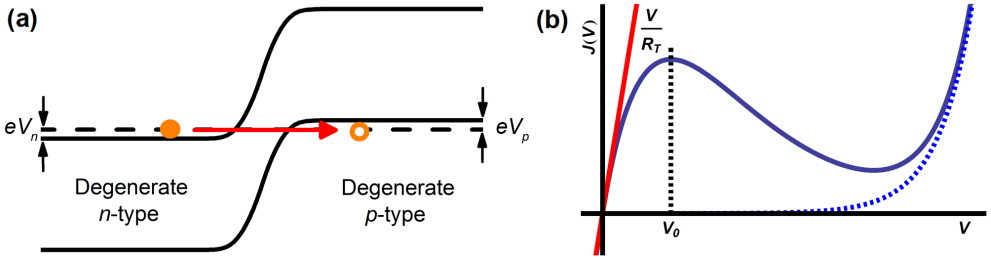


Figure 5.3: (a) The band diagram of an Esaki diode. An Esaki diode is a homojunction which has been degenerately doped on both sides, making band-to-band tunnelling recombination possible. Two important parameters that characterise the diode, are the degeneracies (indicated by eV_n and eV_p) of the n -type and p -type sides of the junction. (b) Empirical J - V characteristic of an Esaki diode (solid blue line) described by Eq. (5.1). The dashed blue line is the current through the diode in absence of the tunnelling recombination transport, i.e., the current predicted from the well-known Shockley equation. The red line is the tangent V/R_t at $V = 0$ and indicates the ohmic behavior of the diode at low bias.

than an ideal resistor. To model the J - V characteristic of an Esaki diode in the circuit, the following empirical relation was used [11]:

$$J_{\text{Esaki}}(V) = \frac{V}{R_t} \exp\left(-\frac{V}{V_0}\right) + J_{0,\text{tunnel}} \left[\exp\left(\frac{eV}{k_B T}\right) - 1 \right]. \quad (5.1)$$

Here, k_B is the Boltzmann constant and T the temperature. The variables R_t (the tunnel resistance under low bias), V_0 (the voltage where J has a local maximum) and $J_{0,\text{tunnel}}$ were used as fitting parameters. As can be seen in Figure 5.3b, an Esaki diode behaves like an ohmic resistance under small forward bias. Under higher bias, the tunnelling recombination current decreases because the conduction band of the n -type side of the junction is eventually located completely above the valence band of the p -type side. This results in a characteristic region where the current through the diode decreases with the bias, as illustrated in Figure 5.3b. This is known as “negative differential resistance”. For even higher bias, the tunnel diode starts to function as an ordinary diode and the second term in Eq. (5.1) starts to dominate, increasing the current again.

The measured J - V data (to be presented in Section 5.4.2) of the tunnelling experiments (described in Section 5.2) has been modelled using different circuits involving an Esaki diode. When fitting the data, it turned out that the tested circuits with the Esaki diode did not significantly improve the fit compared to an ordinary double diode fit. This most likely means that the tunnel junction operates in its ohmic regime so that its contribution cannot be separated from other contributions to the resistance. However, the model

of an Esaki diode is still useful for identifying the factors that can lead to improvement of the tunnelling contact as will be shown in Section 5.3.3.

Because the effect of tunnelling recombination cannot be separated from other series resistance contributions, the well-known double diode equivalent circuit is used to model the measured tunnelling currents and to assess the contact resistance of the stacks. In the double diode fits, the ideality factors of the diodes were kept constant at 1 and 2, so the fits have four free parameters: the dark saturation currents of the diodes J_{01} and J_{02} (with idealities 1 and 2 respectively), the shunt resistance R_{shunt} and the series resistance R_{series} .

5.3.3 Considerations for improving the contact resistance and surface passivation

As discussed in Section 5.3.3, the tunnelling recombination in an Esaki diode decreases strongly when the electrons cannot recombine with holes at the same energy level. The underlying principle is that the recombination is most effective when energy is conserved without the need for involving a third particle (such as a phonon) [11]. Therefore, the alignment of the ZnO conduction band with respect to the *c*-Si valence band is crucial for the charge transport through the tunnelling contact.

From a theoretical perspective, the following factors will be identified that have influence on the alignment of these two bands: the degeneracies of the semiconductors on both side of the junction; the band offsets at the ZnO/Al₂O₃ and Al₂O₃/*c*-Si interfaces and the dielectric properties of the Al₂O₃.

5.3.3.1 Degeneracy of the ZnO and *c*-Si

Assuming that the model of a tunnel diode applies to the tunnelling recombination current through the Al₂O₃ stacks, it is expected that the tunnel resistance decreases with increasing degeneracy (i.e., the distance of the Fermi level above the conduction band, eV_n in Figure 5.3a) of the ZnO, which is related to the carrier concentration n_e . Similarly, it is expected that an increase in the fixed charge density in the Al₂O₃ will decrease the resistance by shifting up the valence band of the *c*-Si with respect to the Fermi level and thus increasing the degeneracy, eV_p , in the *c*-Si.

The effect of the doping in the ZnO on the degeneracy, eV_n , can be measured in various ways. First of all, an increase in eV_n will increase the optical bandgap of the ZnO due to the Burstein–Moss effect. A second method to measure eV_n is by the direct relation between the electron density n_e and the Fermi level. If the bands of the ZnO are assumed to be parabolic, then eV_n is proportional to $n_e^{2/3}$. A more sophisticated approach would take the non-parabolicity of the ZnO bands into account [12], but this requires the determination the non-parabolicity parameter of the ZnO, which is outside of the scope of this paper.

Another method to measure eV_n is through the binding energy of electrons in the ZnO, but this method will not be used in this work.

Both the optical bandgap and the electron density are measured by SE (see Section 5.2). The determination of n_e uses the effective mass (m^*) of the ZnO as input parameter. Generally, m^* depends on the electron concentration (due to the non-parabolicity of the bands), making it difficult to accurately calculate eV_n from n_e . Therefore, the optical bandgap is the preferred measure for eV_n in this work. However, it should be noted that the change in optical bandgap cannot be attributed to the Burstein-Moss shift alone since the bandgap narrows with increasing n_e [13], so care must still be taken in interpreting the data.

5.3.3.2 Band offsets in the ZnO/Al₂O₃/*c*-Si system and the dielectric properties of the Al₂O₃

To investigate the role that the band offsets play in the ZnO/Al₂O₃/*c*-Si system, the band diagram has been calculated for different values of the ZnO/Al₂O₃ and Al₂O₃/*c*-Si band offsets. The red dotted band diagram in Figure 5.2 that is superimposed on the original calculation shows how the band diagram changes when band offsets are used that are less favourable for the alignment of the ZnO conduction band with respect to the *c*-Si valence band. It can be observed that the strength of the electric field in the Al₂O₃ changes significantly, but the positions of the bands in the ZnO and *c*-Si are not very sensitive to the band offsets used. At the interface, the *c*-Si valence band is located 0.27 eV lower than in the original calculation, even though both band offsets have been changed by 0.5 eV. This illustrates that the Al₂O₃ fulfils an important third role besides inducing surface passivation and making the contact selective to holes: as a dielectric it can sustain a high amount of band-bending that facilitates the alignment of the ZnO conduction band and the *c*-Si valence band. It is known that Al₂O₃ layers prepared by plasma ALD have a significant breakdown electric field of $\sim 9 \text{ MV cm}^{-1}$ [14] and for ultrathin ($\sim 1 \text{ nm}$) Al₂O₃ layers, values as high as $\sim 30 \text{ MV cm}^{-1}$ have even been reported [15]. Assuming a breakdown field of $\sim 9 \text{ MV cm}^{-1}$, this means that an energy difference of $\sim 0.9 \text{ eV}$ can exist across a 1 nm layer to accommodate for the alignment of the ZnO bands with the *c*-Si bands. Therefore, it can be concluded that the exact values of the band offsets in the ZnO/Al₂O₃/*c*-Si system will affect the tunnelling resistance through the stack, but that the effect of “misalignment” is diminished by the electric field in the Al₂O₃.

An interesting question is what factors determine the band offsets between materials. Knowing this will help in future research to identify other materials that have a high potential to make a good tunnelling contact. To answer this question, the physical mechanism that causes band offsets needs to be addressed. Often, band offsets are calculated using the Anderson model (also

known as the electron affinity rule) [11, 16], which assumes that the vacuum energy levels of the two materials align across the interface. This assumption leads to the equation $\Delta E_C = e\Delta\chi$, with ΔE_C the conduction band offset and $\Delta\chi$ the difference between electron affinities of the two materials.

However, the Anderson model is not always satisfactory for predicting band offsets and other models have been suggested, such as the Tersoff model [17]. This model states that band offsets arise because allowed states in one semiconductor cause evanescent states in the forbidden gap of the other. This gives rise to a quantum dipole at the interface, which results in band offsets. The Tersoff model asserts that to a good approximation, the quantum dipole tends to align the so-called branch point energies of the materials. Roughly speaking, the branch point energy of a material (which is a bulk property) is the energy level (usually somewhere in the forbidden gap) where the (forbidden) states change from valence band character to conduction band character. This makes the branch point energy a useful material property for the selection of materials in tunnelling stacks.

It should be noted that the Tersoff model has been applied to the ZnO/*c*-Si system [18, 19], but not to the ZnO/Al₂O₃/*c*-Si system and, to the author's knowledge, neither to these materials deposited by ALD.

It can be concluded that, when selecting materials for low resistance tunnelling stacks, both the electron affinity and the branch point energy of the materials are relevant parameters for the optimization of the band alignment of the stack. For an in-depth analysis of the band offsets, detailed calculations of the quantum dipole at the interface are recommended.

5.3.3.3 The effect of amorphous silicon interlayers

It is known that the chemical passivation of Al₂O₃ decreases with thickness for ultrathin films (<5–10 nm) [4], so it is important to investigate ways to improve the chemical passivation of the ultrathin Al₂O₃ that is needed to achieve a good tunnelling resistance. A material that is well known for its excellent chemical passivation, is *a*-Si:H. It is therefore expected that an interlayer of *a*-Si:H between the Al₂O₃ and *c*-Si can complement the surface passivation of the Al₂O₃.

There is, however, another motivation for the insertion of a conductive amorphous interlayer. In the same way that tunnelling recombination is affected by the requirement for energy conservation, momentum should also be considered. In a crystalline material, momentum is a well-defined quantity that is conserved, just like energy. However, in an amorphous material, momentum is not well-defined and it is expected that the insertion of an amorphous interlayer between the *c*-Si and Al₂O₃ will relax the requirement for momentum conservation in the tunnelling recombination mechanism. This would be analogous to the observation that *c*-Si has an indirect bandgap, while the relaxation

Table 5.1: Fitted double diode parameters for the J - V characteristics (shown in Figure 5.5) of the $\text{Al}_2\text{O}_3/\text{ZnO}:\text{Al}$ tunnelling stacks with various aluminium fractions (AF), except for the sample with $\text{AF} = 0.31$.

AF	$E_{\text{G,opt}}$	$J_{01} \left(\frac{\text{mA}}{\text{cm}^2 \text{s}} \right)$	$J_{02} \left(\frac{\text{mA}}{\text{cm}^2 \text{s}} \right)$	$R_{\text{series}} \ (\Omega \text{ cm}^2)$	$R_{\text{shunt}} \ (\Omega \text{ cm}^2)$
0	3.29	$4.96 \cdot 10^{-6}$	$1.82 \cdot 10^{-1}$	2.41	27.2
0.10	3.55	$2.82 \cdot 10^{-5}$	$4.98 \cdot 10^{-1}$	1.74	14.3
0.21	3.69	$9.60 \cdot 10^{-6}$	$3.68 \cdot 10^{-1}$	1.45	26.1

of momentum conservation makes a -Si:H a direct bandgap semiconductor. For this reason, the hypothesis is advanced that a conductive amorphous interlayer such as a -Si:H will reduce the tunnelling resistance.

In principle, the same reduction of contact resistance could be obtained if the conductive amorphous layer is inserted between the ZnO and Al_2O_3 instead of between the c -Si and Al_2O_3 . However, it should be noted that the surface passivation of Al_2O_3 can be damaged by high-energy photons [20]. This makes it more difficult to achieve good surface passivation when plasma-based methods are used for the deposition of the amorphous layer.

5.4 Results and discussion

5.4.1 Passivation of c -Si by $\text{Al}_2\text{O}_3/\text{ZnO}$ stacks

The passivation results of the $\text{Al}_2\text{O}_3/\text{ZnO}$ stacks are presented in Figure 5.4. As can be seen, the SRV of the stacks depends strongly on the thickness of the Al_2O_3 (d_{AlO_x}) which was varied between 1, 3 and 10 nm while the ZnO thickness was kept constant at 40 nm. When the SRV of the stack is compared with that of a single layer of Al_2O_3 , it is observed that the stacks passivate better than the single layers for $d_{\text{AlO}_x} > 3$ nm and worse for $d_{\text{AlO}_x} < 3$ nm. At $d_{\text{AlO}_x} < 3$ nm, the SRV of the stacks is $\sim 700 \text{ cm s}^{-1}$, which is considered too high concerning the goal set in Section 5.1. Another observation is that the passivation of $\text{Al}_2\text{O}_3/\text{ZnO}$ stacks is mostly independent of the doping level of the ZnO.

More results on the passivation performance of the $\text{Al}_2\text{O}_3/\text{ZnO}$ stacks are presented in reference [21].

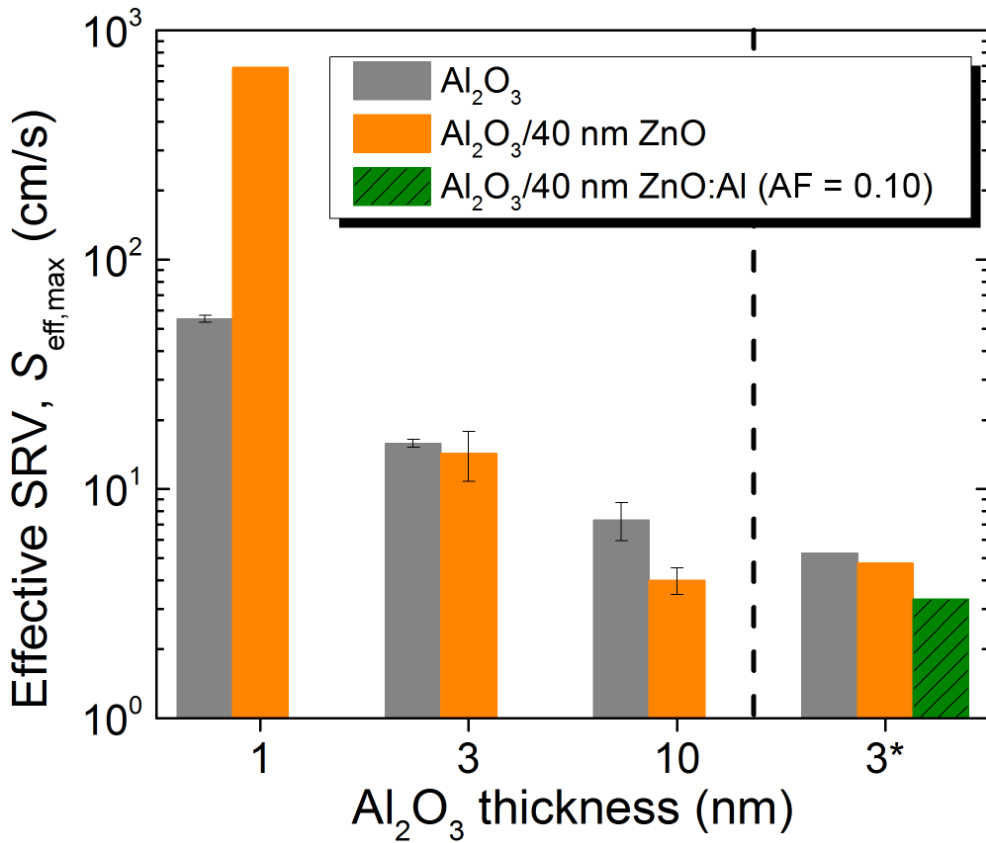


Figure 5.4: Passivation results of Al₂O₃/ZnO stacks on *n*-type silicon. The influence of the Al₂O₃ thickness is shown (left of the dashed line) as well as the influence of ZnO doping (right of the dashed line). Data marked with an asterisk (*) were performed on wafers from a different ingot.

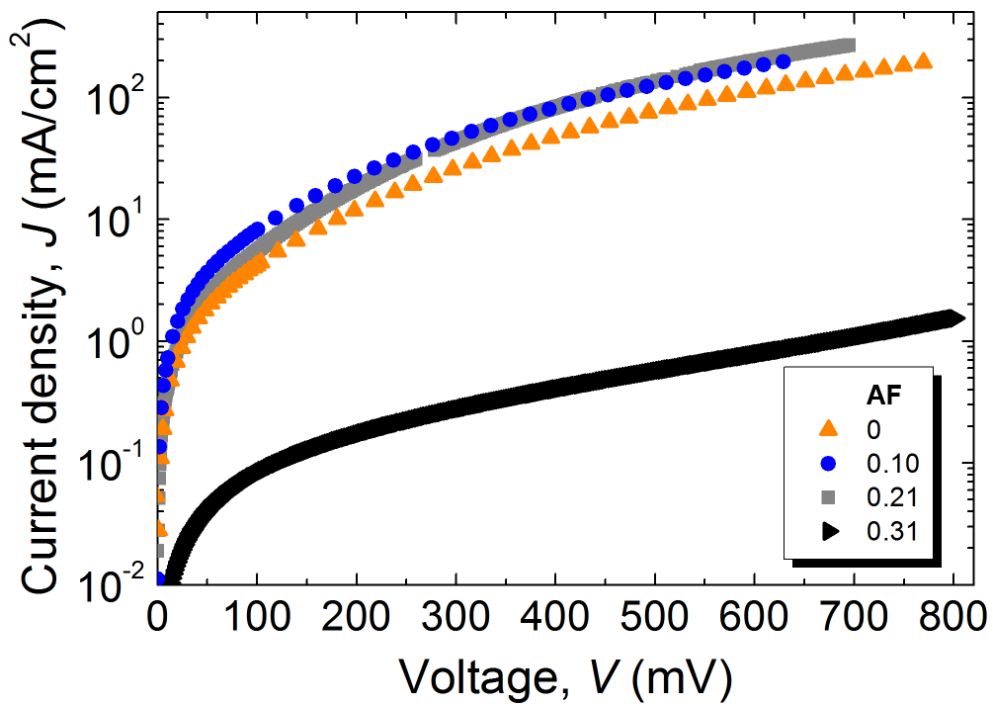


Figure 5.5: Measured J - V characteristics of selected $\text{Al}_2\text{O}_3/\text{ZnO}$ tunnelling samples. The legend shows the aluminium fraction (AF) in the ZnO as measured by XPS.

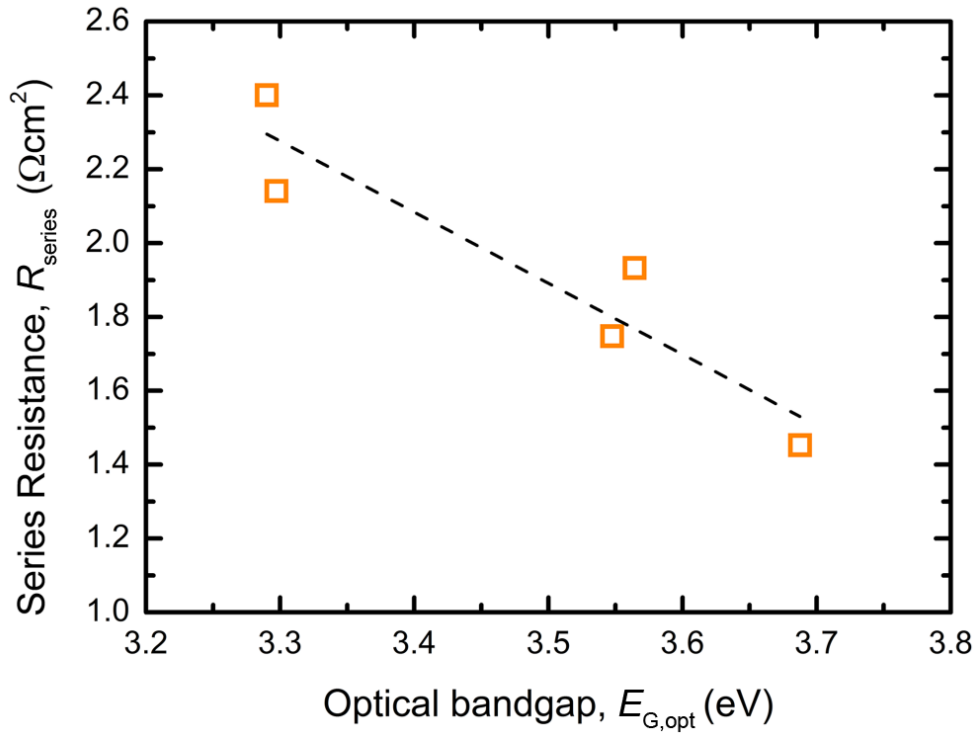


Figure 5.6: The fitted series resistance of the $\text{Al}_2\text{O}_3/\text{ZnO}$ stacks that could be fitted as a function of the optical bandgap of the ZnO. The dotted line is a linear fit through the data points and has a slope of $-1.9 \Omega \text{ cm}^2 \text{ eV}^{-1}$

5.4.2 Charge transport through Al₂O₃/ZnO stacks

Measured J - V curves¹ for stacks with 1 nm Al₂O₃ and 100 nm ZnO are shown in Figure 5.5 and the parameters of the double diode fits of the curves shown in Fig. 5 are in Table 5.1. Figure 5.6 shows the observed trend in the fitted series resistance as a function of the optical bandgap of the ZnO for all measured samples that could be fitted well. First of all, it should be noted that all of the samples reach current densities higher than 43 mA cm⁻² s⁻¹ (i.e., the maximum possible short-circuit current for a single junction c -Si solar cell under AM1.5 illumination) at voltages lower than the operating voltage of a c -Si solar cell. For completeness we mention that J - V measurements have also been carried out on samples with 2 nm of Al₂O₃, but none of these samples reached currents higher than 10 mA cm⁻² s⁻¹. Two observations can be made from Figures. 5.5 and 5.6.

