Norwegian University of Life Sciences
Faculty of Environmental Science and Technology
Department of Mathematical Sciences and Technology

Master Thesis 2015
60 credits

Correlation of Defect Luminescence and Recombination in Multicrystalline Silicon

Guro Marie Wyller

Und Gott sprach:

„Es werde Licht!“

und es ward Licht.

Genesis 1.3

PREFACE

With this thesis I complete my master degree at the University of Life Sciences (NMBU). The thesis is the result of a collaboration between NMBU in Ås, Norway and Fraunhofer ISE in Freiburg, Germany.

Most of the experimental work related to my investigations is done at NMBU. I would like to thank my supervisor, Espen Olsen, for his open office door and for his willingness to give career advice as well as discussing research results and physical phenomena. My gratitude also goes to Andreas Flø and Ingunn Burud for teaching med the hyperspectral imaging setup, and to Torbjørn Mehl for his openness to new ideas regarding this setup and for his help with measurements as well as data processing.

I have been in the lucky situation to spend most of my time during the work with this thesis at Fraunhofer ISE in Freiburg. I would like to thank my supervisor there, Martin Schubert, for his personal kindness and for his willingness to find solutions to administrative challenges as well as for sharing his knowledge regarding the topics of my thesis. I will express my largest gratitude also to Wolfram Kwapil and Florian Schindler for very useful discussions on my experimental results and for useful advice during data analyses as well as the writing process. Your feedback, proof reading and help have definitively increased the level of this thesis significantly.

Further, I would like to thank all my great colleges at Fraunhofer ISE, for making my time in Freiburg unforgettable, both at work and after work. Among them, I would most of all like to thank Freidemann Heinz og Maximilan Pospischil for so often taking me climbing, skiing, biking and running. Without you, my time in Freiburg would not have been what it has been. I will miss you!

Last but not least I will express my thankfulness to my father Torgeir Bruun Wyller for his invaluable help with formatting and proof reading, and to my mother Liv Holtan-Hartwig for her support and care during the last phase of this work.

Nordstrand, May 15, 2015

Guro Marie Wyller

ABSTRACT

Defect related luminescence (DRL) of mc-Si wafers, including the four D line emissions D1-D4, is investigated by hyperspectral photoluminescence (PL) imaging. The background subtraction scheme for the hyperspectral imaging setup is improved in order to obtain enhanced possibilities for comparing the DRL of different samples. In combination with PL based techniques for lifetime and iron imaging, the improved hyperspectral imaging technique is used to compare DRL of n-type and p-type mc-Si material, and to study changes of the DRL spectrum along the height of a crystalline ingot. Further, the correlation between DRL and metallic impurities as well as changes of DRL due to solar cell processing steps are investigated. No differences in the D line emissions that with certainty can be attributed to differences in material type (n/p) are found. We suggest that the spectral shape rather is determined by the recombination mechanism through which the charge carrier lifetime of a sample mainly is limited. In regions with high concentration of iron and other contaminations, we observe reduced intensities of the D3 and D4. It is thus likely that precipitates of iron or other impurities partly supress the D4 and D3 emission intensities.

SAMMENDRAG

Defektrelatert luminescens (DRL) fra mc-Si-wafere, herunder de fire D-emisjonslinjene D1-D4, ble undersøkt ved hjelp av hyperspektral fotoluminescensavbildning (PL-avbildning). Rutinen for bakgrunnssubtraksjon for den hyperspektrale avbildningen ble forbedret for å gi bedre mulighet for å sammenlikne DRL fra ulike prøver. Den forbedrede hyperspektrale avbildningsteknikken ble kombinert med PL-baserte teknikker for avbildning av levetid og jernkonsentrasjon, og benyttet til sammenlikninger av DRL i n-type og p-type mc-Si. Samme teknikk ble også benyttet for å studere hvordan DRL-spekteret varierer med høyden i en støpt krystallblokk. Forholdet mellom DRL og metalliske urenheter samt endringer i DRL knyttet til produksjonsleddene ved fremstilling av solceller ble også studert. Det ble ikke funnet forskjeller i D-emisjonslinjene som med sikkerhet kan knyttes til ulik materialtype (n/p). Vi antar at spekterets egenskaper snarere avhenger av hvilke rekombinasjonsmekanismer som er begrensende for ladningsbærerlevetiden i prøven. Vi fant redusert D3- og D4-intensitet i områder med høy konsentrasjon av jern og andre forurensninger. Det er derfor sannsynlig at utfellinger av jern og andre urenheter delvis hemmer intensiteten av D4 og D3-emisjonene.

CONTENTS

LIST OF ABBREVIATIONS

BCM	Background correction matrix
BCMR	Background correction matrix reference
DRL	Defect related luminescence
EBIC	Electron Beam Induced Current
EL	Electroluminescence
HLI	High-level injection
LLI	Low-level injection
mc-Si	Multicrystalline Silicon
PL	Photoluminescence
SRH	Shockley Read Hall

LIST OF SYMBOLS

Symbol	Interpretation	Unit
E_F	Fermi Level	eV
E_V	Valence band edge	eV
E_C	Conduction band edge	eV
E_G	Energy band gap	eV
E_i	Fermi level in intrinsic semiconductor	eV
N_V	Effective density of states in valence band	cm^{-3}
N_C	Effective density of states in valence band	cm^{-3}
N	Net doping density	cm^{-3}
N_A	Density of acceptors	cm^{-3}
N_A	Density of donors	cm^{-3}
N_t	Density of traps	cm^{-3}
n	Electron density in non-equilibrium	cm^{-3}
n_i	Intrinsic charge carrier density	cm^{-3}
n_0	Electron density in thermal equilibrium	cm^{-3}
p	Hole density in non-equilibrium	cm^{-3}
p_{n0}	Hole density in thermal equilibrium	cm^{-3}
n_{n0}	Electron density in thermal equilibrium	cm^{-3}
p_0	Hole density in thermal equilibrium	cm^{-3}
Δn, Δp	Excess electron density, excess hole density	cm^{-3}
n_1, p_1	SRH-densities	cm^{-3}
σ_n	Capture cross section for electrons	cm^2
σ_p	Capture cross section for holes	cm^2
μ_n	Mobility of electrons	cm^2/Vs
μ_p	Mobility of holes	cm^2/Vs

Symbol	Interpretation	Unit
G_{th}	Rate of thermal generation of charge carriers	$cm^{-3}s^{-1}$
G	Rate of generation of charge carriers	$cm^{-3}s^{-1}$
G_L	Rate of light induced generation of charge carriers	$cm^{-3}s^{-1}$
R_{th}	Rate of recombination of charge carriers in thermal equilibrium	$cm^{-3}s^{-1}$
R_{SRH}	Rate of SRH recombination	$cm^{-3}s^{-1}$
U	Net recombination rate	$cm^{-3}s^{-1}$
β	Radiative recombination constant	$cm^{-3}s$
k	Boltzmann constant	J/K
v_{th}	Thermal velocity	cm/s
D_n	Diffusion coefficient of electrons	cm^2/s
D_p	Diffusion coefficient of holes	cm^2/s

1 INTRODUCTION

Multicrystalline silicon (mc-Si) is currently the dominating material in the silicon photovoltaic market [1]. Due to its low production costs compared to alternative technologies, such as monocrystalline silicon, it is a promising material for solar cells also in the future. The energy conversion efficiency of mc-Si solar cells is, however, lower than that of monocrystalline solar cells. The reduced efficiency is caused by charge carrier recombination due to crystal imperfections such as dislocations, grain boundaries and impurities. Therefore, an extended understanding of the defects in mc-Si may significantly contribute to the improvement of mc-Si solar cells.

Detecting defect related luminescence (DRL) has recently, by various authors, been used as a tool for the characterisation of defects in silicon. Among others, the origin of DRL [2, 3], its dependence on metallic impurities [4, 5] and its variations due to solar cell processing steps [6, 7] have been studied. These studies, however, mostly utilize *either* spatially *or* spectrally resolved techniques. This results in images of the emissions of one wavelength or one wavelength interval, or of photoluminescence spectra of single point or microscopic regions, respectively. The relevance of the measured spectra for entire cell is therefore questionable.

At NMBU in Norway a multispectral imaging setup, allowing for fast and non-destructive detection of DRL spatially as well as spectrally, is developed [8]. Moreover, at Fraunhofer ISE advanced photoluminescence (PL) based techniques for imaging the charge carrier lifetime [9] – which is a good preliminary indicator for the energy conversion efficiency of the final solar cell – as well as the interstitial iron content [10] are available. The combination of these facilities provides a unique possibility of investigating DRL and its relation to lifetime and crystal imperfections. Through this combination, this work aims to contribute to the understanding of crystal imperfections in mc-Si wafers and their effect on carrier lifetime.

In the recent years, the marked share and research activity related to monocrystalline n-type silicon solar cells has increased. It is shown [11], that also for mc-Si, the efficiency potential is larger for n-type than for p-type materials. This is, among others, due to electrical advantages such as reduced impact of most metallic impurities on charge carrier lifetime compared to p-type mc-Si material. It is therefore of interest to investigate the effect of other defects in n-type compared to p-type material. For that purpose, a first aim of this study is to compare DRL of n-type and p-type mc-Si wafers. Further, since the dislocation density, as well as the impurity concentration varies along the height of an ingot [12], the DRL from various ingot position of a p-type mc-Si is investigated as a second topic.

Iron is one of the most important contaminants in mc-Si, its presence often limiting the efficiency of the mc-Si solar cells [13]. As third topic of this work, the impact of iron and other metal impurities on DRL is therefore investigated. In order to minimize the impact of material defects and impurities, and thereby increasing the energy conversion efficiency of finished solar cells, it is of outmost importance to understand how material defects develop during solar cell

processing steps. Thus, as a fourth part of this work, the change of DRL caused by solar cell processing steps is studied.

The remaining chapters in this thesis are organized as follows: Chapter 2 gives an introduction to the semiconductor theory and charge carrier dynamics which is needed for understanding the concept of DRL. Further, a short overview of existing literature related to DRL is given. In chapter 3, the measured samples as well as the applied experimental setups are presented. Theory which is closely related to the measurement techniques is also included. Moreover, this chapter includes measurements related to the hyperspectral measurement setup as such. Chapter 4 is dedicated to the data processing techniques utilized in this work. The description of an enhanced background subtraction procedure for the hyperspectral images, which was developed as a part of this work, plays an important role in this chapter. Further, various ways of analysing and presenting the hyperspectral data is explained, and the reliability of each method is considered. In chapter 5, results related to each of the subtopics to be investigated are presented and discussed. Physical explanations for the observed effects are suggested, and parallels to already described effects are drawn. Finally, in chapter 6, the observations are summarised, and research methods for the further investigation of the suggested hypotheses are proposed.

2 THEORY

This chapter opens, in section 2.1, with a short introduction to the most basic concepts of semiconductors. In section 2.2 and 2.3 a more thorough explanation to charge carrier injection, recombination and lifetime is given. Unless other is mentioned, the theory and equations in these three sections are taken from references [14-17].

Section 2.4 introduces the concept of defects in mc-Si and in section 2.5 some solar cell processing steps that may change the distribution and impact of defects are explained. Finally, in section 2.6, defect related luminescence (DRL) is introduced, and an overview of relevant research results related to this topic is given.

2.1 BASICS OF SEMICONDUCTORS

2.1.1 Intrinsic semiconductors

Due to the short distance between atoms in a crystalline material, the wave functions of the valence electrons in neighbouring atoms will overlap, giving rise to bands of allowed energy states for electrons. Each state may be occupied by exactly one electron. The band gap E_G denotes the distance (in units of energy) between the valence band edge (E_V) and the conduction band edge (E_C). The valence band and the conduction band are the upper band which at T = 0 K is filled with electrons, and the lower band which is not filled with electrons, respectively.

A semiconductor is defined as a material in which the band gap E_G is so small (typically less than a few electron volt [17]) that electrons can be thermally excited from the valence band to the conduction band, leaving unoccupied states, called *holes*, in the valence band. The electrons in the conduction band and the holes in the valence band are denoted *free charge carriers*. In a semiconductor sample at T = 0 K, no thermal excitation will take place, meaning that *all* states in the valence band and *no* states in the conduction band will be occupied by electrons. Thus, no free charge carriers are available, and the sample cannot conduct electricity.

As the temperature is increased, thermal excitation takes place, causing the distribution of electrons within the bands to change. At a certain temperature *T*, given in K, the probability that a state with energy E is occupied by an electron is described by the Fermi-Dirac-Distribution function

$$F(E) = 1/(1 + e^{(E-E_F)/kT}) \qquad \textbf{(2.1)}$$

where *k* is the Boltzmann constant, and E_F denotes the *Fermi level*, meaning the energy level at which the occupation probability exactly equals one half. In an intrinsic semiconductor, meaning a semiconductor without any impurity atoms, the Fermi level is located close to the centre of the band gap. Thus, close to the band edges, the difference $(E\text{-}E_F)$ in the Fermi function is much

larger than the product *kT* (which at room temperature equals 0.026 eV). Thus, the Fermi distribution can be approximated to $(E) \approx e^{-(E-E_F)/kT}$. This is commonly denoted the *Boltzmann approximation*. Based on this approximation, the density of electrons n_0 in the conduction band and holes p_0 in the valence band can be expressed as

$$n_0 = \mathrm{N_C} e^{-(E_C - E_F)/kT} \qquad \textbf{(2.2)}$$

and

$$p_0 = \mathrm{N_V} e^{-(E_F - E_V)/kT} \qquad \textbf{(2.3)}$$

$\mathrm{N_C}$ and $\mathrm{N_V}$ denote the effective density of states in the conduction band and the valence band, respectively. In an intrinsic semiconductor in thermal equilibrium, the density of free electrons n_0 equals that of holes p_0. This density is called the *intrinsic charge carrier density*, and is denoted n_i. In silicon at room temperature n_i approximately equals 1.45×10^{10} $\mathrm{cm^{-3}}$ [14].