First of all, the measured series resistance decreases with increasing ZnO doping. Finite element calculations of the current flow through the ZnO have been used to investigate the influence of the lateral conductivity of the ZnO on the fitted series resistance. It was found that the ZnO increases the effective contact area (assumed to be 1.96 cm in Figure 5.6) by approximately 10–20 % for all samples. It was also found that the observed trend in the tunnelling resistance shown in Figure 5.6 cannot be explained by the change in the ZnO sheet resistance. As can be seen, the decrease of the tunnelling resistance is observed to be approximately linear with the optical bandgap of the ZnO, which corroborates the hypothesis advanced in Section 5.3.3.1 that the ZnO doping improves the alignment between the ZnO conduction band and c -Si valence band and therefore decreases the tunnelling resistance.

5.4.3 Improved c -Si passivation by a -Si:H/Al₂O₃/ZnO stacks

Figure 5.7 shows the passivation results and tunnel resistance of the a -Si:H/Al₂O₃/ZnO stacks, where the thicknesses of the Al₂O₃ and ZnO were kept constant at 1 nm and 100 nm, respectively. The thickness of the a -Si:H was varied between 1.5 nm and 6 nm. The parameters of the double diode fit are listed in Table 5.2. As predicted in Section 5.3.3.3, the surface passivation is already greatly enhanced by the insertion of only 1.5 nm of a -Si:H. This represents a significant reduction in a -Si:H thickness compared to ordinary Si-HJT cells, which typically use ~10 nm or more of a -Si:H [2]. At the same time, the tunnel resistance increases slightly with a -Si:H thickness.

The hypothesis from Section 5.3.3.3 that the insertion of an amorphous silicon interlayer would decrease the tunnelling resistance has not been verified

¹We would like to highlight that the J - V characteristics of several samples were measured shortly after deposition and 1 year later again. The results were very reproducible, and therefore it can be concluded that the electric properties of these stacks are very stable in time.

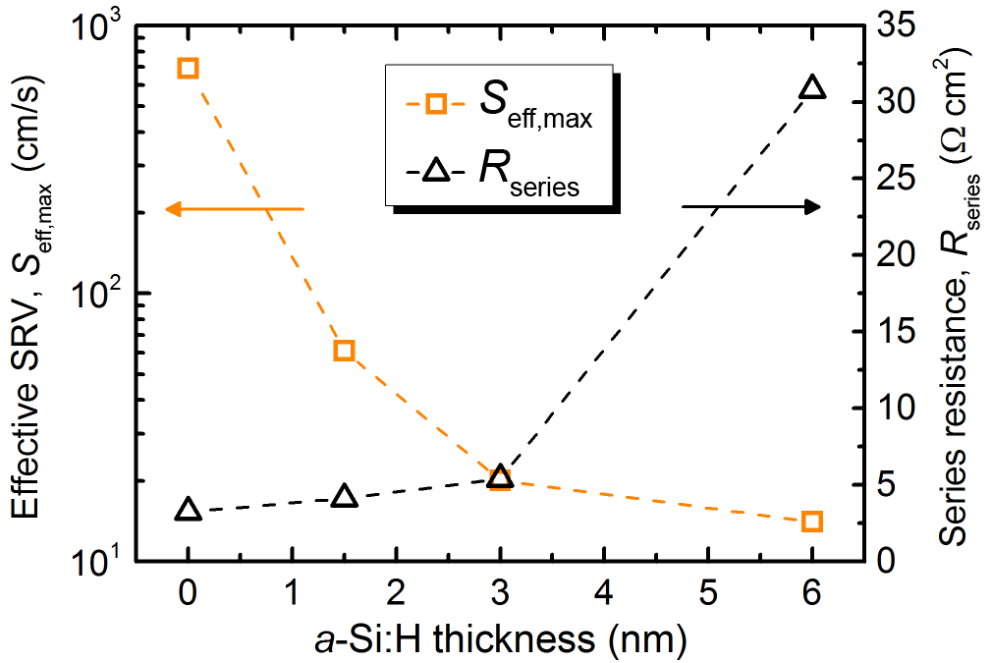


Figure 5.7: The left axis shows the effective SRV of the $a\text{-Si:H}/\text{Al}_2\text{O}_3/\text{ZnO}$ stacks as a function of the $a\text{-Si:H}$ thickness. The passivation has been determined from lifetime samples on p -type silicon. The right axis shows the total series resistance of the tunnelling samples, determined by fitting the J - V characteristic with a double diode model.

Table 5.2: Fitted double diode parameters for the J - V characteristics of the $a\text{-Si:H}/\text{Al}_2\text{O}_3/\text{ZnO:Al}$ tunnelling stacks. The aluminium fraction (AF) of the ZnO:Al is 0.10.

$a\text{-Si:H}$ thickness (nm)	J_{01} ($\frac{\text{mA}}{\text{cm}^2 \text{ s}}$)	J_{02} ($\frac{\text{mA}}{\text{cm}^2 \text{ s}}$)	R_{series} ($\Omega \text{ cm}^2$)	R_{shunt} ($\Omega \text{ cm}^2$)
0	$3.21 \cdot 10^{-6}$	$1.10 \cdot 10^{-1}$	3.32	31.2
1.5	$5.32 \cdot 10^{-7}$	$3.38 \cdot 10^{-2}$	5.57	95.1
3.0	$3.30 \cdot 10^{-8}$	$1.01 \cdot 10^{-2}$	5.39	183
6.0	$8.04 \cdot 10^{-8}$	$1.13 \cdot 10^{-3}$	30.8	180

by the experiments shown in Figure 5.7. Two factors can be advanced that could have played a role in the increased resistance of the samples with the *a*-Si:H interlayer. First is the fact that these samples were annealed at a lower temperature, most likely reducing the amount of fixed charges at the interface [4]. As explained in Section 5.3.3.1, a reduction of interface charges most likely leads to a higher tunnelling resistance. It is also possible that the SiO_x interlayer is thicker in the *a*-Si:H/Al₂O₃/ZnO system due to different oxidation kinetics of *a*-Si:H compared to *c*-Si. More careful analysis of the Al₂O₃/*a*-Si:H interface and oxide thicknesses is required to draw conclusions.

5.5 Conclusions and outlook

The potential of using Al₂O₃/ZnO stacks as contacts on *c*-Si that are simultaneously hole-selective, highly transparent and passivating has been addressed theoretically and experimentally by selected measurements. Charge transport through these stacks is considered to occur by tunnelling recombination, which can be understood by applying the theory of Esaki diodes. This theory demonstrates that the ZnO conduction band should be at a lower energy than the *c*-Si valence band in order to make the tunnelling recombination efficient and to obtain a low contact resistance. Three key elements have been identified that can help to achieve the required band arrangement:

- (i) The negative fixed charges in the Al₂O₃, which cause accumulation of holes at the *c*-Si surface, shifting up the *c*-Si valence band.
- (ii) The doping of the ZnO, which results in a high concentration of electrons, shifting down the ZnO conduction band.
- (iii) The presence of the large bandgap dielectric Al₂O₃, which can sustain a large amount of band bending. This allows the ZnO and *c*-Si bands to align with more freedom than they would have been able to in the absence of the Al₂O₃.

The influence of the doping of the ZnO on the resistance is demonstrated by the experimental observation that the contact resistance decreases with increasing optical bandgap of the ZnO. It has been found that stacks with 1 nm of Al₂O₃ on *p*-type *c*-Si can give contact resistances of $\sim 1.5 \Omega \text{ cm}^2$. It was verified that the electrical properties of the contacts are stable in time for at least 1 year.

It has also been found that the passivation of Al₂O₃/ZnO stacks with 1 nm of Al₂O₃ is insufficient for application in high efficiency solar cells. Therefore, samples have been produced with an *a*-Si:H interlayer between the Al₂O₃ and *c*-Si in order to increase the chemical passivation. It has been demonstrated that such interlayers can reduce $S_{\text{eff,max}}$ to a value as low as 20 cm s^{-1} while

still achieving a contact resistance of $\sim 5 \Omega \text{ cm}^2$. As the *a*-Si:H in these stacks is significantly thinner than in a typical Si-HJT cell, parasitic absorption is expected to be reduced.

For future research, the following research directions are suggested. The *a*-Si:H/ Al_2O_3 interface needs to be better characterized. In particular, the fixed charge density at this interface needs to be understood. It is also suggested to make the Al_2O_3 layer as thin as possible to increase the tunnelling probability, while simultaneously making the density of negative fixed charges as high as possible. To achieve this, the growth of Al_2O_3 on *a*-Si:H during the first few ALD cycles needs to be well-understood and the process tuned for this application.

Another method to increase the tunnelling current through the Al_2O_3 is by engineering defects in the Al_2O_3 at a specific energy level to promote hopping transport through the oxide. Doing so will increase the hole-selectivity of the contact and will most likely also relax the thickness constraints on the Al_2O_3 so that better passivation can be achieved without introducing *a*-Si:H in the stack.

The doping of the ZnO needs to be optimised to make the tunnelling contact as efficient as possible. ALD offers the prospect of using graded doping schemes to optimise the contact interface as well as other relevant parameters of the TCO, such as transparency and sheet resistance. It is also considered important to start fabrication of real solar cells to test the hole-selectivity of the contacts.

Acknowledgements

This work was financially supported by the Dutch Technology Foundation STW (Flash Perspectief Programma). The research of D.G.-A. was funded by a Marie Curie IEF fellowship of the European Community's Seventh Framework Programme (ALD4PV, FP7-MC-IEF-272444). The research of W.M.M.K. has been made possible by the Dutch Technology Foundation STW and the Netherlands Organization for scientific Research (NWO, VICI programma). We acknowledge financial support for this research from ADEM, A green Deal in Energy Materials of the Ministry of Economic Affairs of The Netherlands (www.adem-innovationlab.nl).

References Chapter 5

- [1] Peter Würfel. *Physics of Solar Cells*. Wiley-VCH Verlag GmbH & Co, Weinheim, 2005. ISBN 3-527-40428-7.
- [2] Stefaan De Wolf, Antoine Descoeurdes, Zachary C. Holman, and Christophe Ballif. High-efficiency Silicon Heterojunction Solar Cells: A Review. *green*, 2(1):7–24, January 2012. ISSN 1869-876X. doi: 10.1515/green-2011-0018.
- [3] Bram Hoex, S. B. S. Heil, E. Langereis, M. C. M. van de Sanden, and W. M. M. Kessels. Ultralow surface recombination of *c*-Si substrates passivated by plasma-assisted atomic layer deposited Al₂O₃. *Applied Physics Letters*, 89(4):042112, 2006. ISSN 00036951. doi: 10.1063/1.2240736.
- [4] G. Dingemans and W. M. M. Kessels. Status and prospects of Al₂O₃-based surface passivation schemes for silicon solar cells. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 30(4):040802, 2012. ISSN 07342101. doi: 10.1116/1.4728205.
- [5] S. J. Youn, T. H. Rho, B. I. Min, and Kwang S. Kim. Extended Drude model analysis of noble metals. *physica status solidi (b)*, 244(4):1354–1362, April 2007. ISSN 03701972. doi: 10.1002/pssb.200642097.
- [6] E. Schneiderlöchner, R. Preu, R. Lüdemann, and S. W. Glunz. Laser-fired rear contacts for crystalline silicon solar cells. *Progress in Photovoltaics: Research and Applications*, 10(1):29–34, January 2002. ISSN 1062-7995. doi: 10.1002/pip.422.
- [7] R. Stangl, C. Leendertz, and J. Haschke. Numerical Simulation of Solar Cells and Solar Cell Characterization Methods: the Open-Source on Demand Program AFORS-HET. In Radu D Rugescu, editor, *Solar Energy*, pages 319–352. InTech, 2010. ISBN 978-953-307-052-0.
- [8] Joost Jozef Hendrik Gielis, Bram Hoex, M. C. M. van de Sanden, and W. M. M. Kessels. Negative charge and charging dynamics in Al₂O₃ films on Si characterized by second-harmonic generation. *Journal of Applied Physics*, 104(7):073701, 2008. ISSN 00218979. doi: 10.1063/1.2985906.

- [9] J. B. You, X. W. Zhang, S. G. Zhang, H. R. Tan, J. Ying, Z. G. Yin, Q. S. Zhu, and Paul K. Chu. Electroluminescence behavior of ZnO/Si heterojunctions: Energy band alignment and interfacial microstructure. *Journal of Applied Physics*, 107(8):083701, April 2010. ISSN 00218979. doi: 10.1063/1.3385384.
- [10] Ana Kanevce and Wyatt K. Metzger. The role of amorphous silicon and tunneling in heterojunction with intrinsic thin layer (HIT) solar cells. *Journal of Applied Physics*, 105(9):094507, 2009. ISSN 00218979. doi: 10.1063/1.3106642.
- [11] S.M. Sze and K. Ng Kwok. *Physics of Semiconductor Devices*. John Wiley & Sons, Inc., 2007. ISBN 978-0-471014323-9.
- [12] T. Pisarkiewicz and A. Kolodziej. Nonparabolicity of the Conduction Band Structure in Degenerate Tin Dioxide. *physica status solidi (b)*, 158(1):K5–K8, March 1990. ISSN 03701972. doi: 10.1002/pssb.2221580141.
- [13] Anubha Jain, P. Sagar, and R.M. Mehra. Band gap widening and narrowing in moderately and heavily doped n-ZnO films. *Solid-State Electronics*, 50(7-8):1420–1424, July 2006. ISSN 00381101. doi: 10.1016/j.sse.2006.07.001.
- [14] K. B. Jinesh, J. L. van Hemmen, M. C. M. van de Sanden, F. Roozeboom, J. H. Klootwijk, W. F. a. Besling, and W. M. M. Kessels. Dielectric Properties of Thermal and Plasma-Assisted Atomic Layer Deposited Al₂O₃ Thin Films. *Journal of The Electrochemical Society*, 158(2):G21, 2011. ISSN 00134651. doi: 10.1149/1.3517430.
- [15] H. C. Lin, P. D. Ye, and G. D. Wilk. Leakage current and breakdown electric-field studies on ultrathin atomic-layer-deposited Al₂O₃ on GaAs. *Applied Physics Letters*, 87(18):182904, October 2005. ISSN 00036951. doi: 10.1063/1.2120904.
- [16] R.L. Anderson. Experiments on Ge-GaAs heterojunctions. *Solid-State Electronics*, 5(5):341–351, September 1962. ISSN 00381101. doi: 10.1016/0038-1101(62)90115-6.
- [17] J Tersoff. Theory of semiconductor heterojunctions: The role of quantum dipoles. *Physical Review B*, 30(8):4874–4877, October 1984. ISSN 0163-1829. doi: 10.1103/PhysRevB.30.4874.
- [18] B Höffling, A Schleife, C. Rödl, and F Bechstedt. Band discontinuities at Si-TCO interfaces from quasiparticle calculations: Comparison of two alignment approaches. *Physical Review B*, 85(3):035305, January 2012. ISSN 1098-0121. doi: 10.1103/PhysRevB.85.035305.

- [19] Winfried Mönch. Branch-point energies and the band-structure lineup at Schottky contacts and heterostructures. *Journal of Applied Physics*, 109(11):113724, June 2011. ISSN 00218979. doi: 10.1063/1.3592978.
- [20] H. B. Profijt, P. Kudlacek, M. C. M. van de Sanden, and W. M. M. Kessels. Ion and Photon Surface Interaction during Remote Plasma ALD of Metal Oxides. *Journal of The Electrochemical Society*, 158(4):G88, 2011. ISSN 00134651. doi: 10.1149/1.3552663.
- [21] D. Garcia-Alonso, S. Smit, S. Bordihn, and W. M. M. Kessels. Silicon passivation and tunneling contact formation by atomic layer deposited $\text{Al}_2\text{O}_3/\text{ZnO}$ stacks. *Semiconductor Science and Technology*, 28(8):082002, August 2013. ISSN 0268-1242. doi: 10.1088/0268-1242/28/8/082002.

Chapter 6

Properties of graphene and its potential applications in silicon solar cells

Graphene is a relatively new material with unique electrical and optical properties, the most relevant of which –for the field of solar cells– are its high mobility combined with a low absorption coefficient. It is therefore frequently mentioned in literature as a promising new material for *c*-Si solar cells. In literature several devices can be found that use graphene to form a hole-extracting selective contact for *c*-Si (see, e.g., [1, 2, 3, 4, 5]). The most successful cell so far is the one made by Song *et al.* [5], which achieves a respectable 15.6% efficiency with a V_{OC} of 595 mV. However, most of the research into this type of cell is done by researchers that are first and foremost involved in graphene research rather than photovoltaics (PV) and a comprehensive overview of the strengths and weaknesses of graphene as a material for *c*-Si PV is therefore still absent. The goal of this work is therefore to answer two important questions:

1. Can graphene achieve a high enough conductivity and low enough absorption coefficient to act as a good transparent electrode in *c*-Si cells?
2. Is graphene a suitable material for the formation of selective contacts with *c*-Si?

These questions will be addressed as much as possible from a physics point of view. This means that there will be less emphasis on the various devices and graphene films that have been published in literature, since there are already comprehensive reviews about these topics [3, 6, 7, 8, 9]. Instead, the relevant physics of graphene will be discussed to develop a broad understanding of the possibilities of graphene and to facilitate future research in this area. In Section 6.1 its electronic properties will be reviewed and explained. The conductivity

and transparency of graphene are of special interest here and its potential as a selective electrode will be discussed. Then, in Section 6.2, the nature of *c*-Si-graphene junctions will be explored to see what is necessary to make a good selective contact for a silicon solar cell that utilises graphene, as well as to provide practical tools to perform calculations about these types of cell.

6.1 The electrical and optical properties of graphene

Section 6.1 will give a brief overview of the properties of graphene that are relevant for its application in solar cells. The theory has been adapted from the general theoretical papers on graphene by Falkovsky [10], Castro Neto *et al.* [11], and Abergel *et al.* [12].

6.1.1 The graphene dispersion relation and density of states

Graphene has two defining characteristics that give it its unique electrical and optical properties. These characteristics are the two-dimensional (2D) nature of the material and the unusual relation between energy E and momentum $\hbar k$ of electrons¹ in the material (i.e., its dispersion relation). In most materials the dispersion relation is quadratic, meaning that the electrons in those materials obey a law of the form $E = \hbar^2 k^2 / (2m^*)$, just like classical objects in Newtonian mechanics. In graphene the dispersion relation is very different. For completeness sake, Figure 6.1a shows the energy-momentum relation of graphene over the whole Brillouin zone, but in practice this full dispersion relations is seldom necessary since only the electrons near the Fermi energy are of interest for transport phenomena. As can be seen in the figure, the π and π^* bands (which are basically equivalent to the valence and conduction bands in semiconductor) touch at specific points called the K-points or Dirac points. The momentum at these K-points is denoted by \mathbf{K} . Around these Dirac points the dispersion relation can be approximated to be two touching cones:

$$\epsilon = \pm \hbar v_F k = \pm \hbar v_F \sqrt{k_x^2 + k_y^2}. \quad (6.1)$$

Here, ϵ and $\mathbf{k} = \boldsymbol{\kappa} - \mathbf{K}$ are the energy and momentum measured from the Dirac point (with $\boldsymbol{\kappa}$ the total momentum measured from the Γ -point). The constant $v_F \approx 1 \cdot 10^6 \text{ m s}^{-1}$ is called the *Fermi velocity* of graphene. The shape of the dispersion relation in k -space is frequently referred to as the Dirac cone. The energy at the tip of the cone is called the Dirac energy E_{Dirac} and is located

¹Because the symbol p is also used for hole density, it is preferable to use $\hbar k$ for momentum to avoid confusion. Because of the equivalence between k and momentum, k will be called the momentum and wavenumber interchangeably.

4.5 eV under the vacuum level² [14, 15]. The dispersion relation $\epsilon = \pm \hbar k v_F$ is reminiscent of the energy-momentum relation of photons $E = \hbar k c$. Like photons, electrons in graphene have a constant (i.e., independent of energy) group velocity $d\omega/dk = v_F$, which is why they are often referred to as being *massless*. The constant group velocity also is one of the properties that give graphene its remarkable transport properties since even electrons with very little energy can travel at a high velocity. For comparison, an electron that travels at 10^6 m s^{-1} in vacuum has an energy of 2.8 eV. Unlike photons, though, electrons in graphene are allowed to have energies $\epsilon < 0$.

Just like in any other conductor, each electron occupies a certain volume of phase space that is limited by the uncertainty relation. Since graphene is 2D, the smallest phase space volume that an electron can occupy is given by $(\Delta x \Delta k)^2 = 1/(2\pi)^2$. The electrons have wave functions that extend over the area A of the graphene so that $(\Delta x)^2 = A$. By combining the uncertainty relation with the dispersion relation (6.1) it is possible to find the density of states (DOS, D_G) of graphene. The DOS is measured in units of $\text{cm}^{-2} \text{ eV}^{-1}$ and is given by:

$$D_G(\epsilon) = \frac{2|\epsilon|}{\pi(\hbar v_F)^2}. \quad (6.2)$$

This expression includes the spin degeneracy factor 2. With the DOS given by Eq. (6.2) it becomes possible to relate the electron density to the Fermi energy of the graphene E_{FG} . When the Fermi level is measured from the tip of the Dirac cone, it is denoted by $\epsilon_{FG} = E_{FG} - E_{\text{Dirac}}$. For pure, uncontaminated graphene the charge neutrality point (CNP) of the Fermi level lies at the Dirac point (i.e., at $\epsilon_{FG} = 0$), so that the work function (WF) of intrinsic graphene is $\phi_G = 4.5 \text{ eV}$. In agreement with common conventions, all electrons occupying states with energy $\epsilon > 0$ will be referred to as free electrons (or simply electrons if no confusion arises in a given context) while empty states with energy $\epsilon < 0$ will be called holes. The densities of free electrons and holes are denoted by

²Another frequently quoted value is $E_{\text{Dirac}} = -4.6 \text{ eV}$ (see, e.g., Zhong *et al.* [13]). In this work $E_{\text{Dirac}} = -4.5 \text{ eV}$ will be used, since this seems to be the most accurate experimental value.

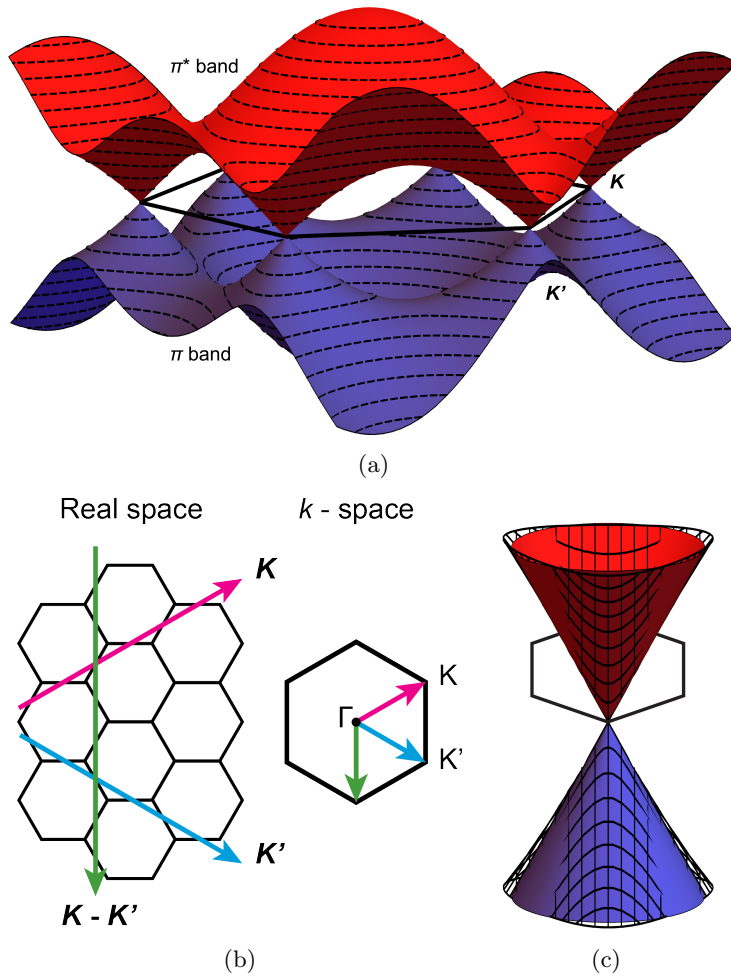


Figure 6.1: (a) The full dispersion relation of graphene as calculated by Eq. (1) by Abergel *et al.* [12]. (b) On the left is shown the real space lattice of graphene. On the right one hexagon from the reciprocal space lattice of graphene is shown with the K and K' -points; the four remaining corners of the hexagon are linear combinations of these two K -points. It should be recalled that a point in k -space represents a momentum or a motion. These motions are drawn in the real space lattice using vectors of corresponding colours. As can be seen, the motions corresponding to the K -points are along the zigzags of atoms that provide the easiest routes for electrons to follow. (c) The full dispersion shown in (a) can be approximated as a cone at the K -points. The cone approximation is drawn as coloured surfaces while the exact dispersion relation is drawn as a wire frame. The energy range drawn here goes from -2 eV to 2 eV . As can be seen, the cone approximation starts to lose its accuracy at around $\pm 1\text{ eV}$.

n_G and p_G (in cm^{-2}) and are given by:

$$\begin{aligned} n_G &= \int_{E_{\text{Dirac}}}^{\infty} D_G(E) f_{E_{FG}}(E) dE \\ &= \frac{2}{\pi} \left(\frac{k_B T}{\hbar v_F} \right)^2 \int_0^{\infty} \frac{x}{1 + \exp\left(x - \frac{\epsilon_{FG}}{k_B T}\right)} dx, \end{aligned} \quad (6.3a)$$

$$\begin{aligned} p_G &= \int_{-\infty}^{E_{\text{Dirac}}} D_G(E) f_{E_{FG}}(E) dE \\ &= \frac{2}{\pi} \left(\frac{k_B T}{\hbar v_F} \right)^2 \int_0^{\infty} \frac{x}{1 + \exp\left(x + \frac{\epsilon_{FG}}{k_B T}\right)} dx, \end{aligned} \quad (6.3b)$$

with $f_{E_{FG}}(E)$ the Fermi distribution function. At $T = 300 \text{ K}$, the prefactor has the value $(2/\pi)(k_B T)^2/(\hbar v_F)^2 = 9.82 \cdot 10^{10} \text{ cm}^{-2}$. Appendix 6.A discusses how the integrals in Eqs. (6.3) can be approximated analytically and provides accurate piecewise approximations for n_G and p_G in Eqs. (6.29). However, for many practical purposes it suffices to approximate the Fermi distribution as a step function to obtain the density of the dominant carriers. This density is denoted by n (i.e., $n = n_G$ when $\epsilon_{FG} > 0$ and $n = p_G$ when $\epsilon_{FG} < 0$) and is given by:

$$n \approx \frac{1}{\pi} \left(\frac{\epsilon_{FG}}{\hbar v_F} \right)^2 \approx 7.35 \cdot 10^{13} \text{ cm}^{-2} \left(\frac{\epsilon_{FG}}{1 \text{ eV}} \right)^2. \quad (6.4)$$

This commonly used approximation has a relative error of $< 10\%$ for $|\epsilon_{FG}| > 5.5 k_B T$ and an error of $< 1\%$ for $|\epsilon_{FG}| > 18.1 k_B T$ compared to the exact integrals.

With $\epsilon_{FG} = 0$ in Eqs. (6.3), the intrinsic carrier density $n_{i,G}$ of graphene can be calculated to be:

$$n_{i,G} = \frac{\pi}{6} \left(\frac{k_B T}{\hbar v_F} \right)^2 = 8.077 \cdot 10^{10} \text{ cm}^{-2}. \quad (6.5)$$

Note that the intrinsic carrier density is of much more interest to semiconductors than it is to graphene. In semiconductors it is possible that both the electrons and holes follow Boltzmann statistics (i.e., when the Fermi level is midgap) so that the law of mass-action is valid: $np = n_i^2$. However, this is not the case for graphene since graphene is always degenerate for at least one of the two carrier types. So due to the fact that it does not have a bandgap there exists no mass-action law for graphene (see also the discussion about intrinsic carrier density in doped semiconductors in Section 3.4).

From the DOS it can be seen that graphene has one forbidden energy level at $\epsilon = 0$, so it can be said that graphene almost has a bandgap. However, a

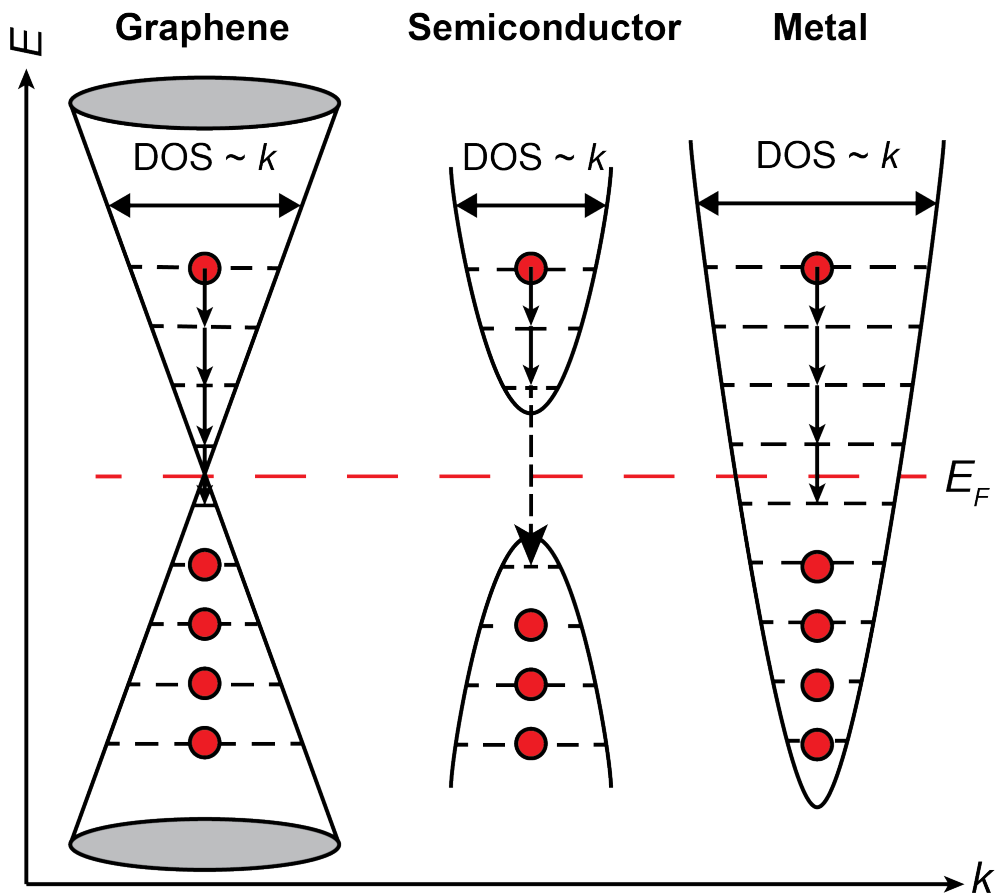


Figure 6.2: The (2D) Dirac cone compared to the (3D) parabolic dispersion relations of a metal and a semiconductor. In all 3 systems, the DOS is proportional to the width of the cone/parabola (though this is a happy coincidence for these particular systems and not a universal law). The Fermi level in the graphene shown in this figure corresponds to the intrinsic level where the numbers of free electrons and holes are equal. In graphene, like in metals, excited electrons can quickly relax back to the lowest available state by passing through intermediate empty states. In semiconductors, excited electrons relax relatively quickly to the bottom of the conduction band. Since the relaxation to the valance band involves a single large jump in energy, this process is much slower than the intraband relaxation time. This is the reason that Fermi level splitting is possible in semiconductors but not in graphene or in metals.

proper semiconductor has a range of forbidden energies, so graphene is sometimes called a semimetal. Like a real metal, graphene cannot sustain Fermi level splitting: any excited electron will quickly decay back through the intermediate states to its equilibrium energy level as shown in Figure 6.2 (see also the discussion of Fermi level splitting in Section 3.2). This means that graphene is unsuited as an absorber material for solar cells. Unlike a metal, though, the Fermi level in graphene is not fixed. In a metal the DOS is very large at the CNP, meaning that a shift in the Fermi level will be accompanied by a very large change in charge in the material. In intrinsic graphene, on the other hand, the DOS vanishes at the CNP so that ϵ_{FG} can change by a significant amount without adding much charge to it. Conversely, adding charge to intrinsic graphene can shift ϵ_{FG} up or down by several tenths of an eV. At room temperature, for example, a Fermi level of $\epsilon_{FG} = \pm k_B T = 25.9$ meV results in a charge of $\mp 1.44 \cdot 10^{11} \text{ cm}^{-2}$. At $\epsilon_{FG} = \pm 10 k_B T = 0.259$ eV the charge is about $\mp 5.1 \cdot 10^{12} \text{ cm}^{-2}$, which is comparable to the fixed charge density in a typical Al_2O_3 passivation film.

The act of shifting ϵ_{FG} away from its intrinsic level, thereby increasing the carrier density, will be referred to as doping. Doping of graphene can be achieved in different ways and often happens unintentionally because the 2D nature of the material makes it very sensitive to influences from its surroundings. For example, fixed charge can be added in the form of adsorbed ions to the material, forcing ϵ_{FG} to shift to restore the charge balance. Nitric acid is a well-known and commonly used adsorbate that can achieve *p*-type doping [16, 7], though it is difficult to keep ions adsorbed to the graphene permanently due to its chemical inertia, making chemical doping of graphene an active field of research [17]. Contaminating species from the environment (such as CO_2 , H_2O and NO_2 [18]) and the processing of the material are also known to cause intentional or unintentional doping [19]. The Fermi level in graphene can also be shifted by putting it on a dielectric substrate (such as a silicon wafer with a thermal oxide film) and then gating it with an external voltage source. When the dielectric is thick enough, the is configuration will behave like an ordinary (i.e., linear) capacitor so that the charge in the graphene can be calculated from the gate voltage in a straightforward manner from the oxide's thickness and dielectric constant [20]. This method can be used to measure the adsorbed charge density and mobility of the graphene [18] and provides a simple method to dope graphene while performing optical measurements [21]. Another possible doping source is the substrate of the graphene itself as will be discussed in Section 6.2.

The question then rises what limits there are to graphene doping. Though this question is difficult to answer, carrier densities over 10^{13} cm^{-2} [22, 23] and up to $5 \cdot 10^{13} \text{ cm}^{-2}$ have been observed in graphene [24]. The Dirac cone approximation for the graphene dispersion relation breaks down at about $|\epsilon_{FG}| = 1$ eV (see also Figure 6.1c) and this corresponds to a carrier density

Table 6.1: Overview of relevant electrical properties of graphene, semiconductors and metals. In practice, metals tend to have a more complex dispersion relation, but for the arguments in this work it is sufficient to regard metals as an ideal degenerate electron gas.

Material	Dimension	Bandgap	Fermi level splitting	E_F	Dispersion
Graphene	2D	Almost	No	Variable	$\hbar v_F k$
Semiconductor	3D	Yes	Yes	Variable	$\frac{\hbar^2 k^2}{2m^*}$
Metal	3D	No	No	Fixed	$\frac{\hbar^2 k^2}{2m^*}$

of $n = 7.4 \cdot 10^{13} \text{ cm}^{-2}$, so this number can be used as an approximate upper limit for graphene doping.

Table 6.1 presents a brief overview of the most important electrical properties of graphene as well as the most important differences between graphene, metals and semiconductors.

6.1.2 Conduction and photon absorption in graphene

For assessing the electrical and optical losses in graphene, the main quantity of interest is the frequency-dependent (dynamic) complex conductivity³ $\sigma(\omega)$. Just like the static (DC) conductance $\sigma_{\text{DC}} = \lim_{\omega \rightarrow 0} \sigma(\omega)$, the dynamic conductivity is the ratio between the in-plane electric field $\mathbf{E}_{\parallel}(\omega)$ and the current density (unit: A m^{-1}) in the graphene $\mathbf{J}(\omega)$:

$$\mathbf{J}(\omega) = \sigma(\omega) \mathbf{E}_{\parallel}(\omega). \quad (6.6)$$

The real part of σ is associated with currents that are in phase with the electric field (which are dissipative) while the imaginary part of σ is associated with out-of-phase currents (which do not dissipate). The frequency dependence of $\sigma(\omega)$ can be measured by spectroscopy (see, e.g., Horng *et al.* [21]; Yan *et al.* [23]; and Weber *et al.* [25]) and at wavelengths relevant for a silicon solar cell, σ is a universal constant $\sigma_{\text{AC}} = e^2/(4\hbar) = 6.09 \cdot 10^{-5} \Omega^{-1}$. Due to the unique conical graphene dispersion relation, this constant is independent of ω ; temperature; the graphene carrier density; and σ_{DC} [10]. The optical

³Because graphene is a 2D material, it is possible to measure the electrical conductivity and resistivity in S and Ω (i.e., the sheet conductance and sheet resistance). There is then no real distinction between conductance and conductivity or between resistance and resistivity and the terms can then be used interchangeably. This is the convention that will be used here. Some authors (such as Weber *et al.* [25]) choose to express the conductivity of graphene in S m^{-1} by normalising to the graphene thickness $d_{\text{G}} = 0.335 \text{ nm}$.

conductivity σ_{AC} can be used to calculate the transmission through a sheet of graphene, but in general this transmission depends also on the design of the cell. Backside reflections from the rear of the cell and the presence of an anti-reflective coating (either on top of or underneath the graphene) can all influence how much light the graphene absorbs and reflects. So while the optical conductivity of graphene is a simple constant, its reflectance and transmittance depend on the context and can in general only be calculated exactly by solving the Fresnel equations. Fortunately, for certain geometries there exist simple approximations. For example, for a sheet of graphene on a bare substrate the transmission can be calculated to be approximately: [25]:

$$\frac{T_{sG}}{T_s} \approx 1 - \pi\alpha \left(\frac{1}{2} + \frac{1}{2(n_s^2 + 1)} \right) = 1 - 0.0229 \left(\frac{1}{2} + \frac{1}{2(n_s^2 + 1)} \right). \quad (6.7)$$

Here, the fraction T_{sG}/T_s is the transmission through the substrate + graphene normalised to the transmission through the substrate in the absence of the graphene; n_s is the substrate refractive index; and $\alpha \approx 1/137.0$ is the fine-structure constant. Thus, for silicon ($n_s \approx 3.44$) the transmission loss in the visible part of the spectrum due to the presence of a sheet of graphene is about 1.2%.

When looking at the dynamic conductivity in the lower part of the spectrum, it is possible that σ shows a feature that resembles a bandgap. This happens if the Fermi level is shifted far away from its intrinsic position (i.e., $|\epsilon_{FG}| \gg k_B T$). In that case there are no allowed interband transitions at $\omega < 2\epsilon_{FG}$ due to a phenomenon called *Pauli blocking* (see Figure 6.3) and should not be misinterpreted as evidence of an electrical bandgap in the graphene. Because the Pauli blocking bandgap is related to ϵ_{FG} , this effect provides a useful method to independently measure the Fermi level and free carrier density in the graphene.

6.1.2.1 The optical Drude response of graphene and DC conductivity

Going to even lower frequencies in the infrared (IR), the optical conductivity shows a pronounced Drude absorption peak due to intraband transitions. These intraband transitions are also referred to as free carrier scattering because they mainly involve energy levels close to the surface of the Fermi sea, which are also the states involved in DC conduction (see Figure 6.3). The Drude contribution to σ (denoted by σ_{Drude}) has the following form [10, 21, 25]:

$$\sigma_{\text{Drude}}(\omega) = \frac{iD}{\pi(\omega + i\tau^{-1})} = \frac{ie^2|\epsilon_{FG}|}{\pi\hbar^2(\omega + i\tau^{-1})} \approx \frac{ie^2v_F\sqrt{n}}{\sqrt{\pi}\hbar(\omega + i\tau^{-1})}. \quad (6.8)$$

Here, e is the elementary charge; τ is the scattering time or inverse scattering rate Γ^{-1} (Γ is also referred to as the Drude broadening parameter or scattering

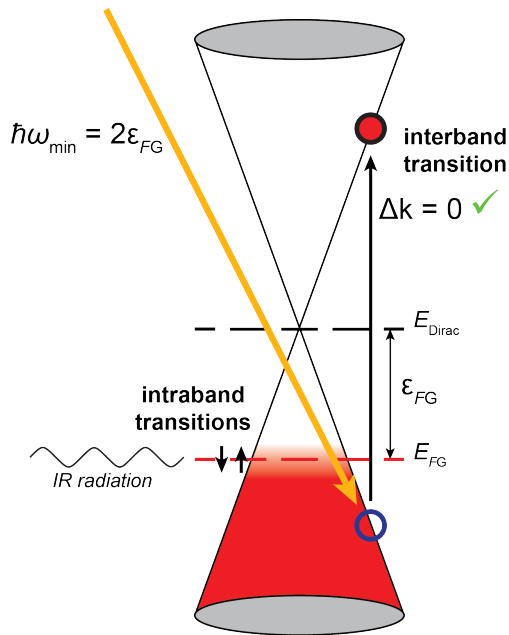


Figure 6.3: Optical transitions in graphene. Interband transitions excite an electron with energy $\epsilon < 0$ to a state $\epsilon > 0$ and are only likely if the \mathbf{k} vector of the electron is conserved. Transitions that do not conserve \mathbf{k} are only possible if the difference in momentum is compensated for by, e.g., a phonon and are therefore significantly less probable. Because of \mathbf{k} conservation, only photons with energy $> 2\epsilon_{FG}$ can be absorbed in an interband transition. Thus, the absorption spectrum of graphene does show characteristics of an optical bandgap, even though it does not have an electrical bandgap. Intraband absorption in graphene is dominated by states close to the Fermi level because those transitions only require small changes in \mathbf{k} . Intraband absorption therefore most pronounced in the IR part of the spectrum and leads to a characteristic Drude-type feature in the absorption spectrum. This Drude peak can be used to measure the carrier density and mobility of the graphene [21].

width). The Drude weight D measures the strength of the interaction between the light and the free carriers⁴. Strictly speaking, Eq. (6.8) is only true when $\epsilon_{FG} \gg k_B T$, but in literature a general Drude form can be found that is also valid at low doping levels [10].