2.1.2 Doped semiconductors

Intentional introduction of impurity atoms in semiconductors is denoted *doping*, and is essential for the application of semiconductors in solar cells. Mostly, an introduced impurity atom will replace a lattice atom in the semiconductor and build covalent atomic bonds to the neighbouring lattice atoms. If the introduced atom has *more* valence electrons than needed for building covalent bonds to its neighbouring atoms, the excess electron will only be loosely bound to the impurity atom. It will thus easily be excited or *donated* to the conduction band, the impurity atom thus being called a *donor*. If, on the other hand, the impurity atom has *less* valence electrons than needed for making bindings to its neighbouring atoms, the atom is denoted an *acceptor*.

A semiconductor in which donors are introduced is called an *n-type* semiconductor. Since the introduction of donor atoms will increase the number of free electrons in the sample, an n-type semiconductor will, as opposed to an intrinsic semiconductor, have a much larger number of electrons in its conduction band than of holes in its valence band. The electrons are therefore said to be its *majority charge carriers*, whereas the holes are its *minority charge carriers*. Similarly, a sample in which acceptors are introduced is called a *p-type* semiconductor. In this case, holes are majority charge carriers and electrons are minority charge carriers. The densities of donor atoms in an n-type semiconductor and of acceptor atoms in a p-type semiconductor are denoted N_D and N_A, respectively.

Because of the introduction of more holes or electrons to the sample, its Fermi level will change. The actual Fermi level $\mathrm{E_F}$ is thus different from the intrinsic Fermi level $\mathrm{E_i}$. When inserting the new Fermi level in equations (2.2) and (2.3), and utilizing that $(E_c - E_F) = (E_C - E_i) - (E_F - E_i)$, the density n_0 of electrons in the conduction band and the density p_0 of holes in the valence band can be expressed as

$$n_0 = n_i e^{(E_F - E_i)/kT} \quad \textbf{(2.4)}$$

and

$$p_0 = n_i e^{(E_i - E_F)/kT} \quad \textbf{(2.5)}$$

From equations (2.4) and (2.5) we see that also in a doped semiconductor, as long as it is in thermal equilibrium, the square of the charge carrier densities equal the square of the intrinsic charge carrier density:

$$n_0 \cdot p_0 = n_i e^{\frac{E_F - Ei}{kT}} \cdot n_i e^{\frac{E_i - E_F}{kT}} = n_i^2 \quad \textbf{(2.6)}$$

At room temperature, for typical doping concentrations below 10^{17} cm^{-3}, one may assume that a semiconductor is *completely ionized*. This means, in the case of a n-type semiconductor, that all the extra electrons introduced by the donor atoms are excited to the conduction band. In a p-type semiconductor, complete ionization implies that all acceptor atoms have introduced one hole to the valence band. The density of doping atoms is in most cases much larger than the intrinsic charge carrier density. Thus, the number of majority charge carriers in thermal equilibrium can be approximate to

$$n_0 \approx N_D \quad \textbf{(2.7)}$$

for an n-type semiconductor and

$$p_0 \approx N_A \quad \textbf{(2.8)}$$

for a p-type semiconductor.

A region of a semiconductor material in which both acceptor and donor atoms are present is said to be partly *compensated*. The electrical effect of the two sorts of doping atoms then partly cancel each other; the extra electrons introduced by the donor atoms will compensate the holes introduced by the acceptor atoms. The impurity type which is present in a larger concentration will thus determine the material type (n or p) of the region, and the number of majority charge carriers is given by

$$p_0 \approx N_A - N_D \quad \text{or} \quad n_0 \approx N_D - N_A \quad \textbf{(2.9)}$$

The mobility μ of the charge carriers in a semiconductor depends on the *total* concentration of impurity atoms, and decreases when this value increases. Thus, the charge carrier mobility is smaller in a compensated than in an uncompensated semiconductor with the same charge carrier density.

2.2 CHARGE CARRIER INJECTION

When a semiconductor sample is subjected to an external excitation source, for instance illumination or forward biasing, the density of free electrons n will increase, causing the sample to leave its state of equilibrium, such that equation (2.6) no longer is fulfilled. The absolute deviation of the actual concentration of electrons n to its equilibrium concentration n_0 is called the *excess carrier concentration* and is denoted Δn. Likewise, the deviation of the concentration of holes p in the valence band to its equilibrium value p_0 is denoted Δp. However, since an electron being excited from the valence band to the conduction band always leaves a hole in the valence band, Δp will always equal Δn.

The process of generating excess charge carriers is referred to as *charge carrier injection.* Based on the magnitude of the number of injected carriers Δn compared to the magnitude of the number of carriers in equilibrium, the *injection level* is determined. An n-type semiconductor is said to be in *low-level injection* (LLI) when $\Delta n \ll n_0$. In the opposite case, if $\Delta n \gg n_0$, the sample is in *high-level injection* (HLI). Likewise, a p-type semiconductor is in low-level injection if $\Delta p = \Delta n \ll p_0$, and in high-level injection if $\Delta p = \Delta n \gg p_0$.

The change of Δn with time in a semiconductor sample is in general given by the continuity equation:

$$\frac{\delta \Delta n}{\delta t} + div(j) = G - R \qquad \textbf{(2.10)}$$

for which G is the rate of generation of charge carriers, R is their recombination rate and j is their current density, given by

$$j = -D_n \cdot grad(\Delta n) - \mu_n n E \qquad \textbf{(2.11)}$$

with D_n and μ_n being the coefficient of diffusion and the mobility of electrons respectively. If we assume that Δn is spatially constant and that no electrical field E influences the sample, equation (2.10) may be simplified to

$$\frac{\delta \Delta n}{\delta t} = G - R \qquad \textbf{(2.12)}$$

In addition to the thermal generation G_{th} always taking place in a semiconductor at T>0K, generation of charge carriers may happen through illumination. Photons from the illumination source will, if their energy E_{ph} is larger than the band gap energy E_G, excite electrons from the valence band to the conduction band when being absorbed, resulting in a light-induced generation rate G_L. The total generation is the sum of the thermal generation and the generation induced by light:

$$G = G_{th} + G_L \qquad \textbf{(2.13)}$$

2.3 RECOMBINATION AND LIFETIME

As soon as a semiconductor sample has left its state of thermal equilibrium, its excess charge carriers will tend to recombine, thus striving to restore the equilibrium state. There are various mechanisms for carrier recombination, each having its individual recombination rate. The recombination mechanisms may be categorised as *intrinsic*, meaning that their presence is unavoidable, even in a perfectly undisturbed and uncontaminated crystal, or as *extrinsic*, meaning that they are related to crystal defects and thus may be avoided. The most important recombination mechanisms are described in the following paragraphs.

2.3.1 Radiative band-to-band recombination

The direct annihilation of an electron from the conduction band and a hole from the valence band, involving the emission of a photon, is called radiative recombination, and can be considered as the opposite process of generation of charge carriers by illumination. This intrinsic recombination process is illustrated in Figure 2.1.a. We can assume that the rate of radiative recombination is proportional to the concentration of particles involved, namely an electron from the conduction band and a hole from the valence band. The following paragraphs describe the implications of this assumption for a semiconductor sample in thermal equilibrium respectively.

Band-to-band recombination in thermal equilibrium

In thermal equilibrium, and without illumination, the number of charge carriers will stay constant, meaning that the recombination rate must exactly balance the generation rate. For an n-type semiconductor, this means that

$$G_{th} = R_{th} = \beta n_{n0} p_{n0} \qquad \textbf{(2.14)}$$

n_{n0} and p_{n0} denote here the equilibrium concentrations of electrons and holes in a n-doped semiconductor sample, and β is a proportionality constant depending on the material. For semiconductors with an indirect band gap, meaning that the conduction band edge and the valence band edge are not related to the same crystal momentum, a band-to-band transition must involve a phonon in addition to the recombining electron-hole-pair. Such a band-to band-transition is thus less probable than the band-to-band transition in a semiconductor with a direct band gap. The proportionality constant β is, accordingly, smaller in indirect semiconductors, such as silicon.

Band-to-band recombination under illumination

As soon as the semiconductor sample is illuminated, the generation rate increase according to equation (2.13), to the value $G = G_{th} + G_L$. Thus, the recombination rate, which still is proportional to the concentration of involved particles, must increases to

$$R = \beta n_n p_n = \beta (n_{n0} + \Delta n)(p_{n0} + \Delta p) \qquad \textbf{(2.15)}$$

We can now define the *net recombination rate U*

$$U \equiv R - G_{th} = G_L \qquad \textbf{(2.16)}$$

meaning the number of charge carriers that recombine per unit time and volume *minus* those that are generated per unit time and volume due to thermal excitation. In steady state, this rate must equal the rate of charge carriers generated by illumination. By substituting the expressions for R and G_{th} from equations (2.14) and (2.15) into equation (2.16), and using that $\Delta n = \Delta p$ we arrive at

$$U = G_L = \beta(n_{n0} + p_{n0} + \Delta p)\Delta p \qquad \textbf{(2.17)}$$

Since we consider an n-type semiconductor, $p_{n0} \ll n_{n0}$. If we further assume low-level injection conditions, implying that $\Delta p \ll n_{n0}$, the expression in equation (2.17) may be simplified to

$$U = G_L = \beta n_{n0}\Delta p \qquad \textbf{(2.18)}$$

Thus, we see that under LLI conditions, the net rate of radiative recombination is proportional to the density n_{n0} of carriers in thermal equilibrium. As explained in section 2.1.1, at room temperature, for typical doping concentrations, this density can be approximated to the density of donor atoms N_D. Similarly, for a p-type semiconductor, the net rate of radiative recombination is proportional to the concentration of acceptor atoms N_A.

From equation (2.18), we further see that the net rate of radiative recombination is proportional to the excess carrier concentration Δp. The reciprocal of the proportionality constant βn_{n0} is called the *lifetime* of the minority carriers.

$$\tau_{p,rad} \equiv \frac{1}{\beta \mathrm{n_{n0}}} \qquad \textbf{(2.19)}$$

Similarly, in a p-type sample

$$\tau_{n,rad} \equiv \frac{1}{\beta \mathrm{p_{p0}}} \qquad \textbf{(2.20)}$$

Assuming that only radiative recombination takes place, the concentration of minority carriers under steady state illumination conditions can now be expressed as

$$\Delta n = G_L \tau_{n,rad} \quad \text{or} \quad \Delta p = G_L \tau_{p,rad} \qquad \textbf{(2.21)}$$

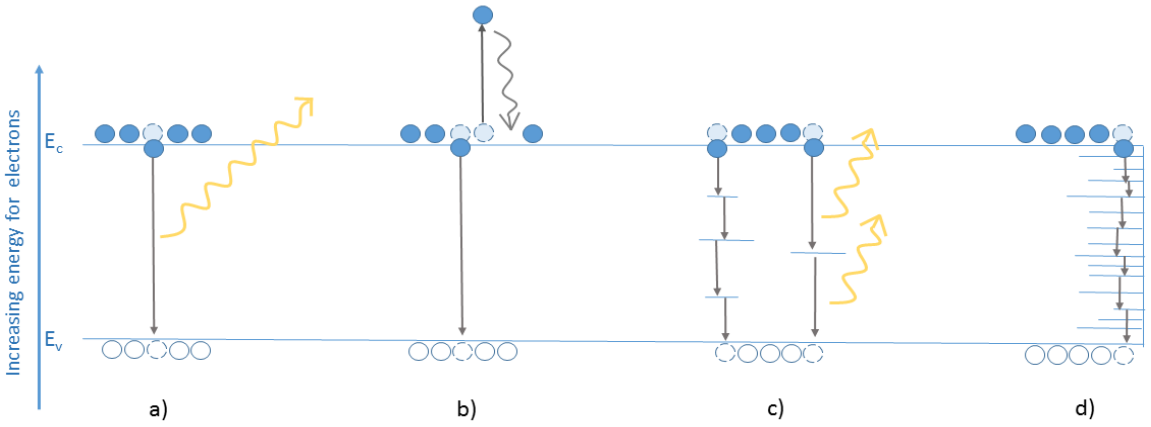


Figure 2.1: *Schematic drawing of the four most important recombination mechanisms for charge carriers in semiconductor materials, including a) Radiative recombination, b) Auger recombination, c) SRH-recombination (recombination via impurities) and d) surface recombination.*

2.3.2 Auger recombination

If the energy released from an annihilating electron-hole pair is transferred to a third charge carrier, rather than being emitted as a photon, Auger recombination takes place (see Figure 2.1.b). Like radiative recombination, Auger recombination is an intrinsic process and cannot be avoided. As for radiative recombination, the rate of Auger recombination is proportional to the density of the particles involved. For energy transfer to a second electron in the valence band, the Auger recombination rate is thus given as

$$R_{Auger}^{nnp} = C_n \cdot n^2 p \qquad (2.22)$$

Similarly, for energy transfer to a second hole in the valence band, the Auger recombination rate is

$$R_{Auger}^{npp} = C_p \cdot n p^2 \qquad (2.23)$$

C_n and C_p are experimentally determined Auger coefficients, for which the most cited values are $C_n = 2.8 \times 10^{-31} cm^6 s^{-1}$ and $C_p = 9.9 \times 10^{-32} cm^6 s^{-1}$, after Dziewior and Schmidt[1] [18]. In a similar way as shown for the radiative band to band recombination above, a lifetime related to the Auger recombination may in low-level injection conditions be defined as

$$\tau_{Auger}^{LLI,n} = \frac{1}{C_n n^2} \text{ and } \tau_{Auger}^{LLI,p} = \frac{1}{C_p p^2} \qquad (2.24)$$

In high-level injection, the corresponding values are

$$\tau_{Auger}^{HLI,n} = \tau_{Auger}^{HLI,p} = \frac{1}{(C_n + C_p)\Delta n^2} \qquad (2.25)$$

[1] These values are found for highly p- or n-doped material; N_A or N_D greater than $5 \times 10^{18} cm^{-6}$.