In the $\omega \rightarrow 0$ limit, the Drude conductivity becomes the static conductivity (i.e., inverse sheet resistance) $\sigma_{DC} = D\tau/\pi$ that measures how well the graphene can be used as a transparent electrode. Horng *et al.* [21] showed that the sheet resistance of graphene can be successfully measured by either four point probe measurements or by optically measuring the Drude weight D : both methods are equally accurate, but the optical method yields information about both the free carrier density and the scattering rate (i.e., about the mobility).

The Drude absorption peak in the IR is of little consequence for the application of graphene in solar cells because the peak is so narrow (typically less than 0.3 eV) that it does not affect photons with energies above the bandgap of silicon. This is in stark contrast to traditional TCOs, in which the Drude free carrier absorption and reflection near the silicon bandgap can be substantial. Traditional TCOs therefore have to compromise between a low resistivity (which requires a high free carrier density) and low absorption/reflection (which requires a low free carrier density). Only by lowering the scattering rate (and increasing the mobility) is it possible to escape this compromise, though the doping of a TCO will always result in ionised impurity scattering, which (together with phonon scattering) imposes an absolute upper limit on the mobility a TCO can achieve. Indeed, for indium oxide, this upper limit has almost been achieved [28, 29]. For graphene, on the other hand, free carrier absorption is of no consequence to a solar cell due to its very low intrinsic scattering rate; only the interband transitions matter.

It is interesting to note that the conductivity in graphene is not proportional to n , like in most materials, but to $|\epsilon_{FG}| \sim \sqrt{n}$. To explain this dependence it is necessary to take a closer look at the Drude model in general. This model is commonly used to describe the conductivity of metals and semiconductors, both degenerate and nondegenerate) and provides an easy method to measure the mobility and carrier density of a material [30]. In the Drude model the (free) electrons are considered to be an ideal gas so that the energy of the electrons is assumed to be purely kinetic: $E = \hbar^2 k^2 / (2m^*)$, with m^* is the effective mass of the electrons in the lattice of the material. If the electrons have an effective scattering rate τ , then the DC conductivity of the material

⁴In practice, it turns out that the measured Drude weight is lower than $e^2 |\epsilon_{FG}| / \hbar^2$ [21]. It would go too far to explain this phenomenon, but the basic idea is that, due to many-body effects, the scattering electrons experience a lower Fermi velocity (about $0.85 \cdot 10^6$ m s⁻¹) than the DOS Fermi velocity [26, 27]

σ_{DC} is⁵ (see, e.g. [32]):

$$\sigma_{\text{DC}} = e\mu n = \frac{e^2\tau}{m^*}n. \quad (6.9)$$

This result is very useful, but it gives the false impression that all free electrons contribute to conduction, which is only the case for a nondegenerate electron gas (such as the free electrons in lowly doped *c*-Si). In a degenerate electron gas, however, only the electrons at the surface of the Fermi sea have the freedom to contribute to conduction. The reason n appears in Eq. (6.9) is because of two separate factors that together result in the proportionality $\sigma_{\text{DC}} \sim n$. Heuristically, σ_{DC} is proportional to the number of electrons at the Fermi surface N_{surf} and to their velocity (i.e. the Fermi velocity v_F):

$$\sigma_{\text{DC}} \sim N_{\text{surf}}v_F. \quad (6.10)$$

By the same familiar argument that was used in Section 6.1.1, the electrons in a metal or semiconductor of volume V are completely delocalised. The uncertainty principle then states:

$$(\Delta x)^3(\Delta k)^3 = V(\Delta k)^3 = \left(\frac{1}{2\pi}\right)^3. \quad (6.11)$$

For an ideal degenerate 3D electron gas with N particles at low temperature, the Fermi wave vector k_F is given by the radius of the k sphere that contains N states. Using Eq. (6.11), k_F can be calculated (taking into account the spin degeneracy of 2):

$$N = 2\frac{(4/3)\pi k_F^3}{\Delta k^3} = \frac{1}{3\pi^2}k_F^3V. \quad (6.12)$$

Thus, the well-known result $k_F = \sqrt[3]{3\pi^2n}$ is obtained. The same argument can be used to show that the number of electrons at the Fermi surface is proportional to the Fermi surface area $4\pi k_F^2$. The velocity of these surface electrons is proportional to the Fermi wave vector, so N_{surf} and v_F scale as:

$$N_{\text{surf}} \sim k_F^2 \sim n^{2/3}, \quad (6.13)$$

$$v_F = \left.\frac{\partial\omega}{\partial k}\right|_{E=E_F} = \frac{\hbar k_F}{m^*} \sim n^{1/3}. \quad (6.14)$$

Hence, $\sigma_{\text{DC}} \sim N_{\text{surf}}v_F \sim n$ as required by the Drude model, but not because all electrons contribute to conduction. To put it strongly: the proportionality $\sigma_{\text{DC}} \sim n$ is merely a coincidence. There is therefore no a-priori reason why

⁵More generally, the dynamic conductivity in the simple Drude model is $\sigma(\omega) = i\sigma_{\text{DC}}/(\omega + i\tau^{-1})$ [31].

$\sigma_{\text{DC}} \sim n$ should be true for graphene, which has a very different kind of dispersion relation. Indeed, due to the linear dispersion relation $\epsilon = \hbar v_F k$ in graphene the Fermi velocity is constant and the number of electrons at the Fermi surface (which is a circle rather than a spherical shell) go as $N_{\text{surf}} \sim k_F \sim |\epsilon_{FG}| \sim \sqrt{n}$. Hence, the conductivity of graphene is also proportional to $|\epsilon_{FG}|$ and \sqrt{n} . To derive Eq. (6.8), it is necessary to resort to calculations based on, e.g., the Boltzmann transport equation such presented in Ibach and Lüth [32], but this is not in the scope of the present work. What is important to remember is that heavy doping of graphene is a somewhat inefficient method to increase the conductivity because of the \sqrt{n} law.

Even though the graphene DC conductivity is no longer proportional to n , it is still possible to define a (n -dependent) mobility for graphene as $\sigma_{\text{DC}} = en\mu$. This mobility is given by:

$$\mu = \frac{\sigma_{\text{DC}}}{en} = \frac{ev_F\tau}{\hbar\sqrt{\pi n}} = \frac{ev_F^2\tau}{\epsilon_{FG}}. \quad (6.15)$$

Mobility values for graphene are commonly quoted in literature, but unfortunately they can be difficult to compare due to the fact that μ is not constant with doping. The scattering rate, on the other hand, is found to be constant with ϵ_{FG} to a reasonable degree of accuracy [33], especially for p -type graphene [21], making it a convenient independent transport parameter.

Eq. (6.15) can also be compared with the common expression $\mu = e\tau/m^*$ to define an effective mass for graphene (m_G) as:

$$\epsilon_{FG} = m_G v_F^2. \quad (6.16)$$

Eq. (6.16) is remarkably simple to remember since it closely resembles the relativistic equation for the equivalence between mass and energy $E = mc^2$. This result can thus be interpreted as if the electrons in the graphene have a mass determined by their energy. The mass relevant for Drude scattering is therefore the mass equivalent to the Fermi energy, since the electrons at the Fermi level are the ones responsible for transport phenomena. In fact, the graphene effective mass has other applications as well. For example, it is also the effective mass that is measured in cyclotron measurements [27]. One should, however, be careful to not overinterpret these ideas. The effective mass m_G has a very different meaning from the usual effective mass m^* defined for ordinary semiconductors, which is directly related to energy-momentum dispersion relation.

6.1.2.2 Prospects of graphene as a transparent electrode for c -Si

For a solar cell, the transmission losses through a sheet of graphene (calculated earlier to be on the order of a percent) are on an acceptable level compared

to other available transparent conductors, but still significant enough that it quickly becomes unattractive to stack multiple sheets of graphene on top of one another (such as done by, e.g., Yan *et al.* [23]). This means that if graphene is used as a transparent electrode, no more than a few monolayers (and preferably just one) can be used to achieve the sheet resistance desired for the cell. It should also be kept in mind that graphene cannot act as an anti reflective coating (ARC) like a transparent conductive oxide (TCO) can, so the graphene will have to be combined with a light trapping scheme; preferably one that is industrially viable. Given the fact that it is no straightforward task to deposit materials on top of graphene (and even if it is possible, those films are likely to degrade the graphene or at least change its properties), realising such a scheme imposes significant practical design limitations. Mie resonators [34] are an attractive option to realise a light trapping scheme on top of graphene, since it does not require the silicon surface to be textured and only needs a relatively limited amount of the cell surface to be covered with extra material to minimise the reflectance.

To assess the conductance potential of a single sheet graphene, it is instructive to compare the Drude theory for graphene with experimental data. In the work by Horng *et al.* [21], the scattering rate $\Gamma = \tau^{-1}$ and the Drude factor D have been measured optically. They found that the scattering rate $\Gamma(\omega)$ is approximately constant for p -type doping, meaning that τ does not depend on the position of the Fermi level ϵ_{FG} . For n -type doping, however, it was also found that $\Gamma(\omega)$ increases significantly with the doping density, making n -type graphene less attractive as a transparent electrode. With these data it becomes possible to address the question if graphene can be a good transparent electrode. For c -Si solar cells, a typical acceptable upper limit for σ_{DC} is about $(100\ \Omega)^{-1} = 1.0 \cdot 10^{-2}\ \text{S}$ [35].

From the work by Horng *et al.* it can be readily verified that -as predicted by theory- the Drude factor scales as $D \sim \sqrt{p}$ for p -doped graphene. In Figure 6.4, their data have been fitted by plotting D^2 against p . However, since D is not a very practical number, it has been divided by $\pi\Gamma_0$, where $\Gamma_0 = 105\ \text{cm}^{-1}$ is approximately the broadening parameter they found. By doing so, the unit of the y axis becomes that of a squared sheet conductance that can be directly compared to the conductance goal $\sigma_{DC}^2 = 1 \cdot 10^{-4}\ \text{S}^2$. As can be seen, the trend is linear and with the obtained fit it is possible to make an estimate for the required hole density p needed to obtain the desired sheet conductance. By extrapolating of the fit (and thus assuming that Γ remains constant), a value $p_{100\ \Omega} = 2.5 \cdot 10^{14}\ \text{cm}^{-2}$ is obtained. In practise, such a high doping density has not been achieved (as discussed previously in Section 6.1.1), so a scattering width of $105\ \text{cm}^{-1}$ is most likely too high to achieve the desired sheet resistance. However, it is known that the substrate of the graphene can affect the scattering of free carriers [36, 33] and SiO_2 (the substrate used by Horng *et al.*) is far from ideal in this regard. Lower scattering rates have been found by, e.g., Yan

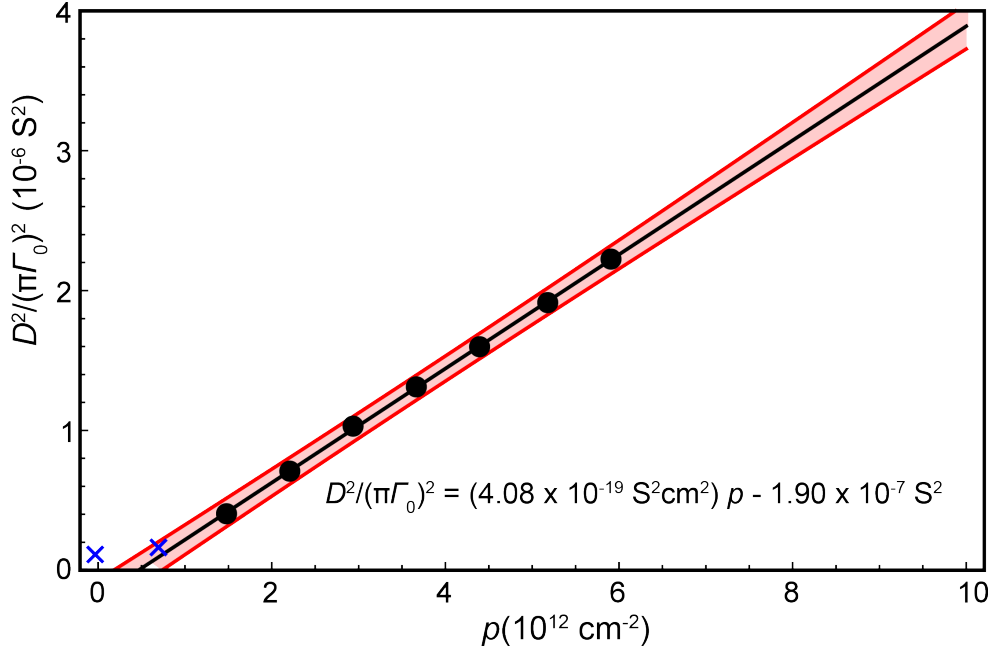


Figure 6.4: Fit of the square of the Drude factor D (expressed as a conductance $(D/(\pi\Gamma_0))^2$) against the hole concentration (assuming a constant scattering rate $\Gamma_0 = 105 \text{ cm}^{-1}$). Data from the work by Horng *et al.* [21]. The blue crosses indicate data points that were excluded from the fit, because at low carrier densities σ_{DC} deviates from the \sqrt{p} law [10]. The bands indicate 99.9% single observation prediction confidence intervals based on the quality of the fit. Note that the authors of the work report an approximate experimental uncertainty of 10% which is not incorporated into the fit.

et al. [23], who report a scattering rate of about 50 cm^{-1} for graphene on a polymer buffer layer. With this value of Γ the critical doping density comes down to $p = 5.6 \cdot 10^{13} \text{ cm}^{-2}$, which is a feasible yet still a significant amount. The question rises how much electron scattering can realistically be suppressed. According to Chen *et al.*, [37], graphene could achieve a sheet resistance of 30Ω at room temperature if all scattering mechanisms other than intrinsic phonon scattering are absent. However, if also surface phonon interaction with an SiO_2 substrate is included this limit shifts up to about 150Ω , showing that the 100Ω goal relevant for solar cells is close to what is theoretically (let alone practically) feasible. Furthermore, the authors indicate that even for pristine exfoliated graphene cleaned and measured under ultra high vacuum conditions they still found a very significant amount of impurity scattering. They attribute this to charged impurities inside of the SiO_2 substrate, the electric fields of which disturb the transport in the graphene. This illustrates that even if it would be possible to process very clean and defect-free graphene on an industrial scale, the substrate underneath the graphene can still substantially degrade its conductance. In other words, if graphene is to be used as a transparent electrode, it is very likely that the rest of the cell needs to be engineered around this idea.

6.2 Junctions of graphene and silicon

Having established the basic properties of graphene, the next step is to review how these properties come into play when a junction is made between graphene and *c*-Si (or another semiconductor similar to silicon). Superficially, such a junction is similar to an ordinary Schottky junction in which a metal fixes the Fermi level at the edge of the silicon and thereby induces a space charge region (SCR) near the interface. However, in graphene the Fermi level is not fixed so the graphene does not only affect the SCR in the *c*-Si, but the latter will, in turn, affect the former as well. Moreover, the behaviour of charge transport between the two materials is expected to be different than in an ordinary Schottky junction due to graphene's unique dispersion relation and DOS.

The simplest way to illustrate the interaction between graphene and silicon is to first look at the equilibrium band diagram, which is the subject of Section 6.2.1. After that, the discussion will be expanded to the non-equilibrium case in Section 6.2.2. Finally, the theory that has been presented will be used to discuss the relevance of graphene-semiconductor (GS) junctions for the field of silicon photovoltaics in Section 6.2.3

6.2.1 Equilibrium conditions

To give a heuristic picture of a GS junction, it is useful to first consider the two systems separately from each other; particularly the positions of the Fermi levels of both systems. It was already discussed in Section 6.1.1 that for intrinsic graphene the Fermi level E_{FG} is at $-\phi_G = -4.5$ eV (measured from the vacuum level). In silicon the equilibrium Fermi level depends on the doping type and amount. If typical $3\ \Omega\text{cm}$ wafers are used, the Fermi levels for n -type and p -type silicon is at $E_F = -4.30$ eV and $E_F = -4.95$ eV respectively. The conduction and valence bands of silicon are at $E_C = -\chi_{\text{Si}} = 4.05$ eV and $E_V = -(\chi_{\text{Si}} + E_{G,\text{Si}}) = 5.17$ eV, with χ_{Si} the silicon electron affinity and $E_{G,\text{Si}} = 1.124$ eV the bandgap. Thus, when graphene is put directly on silicon, the Dirac point is close to the midgap energy of silicon $E_{\text{mid,Si}} = -\chi_{\text{Si}} - E_{G,\text{Si}}/2 = -4.612$ eV. Furthermore, the CNP of graphene is lower than the intrinsic Fermi level of n -type silicon, but higher than that of p -type silicon. When the silicon and the graphene are brought in contact, the difference in intrinsic E_F will result in a temporary flow of electrons (holes) from high (low) E_F to low (high) E_F . Thus, n -type silicon will donate electrons to intrinsic graphene, while p -type silicon will donate holes. This redistribution of electrons and holes will charge the graphene and change its Fermi level, while in the silicon a SCR will form and the bands will bend. Since the intrinsic graphene has a larger free carrier density after the charge redistribution, this effect is often referred to as *substrate doping*. Once equilibrium has been reached, the surface charge in the graphene is mirrored by an opposite charge in the SCR and the Fermi level becomes constant throughout the whole structure so that no more charge flow is possible. For example, it can be calculated that intrinsic graphene on n -type and p -type $3\ \Omega\text{cm}$ silicon has a carrier density of respectively $n_G = 5.8 \cdot 10^{10}\text{ cm}^{-2}$ (corresponding to $\epsilon_{FG} = 10.9$ meV) and $p_G = 1.56 \cdot 10^{11}\text{ cm}^{-2}$ (corresponding to $\epsilon_{FG} = -27.8$ meV).

As can be seen from the example in the previous paragraph, the graphene acts like a surface charge that influences the semiconductor substrate underneath it. The semiconductor, in turn, influences the Fermi level in the graphene and thereby the surface charge in the graphene. The equilibrium band diagram of a junction between a semiconductor and graphene can therefore be calculated by using the facts that the structure is globally charge neutral and that the Fermi level is straight (see also Zhong *et al.* [38]) throughout. The graphene charge consists of the free carrier density $Q_{G,\text{free}} = p_G - n_G$ as well as any adsorbed ionic species. These adsorbed ions will be referred to as chemical doping (as is common practise in literature⁶) and the doping density on the graphene will be denoted by $N_{\text{dop,G}}$, which can be positive or negative and is expressed in units of cm^{-2} . The total graphene charge is $Q_G = Q_{G,\text{free}} + N_{\text{dop,G}}$. It is

⁶Though this terminology is slightly misleading because chemical doping does not necessarily involve chemical bonds but can also be due to physisorption of ions.

assumed here that doping of the graphene is ideal, meaning that the dopants are considered to be adsorbed ions that do not affect the system in any way other than through their charge. It is also assumed that the charge state of the dopants is constant (i.e., the position of the Fermi level does not affect the charge of the dopants).

To calculate an equilibrium band diagram for a general semiconductor device with a graphene layer, it is sufficient to treat the graphene as a variable surface charge that depends on the surface potential and Fermi level of the device. Since most semiconductor modeling tools allow for the insertion of surface charges, this makes it possible to do these calculations without having to model the graphene in more detail. To find the graphene charge Q_G , the following procedure can be iterated:

1. First, a guess is made for the charge Q_{in} and inserted into the device simulator.
2. To check if the guess Q_{in} is correct, the simulation result is used to read off the Fermi level and the position of the conduction band E_C at the surface. Next ϵ_{FG} in the graphene is calculated from these quantities (see also Figure 6.5) to obtain the corresponding surface charge $Q_{\text{out}} = p_G(\epsilon_{FG}) - n_G(\epsilon_{FG}) + N_{\text{dop,G}}$. If the initial guess Q_{in} was correct, then $Q_{\text{in}} = Q_{\text{out}}$ and the desired graphene charge Q_G has been found; otherwise a new guess needs to be made.
3. To find Q_G as quickly as possible, various root finding tricks can be used. For example, one can plot the pairs $\{Q_{\text{in},i}, Q_{\text{out},i}\}$ (where i is the iteration number) in a graph and interpolate to find the best next guess for Q_{in} . From experience it was found that iterating $Q_{\text{in},i+1} = Q_{\text{out},i}$ does not lead to quick convergence (or even to divergence) and is therefore not recommended.