From these equations, and the definitions of the LLI and HLI conditions (see section 2.2), it can be seen that the rate of Auger recombination under LLI conditions should be injection-independent. Assuming $n = N_D$ or $p = N_A$, its magnitude should be proportional to the inverse square of the doping concentration. Under HLI conditions, however, the Auger recombination will be strongly injection dependent, its magnitude being proportional to Δn^3 [15].

2.3.3 SRH-recombination

When defects are present in a semiconductor sample, either in the form of impurities or crystal imperfections, energy levels within the band gap will be introduced. Such energy levels act as recombination centres, since the transition of a charge carrier from a band to the defect level is shorter, and thus more probable, than the band-to-band transition. As illustrated in Figure 2.1.c, recombination through defect levels might lead to the emission of photons with energy smaller than the band gap E_G. However, the energy released from the recombining electron hole-pair might also be emitted as a phonon, the recombination process thus being irradiative.

The recombination of charge carriers through defects is normally denoted SRH-recombination, after the researchers W. Shockley, W. Read [19] and R. Hall [20] who first published the theories regarding this recombination mechanism and its rate. The following, simplified, explanation of their theories mainly follows the structure given in [14].

The transitions of charge carriers between the carrier bands, via energy positions within the band, can schematically be divided in four different processes. These are a) electron capture, meaning the transition of an electron from the conduction band to the trap level, b) electron emission, meaning exactly the opposite of a, c) hole capture, meaning the transition of a hole from the valence band to the trap level, and d) hole emission, meaning exactly the opposite of c. The processes are illustrated in Figure 2.2.

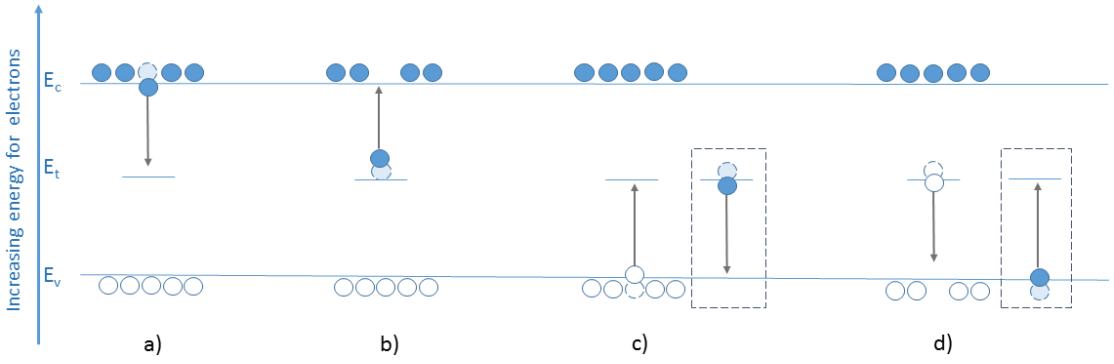


Figure 2.2: *The four basic processes that are involved in recombination through defect levels. a) electron capture, b) electron emission, c) hole capture, d) hole emission. The insets by c) and d) are in order to visualize that a hole capture also can be regarded as an electron moving from the trap level to the valence band, and that an hole emission might be considered as an electron moving from the valence band to the trap level.*

The probability that electron capture (process a in Figure 2.2) shall take place is proportional to the concentration of traps N_t that are *not* occupied by an electron (if the trap already is occupied, a new electron cannot be captured). The capture rate R_a can thus be expressed as

$$R_a \sim n N_t (1 - F) \tag{2.26}$$

For which *n* is the density of free electrons in the sample, and F is the probability that a centre *is* occupied, given by the Fermi distribution function (c.f. section 2.1.1)

$$F(E_t) = \frac{1}{1 + e^{(E_t - E_F)/kT}} \tag{2.27}$$

E_t is the energy level of the trap. The proportionality constant in equation (2.26) can be expressed by the product of the thermal velocity of the electrons in the sample ν_{th} and the capture cross section of the traps for electrons σ_n. Seeing from the reference frame of an electron, the product might be regarded as the volume over which a trap swipes per unit time. The rate of electron capture R_a is thus given by

$$R_a = \nu_{th} \sigma_n n N_t (1 - F) \tag{2.28}$$

Similarly, the rate of electron emission from a trap (process b in Figure 2.2), is proportional to the density of traps that are filled with an electron. The rate can be expressed as

$$R_b = e_n N_t F \tag{2.29}$$

The proportionality constant e_n is called the emission probability. Since, in equilibrium, the rates R_a and R_b must equal each other, e_n may be expressed as

$$e_n = \frac{\nu_{th} \sigma_n n (1 - F)}{F} \tag{2.30}$$

According to equation (2.27), the ratio $(1 - F)/F$ might be expressed as $e^{(E_t - E_F)/kT}$. By additionally using equation (2.4) for n, the emission probability can thus be written as

$$e_n = \nu_{th} \sigma_n n_i e^{(E_t - E_i)/kT} \tag{2.31}$$

In the same way, expressions for the probability of the other transitions in Figure 2.2 and their respective transition rates may be found. By considering that the change of rate of electrons in the conduction band must equal the change of rate of holes in the valence band, we can conclude that the net recombination R_{SRH} caused by the trap levels is

$$R_{SRH} = R_a - R_b = R_c - R_d \tag{2.32}$$

By inserting the expressions for R_a and R_b from equation (2.28) and (2.29) and similar expressions for R_c and R_d and eliminating F, we arrive at

$$R_{SRH} = R_a - R_b = \frac{\nu_{th} \sigma_n \, \sigma_p N_t (pn - n_i^2)}{\sigma_p \left(p + n_i e^{\frac{(E_i - E_t)}{kT}} \right) + \sigma_n \left(n + n_i e^{\frac{(E_t - E_i)}{kT}} \right)} \tag{2.33}$$

Using $n_i = N_V \exp(-(E_i - E_V)/kT) = N_C \exp(-(E_C - E_i)/kT)$ we can reformulate equation (2.33) as follows

$$R_{SRH} = \frac{\nu_{th}\sigma_n\sigma_p \quad N_t(pn - n_i^2)}{\sigma_p\left(p + N_V e^{\frac{(-(E_t - E_V))}{kT}}\right) + \sigma_n\left(n + N_C e^{\frac{(-(E_C - E_t)}{kT}}\right)} \tag{2.34}$$

By defining the so called SRH-densities

$$p_1 = N_V e^{\frac{(-(E_t - E_V))}{kT}} \text{ and } n_1 = N_C e^{\frac{(-(E_C - E_t))}{kT}} \tag{2.35}$$

we finally arrive at

$$R_{SRH} = \frac{\nu_{th}\sigma_n\sigma_p N_t(pn - n_i^2)}{\sigma_p(p + p_1) + \sigma_n(n + n_1)} \tag{2.36}$$

as an expression for the rate of recombination through defect levels in the band gap.

The SRH-densities p_1 and n_1 are statistical factors that, provided that the Fermi level was exactly lying at the energy level E_t of the trap, may be interpreted as the equilibrium densities of holes in the valence band and electrons in the conduction band, respectively. From equations (2.35) and (2.36) we interpret that if the difference in energy between the trap level E_t and the level of the band edges are large, meaning that E_t is close to the middle of the band gap, both the SRH-densities will be small, resulting in an increased recombination rate. Thus, deep trap levels (trap levels close to the middle of the band gap) are more recombination active than shallow trap levels (trap levels close to the band edges).

Based on equation *(2.36)*, the lifetime τ_{SRH} related to the SRH recombination can be expressed as

$$\tau_{SRH} = \frac{\tau_{n0}(p_0 + p_1 + \Delta n) + \tau_{p0}(n_0 + n_1 + \Delta n)}{p_0 + n_0 + \Delta n} \tag{2.37}$$

where τ_{n0} and $\tau_{p0}$are capture time constants for electrons and holes, defined as

$$\tau_{n0} = \frac{1}{N_t\sigma_n\nu_{th}} \text{ and } \tau_{p0} = \frac{1}{N_t\sigma_p\nu_{th}} \tag{2.38}$$

Due to coulomb forces, the capture cross section $\sigma_{n,p}$ of a defect may strongly depend on the charge of the carrier in question. Accordingly, its value for electrons can be very different form that for holes. As will be explained later (section 2.3.5), the lifetime of the *minority* charge carriers is vital to the energy conversion efficiency of solar cells. Since p-type and n-type materials have opposite minority charge carriers, the detrimental effect of one certain defect type can vary enormously between the two material types. The symmetry factor *k* of a defect is defined as the ratio of the capture cross section of the defect for electrons to its capture cross section for holes

$$k = \frac{\sigma_n}{\sigma_p} = \frac{\tau_{p0}}{\tau_{n0}} \quad (2.39)$$

Interstitial iron in silicon has capture cross section values $\sigma_p = 6.8 \cdot 10^{-17} cm^2$ and $\sigma_n = 4.0 \cdot 10^{-14} cm$ [21]. The symmetry factor is accordingly close to 600, illustrating that the influence of interstitial iron on lifetime is much larger in p-type than in n-type silicon. Similar features are seen also for other transition metals.

Obviously, SRH-recombination may be reduced by reducing the concentration of impurities and crystal imperfections in a solar cell. Different types of defects in multicrystalline silicon and ways of reducing their impact will be described in section 2.4 and 2.5 respectively.

2.3.4 Surface recombination

At a sample surface, the discontinuity of the crystal lattice introduces a large number of unsaturated atomic bindings, often described as *dangling bonds*. These causes energy states in the band gap, which, in turn, results in enhanced recombination activity. Unlike radiative recombination and Auger recombination, the surface recombination may be reduced. Such a reduction can be achieved by surface passivation, for instance by a thermally grown layer of SiO_2. Atoms from this layer tend to saturate the unsaturated bindings at the surface of the silicon crystal, thus reducing the number of energy states within the band gap.

Additionally, the density of minority charge carriers close to the surface - and thus the surface recombination rate – may be reduced by the use of *field effect passivation*. In p-type silicon, such a passivation is normally achieved by the application of a layer of Al_2O_3, carrying a relatively large fixed negative charge. Due to coulomb forces, the negatively charged minority carriers will therefore be repelled [16]. For n-type surfaces, an equivalent effect is achieved through the application of a SiN_x layer. This layer carries fixed positive charge repelling the positively charged holes, which in n-type silicon are minority charge carrier. Moreover, like SiO_2, SiN_x also tends to saturate unsaturated bindings at the sample surface.

2.3.5 Effective carrier lifetime

In a semiconductor sample, the four recombination mechanisms described above may occur simultaneously, and in combination with other recombination mechanisms not described here. They will result in a total recombination rate R_{eff} equal to the sum of the four individual recombination rates:

$$R_{eff} = R_{rad} + R_{Aug} + R_{SRH} + R_{surface} + R_{other} \quad (2.40)$$

For each recombination mechanism, a corresponding lifetime can be designated, which is inversely proportional to the recombination rate. The effective lifetime of the sample τ_{eff} is thus given by

$$\frac{1}{\tau_{eff}} = \frac{1}{\tau_{SRH}} + \frac{1}{\tau_{Auger}} + \frac{1}{\tau_{Rad}} + \frac{1}{\tau_{surface}} + \frac{1}{\tau_{other}} \quad (2.41)$$

Since the various recombination mechanisms react differently to changes in injection level, their relative impact on the total lifetime also changes as a function of Δn. Figure 2.3 shows simulated

values for the lifetimes related to various recombination channels in the silicon bulk as well as the effective lifetime τ_{eff} in dependency of Δn.

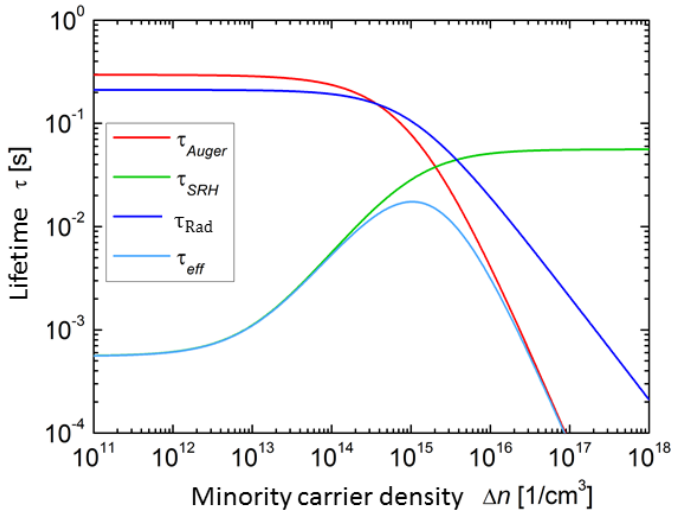


Figure 2.3: *Simulation of lifetimes related to different recombination mechanisms, as well as the effective lifetime, as a function of injection level. The simulation was done for p-doped silicon, with the following simulation parameters: p_0=1·10^{15} cm^{-3}, T = 300 K, defect level at E_t - E_V=0,4 eV, $\boldsymbol{\sigma_n}$=3,6·1015 cm2, k=100, density of empty traps: n_t=5·1012 cm-3. Reprinted from [22].*

Importance of minority carrier lifetime for the energy conversion efficiency of a solar cell

In an operating solar cell, charge carriers are exited due to illumination. Ideally, all free charge carriers produced by the illumination should be conducted to an external circuit, in which their potential energy can be used to preform electrical work. This must happen before they recombine within the cell. High recombination rates (i.e. short lifetimes) will, in other words, reduce the energy output from the solar cell. At normal doping concentration, the density of majority charge carriers is many orders of magnitude larger than the density of minority charge carriers. Therefore, even if the absolute rate of recombination of the two sorts of charge carriers must equal, the density of majority charge carries can be considered constant. The density of minority charge carriers, on the other hand, is highly impacted by recombination. The *minority charge carrier lifetime* is thus of outmost importance for the energy conversion efficiency of the solar cell.

2.4 DEFECTS IN MULTICRYSTALLINE SILICON

Generally, a crystalline defect is a region of a crystal in which the arrangement of atoms differs much from that of a perfect crystal lattice[23]. In this section, a short introduction to the various types of defects is given.

2.4.1 Point defects

Intrinsic point defects are the absence of a Si atom or the presence of an extra Si atom in the crystal lattice. Due to thermal excitation, in a silicon crystal at T > 0K, some Si atoms will leave their lattice position, thus creating a *vacancy* in the crystal lattice. The Si atom itself might diffuse through the crystal, as an *interstitial* point defect. Intrinsic point defects are, as indicated by the name, unavoidable and occur even in perfect crystals at temperatures above 0K [16, 23].

However, contamination atoms of other species such as iron, chromium, oxygen and carbon might be present in a silicon crystal, forming *extrinsic* point defects. They are, to some extent, present already in the silicon feedstock, but can also be introduced during the solar cell processing, especially in the crystallisation step [24]. Some contaminants (e.g. iron and chromium) exist in larger concentrations close to the ingot edges, due to solid-state in-diffusion from the crucible concentrations [25]. Contamination atoms may substitute Si atoms in the crystal lattice or be situated between the lattice atoms, thus being classified as *substitutional* or *interstitial* point defects, respectively [16, 23]. Figure 2.4 gives an overview of some metal contaminations in silicon, and the trap levels E_t that they introduce. Some impurity atoms, can exist in different electrical states, for example by donating a varying number of electrons, or by building complexes with various materials also being present in the sample. Thus, for one contamination species, various trap level may be introduced. As will be further explained, iron can, for example, exist interstitially or in complexes with boron, thus causing two different trap levels.

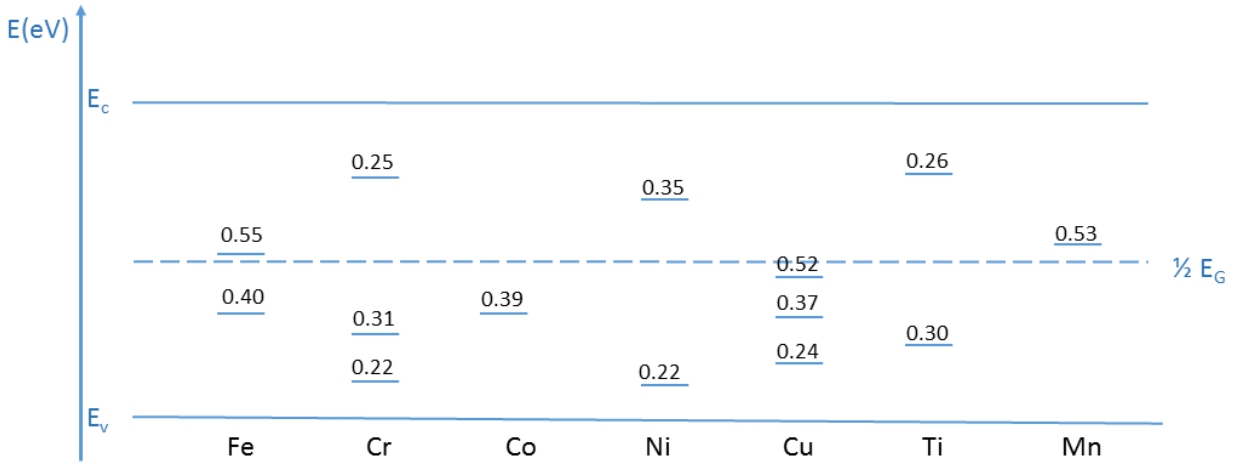


Figure 2.4: *Trap levels in of selected transition metals in silicon. For the levels in the upper half of the band gap, the numbers give the difference in energy between the conduction band edge EC and the trap level Et . For the levels in the lower half of the band gap, the numbers give the difference in energy between the trap level Et and the valence band edge EV. The numbers are taken from [26] and [27].*

Dopants such as phosphorus and boron are also, in principle, substitutional point defects. However, because of their shallow defect levels, their recombination activity is low compared to most other defects.

Metastable point defects

Some extrinsic point defects may exist in two or more electrically different states, and are thus called *metastable point defects.* Iron, for example, can exist as interstitial atoms in the silicon crystal (denoted Fe_i). Like other transition metals, it will tend to donate an electron to the material, thus being a positively charged ion. In boron-doped silicon, coulomb forces will cause an attraction between positively charged iron ions and negatively charged acceptor ions giving rise to donator-acceptor-pairs (denoted FeB). Interstitial iron in a boron-doped silicon wafer can, in other words, exist either in Fe_i state or in FeB state [16].

The bonding energy of the FeB pair is 0.66 eV. Accordingly, the donator-acceptor-pairs might reversibly be split by a corresponding increase in temperature. By lower temperatures, through diffusion, the FeB pairs will tend to reassociate. The same effect is seen for chromium forming CrB pairs in boron-doped silicon. A specialty for iron, however, is that the pairs may also be split through illumination or by another sort of charge carrier injection, for example through the application of a forward-biased voltage [16, 28].

Since the trap level E_t and the capture cross sections in the Fe_i state differ from those in the FeB state, also the corresponding recombination rates and the carrier lifetimes including their dependence on excess carrier density will differ.

2.4.2 Structural defects

Like point defects, also structural defects result in the introduction of energy levels within the band gap, thus leading to enhanced recombination [27]. Structural defects include *grain boundaries* and *line defects*, also known as *dislocations*. The latter may be categorized as *edge dislocations* and *screw dislocations*. Also combinations of the two exist [24].

Edge dislocations

An *edge dislocation* (Figure 2.5a) can be imagined as removing half of a lattice plane from a crystal, then pushing the remaining atomic layers together in order to preserve the regular crystal structure everywhere except along the edge of the removed half plane. As result, the neighbouring lattice planes will have to bend, in order to fill the empty space from which the half plane was removed [23]. Thus, in the vicinity of the half plane edge, tensions in and between the lattice planes will occur.

Screw dislocations

In a similar manner, a *screw dislocation* (Figure 2.5b) can be imagined as making a horizontal cut halfway into a crystal lattice, then pushing the upper half of the crystal to the right a distance equal to one lattice vector. Any atom in the bottommost atomic layer above the face of the cut will then be re-joined with the atom in the layer just below this face that originally was placed one unit length to the right of it. As for the edge dislocation, the body will then preserve its

regular crystalline structure except for the line along the edge of the cut, and shear forces in this region will occur [23].

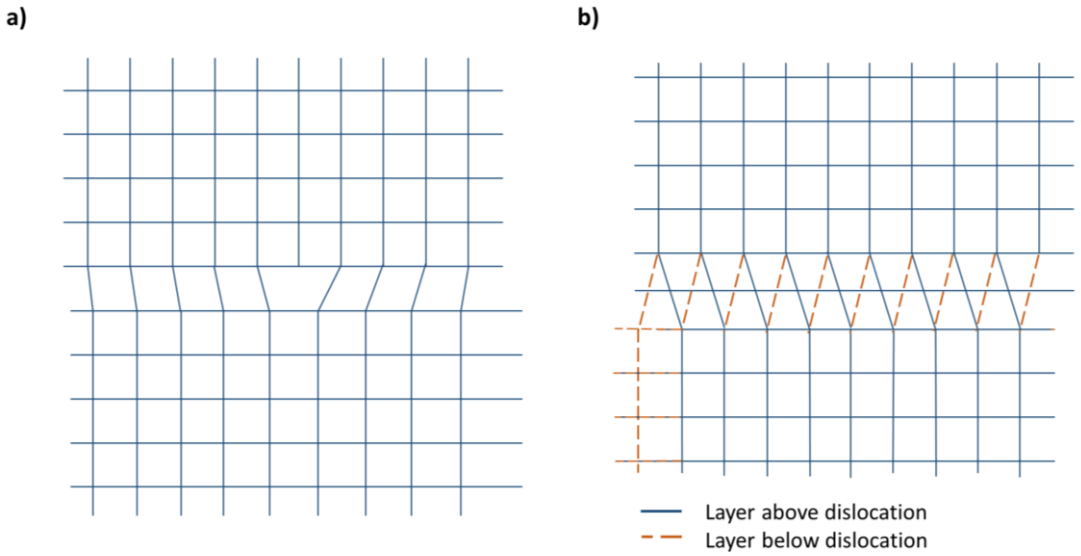


Figure 2.5: *Simplified sketch illustrating the concepts of a) edge dislocation and b) screw dislocation.*

Origin and evolution of dislocations

The dislocations originate from mechanical tension in the crystal during its solidification and the following cooling process. As long as the temperature is above a certain level, the dislocations are mobile, and may move along 12 different directions in crystal, all parallel to the {111}-planes, e.g. along all the atomic planes in the crystal. When a dislocation moves, new tensions in the crystal occur, giving rise to new dislocations. This process is called dislocation multiplication and may result in crystal regions with very high dislocation density compared to other crystal regions. Such regions are called dislocation clusters, and – as opposed to single dislocations – they spread mainly in the growth direction of the crystal [24]. Thus, the same pattern of dislocation clusters may be identified on wafers from block positions close to each other.

The thermal stress caused by crystal solidification can also cause parts of a crystal to be pushed and slide in different directions compared to each other, introducing so-called slip lines or slip dislocations [29].

Grain boundaries

Grain boundaries are the planes at which two crystal regions of different lattice orientation meet. Their characteristics and recombination activity depends, among other factors, on the angular deviation of crystal structure of the neighbouring crystal grains. Highly symmetrical grain boundaries, for which the change of lattice direction between the neighbouring grains only includes a 60˚ rotation around an axis perpendicular to the {111}planes, are called twins [24]. By very small angles between the crystal directions of subsequent grains (small-angle grain boundaries), the difference in direction is often compensated with a large number of dislocations close to each other. There are also theories that broken bonds exist at the boundaries. Thus, the recombination activity of such grain boundaries is mostly reported to be strong [30].

2.4.3 Precipitates

Precipitates are clusters of numerous impurity atoms in a crystal lattice [16]. In a silicon wafer, interstitial metal atoms (e.g. iron) are often present at concentrations above their solubility limit. These atoms have high diffusivities, which at elevated temperatures is further increased. Thus, at high temperatures, due to their oversaturation, metal tend to conglomerate (i.e. precipitate) in certain regions. This precipitation will often take place in the vicinity of a lattice defect, such as dislocations, grain boundaries or other impurity precipitates [31]. The defects around which the impurity atoms gather, are said to be *decorated* with impurities. Through precipitation, the total solute concentration of the impurity atoms and thus their mean recombination activity will decrease [16]. On the other hand, according to Kveder et al. [32], the recombination activity of dislocations increases strongly if they are decorated with metal impurities. They suggest that the shallow levels introduced by strain fields may exchange electrons and holes with deeper levels introduced by transition metals decorating the dislocations. Thus, even very small concentrations of impurity atoms may enhance the recombination activity of dislocation clusters dramatically.