Alternatively, if the program does not support surface charge boundary conditions (BCs), then one can guess the surface Schottky barrier Φ_{in} (i.e., the value of $E_C - E_F$ at the surface). From the simulation results it is then possible to deduce the surface charge from the electric field at the surface by:

$$Q_{\text{out}} = -\frac{\epsilon_r \epsilon_0 \mathbf{E} \cdot \hat{\mathbf{n}}}{e}, \quad (6.17)$$

with $\mathbf{E} = -\nabla\phi$ the electric field and ϕ the electrostatic potential (which is defined as $\phi = -E_{\text{vac}}$, i.e., minus the vacuum energy level) at the surface; ϵ_0 and ϵ_r the vacuum permittivity and relative dielectric constant; and $\hat{\mathbf{n}}$ the unit outward normal vector⁷. From the obtained value of Q_{out} the graphene Fermi

⁷On a 1D interval $[0, L]$, $\hat{\mathbf{n}}(L) = 1$ and $\hat{\mathbf{n}}(0) = -1$

level and energy barrier Φ_{out} can be calculated and compared with Φ_{in} . The procedure is then iterated until the stabilising value $\Phi_{\text{in}} = \Phi_{\text{out}}$ has been found.

Sometimes the only quantities of interest are the amount of band bending in the substrate and the carrier density of the graphene. In those situations the integrated Poisson-Boltzmann (PB) equation (Eq. (A.13) in Appendix A.2) provides a simple method for finding the correct charge balance in the structure. The amount of charge in the semiconductor and the graphene can be considered functions of the band bending $\Delta\phi$ and the neutrality requirement can then be used to find the value of $\Delta\phi$ that satisfies:

$$Q_G(\Delta\phi) + Q_{\text{substrate}}(\Delta\phi) = 0. \quad (6.18)$$

This equation can, in general, only be solved approximately by a numerical root finding algorithm, but since the value of $\Delta\phi$ is usually between -1 eV and 1 eV the root should be easy to find for a computer.

Examples are given in Figures 6.5 and 6.6, which show equilibrium band diagrams of graphene on n -type silicon that were obtained using the integrated PB equation. In Figure 6.5 the graphene has been chemically doped p -type by adsorbed negative ions ($N_{\text{dop,G}} = -5 \cdot 10^{12} \text{ cm}^{-2}$), the charges of which are mirrored by positive charges in both the graphene and the silicon. The hole density in the graphene is $p_G = 4.9 \cdot 10^{12} \text{ cm}^{-2}$ here, so almost all of the negative charge is mirrored in the graphene. In Figure 6.6 a thin Al_2O_3 film has been inserted between the graphene and c -Si and this material is known to have a high density of fixed negative charge Q_f when deposited on silicon [39, 40]. These charges can, in principle, also be used as a stable source of doping for graphene. In Figure 6.6 the assumed charge density is $Q_f = -5 \cdot 10^{12} \text{ cm}^{-2}$, the same value as that for $N_{\text{dop,G}}$ in Figure 6.5. However, the presence of the dielectric and the closer proximity of the fixed charges to the c -Si result in a lower doping of the graphene of $p_G = 4.1 \cdot 10^{12} \text{ cm}^{-2}$ and more mirror charges in the c -Si SCR.

One could ask how effective this fixed charge doping scheme for graphene can be. Figure 6.7 shows a graph of the carrier density in graphene as a function of the fixed charge in the dielectric for a similar configuration as in Figure 6.6. The silicon here is assumed to be an n -type $3 \Omega \text{ cm}$ wafer. As can be seen, the doping efficiency decreases sharply with the thickness of the dielectric (which has been expressed as equivalent SiO_2 thickness or EOT) and with the fixed charge density, making it difficult to obtain carrier densities greater than $5 \cdot 10^{12} \text{ cm}^{-2}$ for typical fixed charge densities $|Q_f| < 1 \cdot 10^{13} \text{ cm}^{-2}$.

6.2.2 Non-equilibrium conditions

When non-equilibrium operating conditions are considered, there is no longer a single, constant Fermi level throughout the whole device. In this situation, it

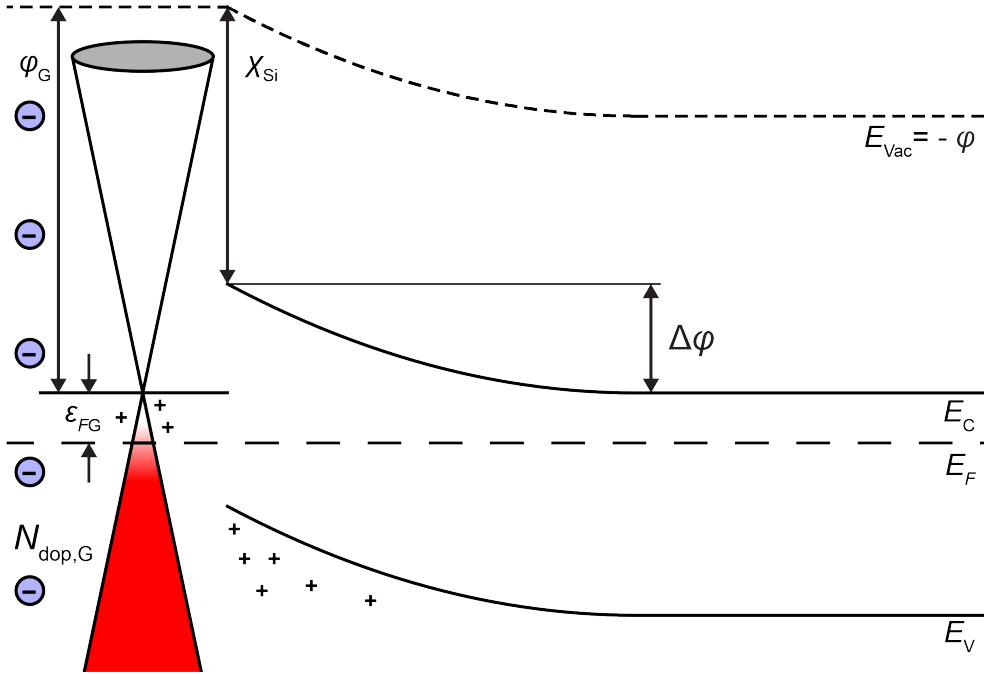


Figure 6.5: Equilibrium band diagram of graphene on silicon calculated using the integrated PB equation. The negative charge of the physisorbed ions (left, $N_{\text{dop,G}} = -5 \cdot 10^{12} \text{ cm}^{-2}$) is compensated by an induced positive charge in the graphene and silicon. It should be noted that the graphene is represented by its Dirac cone in k -space; a convention often found in literature. This makes the graphene readily identifiable and conveniently shows the location of the Dirac point. The hole density in the graphene is $p_G = 4.9 \cdot 10^{12} \text{ cm}^{-2}$ in this situation, so the negative charge is mirrored mainly in the graphene.

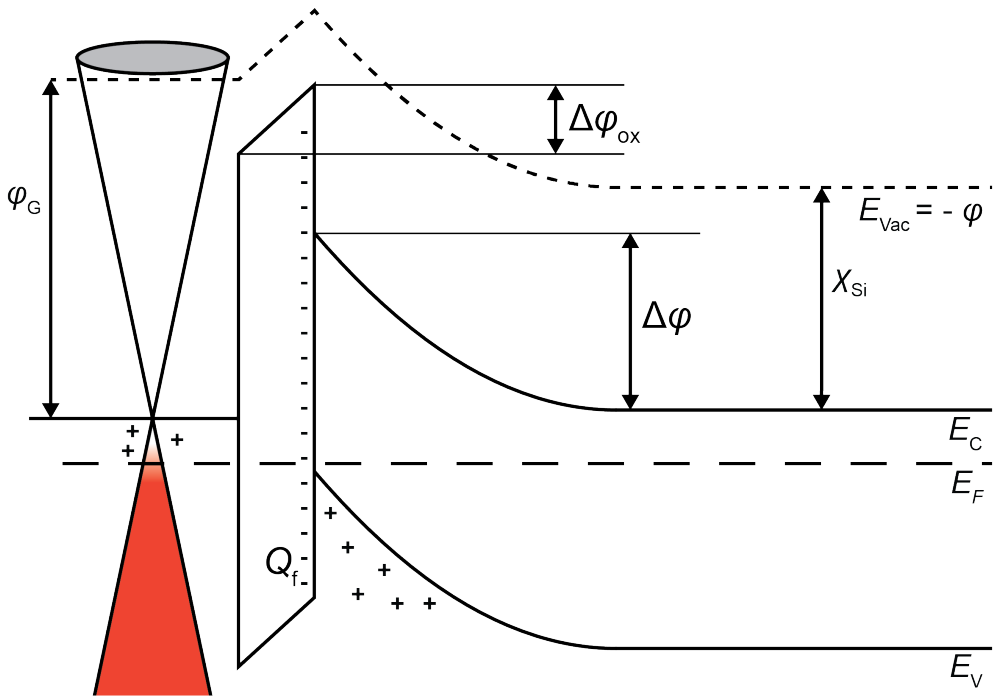


Figure 6.6: Equilibrium band diagram of graphene on n -type silicon ($3\ \Omega\text{ cm}$) with Al_2O_3 interlayer. In this situation, there is no negative charge present of the physisorbed ions, but now the fixed charges at the $c\text{-Si-Al}_2\text{O}_3$ interface ($-5 \cdot 10^{12}\ \text{cm}^{-2}$) act as doping for the graphene. The thickness of the Al_2O_3 influences how the potential drop across the oxide $\Delta\phi_{\text{ox}}$, which in turn determines what fraction of Q_f will be mirrored in the graphene and what fraction will be mirrored in the $c\text{-Si}$. Generally, the thicker the oxide (and the lower its ϵ_r), the greater the fraction of Q_f that will be mirrored in the $c\text{-Si}$. In this situation, the EOT of the oxide is 2 nm and the resulting carrier density in the graphene is $p_G = 3.5 \cdot 10^{12}\ \text{cm}^{-2}$. Note how the $c\text{-Si}$ valence band has better alignment here with the graphene Fermi level than in Figure 6.5 due to the potential drop across the oxide layer.

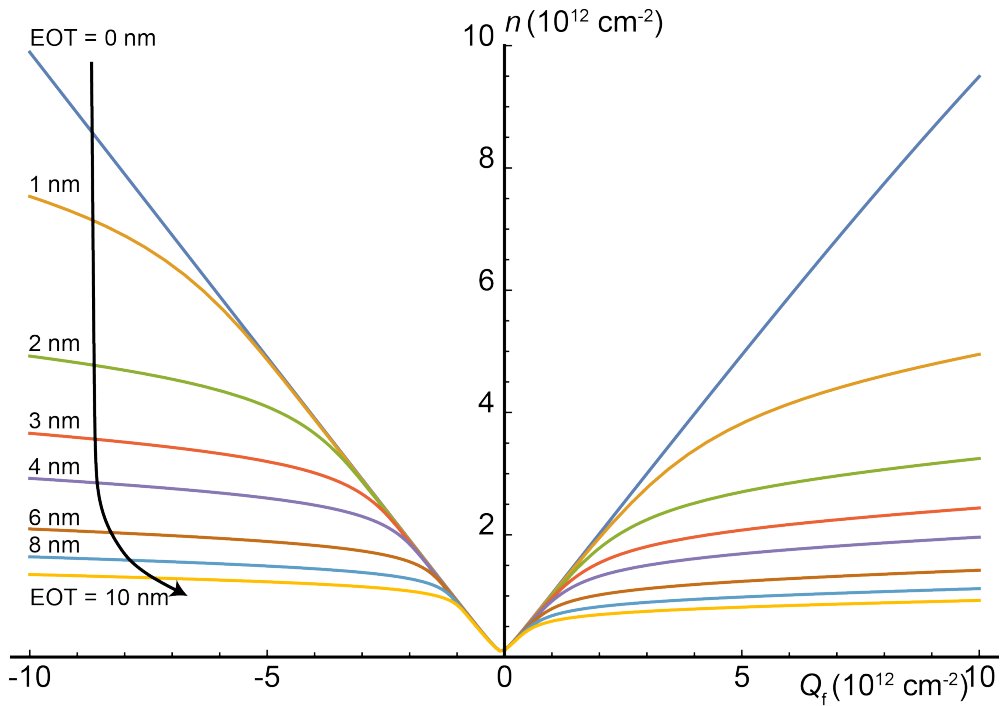


Figure 6.7: The graphene doping density due to the fixed oxide charge as a function of Q_f . The oxide thickness was varied from an EOT of 0 nm to 10 nm. As can be seen, a thick oxide reduces the doping efficiency of the fixed charges. The asymmetry around $Q_f = 0$ is due to the fact that the Dirac point is not precisely midgap, which makes it easier to induce p -type doping in the graphene. There is also a very minor second effect that stems from the fact that the substrate is an n -type wafer: it is easier for the fixed charges to nullify the (substrate induced) n -type doping in the graphene than it is to increase it. It is this effect that shifts the minimum carrier density to a point slightly left of the y -axis.

is generally required to solve the complete set of coupled drift-diffusion equations (Eqs. (A.5) in Appendix A.1) in the device. The graphene can then be described by applying the correct BCs for this set of equations. Here, a 1D transport problem on the interval $[0, L]$ will be considered with the graphene located on the left boundary $x = 0$. Generally, the transport equations in the semiconductor require three BCs at $x = 0$ to provide a unique solution: one BC for the electrostatic potential ϕ and two BCs for the transport equations of the electrons and holes. In addition, a fourth BC is required to fix the Fermi level in the graphene E_{FG} (or $\epsilon_{FG} = E_{FG} + \phi(0) + \phi_G$). Thus, the graphene boundary consists of a system of four equations with four unknowns. As was discussed in Section 6.2.1, the whole system should be globally charge neutral, meaning that one of the equations that should be imposed at $x = 0$ has to be the following one:

$$-\nabla\phi \cdot \hat{\mathbf{n}} = -\frac{\epsilon Q_G(\epsilon_{FG})}{\epsilon_0\epsilon_r} \quad (6.19)$$

Eq. (6.19) is not an ordinary Neumann BC for ϕ (i.e., a BC that prescribes the normal component of $\nabla\phi$), since it also involves the unknown ϵ_{FG} . Eq. (6.19) is therefore more accurately described as a nonlinear mixed BC.

A second BC is obtained by specifying the bias voltage over the device. For example, it can be assumed that the graphene is grounded while the other terminal of the device is biased. The easiest way to implement a bias is by setting Dirichlet BCs for $\phi(0)$ and $\phi(L)$, but it should be kept in mind that true the bias over the device V_B is determined by the Fermi level in the graphene E_{FG} and the Fermi level in the metal at the back contact $E_F(L)$ (i.e., $V_B = E_F(L) - E_{FG}$). Since the charge in the graphene and ϵ_{FG} can change with the voltage over the device, it is important to calculate the correct bias if the simulation controls the voltage by varying $\phi(0)$ and $\phi(L)$. This is especially an issue for near-intrinsic graphene in which the Fermi level is very sensitive to the bias over the device. In doped graphene, the DOS is nonzero at the position of the Fermi level, making ϵ_{FG} less sensitive to changes in Q_G . Thus, the larger $|\epsilon_{FG}|$, the more graphene behaves like a metal and the less the graphene Fermi level is influenced by the operating conditions of the cell.

Two BCs remain to be specified at $x = 0$ and at this point different choices can be made depending on the current transport model that is assumed at the GS interface. Here, two different models will be discussed: (a) the continuous quasi Fermi level (CQFL) model; and (b) the thermionic emission (TE) model.

In the CQFL model it is assumed that the electron and hole quasi Fermi levels merge together continuously in the graphene:

$$E_{Fn}(0) = E_{FG}, \quad (6.20)$$

$$E_{Fp}(0) = E_{FG}. \quad (6.21)$$

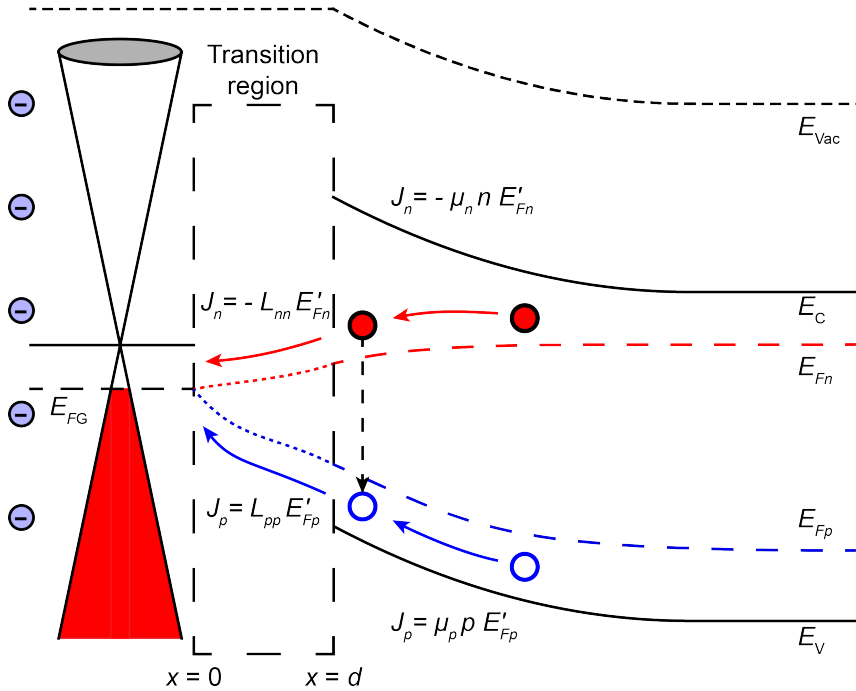


Figure 6.8: Sketch of a non-equilibrium band diagram of graphene on a semiconductor. The dashed boxed rectangle represents the transition region that connects domain where ordinary semiconductor physics applies from the graphene. In the transition region, E_{Fn} and E_{Fp} are continuous, but the transport coefficients L_{nn} and L_{pp} (see Section 2.6 in Chapter 2) are in principle unknown. This makes an exact description of this region difficult, but by introducing empirical parameters it is possible to incorporate the transition region through modified BCs on $E_{Fn,p}$, such as the TE BCs. These modified BCs will account for the resistance of the transition region and generally make $E_{Fn,p}$ discontinuous at the graphene interface, but these discontinuities are understood to be artifacts of the modified description of the interface rather than fundamental properties of $E_{Fn,p}$.

The advantage of the CQFL model is that it is straightforward to implement, but it has the drawback that it ignores the resistance associated with the graphene/semiconductor interface (see Chapter 3, Section 3.3). This makes the CQFL suitable to simulate cell designs in which the graphene interface is not the (main) selective element on that side of the cell (i.e., when its primary function is to act as a transparent electrode).

The TE model does not assume that E_{Fn} and E_{Fp} have to be continuous across the interface, so generally $E_{Fn} \neq E_{FG}$ and $E_{Fp} \neq E_{FG}$. As was discussed in Chapter 3, Section 3.3, the TE model can be interpreted as an extension of the CQFL model that includes an unknown transition region in between the graphene and the semiconductor. The transition region starts at $x = 0$ in the graphene and ends at $x = d$, as shown in Figure 6.8. In principle,

it would be possible to explicitly include the transition region in the transport physics and in this case, the Fermi potentials would be continuous everywhere as shown in the figure. However, since the exact nature of the transition region is unknown and not of great interest, it is integrated out and included only implicitly by modifying the BCs. This then results in the TE model, which uses empirical parameters to characterise the transition region.

For notational simplicity, we will only consider electron currents in the next paragraph with the understanding that it follows the same line for holes. As was discussed in Section 3.3, the TE model for an ordinary metal-semiconductor (MS) interface asserts that the TE current electron current J_{TE} is given by:

$$eJ_{\text{M,S}} = \frac{4\pi em^*}{h^3} (k_{\text{B}}T)^2 \exp\left(-\frac{\Phi_{\text{B}}}{k_{\text{B}}T}\right) = A^*T^2 \exp\left(-\frac{\Phi_{\text{B}}}{k_{\text{B}}T}\right), \quad (6.22a)$$

$$eJ_{\text{S,M}} = A^*T^2 \frac{n}{N_{\text{C}}}, \quad (6.22b)$$

$$\begin{aligned} eJ_{\text{TE}} &= A^*T^2 \left[\exp\left(\frac{E_{\text{F,met}} + \phi + \chi}{k_{\text{B}}T}\right) - \exp\left(\frac{E_{\text{F,surf}} + \phi + \chi}{k_{\text{B}}T}\right) \right] \\ &= A^*T^2 \frac{n}{N_{\text{C}}} \left(\exp\left(\frac{E_{\text{F,met}} - E_{\text{F,surf}}}{k_{\text{B}}T}\right) - 1 \right). \end{aligned} \quad (6.22c)$$

Note that here the current is defined as a particle current rather than a charge current, hence appearance of the elementary charge e in front of the currents. In Eqs. (6.22), $J_{\text{M,S}}$ ($J_{\text{S,M}}$) is the current from the metal into the semiconductor (and vice versa); $\Phi_{\text{B}} = \Phi_{\text{M}} - \chi$ is the Schottky barrier height (with Φ_{M} the metal work function and χ the semiconductor electron affinity); n and N_{C} are the electron density and effective density of states in the semiconductor at the interface; m^* is the effective mass of the electrons in the semiconductor; $E_{\text{F,met}}$ and $E_{\text{F,surf}}$ are the Fermi level in the metal and the electron Fermi level in the semiconductor; and A^* the effective Richardson constant. The (less common) form of the TE current across a MS interface in Eq. (6.22c) emphasises that the difference in Fermi level $\Delta E_{\text{F}} = E_{\text{F,met}} - E_{\text{F,surf}}$ across the interface drives the current and that A^* is a measure of the resistance of the interface.