2.4.4 Effect of temperature on the recombination activity of defects

The recombination activity of defect depend strongly on temperature [33]. When, for instance, the trap level introduced by a defect is close to the conduction band edge, an electron being captured can, if the temperature is sufficiently high, be thermally excited back to carrier band (c.f. Figure 2.2.). Thus, the probability that the electron recombines with a hole decreases. If, on the other hand, the temperature is lowered, the probability of a thermal excitation from the trap level to the conduction band decreases. Accordingly, the probability that a trapped electron instead recombines with a hole from the valence band will increase. The recombination activity of the trap will thereby increase. A similar statement holds true for trap levels close to the valence band.

2.5 Processes that change the distribution and impact of crystal defects

The active use of precipitation and similar processes in order to enhance material quality is called *defect engineering*. In this subsection, some defect engineering techniques that are relevant for the experiments done in this work are introduced. The section is not intended to give a full overview of processes that may change the distribution or impact of crystal defects. Unless other references are given, the information is taken from [16].

2.5.1 Gettering

Gettering refers to the removal or deactivation of defects in silicon. There are in principle two ways of achieving such an effect: By *external gettering*, contaminations are either totally removed from the semiconductor material or brought to regions of the material in which their detrimental effect is smaller. *Internal gettering*, on the other hand, denotes the precipitation of dissolved contamination atoms, resulting in a lower dissolved concentration and thus typically a lowered recombination activity.

External gettering

External gettering is, for photovoltaic applications, often achieved through the diffusion of a phosphorous layer (P-diffusion) on the wafer sample. In the phosphorous diffused region (the emitter), the solubility for contamination species will be higher than in the remaining material.

Consequently, a net diffusion of contamination atoms in the direction of the emitter will take place. A large part of the contamination atoms diffuse from the bulk material, in which they typically are detrimental to the local lifetime, to the diffused layer, in which the doping concentration mostly is so high that the local lifetime anway is limited by Auger recombination, rather than by the high impurity concentration [24]. The gettering process may, therefore, drastically increase the lifetime of the sample. In order to allow for measurements of the bulk lifetime, the emitter layer is for research purposes often totally removed from the sample through the use of an etchant after the gettering process.

Internal gettering

Internal gettering might be achieved through a controlled increase and subsequent decrease of the sample temperature to a carefully chosen value. The solubility of contamination atoms, in a semiconductor sample, depends strongly on its temperature. In solar grade silicon, the concentration of iron is normally in the range 10^{13}-10^{14} cm $^{-3}$. Thus, in order for all the iron atoms to possibly exist in a dissolved form, a temperature 800-1000˚C is needed. By such temperatures, the mobility of the impurities will also be high, resulting in an efficient diffusion process. A subsequent decrease of the process temperature decreases the solubility and results in areas with oversaturation. Thus, contamination atoms will tend to accumulate at precipitation nuclei such as crystal defects and grain boundaries. This process is called relaxation, and ensures a decrease in the concentration of resolved impurities in the vicinity of the precipitate. The concentration gradient around the precipitate is, in turn, a driving force for continued diffusion towards the developing precipitate.

The precipitation rate, and the total effect of the gettering process, depends also on type of contamination material and on the density of precipitation nuclei. For multicrystalline material – in which the density of defects (and thus of precipitation nuclei) is multiples higher than in monocrystalline silicon – the potential of improvement in material quality is accordingly large.

2.5.2 Passivation

Passivation is another method that strives to reduce the impact of defects in silicon. Some methods for passivation of sample surfaces were already described in section 2.3.4. However, defects in the crystal bulk may also be passivated. One way of achieving this is through the application of a layer of silicon nitride (SiN_x:H), followed by high temperature treatment. The high temperatures causes hydrogen from the applied layer to diffuse into the silicon sample and saturate unsaturated bindings related to dislocations and grain boundaries in the sample bulk that unless would be active carrier recombination sites. The impact of metallic defects, such as Fe_i, on a sample may also be reduced through H-passivation [34]. The mechanisms causing this positive effect is, however, still unclear.

2.6 DEFECT RELATED LUMINESCENCE (DRL)

Defect related luminescence (DRL) is a common name for all photons with energies lower than the bandgap energy E_G emitted from semiconductor samples when subjected to charge carrier injection. The origin of various categories of DRL, their dependency on temperature and their possible changes during solar cell processing have been extensively discussed during the last 30 years. The research results are partly contradictory. In this subsection, the principle of measuring DRL as well as well as band-to-band luminescence is briefly explained. Further, the section will give an overview of the DRL related research results that are of highest relevance for this work.

2.6.1 Detecting DRL

DRL is often measured by techniques based on photoluminescence (PL) – literally meaning light emission due to excitation by light – or electroluminescence (EL) – meaning light emission due to electrical excitation. Both methods utilizes the principle that charge carriers are excited to higher energy states due to an external excitation source (e.g. laser illumination or forward biasing). As the charge carriers subsequently return to their original state, they emit photons that can be detected by some sort of light senor (e.g. by a camera). Phonons may be emitted both from direct band-to-band recombination of electron-hole pairs, and due to recombination via defect levels in the band gap, the signal being denoted PL/EL and DRL respectively. From an ideal semiconductor without any defect levels in the band gap, no DRL would be detected.

Many of the methods for detection of DRL applies liquid nitrogen (e.g.[3, 35]) or helium (e.g.[36],) in order to cool the sample to cryogenic temperatures. When a sample is cooled, the amount of phonons available in the sample will decrease. This further means that the free charge carriers in the sample will tend to occupy the states that are related to the smallest possible amount of energy, e.g. the states close to the band edges. Thus, the recorded luminescence spectra will have sharp peaks, corresponding to the energy difference between the conduction band and the valence band, or to charge carrier transitions due to other energy levels. At higher temperatures, however, the charge carriers will tend to occupy states farther away from the band edges. Moreover, the probability that a recombining charge carrier interferes with a phonon, thereby absorbing or emitting energy, increases. Thus the recorded energy peaks will be less distinct.

Originally, the DRL signals have been recorded with various point based methods, meaning that only spectra from single points or microscopic regions could be recorded. The spectra have often been convoluted by a Gaussian fit algorithm, for example in [37] and [38], in order to resolve the individual peaks contributing to the signal. In 2007, due to advances of the high-sensitivity near-infra-red (NIR) camera technique, it became possible to obtain spatially resolved, but spectrally integrated images of sub bandgap wavelengths [39]. Thus, spatially resolved images of the defect band at about 0.8 eV could be obtained. Such imaging techniques are applied for example in [40] and [41].

2.6.2 Origin and temperature dependence of DRL

Emissions related to crystal defects in deliberately deformed silicon was firstly found by Drozdov et al. [36] in 1976, on n-type Si samples immersed in liquid helium (4.2K). The growing technique of the sample is not mentioned in the article. Due to the assumption that the observed emission lines were due to dislocations, they were called D-lines, and were labelled D1-D4, according to their energy level. The respective energy levels of the lines were reported to be as follows: D1 – 0.812eV, D2 – 0.875eV, D3 – 0.934eV, and D4 – 1.000eV.

In 1985, Sauer et al. [42] examined the D-lines and found that their emission intensity and spectral shape depends on the dislocation density, supporting the theory that the D-line emissions originate from dislocations. They further reported that the D-lines tended to appear in pairs of D1/D2 and D3/D4. It was suggested that the D3/D4 pair is related to the dislocations themselves, while the D1/D2 pair probably originates from deformation-produced point defects around the dislocations.

In 1983, Tajima and Mathsustia [43] found emissions similar to the D-lines in Czochralski-grown silicon (CZ Si) after annealing at 1000˚. Because CZ Si contains interstitial oxygen at higher concentrations than its solubility, such an annealing will lead to precipitation of oxygen, similar to the internal gettering process described in section 2.5.1. The oxygen precipitates will, in turn, lead to strain fields that cause dislocations in the silicon crystal [39]. The finding in [43] however pointed out the possible correlation between D-line emissions and oxygen precipitates.

Later on, the D-lines have been studied also in mc-Si, and their origin, relative spatial distributions and temperature dependence has been extensively discussed. In 2000, Ostapenko et al.[38] investigated the defect emissions in mc Si samples, and found that a defect emission with its maximum at about 0.8eV was present at room temperature. Further, they studied the four D-lines at 4.2 K and found that among them, only D1 could be tracked at increasing temperatures and thus being linked to the 0.8 emission band found at room temperatures. The shift in energy between the maximum intensity of the D1 emission (0.812eV), found at 4.2K and that of the defect band (0.8K) found at room temperature was attributed to the dependence of the band gap energy E_G on temperature. The fact that only D1 could be detected at room temperatures is in agreement with findings done by Kveder et al.[44] already in 1995. This fact has proven to have practical applications in material characterisation. Intensity mappings of the BB-band and the defect band are done, for example in [45], [6] and [2], and an inverse correlation between the two bands have been observed.

In their paper from 2000 Ostapenko et al. suggest that the 0.8 eV defects emission band is due to electrically active dislocation networks, interacting with point defects, impurities or their clusters. They state that oxygen precipitates and transition metals (Fe, Ni and Cr) are most likely to be involved, whereas other contamination species, such as carbon, may also contribute. This has been acknowledged among others by Arguirov [37] and by Tajima et al.[2], both stating that the D1/D2 pair is caused by secondary defects such as oxygen or metallic precipitates decorating dislocation clusters.

Tajima et al. further find evidence for the generally accepted idea [2, 33, 39], that the D3/D4 pair mainly is caused by the intrinsic nature of silicon (e.g. from crystal defects that are present independently of contamination). This hypothesis is partially based upon measurements suggesting that the intensities of the D3/D4 emissions are higher directly on sub-grain boundaries, whereas those of the D1/D2 emissions are higher around sub-grain boundaries [2, 46]. Along sub-grain boundaries with high concentrations of oxygen precipitates, a reduced intensity of D3 and D4 is observed. The authors thus speculate that the emissions caused by intrinsic properties of the dislocations are partly supressed due to such precipitates [2].

It has been suggested [33] that D3 is the phonon replica of D4. The term *phonon replica* denotes a transition for which a part of the energy released from the recombining electron-hole-pair is transmitted to a phonon instead of being emitted as a photon. The phonons can attain only certain energy values, defined by the vibration moduli of the crystal atoms. Therefore, as long as the temperature is kept constant, the energy shift between a transition line and its phonon replica must be constant, regardless of location [7]. In a recent publication [3], however, Nguyen et al. found that the intensities as well as the peak energies of D3 and D4 varies differently as a function of the distance from grain boundaries. Moreover, Burud et al. [47] found large differences in the intensity ratio between the D3 and D4 emissions spatially on the wafer. These observations both contradict the idea that D3 is the phonon replica of D4.

2.6.3 Dependendence of DRL on metallic impurties

In 2009 Gundel et al. [4] conducted room temperature measurements of the DRL of n-type float-zone wafers, deliberately contaminated with iron and copper. They found that the relative intensity of the defect PL increased in iron contaminated samples compared to uncontaminated samples. At points with iron precipitates, however, the peak of the broad defect PL band shifted to longer wavelengths. In samples contaminated with copper, the defect PL signal was supressed. They also report a shift of the maximum defect PL to higher energies for samples containing iron precipitates.

Four years later, Flø et al. [35] spatially compared the D line emission, recorded by multispectral imaging, with iron content, found by FeB pair splitting (see section 3.4.1). They report correlation between the distribution of D4 emission line and iron contamination.

In two recent publications [3, 7] Nguyen et al. studied p-type boron-doped mc-Si samples in their as-cut state, after phosphorus gettering and after implantation of iron with subsequent annealing. The gettering step is supposed to reduce the mean iron concentration. However, because some internal gettering might take place in parallel to the external getting process, the iron concentration close to the grain boundaries might be higher in the gettered than in the as-cut sample. The deliberately iron contaminated sample, however, should have even higher iron concentrations. The authors report that D1 was significantly higher in the gettered sample than in the as-cut sample, but did not change from the as-cut to the iron contaminated state. They therefore conclude that the D1 intensity increases when the metal impurities are removed from the sub-grain boundaries. The authors further report that no changes in the intensities of D3 and D4 between the samples were observed.

2.6.4 Changes of DRL during solar cell processing

In a recent publication, Johnston et al. [6] investigated the DRL of wafers at different processing steps, from as-cut wafer to finished solar cell. They found that during the process of antireflection coating (ARC), the sub-bandgap PL (subPL) signal (including energies in the interval 0.73-0.92eV) changes dramatically, whereas the band-to-band PL stays more or less unchanged. For the subPL signal, they found that some regions change from being bright at the diffused and etched wafer to being dim after the subsequent ARC procedure. Other regions, however, give only a dim subPL signal at the diffused an etched step, but a bright signal after the ARC step. During the other processing steps, the relative intensities of the subPL signal stay more constant, see Figure 2.6.

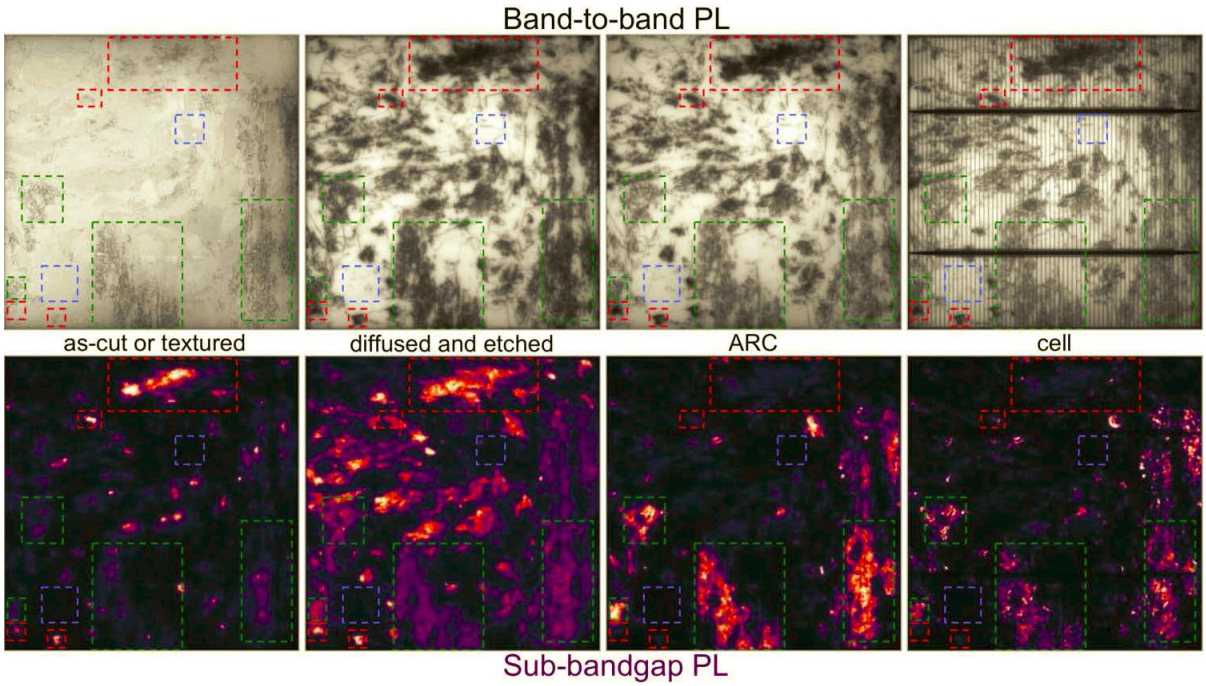


Figure 2.6*: Change of subPL intensity as found by Johnston et al. From the diffused and etched step to the antireflection coated step, some (marked with red) turn from bright to dim, while regions other regions (marked with green) change from dim to bright. Defect-free regions are marked with blue. Reprinted from [6].*

The ARC process step used for the investigation consists of depositing silicon nitride (SiN_x) for 30 minutes at 450˚C. Johnston et al. assume that the substantial change in subPL intensity is induced by the temperature related to the ARC processing step, and could thus be caused also by equivalent heating [6]. In another study [41], it is shown that the inversion of the subPL may be reverted by temperatures above 500˚C.

Arguirov [37] also investigated the effect of PDG (phosphorus diffusion gettering, P-gettering) on DRL, by comparing luminescence spectra of as-grown samples and samples subjected to P-gettering at various temperatures from 80 to 300K. Both samples were of n-type EFG Si (Edge-defined Film-fed Growth silicon). In the as-cut samples, he found a distinct peak at about 0.83 eV (probably being a mix of D1 and D2) that was visible also at higher temperatures. D3 and D4 had relatively low intensities in these samples. In the P-gettered sample, however, he found much higher intensities of D3 and D4 whereas the D1/D2 peak was totally absent. Since P-gettering

probably reduces the concentrations of metal and oxygen contaminations, this result supports the theory that D1/D2 are due to such contaminations interacting with dislocations, whereas D3/D4 rather are caused by the dislocations themselves.

3 EXPERIMENTAL

This chapter starts with a description of the analysed samples (section 3.1), followed by a presentation of the utilized measurement techniques (section 3.2 - 3.5). Besides describing the measurement setups and procedures, the three latter sections also include explanations of the theoretical background for the measurement methods as well as measurement procedure optimizations done as a part of the work.

3.1 SAMPLES AND THEIR PROCESSING

The motivation for this work was to study differences of DRL between n- and p-type wafers, the change of DRL along the height of an ingot, the impact of iron and other metal impurities on DRL, and the change of DRL due to solar cell processing. For this purpose, multicrystalline silicon samples, solidified and processed at Fraunhofer ISE, are chosen. Table 3.1 gives an overview of all investigated samples, including their doping concentrations, their ingot positions, and the processing steps they have been through. The processes will be more precisely described in the following.

As shown in the table, the samples are taken from three different ingots. Two of them are one p-type and one n-type ingot solidified in identical crucibles and, as far as possible, with the same process parameters, in order to allow a comparison of p- and n-type characteristics. These ingots are in the following denoted the *twin* ingots. The third ingot, also made with the comparison of p- and n-type as motivation, switches from n-type to p-type and then back to n-type along its direction of solidification, and is accordingly denoted *npn ingot*. The details on silidification and wafer processing of the three ingots given in this section are taken from [48] and [49] . These are the works for which the p- and n-type ingots and the npn-ingot, respectively, originally were prepared.

3.1.1 Wafers from the twin ingots

The twin ingots were solidified in a G2 crucible, meaning that 2x2 bricks with an area of 156 mm x 156 mm can be cut from the ingot. In this case, however, the ingots were sawn to three 156 mm x 156 mm bricks. The bricks from which the wafers characterised in this work originate were placed centrally along one of the ingots edges, 2 cm away from the rims (see Figure 3.1). Later, the wafers sawn from the block were cut to dimensions of 125 mm x 125 mm. The position of these wafers on the brick is highlighted in the figure.

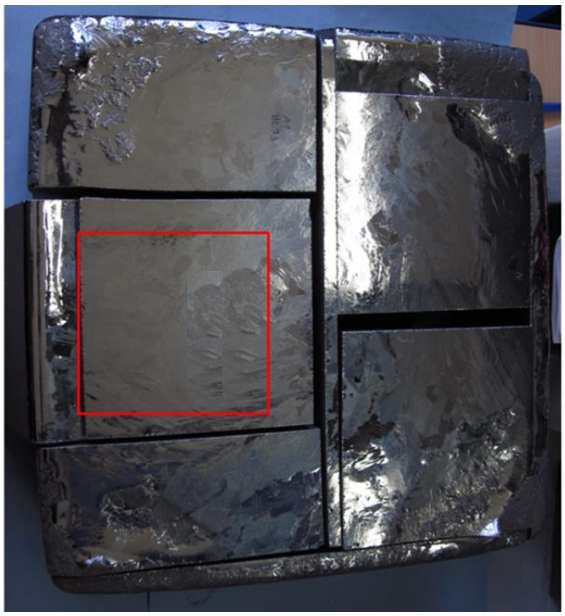


Figure 3.1: *Photo showing how the cast ingots were sawn to three bricks. The position of the 125 mm x125 mm wafers investigated in this work is highlighted. Reprint from [48].*

During the solidification process, in-diffusion of iron and other impurities from the crucible to the edge regions of the ingot takes place. As seen from the figure, the part of the ingot that was closest to the crucible during the casting process is not included in the brick. However, the region of the ingot affected by in-diffusion is supposed to be significantly wider than the removed rim. The samples thus allows for an investigation of the effect of in-diffused iron and other metal contaminations[2].

In order to allow for an investigation of the effect of solar cell processing on DRL, wafers from the n-ingot as well as the p-ingot, have been through various processing steps. The process through which the various wafers have been, are shortly described in the following paragraph. In order to allow a comparison of the effect of the different processing steps on n- type and p-type wafers, samples from both ingots have been exposed to exactly the same processing conditions.

The wafers are here given names that indicate from which ingot and which ingot height the wafer originates, as well as the processing step(s) it has been through (see Table 3.1).

I) *As-cut* wafers have not been subjected to any subsequent treatment after being sawn from the brick. These wafers are given the name "as-cut" in addition to the name indicating their ingot position (see Table 3.1).

II) *Passivated* wafers are after sawing treated with a chemical polishing (CP) etchant in order to smoothen the wafer surface and remove saw damages. In this etching process, approximately 15 µm of the silicon material on each side of the wafer are removed.

[2] The n-type wafers are oriented as shown in Figure 3.1 (with the in-diffusion region at their left edge). The p-type samples, however, where rotated 180º before the labelling, meaning that they have the in-diffusion region on their right edge.

Subsequently, they are passivated (see the following paragraph). Since these wafers have not been through any other high-temperature process except from the passivation, they are in this thesis sometimes referred to as "*initial*" rather than "*passivated*". Since this wafers are in their initial or "neutral" state, they are in Table 3.1 given no name extension

III) Wafers being subjected to *phosphorous diffusion (P-diffusion)* were after the initial CP treatment placed in an oven at 800ºC for one hour. In the oven, $POCl_3$ gas serves as phosphorous source for the diffusion. Subsequently, the wafers were left in a hydrofluoric acid (HF) etchant for five minutes in order to remove the phosphorus silicate glass (PSG). As explained in section 2.5.1, the phosphorus diffusion leads to the removal of metal contaminations from the wafer bulk (external gettering). These wafers are in Table 3.1 given the name extension "P".

IV) Following the phosphorous diffusion, *H-passivation* (see section 2.5.2) was achieved through deposition of a 70 nm thick layer of SiN_x on each side of the wafer and a subsequent heating to 815ºC for approximately 15 s. The high-temperature step causes hydrogen from the SiN_x layer to diffuse into the sample. Finally, the SiN_x layer was removed by a HF etchant. These wafers are in Table 3.1 denoted "P-H".

Following these processing steps, all samples (except the as-cut wafers) were surface passivated. Due to the different charge types of the minority charge carriers, the processes to which n- and p-type samples were subjected at this point differ. The n-type wafers were at both sides coated with a 60 nm SiN_x layer, whereas the p-type samples were coated with a 10 nm Al_2O_3 layer. Because this aluminium oxide layer is thin and damageable, it is additionally covered with a 70 nm nitride layer, denoted *capping*. For the activation of the Al_2O_3 passivation layer on the p-type samples, a temperature step of 400ºC is necessary. In order to increase the comparability of the n- and p-type samples, also the n-type wafers where exposed to the same temperature step.

3.1.2 Wafers from the npn-ingot

The npn-ingot was solidified in a G1 crucible, meaning that one brick of dimensions 156 mm x 156 mm can be cut from the ingot, with phosphorus as its basic doping material. During its solidification, extra doping material (boron and phosphorus) was added to the fluid silicon in order to achieve the change between the doping types. In a solidifying crystal, both crystal grains and dislocation clusters mostly grow in the solidification direction of the ingot. The concentration and distribution of contamination materials is approximately the same for close ingot positions. Thus, having both n-type and p-type wafers from neighbouring positions of the same ingot allows for extraordinary possibilities for comparison of n- and p- type material. One should, however, note that since both acceptor- and donor atoms are present in some parts of the ingot (overcompensation, see section 2.1.2), some characteristics (e.g. mobility) for this material may differ from that of uncompensated n- and p- material.

From the npn-brick, 125 mm x 75 mm wafers were investigated. In order to enable the characterisation of in-diffusion influences, also these wafers were cut from a region close to the ingot edge. Like the wafers from the n- and p-type ingots, the wafers from the npn- ingot were subjected to a CP treatment in order to remove saw damage, before their surfaces were

passivated. The passivation processes were, both for n- and p-type wafers, equal to those described in the previous section. However, in this case the n-type samples were *not* subjected to the temperature step needed to activate the Al_2O_3 passivation of the p-type samples. Neither P-diffusion nor H-passivation were done.

Table 3.1: *Overview of all samples investigated in this work, their ingot positions, the processing steps they have been through and their net doping concentration. p-type samples are marked with blue colour, and n-type with pink.*

Sample name used in this thesis	Ingot	Block position			Processing steps				$N_{A/D}$ [cm^{-3}]
		Bottom	Central	Top	Initial HF-treatment	P-diffusion	H-passivation	Surface passivation	
p-twin-bottom	p-twin	■			■			■	1.4×10^{16}
p-twin-bottom-P		■			■	■		■	1.4×10^{16}
p-twin-bottom-P-H		■			■	■	■	■	1.4×10^{16}
p-twin-central			■		■			■	1.4×10^{16}
p-twin-central-P			■		■	■		■	1.4×10^{16}
p-twin-central-P-H			■		■	■	■	■	1.4×10^{16}
p-twin-central-as-cut			■						1.4×10^{16}
p-twin-bottom				■	■			■	1.6×10^{16}
p-twin-bottom-P				■	■	■		■	1.6×10^{16}
p-twin-bottom-P-H				■	■	■	■	■	1.6×10^{16}
n-twin-central	n-twin		■		■			■	2.3×10^{15}
n-twin-central-P			■		■	■		■	2.3×10^{15}
n-twin-central-as-cut			■						2.3×10^{15}
n-npn	npn		■		■			■	9.3×10^{15}
p-npn			■		■			■	1.24×10^{16}

3.2 CHARACTERISATION METHODS – AN OVERVIEW

Different measurement techniques, all based on photoluminescence, are utilized in this work. All the applied techniques therefore have in common that they require an *excitation source* providing photons that may be absorbed by the sample under investigation, and some kind of *light sensor* (either in the form of a camera or a photodiode) capable of detecting the photons being emitted from the sample. The measurement techniques differ, however, regarding the time dependency and the spatial distribution of the laser, and regarding the energy interval being detected by the light sensor. Thus, different physical characteristics of the investigated samples can be measured and compared.