The particular T^2 form of the TE prefactor $J_0 = A^*T^2$ is due to the fact that in the derivation of the TE formulae (see, e.g. [32]) it is assumed that the electrons in the metal form an ideal degenerate 3D electron gas with a quadratic dispersion relation $\epsilon = \hbar^2 k^2 / (2m^*)$. From that assumption one then calculates the number of electrons moving toward the interface that have enough energy to cross the barrier Φ_{B} without tunnelling. In practice, the value of A^* can be different than the form derived from the ideal gas assumption (shown in Eq. (6.22a)), since the dispersion relation in a metal is usually not perfectly parabolic over the energy range that needs to be considered. In addition, quantum mechanical effects can lead to reflection of carriers that classically would cross the barrier (i.e. the opposite of tunnelling). For this reason A^*

should be considered as an empirical parameter that needs to be measured by experiment.

It should be clear that the traditional TE model cannot be applied to graphene straight away, since graphene has a linear dispersion relation rather than a quadratic one. Moreover, in graphene the electrons move in a 2D plane and have little tendency leave the plane by moving in the third, perpendicular dimension. It is therefore instructive to first consider thermionic emission from graphene in into the vacuum, as was done by Liang *et al.* [15]. They assume that the electrons in the graphene are confined by a finite square quantum well in the out-of-plane direction and find the following TE expression for vacuum emission:

$$eJ_{\text{TE,Vac}} = \beta T^3 \exp\left(-\frac{\phi_G}{k_B T}\right) = \frac{ek_B^3 T^3}{\pi \hbar^3 v_F^2} \exp\left(-\frac{\phi_G}{k_B T}\right), \quad (6.23)$$

$$\beta T^3 = 0.0285 A_0 T^2 = 3.078 \cdot 10^5 \text{ A cm}^{-2} \quad (\text{at } T = 300 \text{ K}).$$

Here, $J_{0,\text{Vac}}$ is the vacuum emission TE prefactor and $A_0 = 4\pi ek_B^2 m_e/h^3$ (with m_e the electron mass) is the unmodified Richardson constant. As can be seen, the exponential dependence on the work function from the usual Richardson law is retained because the electrons that have sufficient energy to escape the graphene satisfy Boltzmann statistics. However, the $\sim T^2$ law has now been replaced by a $\sim T^3$ law and there is no mention of an electron mass (as is appropriate for graphene). Furthermore, the thermionic emission from graphene at low temperatures (i.e., room temperature and lower) is significantly reduced compared to metals.

The next question is whether Eq. (6.23) is also valid for junctions between graphene and semiconductors. Experimental data by Sinha *et al.* [41] show that the TE current under forward bias can be significantly lower for graphene/silicon Schottky junctions than for similar metal-based devices as long as the interface is kept very clean. The authors propose their own Landauer model based on the assumption that electrons with a large enough energy will leave the graphene with a certain frequency τ^{-1} . They performed temperature dependent characterisation of the GS junctions in an attempt to prove their hypothesis, but unfortunately the current was found to be dominated by the temperature dependence in the exponent $\exp(-\Phi_B/(k_B T))$ so that their data could not distinguish between the Schottky model and the Landauer model (i.e., both models could fit the data equally well). The exact form of the graphene TE law should therefore be considered as an ongoing area of research. Regardless, at $T = 300 \text{ K}$ the authors obtain an effective Richardson constant of $A^* = 1.7 \cdot 10^{-4} A_0$ for their best device operating under forward bias (Figure 1b in their article), meaning that their devices have even less thermionic emission than the vacuum TE current predicted by Liang *et al.* The results from Sinha *et al.* demonstrate that graphene/*c*-Si junctions have significantly different characteristics compared to metal/*c*-Si Schottky junctions in that the 2D

nature of the material makes it more difficult for charge carriers to cross from a semiconductor to graphene. In addition, part of the low TE current of GS junctions might be explained by the low DOS of graphene at the *c*-Si midgap energy. At metal-semiconductor interfaces, the metal induced a large density of defect states called MIGS (Metal Induced Gap States, see also Section 3.3) in the semiconductor forbidden gap. These MIGS lead to extra recombination at the interface, which can also be regarded as modification of the interface resistance (as was demonstrated in Section 3.6 in Chapter 3). It is likely that the low DOS of graphene limits the number of induced defect states, which would translate to a lower effective TE current.

Thus, from a practical point of view, graphene can be considered as a metal as long as it is kept in mind that: a) the prefactor for the thermionic emission law for graphene can be much lower than the one commonly used for metals (and can also follow a different temperature dependence); and b) the TE barrier height of graphene is not constant but depends on the substrate, the graphene doping and the bias of the device⁸.

6.2.3 Graphene as a carrier selective contact

As was already briefly discussed in Section 3.3, the low A^* of GS junctions makes them a potential candidate for a selective contact, unlike regular metal Schottky junctions which are simply too lossy. In a GS junction, the interface is the primary carrier selective element, rather than the silicon SCR as is often believed. Based on the value $A^* = 1.7 \cdot 10^{-4} A_0$, a GS junction could reach a V_{OC} of 750 mV (under AM1.5 illumination and if the GS junction is the limiting factor of the cell) for a barrier height $\Phi_B > 1.026$ eV. Assuming that this value of A^* is the same for both electrons and holes, such a barrier requires a graphene electron density of $n_G > 9.29 \cdot 10^{12} \text{ cm}^{-2}$ (for an electron-selective contact) or hole density of $p_G > 2.46 \cdot 10^{13} \text{ cm}^{-2}$ (for a hole-selective contact). These are carrier densities that are well within the physically feasible limits, as long as a stable doping scheme for graphene can be found that is suitable for solar cell applications.

Doping control of the graphene can be seen to be of great importance for making a selective contact out of a GS junction. The doping of the graphene tunes the barrier height Φ_B and thereby the selectivity of the GS interface to either electrons or holes. It can be calculated that the respective graphene doping densities that would achieve the $V_{OC} = 750$ mV limit are $N_{\text{dop,G}} > 9.73 \cdot 10^{12} \text{ cm}^{-2}$ (electron-selective contact) and $N_{\text{dop,G}} < -2.50 \cdot 10^{13} \text{ cm}^{-2}$ (hole-selective contact)⁹. The required doping densities are slightly higher

⁸This last effect is demonstrated convincingly by Sinha *et al.* for the reverse bias of their devices [41].

⁹The Landauer model proposed by Sinha *et al.* [41] yields slightly different values than the assumption $A^* = 1.7 \cdot 10^{-4} A_0$ (which is used in this work for simplicity) because in

than the required carrier densities because the SCR in the *c*-Si also needs to be compensated for by the doping.

Doping control of the graphene is also very important to effectively collect the majority carriers. This is illustrated in Figure 6.9, where two hole-selective GS contacts have been simulated under max power conditions. Figure 6.9a shows a band diagram of the GS contact with no graphene doping, while in Figure 6.9b the doping is $N_{\text{dop,G}} = -2.50 \cdot 10^{13} \text{ cm}^{-2}$. An important difference between these two contacts is the position of the graphene Fermi level E_{FG} : in the doped contact, E_{FG} aligns neatly with E_{Fp} in the *c*-Si, meaning that the holes are collected virtually without loss. This, incidentally, also shows that the low A^* value of the GS junction does not lead to a collection problem for the majority carriers. On the other hand, in the undoped contact there is a major mismatch of 0.41 eV between E_{FG} and E_{Fp} , leading to significant collection losses even though E_{Fp} is very straight in the *c*-Si. In this case, there are significant ohmic losses in the transition region between the graphene and *c*-Si. The conclusion that should be drawn from these two diagrams is that graphene doping is not just important to make the graphene selective: even if another source of selectivity is already present (say, an amorphous silicon heterojunction), then it is still important that the graphene's Fermi level aligns with the majority carrier Fermi level in the cell to prevent collection losses.

6.3 Conclusions

In this chapter, the physics of graphene were discussed that are relevant to the field of silicon PV. The unique transport properties of graphene stem from its linear dispersion relation $|\epsilon| = \hbar v_{\text{F}} k$, which gives the electrons a constant group velocity so that even carriers with little energy can still contribute significantly to charge transport. As a transparent top electrode in *c*-Si solar cells, graphene shows promise because in principle it should be possible to achieve a sufficiently low sheet resistance with just a single sheet. However, to obtain a graphene sheet that is conductive enough, it is necessary to eliminate/minimise as many scattering processes as possible, including substrate scattering. Doping of graphene also plays an important role in lowering its sheet resistance, but it should be kept in mind that in graphene the conductance is proportional to the square root of the carrier density, meaning that increasing the doping density has diminishing returns. Furthermore, the optical properties of graphene make the material unsuited to be stacked several times since this will quickly lead to a prohibitive amount of parasitic absorption. On the bright

the Landauer model A^* also depends on Φ_{B} . Using their obtained value for the graphene escape timescale $\tau = 4.62 \cdot 10^{-11} \text{ s}$, the maximum required graphene doping density for a V_{OC} of 750 mV becomes $N_{\text{dop,G}} > 1.26 \cdot 10^{13} \text{ cm}^{-2}$ (electron-selective contact) and $N_{\text{dop,G}} < -2.93 \cdot 10^{13} \text{ cm}^{-2}$ (hole-selective contact)

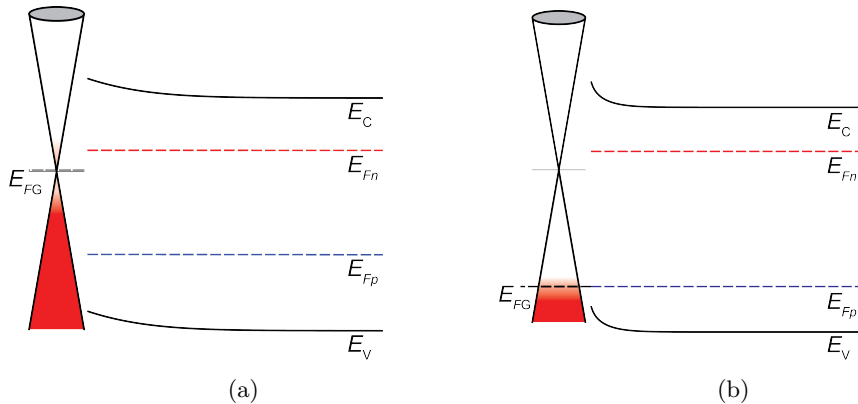


Figure 6.9: Calculated band diagrams under max-power conditions of hole-selective GS contacts on n -type ($3 \Omega \text{ cm}$) silicon using the effective Richardson constant $A^* = 1.7 \cdot 10^{-4} A_0$. In (a) the graphene doping density $N_{\text{dop,G}}$ is zero; in (b) it is $N_{\text{dop,G}} = -2.50 \cdot 10^{13} \text{ cm}^{-2}$. A keen eye can see that in (a) the graphene is slightly n -type in the situation without chemical doping due to the n -type silicon substrate doping. Both diagrams are zoomed-in on the GS junction: the rest of the cell is assumed ideal, meaning that there is no bulk recombination and that the electron-selective contact is perfectly insulating to holes. It should be noted that the cell in (b) indeed achieves a V_{OC} of 750 mV, as predicted. Simulations were performed in COMSOL Multiphysics[®].

side, though, graphene's very high mobility makes sure that free carrier absorption/reflection is completely negligible for solar cell applications, so one does not have to worry about the balance between parasitic absorption/reflection and DC conductance like in a traditional TCO.

The electrical contact between graphene and a semiconductor is somewhat similar to a traditional MS Schottky contact. What makes graphene unique is that its barrier height is variable and can depend on many different factors; the most important of which is the graphene doping density. Furthermore, the TE law that describes charge transport between the semiconductor and graphene is expected to have a different temperature dependence than the T^2 law that applies to MS junctions. The exact form of this TE law for GS junctions is still uncertain, but experimental evidence suggests that the effective Richardson constant of a GS contact can be made so low that it could potentially be used as a carrier-selective contact for a high-efficiency c -Si solar cell. To achieve such a selective contact, it is vital to use highly doped graphene, since the doping will both reduce the minority carrier leakage current as well as the majority carrier ohmic collection losses.

Appendix

6.A Approximate expressions for the graphene carrier densities

For some applications it is useful to have approximate expressions for the relation between the Fermi level in the graphene ϵ_{FG} and the carrier densities. It is therefore useful to consider two limits for n_G and p_G , namely the limit where $|\epsilon_{FG}|$ is large compared to $k_B T$ and the limit where $|\epsilon_{FG}|$ is on the same order of magnitude (or smaller than) as $k_B T$. First, suppose that ϵ_{FG} is large and positive so the electrons dominate in the graphene. It is then possible to approximate the Fermi distribution function for the electrons $1/(1+e^{(\epsilon-\epsilon_{FG})/k_B T})$ by a step function, but a significantly more accurate approximation is obtained by using the Sommerfeld expansion (see, e.g., Le Bellac [31]). This expansion is given by:

$$\int_0^\infty \frac{D_G(\epsilon)}{1 + \exp\left(\frac{\epsilon - \epsilon_{FG}}{k_B T}\right)} d\epsilon = \int_0^{\epsilon_{FG}} D_G(\epsilon) d\epsilon + \frac{\pi^2}{6} k_B T^2 D'_G(\epsilon_{FG}) + \mathcal{O}\left((k_B T)^4 D_G^{(3)}(\epsilon_{FG})\right), \quad (6.24)$$

and provides an extra correction term for the thermal broadening of the Fermi distribution. In general, the Sommerfeld expansion can provide more than one correction terms, depending on how many times the DOS can be differentiated at $\epsilon = \epsilon_{FG}$. However, since the DOS of graphene is linear, only one correction term can be obtained this way. The fact that there are no more correction terms does not mean that the Sommerfeld approximation is exact here. Instead, the discontinuity in the derivative of $D_G(\epsilon)$ at $\epsilon = 0$ is the reason that the Sommerfeld formula cannot give more terms. Applying the expansion to n_G for the case $\epsilon_{FG} \gg k_B T$ gives:

$$n_G(\epsilon_{FG}) \approx \frac{2}{\pi} \left(\frac{k_B T}{\hbar v_F}\right)^2 \left(\frac{1}{2} \left(\frac{\epsilon_{FG}}{k_B T}\right)^2 + \frac{\pi^2}{6}\right). \quad (6.25)$$

Furthermore, for $\epsilon_{FG} \gg k_B T$ the distribution function for holes $1/(1+e^{(\epsilon+\epsilon_{FG})/k_B T})$ can be approximated as an exponential by neglecting the

1 in the denominator, so that the approximation for p_G becomes:

$$p_G(\epsilon_{FG}) \approx \frac{2}{\pi} \left(\frac{k_B T}{\hbar v_F} \right)^2 \exp \left(-\frac{\epsilon_{FG}}{k_B T} \right) \quad (6.26)$$

Of course, when $\epsilon_{FG} \ll -k_B T$ and the holes dominate, the same argument can be used to show that:

$$n_G(\epsilon_{FG}) \approx \frac{2}{\pi} \left(\frac{k_B T}{\hbar v_F} \right)^2 \exp \left(\frac{\epsilon_{FG}}{k_B T} \right), \quad (6.27a)$$

$$p_G(\epsilon_{FG}) \approx \frac{2}{\pi} \left(\frac{k_B T}{\hbar v_F} \right)^2 \left(\frac{1}{2} \left(\frac{\epsilon_{FG}}{k_B T} \right)^2 + \frac{\pi^2}{6} \right). \quad (6.27b)$$

The Sommerfeld and exponential approximations have a relative error of $< 1\%$ for $|\epsilon_{FG}| > 3k_B T$.

On the other hand, when $|\epsilon_{FG}|$ is small (i.e. on the order of a few $k_B T$), it is possible to expand n_G and p_G in a Taylor series around $\epsilon_{FG} = 0$:

$$n_G = \frac{2}{\pi} \left(\frac{k_B T}{v_F \hbar} \right)^2 \left[\frac{\pi^2}{12} + \ln(2) \frac{\epsilon_{FG}}{k_B T} + \frac{1}{4} \left(\frac{\epsilon_{FG}}{k_B T} \right)^2 + \frac{1}{24} \left(\frac{\epsilon_{FG}}{k_B T} \right)^3 - \frac{1}{960} \left(\frac{\epsilon_{FG}}{k_B T} \right)^5 + \mathcal{O} \left(\left(\frac{\epsilon_{FG}}{k_B T} \right)^7 \right) \right], \quad (6.28a)$$

$$p_G = \frac{2}{\pi} \left(\frac{k_B T}{v_F \hbar} \right)^2 \left[\frac{\pi^2}{12} - \ln(2) \frac{\epsilon_{FG}}{k_B T} + \frac{1}{4} \left(\frac{\epsilon_{FG}}{k_B T} \right)^2 - \frac{1}{24} \left(\frac{\epsilon_{FG}}{k_B T} \right)^3 + \frac{1}{960} \left(\frac{\epsilon_{FG}}{k_B T} \right)^5 + \mathcal{O} \left(\left(\frac{\epsilon_{FG}}{k_B T} \right)^7 \right) \right], \quad (6.28b)$$

In this expansion, all even orders > 2 are zero.

As can be seen, there are three approximations for n_G and p_G : the Sommerfeld expressions (6.25) and (6.27b) for large densities; the exponential expressions (6.26) and (6.27a) for small densities; and the Taylor expressions (6.28) for intermediate densities. These three approximations can be patched together to obtain piecewise formulae for n_G and p_G . However, these piecewise functions can only be continuous if one uses a third order (rather than a fifth order) Taylor series to patch the Taylor expression with the Sommerfeld expression. Therefore, a reasonable approximation for n_G and p_G that works

for any value of $x = \epsilon_{FG}/k_B T$ is:

$$n_G = \frac{2}{\pi} \left(\frac{k_B T}{v_F \hbar} \right)^2 \begin{cases} \exp(x) & x < -1.97 \\ \frac{\pi^2}{12} + x \log(2) + \frac{x^2}{4} + \frac{x^3}{24} - \frac{x^5}{960} & -1.97 \leq x < 0 \\ \frac{\pi^2}{12} + x \log(2) + \frac{x^2}{4} + \frac{x^3}{24} & 0 \leq x < 2.50 \\ \frac{x^2}{2} + \frac{\pi^2}{6} & 2.50 \leq x \end{cases}, \quad (6.29a)$$

$$p_G = \frac{2}{\pi} \left(\frac{k_B T}{v_F \hbar} \right)^2 \begin{cases} \frac{x^2}{2} + \frac{\pi^2}{6} & x < -2.50 \\ \frac{\pi^2}{12} - x \log(2) + \frac{x^2}{4} - \frac{x^3}{24} & -2.50 \leq x < 0 \\ \frac{\pi^2}{12} - x \log(2) + \frac{x^2}{4} - \frac{x^3}{24} + \frac{x^5}{960} & 0 \leq x < 1.97 \\ \exp(-x) & 1.97 \leq x \end{cases}. \quad (6.29b)$$

The relative errors of the piecewise approximations in Eqs. (6.29a) and (6.29b) have local maxima at the patching points $x = \mp 1.97$ (3.4% error) and $x = \pm 2.50$ (1.7% error). Asymptotically, the relative errors converge to zero in both directions when $x \rightarrow \pm\infty$. If greater accuracy is required, it is always possible to include more terms in the Taylor series and recalculate the patching points.