In more detail, photoluminescence measurements done with a hyperspectral camera giving spectrally and spatially resolved images of the luminescence signal of wafers, were compared to lifetime maps, maps of charge carrier density, and maps of the interstitial iron concentration of the same wafers, all made by PL-imaging calibrated by a harmonically modulated PL technique [9][3]. Figure 3.2 gives a simplified overview of the performed measurements and analyses.

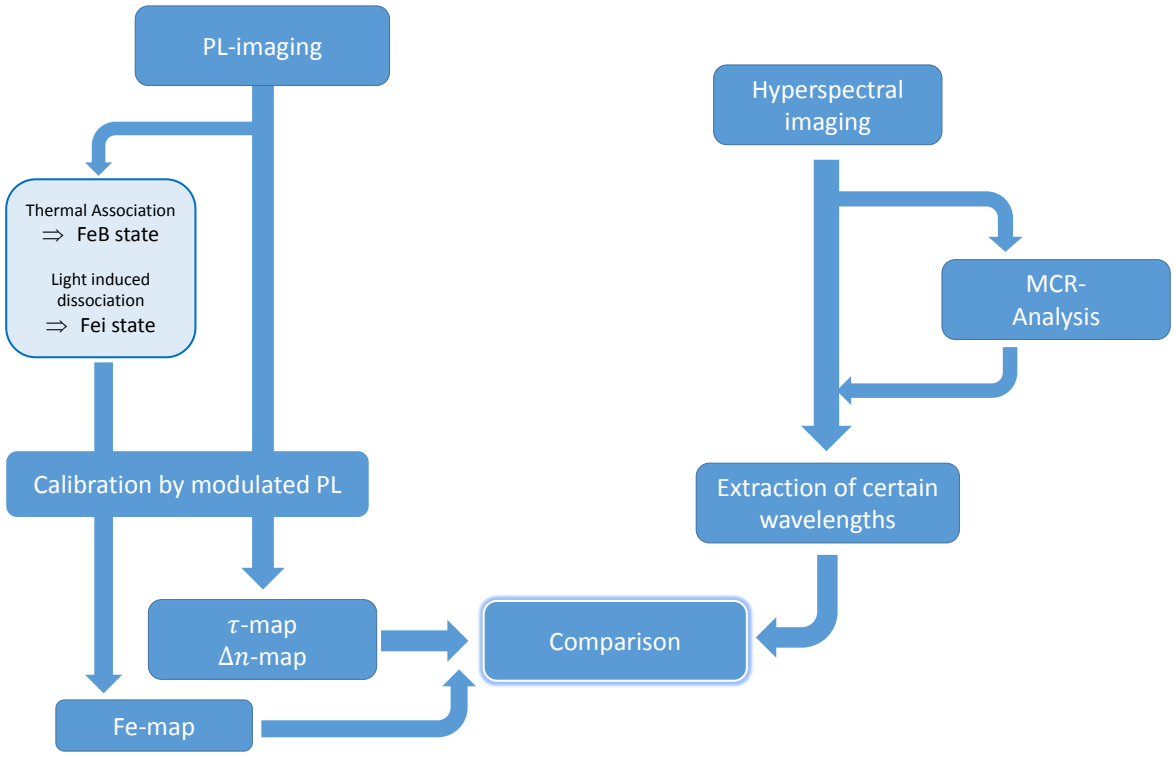


Figure 3.2: *Overview of the measurements done in this work. All samples were subjected to PL-imaging calibrated by modulated PL. The p-type samples were additionally subjected to Fe-Imaging (see section 3.4). Hyperspectral images of all samples were made, and certain wavelength intervals, extracted by MCR and other methods were, compared to the results of the PL measurements.*

[3] This calibration technique was previously called quasi-steady-state-photoluminescence (QSSPL).

3.3 PHOTOLUMINESCENCE IMAGING CALIBRATED BY MODULATED PL

3.3.1 PL imaging

As excitation source for Photoluminescence imaging, done at Fraunhofer ISE, a 790 nm laser from LIMO was used. The laser illuminates the sample homogeneously, and its intensity can be adjusted to values ranging from ~1/1000 sun to approximately 1.6 suns. PL-images for this work were mainly taken at constant laser illuminations of 1 sun, corresponding to a photon flux of 2.55 x 10^{17} $cm^{-2}s^{-1}$.

In order to record the images, a Si CCD Camera named Pixis 1024 BR from Princeton Instruments was employed. In front of the camera, a 950 nm long-pass filter and a 978 nm long-pass Razor Edge filter were placed, both in order to hinder the entrance of reflections of the laser light into the camera. Considering the measured Quantum Efficiency of the camera [50], and the characteristics of the applied filters, the detection interval of the camera can be estimated to 978-1078 nm, corresponding to 1.15 -1.27 eV.

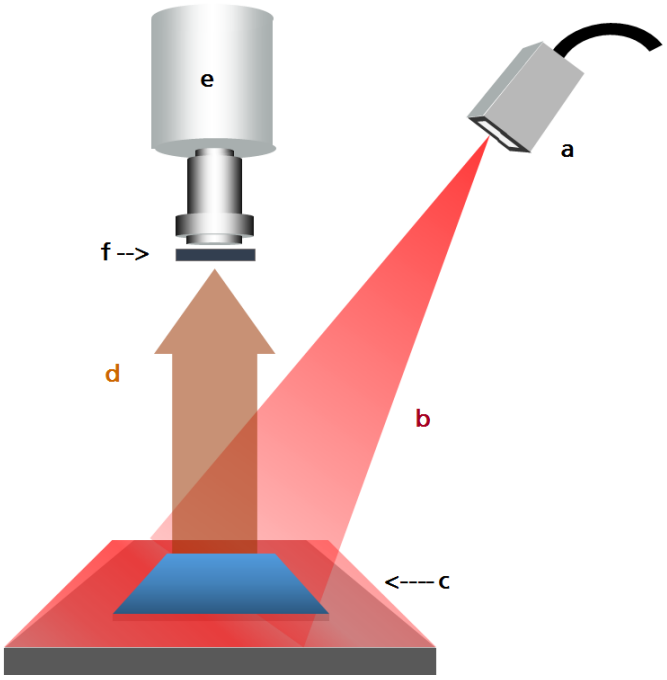


Figure 3.3: *PL imaging setup . The excitation laser (a) and its optics provide a homogeneous illumination profile (b) on the wafer sample (c). The photoluminescence signal (d) from the wafer is detected by a Si CCD camera (e). Long-pass filters (f) prevent the penetration of reflections of the laser light into the camera. After [50].*

3.3.2 Calibration by harmonically modulated PL

The PL images were calibrated using a harmonically modulated photoluminescence technique developed at Fraunhofer ISE [9]. The setup for this technique, schematically shown in Figure 3.4, includes a 790 nm time modulated laser that homogenously illuminates the wafer being

investigated, and of two photodiodes. The laser modulation frequency can be adjusted to values between 4 Hz and 160 Hz. Of the two photodiodes, one measures the time dependent irradiation intensity of the laser while the other measures the time dependent luminescence signal from the active area of the wafer, caused by the laser illumination. The latter of the two is shielded by a long pass filter, ensuring that no light from the laser is detected. A metal mask with a circular opening (Ø 45mm) is placed over the wafer, defining its active area. The currents from the two photodiodes are amplified, and read into a computer.

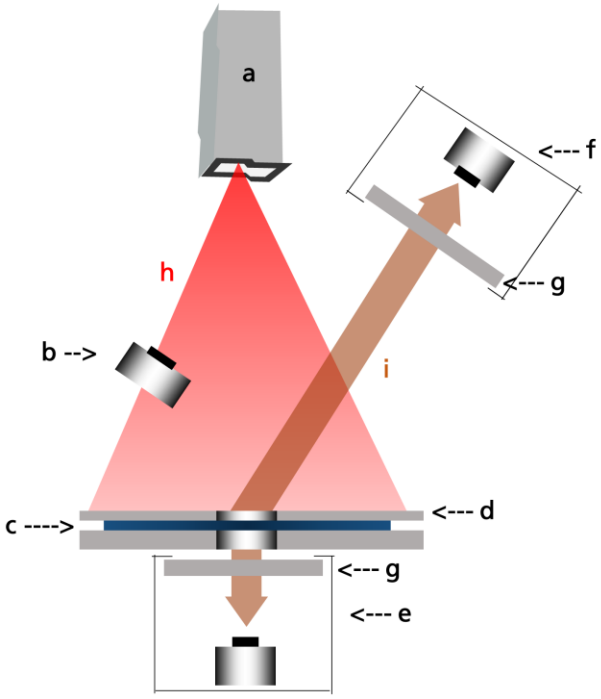


Figure 3.4: *Setup for the modulated PL calibration. The time modulated laser (a) illuminates the sample homogeneously. A photodiode (b) detects the time modulation of the laser light. The sample (c) is covered by a mask (d), allowing the laser light to hit only a circular area (Ø 45mm) centrally placed on the wafer. The active area defined by the mask emits a photoluminescence signal that is detected by a second photodiode placed either below the sample (transmission mode)(e) or above the sample(reflection mode)(f). The second photodiode is sheltered by a long-pass filter (g) preventing transmitted/reflected laser light to be detected by the photodiode. In this work, transmission geometry was used. The laser excitation light (h) and the photoluminescence signal (i) are also indicated. After [51].*

The following brief explanation of how the lifetime is obtained from the two amplified signals is mainly based on [9].

In accordance with equation (2.15), the recorded photoluminescence signal $\Phi(t)$ at a given time *t* can be expressed as

$$\Phi(t) = A\Delta n(t)(\Delta n(t) + N) \qquad \textbf{(3.1)}$$

The factor A then includes the coefficient of radiative recombination β as well as optical properties related to the sample and the measurement setup. N is the net dopant concentration of the sample. Using equation (2.21), this can be rewritten as

$$\Phi(t) = A\big[\mathrm{G}(t) \cdot \tau\big(\Delta n(t)\big)\big]\big[\mathrm{G}(t) \cdot \tau\big(\Delta n(t)\big) + N\big] \qquad \textbf{(3.2)}$$

Thus, for each time *t*, if both A and N were known, a lifetime τ could be calculated as the positive solution of equation (3.2). However, since the factor A is unknown, a calibration procedure is needed.

Equation (2.12), describing the density Δn of excess charge carriers under steady state illumination, might be generalized to the case of a time dependent generation rate G(t). It then reads

$$\frac{\delta \Delta n(t)}{\delta t} = G(t) - U(t) \qquad \textbf{(3.3)}$$

Inserting $\Delta n(t)/\tau(\Delta n(t))$ for the net recombination rate *U*(*t*) (cf. equations (2.18) and (2.21)) and solving for the injection dependent lifetime yields

$$\tau\big(\Delta n(t)\big) = \frac{\Delta n(t)}{G(t) - \delta \Delta n(t)/\delta t} \qquad \textbf{(3.4)}$$

Finding the correct injection dependent lifetime of the active area of the wafer, based on a mean of a large amount of modulation cycles, may now be explained as follows: With the sinusoidally modulated generation rate G(t) in combination with the measured time dependent PL signal $\Phi(t)$, and assuming a certain factor A, expressions for $\Delta n(t)$ and τ from equations (3.1) and (3.2) can be inserted in equation (3.4). For each laser modulation cycle, Δn undergoes an increase followed by a decrease, and thus attain any value in the interval $< \Delta n_{min}, \Delta n_{max} >$ twice. The factor A from equations (3.1) and (3.2) can therefore be found through an iterative algorithm which eliminates the hysteresis of the trajectory of equation (3.4) for an entire cycle, meaning basically that the value for $\tau(\Delta n)$ for an increasing value of Δn must be the same as $\tau(\Delta n)$for a decreasing value of Δn. Thus, this self-consistent solution of equation (3.5) converts the time-dependent PL signal $\Phi(t)$ into an injection-dependent lifetime $\tau(\Delta n)$.

One may think that knowledge of the net dopant concentration N is necessary when searching the injection dependent lifetime with this method. However, under low-injection-conditions ($\Delta n \ll N$), equation (3.1) can be simplified to

$$\Phi(t) = A' \Delta n(t) \qquad \textbf{(3.5)}$$

where $A' = AN$. This constant is thus the only unknown of the system. Simplified, calibration by modulated PL means finding the constant A' through the iterative process explained in the previous paragraph. Thereby, for any pixel i in a photoluminescence image with a recorded PL

pixel value ϕ_i, and a known generation rate G, the injection level Δn_i and the lifetime τ_i can be calculated.[4]

3.4 IMAGING OF INTERSTITIAL IRON CONCENTRATION

3.4.1 Measurement principle

When a metastable defect (see section 2.4.1) changes from one state to another, the lifetime related to the defect will also change. Therefore a spatially resolved image of the concentration of the defect can be found through a comparison of lifetime measurements done before and after a transition between two states. Zoth and Bergholz [52], in 1990, were the first to utilize this principle for measuring iron concentration in silicon wafers. Almost two decades later Macdonald et al. [53] where the first to apply PL for iron imaging. The expressions explained in the following and utilized in this work, are based on their publication.