References Chapter 6

- [1] Xinming Li, Hongwei Zhu, Kunlin Wang, Anyuan Cao, Jinqian Wei, Chunyan Li, Yi Jia, Zhen Li, Xiao Li, and Dehai Wu. Graphene-on-silicon Schottky junction solar cells. *Advanced materials (Deerfield Beach, Fla.)*, 22(25):2743–8, July 2010. ISSN 1521-4095. doi: 10.1002/adma.200904383.
- [2] Xiaochang Miao, Sefaattin Tongay, Maureen K Petterson, Kara Berke, Andrew G Rinzler, Bill R Appleton, and Arthur F Hebard. High efficiency graphene solar cells by chemical doping. *Nano letters*, 12(6):2745–50, jun 2012. ISSN 1530-6992. doi: 10.1021/nl204414u.
- [3] Yu Ye and Lun Dai. Graphene-based Schottky junction solar cells. *Journal of Materials Chemistry*, 22(46):24224, 2012. ISSN 0959-9428. doi: 10.1039/c2jm33809b.
- [4] Tongxiang Cui, Ruitao Lv, Zheng-Hong Huang, Shuxiao Chen, Zexia Zhang, Xin Gan, Yi Jia, Xinming Li, Kunlin Wang, Dehai Wu, and Feiyu Kang. Enhanced efficiency of graphene/silicon heterojunction solar cells by molecular doping. *Journal of Materials Chemistry A*, 1(18):5736, 2013. ISSN 2050-7488. doi: 10.1039/c3ta01634j.
- [5] Yi Song, Xinming Li, Charles Mackin, Xu Zhang, Wenjing Fang, Tomás Palacios, Hongwei Zhu, and Jing Kong. Role of Interfacial Oxide in High-Efficiency GrapheneSilicon Schottky Barrier Solar Cells. *Nano Letters*, 15(3):150225082743008, 2015. ISSN 1530-6984. doi: 10.1021/nl505011f.
- [6] Xiao Huang, Zhiyuan Zeng, Zhanxi Fan, Juqing Liu, and Hua Zhang. Graphene-based electrodes. *Advanced materials (Deerfield Beach, Fla.)*, 24(45):5979–6004, nov 2012. ISSN 1521-4095. doi: 10.1002/adma.201201587.
- [7] Gunho Jo, Minhyeok Choe, Sangchul Lee, Woojin Park, Yung Ho Kahng, and Takhee Lee. The application of graphene as electrodes in electrical and optical devices. *Nanotechnology*, 23(11):112001, mar 2012. doi: 10.1088/0957-4484/23/11/112001.

- [8] Nathan O Weiss, Hailong Zhou, Lei Liao, Yuan Liu, Shan Jiang, Yu Huang, and Xiangfeng Duan. Graphene: an emerging electronic material. *Advanced materials (Deerfield Beach, Fla.)*, 24(43):5782–825, nov 2012. ISSN 1521-4095. doi: 10.1002/adma.201201482.
- [9] Xinming Li, Zheng Lv, and Hongwei Zhu. Carbon/Silicon Heterojunction Solar Cells: State of the Art and Prospects. *Advanced Materials*, 27(42): 6549–6574, nov 2015. ISSN 09359648. doi: 10.1002/adma.201502999.
- [10] L. a. Falkovsky. Optical properties of graphene. *Journal of Physics: Conference Series*, 129:012004, oct 2008. ISSN 1742-6596. doi: 10.1088/1742-6596/129/1/012004.
- [11] a. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and a. K. Geim. The electronic properties of graphene. *Reviews of Modern Physics*, 81(1):109–162, jan 2009. ISSN 0034-6861. doi: 10.1103/RevModPhys.81.109.
- [12] D.S.L. Abergel, V. Apalkov, J. Berashevich, K. Ziegler, and Tapash Chakraborty. Properties of graphene: a theoretical perspective. *Advances in Physics*, 59(4):261–482, jul 2010. ISSN 0001-8732. doi: 10.1080/00018732.2010.487978.
- [13] Haijian Zhong, Ke Xu, Zhenghui Liu, Gengzhao Xu, Lin Shi, Yingmin Fan, Jianfeng Wang, Guoqiang Ren, and Hui Yang. Charge transport mechanisms of graphene/semiconductor Schottky barriers: A theoretical and experimental study. *Journal of Applied Physics*, 115(1):013701, January 2014. ISSN 0021-8979. doi: 10.1063/1.4859500.
- [14] Kun Xu, Caifu Zeng, Qin Zhang, Rusen Yan, Peide Ye, Kang Wang, Alan C. Seabaugh, Huili Grace Xing, John S. Suehle, Curt A. Richter, David J. Gundlach, and N. V. Nguyen. Direct Measurement of Dirac Point Energy at the Graphene/Oxide Interface. *Nano Letters*, 13(1):131–136, jan 2013. ISSN 1530-6984. doi: 10.1021/nl303669w.
- [15] Shi-jun Liang and L. K. Ang. Electron Thermionic Emission from Graphene and a Thermionic Energy Converter. *Physical Review Applied*, 3:1–8, 2015. ISSN 2331-7019. doi: 10.1103/PhysRevApplied.3.014002.
- [16] Amal Kasry, Marcelo A. Kuroda, Glenn J. Martyna, George S. Tulevski, and Ageeth A. Bol. Chemical Doping of Large-Area Stacked Graphene Films for Use as Transparent, Conducting Electrodes. *ACS Nano*, 4(7): 3839–3844, jul 2010. ISSN 1936-0851. doi: 10.1021/nn100508g.
- [17] Qing Tang, Zhen Zhou, and Zhongfang Chen. Graphene-related nanomaterials: tuning properties by functionalization. *Nanoscale*, 5(11):4541, 2013. ISSN 2040-3364. doi: 10.1039/c3nr33218g.

- [18] F Schedin, a K Geim, S V Morozov, E W Hill, P Blake, M I Katsnelson, K S Novoselov, and S V Morozov E W Hill P Blake M I Katsnelson & K S Novoselov F. Schedin A.K. Geim. Detection of individual gas molecules adsorbed on graphene. *Nature materials*, 6(9):652–5, sep 2007. ISSN 1476-1122. doi: 10.1038/nmat1967.
- [19] Y.-W. Tan, Y. Zhang, K. Bolotin, Y. Zhao, S. Adam, E. H. Hwang, S. Das Sarma, H. L. Stormer, and P. Kim. Measurement of Scattering Rate and Minimum Conductivity in Graphene. *Physical Review Letters*, 99(24):246803, dec 2007. ISSN 0031-9007. doi: 10.1103/PhysRevLett.99.246803.
- [20] K. S. Novoselov. Electric Field Effect in Atomically Thin Carbon Films. *Science*, 306(5696):666–669, oct 2004. ISSN 0036-8075. doi: 10.1126/science.1102896.
- [21] Jason Horng, Chi Fan Chen, Baisong Geng, Caglar Girit, Yuanbo Zhang, Zhao Hao, Hans a. Bechtel, Michael Martin, Alex Zettl, Michael F. Crommie, Y. Ron Shen, and Feng Wang. Drude conductivity of Dirac fermions in graphene. *Physical Review B - Condensed Matter and Materials Physics*, 83(16):1–5, 2011. ISSN 10980121. doi: 10.1103/PhysRevB.83.165113.
- [22] Henry Medina, Yung Chang Lin, Dirk Obergfell, and Po W. Chiu. Tuning of charge densities in graphene by molecule doping. *Advanced Functional Materials*, 21(14):2687–2692, jul 2011. ISSN 1616301X. doi: 10.1002/adfm.201100401.
- [23] Hugen Yan, Xuesong Li, Bhupesh Chandra, George Tulevski, Yanqing Wu, Marcus Freitag, Wenjuan Zhu, Phaeton Avouris, and Fengnian Xia. Tunable infrared plasmonic devices using graphene/insulator stacks. *Nature Nanotechnology*, 7(5):330–334, apr 2012. ISSN 1748-3387. doi: 10.1038/nnano.2012.59.
- [24] A Das, S Pisana, B Chakraborty, S Piscanec, S K Saha, U V Waghmare, K S Novoselov, H R Krishnamurthy, a K Geim, a C Ferrari, and a K Sood. Monitoring dopants by Raman scattering in an electrochemically top-gated graphene transistor. *Nature Nanotechnology*, 3(4):210–215, apr 2008. ISSN 1748-3387. doi: 10.1038/nnano.2008.67.
- [25] J. W. Weber, a. a. Bol, and M. C. M. van de Sanden. An improved thin film approximation to accurately determine the optical conductivity of graphene from infrared transmittance. *Applied Physics Letters*, 105(1):013105, jul 2014. ISSN 0003-6951. doi: 10.1063/1.4889852.
- [26] Saeed H. Abedinpour, G. Vignale, A. Principi, Marco Polini, Wang-Kong Tse, and A. H. MacDonald. Drude weight, plasmon dispersion, and ac

- conductivity in doped graphene sheets. *Physical Review B*, 84(4):045429, jul 2011. ISSN 1098-0121. doi: 10.1103/PhysRevB.84.045429.
- [27] M Orlita, I Crassee, C Faugeras, A B Kuzmenko, F Fromm, M Ostler, Th Seyller, G Martinez, M Polini, and M Potemski. Classical to quantum crossover of the cyclotron resonance in graphene: a study of the strength of intraband absorption. *New Journal of Physics*, 14(9):095008, sep 2012. ISSN 1367-2630. doi: 10.1088/1367-2630/14/9/095008.
- [28] B. Macco, Y. Wu, D. Vanhemel, and W. M. M. Kessels. High mobility $\text{In}_2\text{O}_3\text{:H}$ transparent conductive oxides prepared by atomic layer deposition and solid phase crystallization. *physica status solidi (RRL) - Rapid Research Letters*, 8(12):987–990, dec 2014. ISSN 18626254. doi: 10.1002/pssr.201409426. URL <http://doi.wiley.com/10.1002/pssr.201409426>.
- [29] Bart Macco, Harm C. M. Knoops, and Wilhelmus M. M. Kessels. Electron Scattering and Doping Mechanisms in Solid-Phase-Crystallized $\text{In}_2\text{O}_3\text{:H}$ Prepared by Atomic Layer Deposition. *ACS Applied Materials & Interfaces*, 7(30):16723–16729, aug 2015. ISSN 1944-8244. doi: 10.1021/ac-sami.5b04420.
- [30] Harm C. M. Knoops, Bas W. H. van de Loo, Sjoerd Smit, Mikhail V. Ponomarev, Jan-Willem Weber, Kashish Sharma, Wilhelmus M. M. Kessels, and Mariadriana Creatore. Optical modeling of plasma-deposited ZnO films: Electron scattering at different length scales. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 33(2):021509, mar 2015. ISSN 0734-2101. doi: 10.1116/1.4905086.
- [31] Michel Le Bellac, Fabrice Mortessagne, and G. George Batrouni. *Equilibrium and Non-Equilibrium Statistical Thermodynamics*. Cambridge University Press, Cambridge, 2004. ISBN 0 521 82143 6.
- [32] Harald Ibach and Hans Lüth. *Solid-State Physics*. Springer-Verlag Berlin Heidelberg New York, Berlin, Heidelberg, 3 edition, 2003. doi: Phys-RevB.83.165113.
- [33] I-Tan Lin and Jia-Ming Liu. Terahertz Frequency-Dependent Carrier Scattering Rate and Mobility of Monolayer and AA-Stacked Multilayer Graphene. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(1):122–129, jan 2014. ISSN 1077-260X. doi: 10.1109/JSTQE.2013.2263116.
- [34] P Spinelli, M.A. Verschuuren, and Albert Polman. Broadband omnidirectional antireflection coating based on subwavelength surface Mie res-

- onators. *Nature Communications*, 3:692, feb 2012. ISSN 2041-1723. doi: 10.1038/ncomms1691.
- [35] Stefaan De Wolf, Antoine Descoeur, Zachary C. Holman, and Christophe Ballif. High-efficiency Silicon Heterojunction Solar Cells: A Review. *green*, 2(1):7–24, January 2012. ISSN 1869-876X. doi: 10.1515/green-2011-0018.
- [36] Damon B. Farmer, Hsin-Ying Chiu, Yu-Ming Lin, Keith A. Jenkins, Fengnian Xia, and Phaeton Avouris. Utilization of a Buffered Dielectric to Achieve High Field-Effect Carrier Mobility in Graphene Transistors. *Nano Letters*, 9(12):4474–4478, dec 2009. ISSN 1530-6984. doi: 10.1021/nl902788u.
- [37] Jian-hao Chen, Chaun Jang, Shudong Xiao, Masa Ishigami, and Michael S Fuhrer. Intrinsic and extrinsic performance limits of graphene devices on SiO₂. *Nature Nanotechnology*, 3(4):206–209, apr 2008. ISSN 1748-3387. doi: 10.1038/nnano.2008.58.
- [38] Haijian Zhong, Zhenghui Liu, Gengzhao Xu, Yingmin Fan, Jianfeng Wang, Xuemin Zhang, Liwei Liu, Ke Xu, and Hui Yang. Self-adaptive electronic contact between graphene and semiconductors. *Applied Physics Letters*, 100(12):122108, 2012. ISSN 00036951. doi: 10.1063/1.3696671.
- [39] Bram Hoex, S. B. S. Heil, E. Langereis, M. C. M. van de Sanden, and W. M. M. Kessels. Ultralow surface recombination of *c*-Si substrates passivated by plasma-assisted atomic layer deposited Al₂O₃. *Applied Physics Letters*, 89(4):042112, 2006. ISSN 00036951. doi: 10.1063/1.2240736.
- [40] G. Dingemans and W. M. M. Kessels. Status and prospects of Al₂O₃-based surface passivation schemes for silicon solar cells. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 30(4):040802, 2012. ISSN 07342101. doi: 10.1116/1.4728205.
- [41] Dhiraaj Sinha and Ji Ung Lee. Ideal Graphene/Silicon Schottky Junction Diodes. *Nano Letters*, 14(8):4660–4664, aug 2014. ISSN 1530-6984. doi: 10.1021/nl501735k.

Appendix A

Mathematical formulation of the charge transport in solar cells

A.1 The drift-diffusion equations

The mathematical description of a *c*-Si solar cell is based on a set of assumptions that are summarised in the points (i) till (v) below. These assumptions lead to the 7 numbered equations (A.1) till (A.4) (also below): the Poisson equation; two continuity equations; two equations about carrier statistics; and two current-force relations.

- (i) The operation of a solar cell can be described by considering it as a system of positive and negative charge carriers that can move freely through a solid, while simultaneously also being confined to said solid. It is assumed that this solid is not subjected to any net external forces or torques and the reference frame of the solid will always be used to describe any movement of the free carriers inside. A typical example of such a system includes free electrons (negative charge carriers, e^-) and holes (positive charge carriers, h^+) that move freely inside of a *c*-Si wafer, but cannot escape it except through (metal) contacts connected to the wafer.

The electrons and holes are characterized by their densities n and p , which can be dependent on position x and time t . In this work the x and t dependencies will not be written explicitly unless extra attention needs to be called to them. If electrons and/or holes (locally) exhibit concerted net motion, this motion will be called current and will be represented by a vector field \mathbf{J}_n (for electrons) or \mathbf{J}_p (for holes). In this work, $\mathbf{J}_{n,p}$ will always denote particle currents (as opposed to charge currents) and are measured in units of $[\text{Length}^{-2} \cdot \text{Time}^{-1}]$. Note that x can refer to

coordinates of one, two or three dimensions: only vector fields such as \mathbf{J} will be printed in bold.

- (ii) Since the discussion of the solar cell focusses on charged particles, electromagnetic fields play an important role in the description. In solar cells, magnetic fields are too weak to play a role in the operation of the device and this makes it possible to reduce the Maxwell equations to a single equation for the electrostatic potential ϕ , called the Poisson equation:

$$\nabla \cdot (\epsilon \nabla \phi) = -e^2 (p - n + N_{\text{dop}}) = -e \rho_{\text{net}}. \quad (\text{A.1})$$

Here, $\epsilon = \epsilon_0 \epsilon_r$ is the dielectric constant and N_{dop} is the background doping concentration. Note that ϕ is taken in units of energy here, which is more convenient in a thermodynamic context because it avoids the need to multiply ϕ by e to obtain an energy. Nevertheless, ϕ will be referred to as electrostatic energy or potential interchangeably.

- (iii) The continuity equations for electrons and holes are given by:

$$\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{J}_n = G - R_n, \quad (\text{A.2a})$$

$$\frac{\partial p}{\partial t} + \nabla \cdot \mathbf{J}_p = G - R_p. \quad (\text{A.2b})$$

The time derivatives in Eqs. (A.2) are important when discussing certain solar cell characterisation experiments such as transient photoconductance decay or PCD (i.e., lifetime testing). However, in this work the main focus is on the steady-state operation of a solar cell when it is producing net power, so then time derivatives are assumed to be zero: $\partial n / \partial t = \partial p / \partial t = 0$. In the steady state it is only necessary to assume $R_n \neq R_p$ if the trapped charges are still mobile and can hop to nearby defects. Here, it will be assumed that trapped charges are completely fixed, so that the simplification $R_n = R_p = R$ can be used. The term $G - R$ will be frequently abbreviated by G_{net} .

- (iv) The relation between the carrier density and the potentials, i.e., the carrier statistics. If not stated otherwise, it is assumed that Boltzmann statistics apply, so that:

$$\begin{aligned} n &= N_{\text{C}} \exp\left(-\frac{E_{\text{C}} - E_{\text{Fn}}}{k_{\text{B}}T}\right) = N_{\text{C}} \exp\left(\frac{E_{\text{Fn}} + \phi + \chi}{k_{\text{B}}T}\right) \\ &= N_{\text{C}} e^{\mu_n^{\text{c}} / (k_{\text{B}}T)}, \end{aligned} \quad (\text{A.3a})$$

$$\begin{aligned} p &= N_{\text{V}} \exp\left(-\frac{E_{\text{Fp}} - E_{\text{V}}}{k_{\text{B}}T}\right) = N_{\text{V}} \exp\left(-\frac{E_{\text{Fp}} + \phi + \chi + E_{\text{G}}}{k_{\text{B}}T}\right) \\ &= N_{\text{V}} e^{\mu_p^{\text{c}} / (k_{\text{B}}T)}, \end{aligned} \quad (\text{A.3b})$$

or equivalently:

$$E_{Fn} = k_B T \ln \left(\frac{n}{N_C} \right) - \phi - \chi,$$

$$E_{Fp} = -k_B T \ln \left(\frac{p}{N_V} \right) - \phi - \chi - E_G.$$

Here, χ is the electron affinity, $E_{C,V}$ the energies of the conduction/valence band edges, and $\mu_{n,p}^c$ the chemical potentials. Sometimes the use of quasi Fermi energies $E_{Fn,p}$ causes a needless confusion of signs, in which case it is more convenient to work with the electrochemical potentials $\eta_{n,p}$. These potentials measure the free energy of free electrons and holes and are given by:

$$\eta_n = E_{Fn} = \mu_n^c - \phi - \chi,$$

$$\eta_p = -E_{Fp} = \mu_p^c + \phi + \chi + E_G.$$

Since free electrons and holes are always generated and annihilated in pairs, another frequently occurring quantity is the free energy per e^-h^+ pair, $\Delta\eta_{eh}$:

$$\Delta\eta_{eh} = \eta_n + \eta_p = E_{Fn} - E_{Fp} = E_G + k_B T \ln \left(\frac{np}{N_C N_V} \right).$$

If Fermi-Dirac statistics are required, the exponential functions in Eqs. (A.3) have to be replaced by the Fermi-Dirac integral $F_{1/2}$:

$$F_{1/2}(x) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\sqrt{y}}{\exp(y-x) + 1} dy.$$

Note that the argument x is dimensionless here. The factor $2/\sqrt{\pi}$ is included such that in the limit of large negative x , $F_{1/2}(x)$ approaches the exponential function and Boltzmann statistics are recovered:

$$\lim_{x \rightarrow -\infty} \frac{F_{1/2}(x)}{\exp(x)} = 1.$$

Like the exponential function, $F_{1/2}$ has an inverse, which will be denoted by $F_{1/2}^\leftarrow$ and which can be considered to be the Fermi-Dirac equivalent of the logarithm. Much like the logarithm, $F_{1/2}^\leftarrow(x)$ is only well-defined for arguments $x > 0$ (as long as only the real numbers are considered).

A common way to evaluate the function $F_{1/2}$ is by using the Aymerich-Humet approximation [1]:

$$F_{1/2}(x) \approx \frac{1}{\exp(-x) + 3\sqrt{\frac{\pi}{2}} \left[x + 2.13 + \left(|x - 2.13|^{2.4} + 9.6 \right)^{\frac{5}{12}} \right]^{-\frac{3}{2}}}.$$

The Aymerich-Humet approximation makes it possible to directly calculate the carrier densities from the potentials without the need to perform numerical integration. However, there is no closed form for the inverse of the approximation, so to calculate E_{Fn} or E_{Fp} from the carrier densities n and p , it is necessary to use a numerical root-finding method to invert the equation.

- (v) The current-force relation for \mathbf{J}_n and \mathbf{J}_p . The local balances for electrons and holes have been specified in Eqs. (A.2), but these are not enough to arrive at a closed set of transport equations since they only specify the divergence of the vectors $\mathbf{J}_{n,p}$. Additionally, if in Eqs. (A.2) the recombination R depends on n and p (as is almost always the case), it is necessary to specify how $\mathbf{J}_{n,p}$ relate to the other variables. This is the point where solar cell physics become truly empirical: all of the previous equations (i.e., Eqs. (A.1)) through (A.3b) are fundamental or a limiting case of a fundamental law. The Poisson equation follows from Maxwell's equations (with zero magnetic field); the electron and hole balances follow from conservation of particles and charge; and the carrier statistics can be traced back to quantum mechanical principles.

Solar cell physics commonly only considers the diagonal elements of the transport matrix L (see Eqs. (2.55), Section 2.6) and this convention will also be followed here, unless stated otherwise. The transport coefficients are expressed as a product of carrier density and mobility $\mu_{n,p}$, though generally the mobilities are not constant but nonlinear functions of the carrier and doping densities. Thus, $\mathbf{J}_{n,p}$ are given by:

$$\mathbf{J}_n = -\frac{n\mu_n}{e}\nabla E_{Fn} = -\frac{\sigma_n}{e^2}\nabla E_{Fn}, \quad (\text{A.4a})$$

$$\mathbf{J}_p = \frac{p\mu_p}{e}\nabla E_{Fp} = \frac{\sigma_p}{e^2}\nabla E_{Fp}. \quad (\text{A.4b})$$

Here, $\sigma_{n,p}$ are the electron and hole conductivities of the material expressed in $\Omega^{-1}\text{m}^{-1}$. It should be noted that the proportionality between carrier density and conductivity is not entirely self-evident as it may appear to be. Indeed, in the degenerate regime most carriers are too deep in the Fermi sea to contribute to conduction. For degenerate semiconductors the conductivity is only proportional to the carrier density if the bands are parabolic and demonstration of this proportionality requires a more detailed analysis based on the Boltzmann transport equation (see, e.g., Ibach & Lüth [2]). For non-standard conductors (such as graphene) it is therefore always recommended to check if it is permissible to write the conductivity as a product of density and mobility.

Equations (A.1) through (A.4) are the seven basic physical equations for the charge transport in a solar cell. Together, the equations can be solved to

obtain the seven unknowns ϕ , n , p , E_{Fn} , E_{Fp} , \mathbf{J}_n and \mathbf{J}_p . These equations have to be supplemented with models that specify the generation and recombination rates G and R ; the mobilities μ_n and μ_p ; and the band gap E_G and electron affinity χ .

It is customary to reduce the seven basic equations (A.1) through (A.4) by substitution. The most common way to achieve this, is by eliminating $E_{Fn,p}$ and $\mathbf{J}_{n,p}$, which results in:

$$\nabla \cdot (\epsilon \nabla \phi) = -e^2 \rho_{\text{net}}, \quad (\text{A.5a})$$

$$\nabla \cdot \left(-\frac{n\mu_n}{e} \nabla \left[k_B T F_{1/2}^{\leftarrow} \left(\frac{n}{N_C} \right) - \phi - \chi \right] \right) = G_{\text{net}}, \quad (\text{A.5b})$$

$$\nabla \cdot \left(\frac{p\mu_p}{e} \nabla \left[-k_B T F_{1/2}^{\leftarrow} \left(\frac{p}{N_V} \right) - \phi - \chi - E_G \right] \right) = G_{\text{net}}. \quad (\text{A.5c})$$

Equations (A.5b) and (A.5c) can be simplified further if Boltzmann statistics are assumed and if $\nabla k_B T$ is assumed to be negligible. Then $F_{1/2}^{\leftarrow}$ can be replaced by the natural logarithm and the familiar drift-diffusion equations are obtained:

$$\nabla \cdot \left(-D_n \nabla n + \frac{n\mu_n}{e} \nabla [\phi + \chi] \right) = G_{\text{net}}, \quad (\text{A.6a})$$

$$\nabla \cdot \left(-D_p \nabla p - \frac{p\mu_p}{e} \nabla [\phi + \chi + E_G] \right) = G_{\text{net}}, \quad (\text{A.6b})$$

with $D_{n,p} = k_B T \mu_{n,p} / e$ the diffusion constants.

A.2 The Poisson-Boltzmann equation

The Poisson-Boltzmann (PB) equation is a useful approximation to the full solar cell transport equations presented in Appendix A.1. It can be used to gain quick insights into the shape of band structures without the need to calculate the currents flowing through the device. A typical example of an application of the PB equation can be found in the work by Girisch [3], where it is used to calculate the amount of induced band bending in silicon due to a fixed interface charge in a dielectric (SiO_2 in this case) that has been grown on the silicon.

The derivation of the PB equation uses the approximation that the quasi Fermi levels $E_{Fn,p}$ are constant throughout the region of interest; it is obtained by simply substituting the equations of the carrier statistics (Eqs. (A.3)) into the Poisson equation (Eq. (A.1)):

$$\phi'' = -\frac{e^2}{\epsilon} \left[N_V F_{1/2} \left(-\frac{\Phi_p + \phi}{k_B T} \right) - N_C F_{1/2} \left(\frac{\Phi_n + \phi}{k_B T} \right) + N_{\text{dop}} \right], \quad (\text{A.7})$$

$$\Phi_n = E_{Fn} + \chi, \quad (\text{A.8})$$

$$\Phi_p = E_{Fp} + \chi + E_G. \quad (\text{A.9})$$

Here, the quantities $\Phi_{n,p}$ have been introduced to simplify the notation. In equilibrium conditions where $E_{Fn} = E_{Fp} = E_F$ the PB equation is exact, making it useful for calculating equilibrium band diagrams. In non-equilibrium situations, on the other hand, the gradients in E_{Fn} and E_{Fp} can be estimated from the typical currents that run through the system and the typical electron and hole densities. For many common devices (such as a *c*-Si solar cells working at max-power point under AM1.5 illumination) it is found that the quasi Fermi levels are approximately constant, so the potential applications of the PB equation extend beyond simple equilibrium systems. Moreover, in many cases it can even be used in situations where one of the quasi Fermi levels is not constant, such as in selective contacts. Consider, e.g., again the hole selective membranes from Chapter 3: in such selective membrane, E_{Fn} is not constant (as seen in Figures 3.2 and 3.6), but since the electron density is so low anyway, the value of E_{Fn} is immaterial when calculating the charge density. So for selective contacts it can safely be assumed that E_{Fn} is simply equal to its bulk value ($E_{Fn}(0)$ in the figures) everywhere.

Compared to the three coupled, non-linear, drift-diffusion equations (A.5), the PB equation (A.7) provides a simpler alternative for the calculation of band diagrams. Furthermore, in situations where it is necessary to evaluate the full set of transport equations, the PB equation can still provide a good initial guess of the solution for iterative numerical solvers (such as a Newton solver).

If the material parameters (N_{dop} , χ , E_G , $N_{C,V}$) are constant in the interval $0 \leq x \leq d$, it is also possible to integrate the PB equation analytically. As an example, take $x = 0$ as a point somewhere in the bulk of a solar cell and $x = L$ as the edge of the cell where a fixed-charge dielectric such as Al_2O_3 is present. The total surface charge in the dielectric is Q_{Fixed} (in cm^{-2}). This surface charge induces a space charge region (SCR) in the semiconductor with an equal and opposite amount of charge $Q_{\text{SCR}} = -Q_{\text{Fixed}}$. The boundary conditions are then $\phi(0) = 0$ (which is simply a choice of the reference potential) and $\phi'(0) = 0$ (meaning that in the bulk of the cell the electric field has been shielded by the mobile charges). To get rid of ϕ'' , the PB equation is first multiplied by ϕ' and then it is integrated from 0 to L . For the left hand side this yields:

$$\int_0^L \phi' \phi'' dx = \frac{1}{2} \phi'(L)^2 \quad (\text{A.10})$$

Since $-\phi'/e$ is the electric field, it is related to the total charge Q_{SCR} in the SCR by:

$$Q_{\text{SCR}} = -\frac{\epsilon}{e^2} \phi'(L). \quad (\text{A.11})$$

Integration of the right hand side of the PB equation (after multiplication by ϕ') requires a primitive function of $F_{1/2}$, which is denoted by $F_{1/2}^{(-1)}$ (the (-1)

should be interpreted as negative differentiation):

$$\frac{d}{dx} F_{1/2}^{(-1)}(x) = F_{1/2}(x). \quad (\text{A.12})$$

In the regime of Maxwell-Boltzmann statistics where $F_{1/2}(x) = \exp(x)$, the simplest choice for $F_{1/2}^{(-1)}(x)$ is, of course, the exponential function $\exp(x)$ again. In the regime of Fermi-Dirac statistics, the easiest way to proceed is to tabulate the integral of $F_{1/2}(x)$ for fast numerical evaluation. The integrated form of the PB equation is then:

$$Q_{\text{SCR}}^2 = \frac{2k_{\text{B}}T\epsilon}{e^2} \left[N_{\text{V}} \left\{ F_{1/2}^{(-1)} \left(-\frac{\Phi_p + \phi(L)}{k_{\text{B}}T} \right) - F_{1/2}^{(-1)} \left(-\frac{\Phi_p}{k_{\text{B}}T} \right) \right\} + N_{\text{C}} \left\{ F_{1/2}^{(-1)} \left(\frac{\Phi_n + \phi(L)}{k_{\text{B}}T} \right) - F_{1/2}^{(-1)} \left(\frac{\Phi_n}{k_{\text{B}}T} \right) \right\} - \frac{\phi(L)N_{\text{dop}}}{k_{\text{B}}T} \right]. \quad (\text{A.13})$$

Thus, Eq. (A.13) relates the surface potential $\phi(L)$ to the total amount of space charge Q_{SCR} in the semiconductor. The equation cannot be analytically inverted (even in the Maxwell-Boltzmann regime) to give $\phi(L)$ as a function of Q_{SCR} , so a numerical root-finding algorithm is required if the surface potential is the unknown. Brent's method [4] is a robust algorithm that is well-suited for this task. When solving Eq. (A.13), care should be taken that the sign for $\phi(L)$ (which should be opposite of the sign of Q_{SCR}) is correct.

References for Appendices A.1 and A.2

- [1] X. Aymerich-Humet, F. Serra-Mestres, and J. Millán. A generalized approximation of the Fermi-Dirac integrals. *Journal of Applied Physics*, 54(5):2850–2851, 1983. ISSN 00218979. doi: 10.1063/1.332276.
- [2] Harald Ibach and Hans Lüth. *Solid-State Physics*. Springer-Verlag Berlin Heidelberg New York, Berlin, Heidelberg, 3 edition, 2003. doi: Phys-RevB.83.165113.
- [3] R.B.M. Girisch, R.P. Mertens, and R.F. De Keersmaecker. Determination of Si-SiO₂/ interface recombination parameters using a gate-controlled point-junction diode under illumination. *IEEE Transactions on Electron Devices*, 35(2):203–222, 1988. ISSN 00189383. doi: 10.1109/16.2441.
- [4] R. P. Brent. An algorithm with guaranteed convergence for finding a zero of a function. *The Computer Journal*, 14(4):422–425, apr 1971. ISSN 0010-4620. doi: 10.1093/comjnl/14.4.422.

Summary

Passivating contacts for silicon photovoltaics: Solar cells designed by physics

The central theme of this dissertation is to investigate how the knowledge of solar cell physics can be used to contribute to the improvement of solar cell designs. To answer this question, the traditional semiconductor physics that are commonly used to describe photovoltaic devices have been reframed in a thermodynamic language, since this makes it possible to accurately quantify all efficiency losses in a solar cell. To make a truly thermodynamic description of solar cells possible, the field of non-equilibrium thermodynamics was reviewed and applied to the area of solar cell physics. A discussion was presented of the differences and similarities between the thermodynamic approach and the traditional semiconductor description of solar cells.

From the thermodynamic view of a solar cell, the local entropy generation emerged as a new figure of merit to assess the performance of the cell. Where traditional solar cell performance parameters only provide global measures, the entropy generation rate provides local information based on all internal loss mechanisms (such as recombinative and ohmic losses) and measures their effect all on the same scale. This local information was used to develop a variational optimization method that can directly predict the most efficient cell design consistent with the physical limitations of the cell (such as the recombination mechanisms and the doping dependence of the mobility). The variational method was illustrated by showing how it can be used to formulate a system of differential equations that determine the optimal shape for the doping profile in a *c*-Si homojunction cell in complete generality.

One of the most important sources of entropy generation in *c*-Si homojunction solar cells is the recombination at the metal contacts. This loss mechanism was discussed in detail and several strategies to reduce contact losses were categorised and discussed. These strategies are summarised with the term selective membranes and selective contacts. Key features for high quality selective membranes were identified and discussed from a thermodynamic point of view.

The most successful type of selective membrane found in literature is based

on hydrogenated amorphous silicon (a -Si:H). The main drawback of this type of membrane is its very substantial parasitic absorption of light in the a -Si:H. To reduce the parasitic absorption of a -Si:H, a new type of selective contact based on Al_2O_3 was introduced. The passivation and contact performance of this contact was investigated experimentally. It was concluded that Al_2O_3 cannot fully replace a -Si:H yet, however it does make the use of thinner a -Si:H films feasible to reduce the parasitic absorption.

Graphene, a material well known for its high transparency and conductivity, was also investigated for its potential as a transparent electrode and as a selective contact. The physics of graphene that are relevant for solar cell application were reviewed to assess its viability as a contact material c -Si solar cells. It was concluded that graphene shows potential as a selective contact because it is significantly less recombination active than an ordinary metal. However, graphene is only usable in solar cells if a stable doping scheme can be found to engineer its work function.

In conclusion, the work in this dissertation explores how the physics underlying the operation of solar cells can be used not just to describe the photovoltaic conversion process (which current semiconductor physics are already very successful at), but also how to predict the optimal design of solar cells.

List of publications

- [1] R M W van Bijnen, S. Smit, K A H van Leeuwen, E J D Vredenburg, and S J J M F Kokkelmans. Adiabatic formation of Rydberg crystals with chirped laser pulses. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 44(18):184008, sep 2011. ISSN 0953-4075. doi: 10.1088/0953-4075/44/18/184008.
- [2] D. Garcia-Alonso, S. Smit, S. Bordihn, and W. M. M. Kessels. Silicon passivation and tunneling contact formation by atomic layer deposited $\text{Al}_2\text{O}_3/\text{ZnO}$ stacks. *Semiconductor Science and Technology*, 28(8):082002, August 2013. ISSN 0268-1242. doi: 10.1088/0268-1242/28/8/082002.
- [3] S. Smit, D. Garcia-Alonso, S. Bordihn, M. S. Hanssen, and W. M M Kessels. Metal-oxide-based hole-selective tunneling contacts for crystalline silicon solar cells. *Solar Energy Materials and Solar Cells*, 120(PART A):376–382, January 2014. ISSN 09270248. doi: 10.1016/j.solmat.2013.06.016.
- [4] Benedicte Demaurex, Johannes P. Seif, Sjoerd Smit, Bart Macco, W. M. M. Erwin Kessels, Jonas Geissbuhler, Stefaan De Wolf, and Christophe Ballif. Atomic-Layer-Deposited Transparent Electrodes for Silicon Heterojunction Solar Cells. *IEEE Journal of Photovoltaics*, 4(6):1387–1396, nov 2014. ISSN 2156-3381. doi: 10.1109/JPHOTOV.2014.2344771.
- [5] B Macco, D Deligiannis, S Smit, R a C M M van Swaaij, M Zeman, and W M M Kessels. Influence of transparent conductive oxides on passivation of a-Si:H/c-Si heterojunctions as studied by atomic layer deposited Al-doped ZnO. *Semiconductor Science and Technology*, 29(12):122001, dec 2014. ISSN 0268-1242. doi: 10.1088/0268-1242/29/12/122001.
- [6] Sjoerd Smit and W. M. M. Kessels. Variational method for the minimization of entropy generation in solar cells. *Journal of Applied Physics*, 117(13):134504, 2015. ISSN 0021-8979. doi: 10.1063/1.4916787.
- [7] Harm C. M. Knoop, Bas W. H. van de Loo, Sjoerd Smit, Mikhail V. Ponomarev, Jan-Willem Weber, Kashish Sharma, Wilhelmus M. M. Kessels, and Mariadriana Creatore. Optical modeling of plasma-deposited ZnO films:

Electron scattering at different length scales. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 33(2):021509, mar 2015. ISSN 0734-2101. doi: 10.1116/1.4905086.

- [8] Benjamin L. Williams, Sjoerd Smit, Bas J. Kniknie, Klaas J. Bakker, Wytze Keuning, W. M. M. Kessels, Ruud E. I. Schropp, and M. Creatore. Identifying parasitic current pathways in CIGS solar cells by modelling dark J-V response. *Progress in Photovoltaics: Research and Applications*, 23(11): 1516–1525, nov 2015. ISSN 10627995. doi: 10.1002/pip.2582.

Acknowledgements

And now for the most important part of this work, as people tend to call the acknowledgements in humorous vein. Then again: what would life be like if no one would go through the effort to show appreciation to the people around them? So in the grand scheme of things, the acknowledgements may indeed be the most important part of the thesis when it comes right down to it. It is my own opinion that expressing gratitude is not one of my strong points and I probably fell short in that particular area at various moments during my PhD (and indeed in my life in general). Hopefully the words I'm about to write down will go some way towards correcting for that shortcoming, though one cannot shake the feeling that words on paper are ultimately little more than inadequate substitutes for the real thing.

First and foremost I'd like to thank everyone I met during my time at **PMP**. I know that such a catch-all phrase can come off as somewhat cheap, but PMP has been such a nice and stimulating group to work in over the last four years. I really mean it when I say I'd like to thank each and every one of you (and I'm picturing many of your faces as I write this) for your contribution to the atmosphere of the group. Hopefully I will be able to thank many of you in person when I'm back in Eindhoven for my defence, but inevitably there will be (ex-)PMP members left I will not be able to see. If you are one of these people, hopefully you will read this some time in the future and know that I'm happy that I met you and worked alongside of you.

Of course, there are some PMP members I got to know better than others. **Erwin**, thank you for supporting my work from the beginning to the very end, even when I just went off on my own to chase ideas that were rather far removed from the original definition of the project. Your confidence in my abilities helped me so much to gain that confidence for myself. I couldn't have wished for a better mentor. **Jeanne** and **Lianne**, both of you have helped me with so many administrative chores it would have been impossible to keep count. Your patience often seemed unending and I can only hope that I haven't tested it too frequently. **Cristian**, **Wytze**, **Janneke**, **Joris**, **Ries**, **Caspar**, and **Bernadette** (not particularly in that order): I think we can all agree that my experimental side in physics was never as strong as my theoretical side, but it certainly improved a lot thanks to your help and patience. **Bart**,

Bas, Jimmy, and Harm: my fellow PV friends and connoisseurs. Thank you so much for all of the discussions and fruitful cooperation. Om maar niet te spreken van het gezwam tijdens de koffiepauzes, lunch and de reizen naar de SiPV :). Door jullie was het allemaal zoveel leuker. **Diana**, thank you for being such a nice and helpful colleague and friend and for your company during the many train rides we made together. **Stefan Bordihn**, I'm grateful for your help in getting my first experimental results. Your input was always valuable.

Special thanks go to everyone in the **FLASH** project and at **STW** who helped support this project. I have good memories of the cooperation with the people at **ECN**, particularly **Paula** and **Petra** for their experimental support and **Gaby** for our theoretical discussions. I would also like to send my thanks to **Bénédicte** and **Johannes** for their help in the joint experiments between PMP and EPFL.

After almost eleven years, this is (for the moment at least) my goodbye to the **Eindhoven University of Technology**. I grew more attached to it than I ever thought I would and I would like to thank everyone involved and everyone I met for making this the great and “gezellige” university it has been during my time here.

Maaike, heel erg bedankt voor je steun, je liefde, en voor het nieuwe avontuur in Oxford waar je me in meegetrokken hebt.

Pa, ik weet dat ik veel van mijn sterke en zwakke punten van jou heb. Zoals je soms hopeloze pogingen om dankbaarheid naar mij te tonen en omgekeerd, maar ook de mateloze interesse in techniek die we delen. Weet alsjeblieft hoe veel het voor me betekent wat je allemaal voor me gedaan hebt. En nog veel belangrijker: hoe belangrijk het voor me is dat je uiteindelijk ervoor hebt gekozen om je eigen leven weer op de rails te krijgen. Zelfs al gaat het niet altijd perfect tussen ons, je bent en blijft mijn lieve onmisbare pappa.

Finally, a written work is nothing without a **reader**. Whoever you are reading this, I thank you for giving some of your time to the thoughts I tried to solidify in this thesis.

Curriculum Vitae

09-10-1986	Born in Waalwijk, the Netherlands
1999-2005	Secondary education, Willem van Oranje College, Waalwijk, the Netherlands
2005-2009	Mixed propaedeutics in Applied Mathematics and Applied Physics, Eindhoven University of Technology, Eindhoven, the Netherlands Bachelor of Science in Applied Physics, Eindhoven University of Technology, Eindhoven, the Netherlands
2009-2010	Master of Science, Education and communication (unfinished), Eindhoven University of Technology, Eindhoven, the Netherlands Teaching internship at Koning Willem II college, Tilburg, the Netherlands
2009-2011	Master of Science in Applied physics, Eindhoven University of Technology, Eindhoven, the Netherlands Traineeship at Phillips Research, Eindhoven, the Netherlands Master thesis project at the Coherence and Quantum Technology group, Department of Applied Physics, Eindhoven University of Technology, the Netherlands
2011-2016	PhD. student at the Plasma & Materials Processing group, Department of Applied Physics, Eindhoven University of Technology, the Netherlands
Present	Technical Consultant at Wolfram Research Europe Ltd., Long Hanborough, Oxfordshire, United Kingdom

To contact the author, feel free to send an e-mail to s.smit1986@gmail.com