For the case of iron, one lifetime measurement of a wafer before dissociation and one measurement after dissociation of FeB pairs are considered. For every pixel of the image, two effective lifetimes τ_0 (before dissociation) and τ_1 (after dissociation) are assigned. The two measurements are done at the same generation rate G. Due to the difference in lifetime, the injections $\Delta n_0 = G\tau_0$ and $\Delta n_1 = G\tau_1$, cf. Equation (2.21), are therefore different. As explained in Equation (2.41), the two lifetimes can be expressed as function of the lifetimes related to different recombination mechanisms:

$$\frac{1}{\tau_{eff\,0}} = \frac{1}{\tau_{FeB}} + \frac{1}{\tau_{Auger\,0}} + \frac{1}{\tau_{OTHER\,0}} \qquad \textbf{(3.6)}$$

$$\frac{1}{\tau_{eff\,1}} = \frac{1}{\tau_{Fei}} + \frac{1}{\tau_{Auger\,1}} + \frac{1}{\tau_{OTHER\,1}} \qquad \textbf{(3.7)}$$

$\tau_{Auger\,0}$ and $\tau_{Auger\,1}$ denote here lifetimes related to Auger recombination at the two different injection levels, whereas $\tau_{OTHER\,0}$ and $\tau_{OTHER\,1}$ are the lifetimes related to all other recombination mechanisms, also including radiative recombination, surface recombination and SRH recombination caused by all other defects than iron. By assuming that the lifetimes related to other mechanisms are equal before and after the dissociation, the last term cancels out if we subtract equation (3.7) from equation (3.6). Further, by using the Shokley-Read-Hall expressions (equation *(2.37)*) for the lifetimes related to recombination due to Fe_i and FeB, accompanied by

[4] In multicrystalline wafers, spatial variations of lifetime and injection level are large. Thus, to find the calibration factor A', it must be considered that areas with different lifetime contribute differently to the overall PL signal recorded in the HML measurement. A profound explanation of the applied averaging procedure used for that case can be found in [9].

some simplifications and rearrangements [53], we find the following expression for the interstitial iron concentration $[Fe_i]$:

$$[Fe_i] = C^{Fe}(\Delta n)\left(\frac{1}{\tau_1} - \frac{1}{\tau_0} - \frac{1}{\tau_{Auger\,1}} + \frac{1}{\tau_{Auger\,0}}\right) \quad \textbf{(3.8)}$$

The factor $C^{Fe}(\Delta n)$ is denoted C-factor and is given by

$$C^{Fe}(\Delta n) = \left(\frac{1}{\chi^{Fei} - \chi^{FeB}}\right) \quad \textbf{(3.9)}$$

Where the coefficients χ^{Fei} and χ^{FeB} are

$$\chi^{Fei} = \frac{v_{th}(N_A + \Delta n_1)}{\frac{1}{\sigma_n^{Fei}}\left(N_A + p_1^{Fei} + \Delta n_1\right) + \frac{\Delta n_1}{\sigma_p^{Fei}}} \quad \textbf{(3.10)}$$

and

$$\chi^{FeB} = \frac{v_{th}(N_A + \Delta n_0)}{\frac{1}{\sigma_n^{FeB}}(N_A + \Delta n_0) + \frac{1}{\sigma_p^{FeB}}(n_1^{FeB} + \Delta n_0)} \quad \textbf{(3.11)}$$

v_{th} is here the thermal velocity of electrons, N_A is the total acceptor concentration and $\Delta n_{0/1}$ is the excess carrier concentration before and after dissociation, respectively. Further, $\sigma_{n/p}^{FeB/Fei}$ is the capture cross section for electrons/holes in the FeB/Fe_i state. p_1^{Fei} and n_1^{FeB} are the SRH densities of holes for Fe_i and of electrons for FeB, as defined by equation *(2.35)*.

3.4.2 Incomplete association/dissociation

When regarding the electrical state of an iron contaminated silicon sample for which the iron concentration should be measured, its *associated* and *dissociated* states denote the state of the entire sample after being subjected either to thermal association or to thermal dissociation by illumination. These terms are, in other words, used to describe the state of a sample on a macroscopic level. The terms *FeB state* and *Fe_i state*, however, describe the state of one specific iron atom on a microscopic level.

The expression given in equation (3.8) is correct only if absolutely all resolved iron atoms in the sample are paired with boron (i.e. are in the FeB state) when the sample is in its associated state. Likewise, when the sample is in the dissociated state, all the iron atoms must be present in the Fe_i state [16]. If this is not the case, the FeB pairs present in the dissociated state and the Fe_i atoms present in the associated state will work as extra recombination paths in addition to those included in equation (3.8). In the associated state, for example, many charge carriers will

recombine due to the many FeB-pairs present in the sample. Some will, however, also recombine via the few present Fe_i atoms.

This case, and the similar but opposite mechanism in the dissociated state was considered by Schubert et al. [10] in 2011. The following explanation of it is mainly based on [16]. Schubert et al. introduced the factors f_{Ass} giving the fraction of the iron atoms that are in the FeB state when the sample is in its associated state, and f_{Dis} giving the fraction of the iron atoms that are in Fe_i state when the sample is in its dissociated state. Accordingly, the fraction of the atoms being in the opposite state will be $(1 - f_{Ass})$ and $(1 - f_{Dis})$ for the associated and the dissociated state, respectively. The SRH-lifetimes for the two states can thus be expressed as

$$\frac{1}{\tau_{SRH}^{Ass}} = f_{Ass}\left(\frac{1}{\tau_{SRH}^{FeB}}\right) + (1 - f_{Ass})\left(\frac{1}{\tau_{SRH}^{Fei}}\right) \qquad \textbf{(3.12)}$$

$$\frac{1}{\tau_{SRH}^{Dis}} = f_{Dis}\left(\frac{1}{\tau_{SRH}^{Fei}}\right) + (1 - f_{Dis})\left(\frac{1}{\tau_{SRH}^{FeB}}\right) \qquad \textbf{(3.13)}$$

As a consequence, the C-factor needs to be modified, and is then given by

$$C^{Fe}(\Delta n) = \frac{1}{f_{Ass} \cdot \chi_{Ass}^{FeB} + (1 - f_{Ass}) \cdot \chi_{Ass}^{Fei} - f_{Dis} \cdot \chi_{Dis}^{Fei} - (1 - f_{Dis}) \cdot \chi_{Dis}^{FeB}} \qquad \textbf{(3.14)}$$

In other words the C-factor, and thus the calculated iron concentration, depends on the degree to which the resolved iron atoms are associated or dissociated to/from present boron atoms. Since the coefficients χ depend, among others, on the injection of minority charge carriers (cf. equations (3.10) and (3.11)), they are different for the associated and the dissociated state. This is underlined by the subscripts in equation (3.14).

For the iron concentration measurements, incomplete association/dissociation of iron is considered, by using the expression for the C-factor given in equation (3.14).

3.4.3 Injection dependence of the lifetimes in Fe_i- and FeB state

Since the defect parameters (capture cross sections $\sigma_{n/p}$ and trap level E_t) of Fe_i-ions and FeB-pairs are different, the dependence of the SRH-lifetimes of the two states on charge carrier injection will differ. For a specific minority carrier density, is the lifetime of the Fe_i state equal to the lifetime of the FeB state. This point is called the *cross-over point*, and the according minority carrier density is denoted $\Delta n_{crossover}$ [15]. An expression for $\Delta n_{crossover}$ can be found analytically by equating the expressions for the SRH-lifetimes of the Fe_i state and the FeB state (equation (2.37)), and solving for Δn [15]. As result, one obtains an expression showing that $\Delta n_{crossover}$ is independent of defect concentration. It depends, however, strongly on the defect parameters of the two states, and to some extent on the doping concentration. According to

Macdonald et al.[54], the cross-over point for 1Ωcm B-doped material occurs at minority carrier density of approximately $1 \cdot 10^{14}\ cm^{-3}$. Figure 3.5 shows a simulation of the injection dependency of the Fe_i and FeB lifetimes. As can be seen, the lifetime before dissociation (i.e. in the FeB state) is higher than after dissociation (i.e. in the Fe_i state), for excess carrier densities *below* the cross-over point. At an excess carrier density approximately one order of magnitude below the cross-over point, the lifetime of the dissociated state increases, finally ending up at a higher value than that of the associated state.

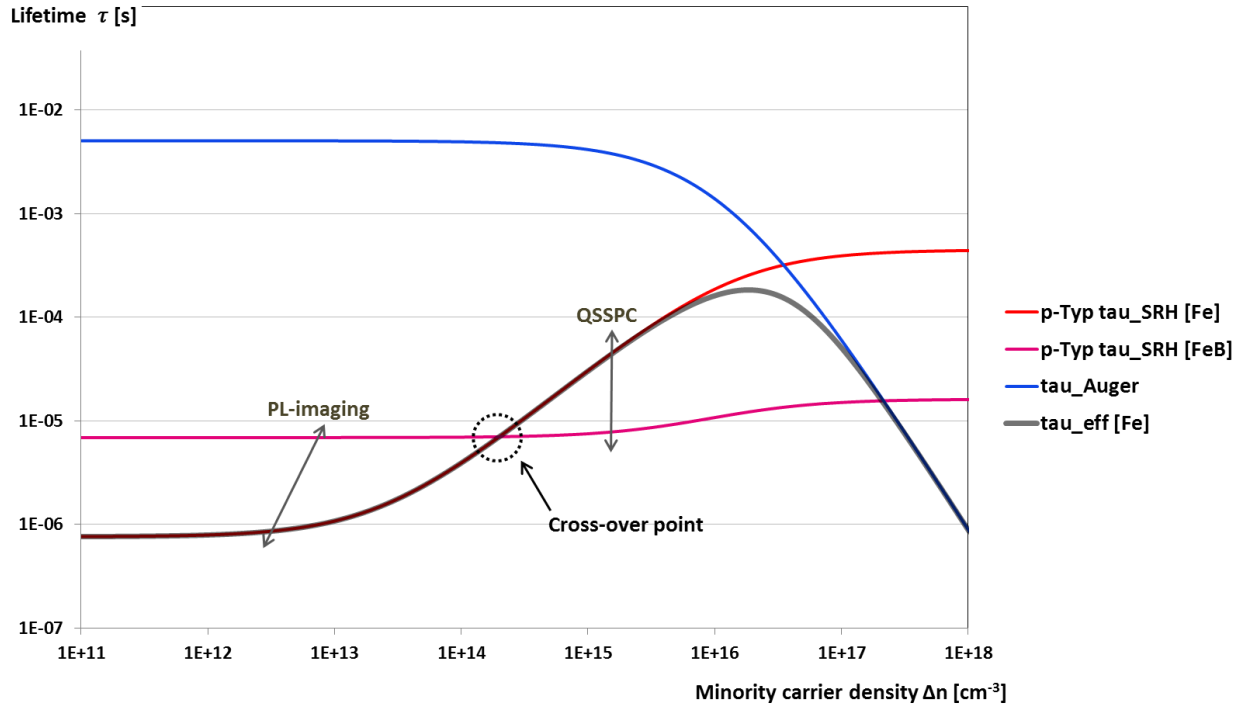


Figure 3.5: *Simulation of the dependence of SRH-lifetimes on excess carrier density. The simulation is done with a pre-programmed Excel-simulation sheet made by D. Macdonald (ANU), K. MacIntosh (ANU), K. Bothe (ISFH). The following simulation parameters for p-type Si are used:* $\boldsymbol{N_A = 1,4 \cdot 10^{16}}$, $\boldsymbol{N_t = 3 \cdot 10^{12}}$. *The cross-over point, meaning the injection level for which* $\boldsymbol{\tau_{Fei} = \tau_{FeB}}$, *is highlighted. The arrows indicate the injection levels at which the* Fe_i *and FeB measurements for different measurement methods typically are done. After [53].*

Since the calculation of the iron concentration is based on the difference of the lifetime before and after dissociation, it is of great importance for the accuracy of the result that the measurements are not done close to the crossover point. Close to this point, the change in lifetime upon dissociation is very small, and the the measurement's sensitivity for iron is accordingly weak. In order to avoid excess carrier densities close to the crossover point, different generation rates must be chosen for samples with different lifetimes [53]. The arrows in Figure 3.5 illustrate typical injection levels for PL-imaging and QSSPC (quasi steady state photoconductance) measurements. Note that for PL-imaging, when the images before and after dissociation are done at the same generation rate, the excess carrier density of the two images will differ, due to the difference in lifetime between the two states.

3.4.4 Measurement procedure

In order to accomplish a spatially resolved iron measurement, PL images of the sample in its FeB and Fe_i state must be made. Figure 3.6 schematically shows the measurement procedure, and how the FeB and Fe_i states are prepared. For the optimisation of the measurements, several factors must carefully be regarded. The most important of them are listed below.

Firstly, *thermal dissociation* of FeB pairs during the preparation of the associated state must be avoided. In order to achieve a short association time for FeB pairs, the mobility of interstitial iron ions, and thus the temperature of the sample must be increased. However, if an FeB pair is supplied with an amount of thermal energy higher than the bonding energy between the two atoms, they will break apart. Thermal dissociation will in other words take place [16]. Since this process happens in parallel with the association of FeB pairs, the percentage of iron in the sample being present as FeB will reach an equilibrium state of a level that depends upon temperature and doping concentration [16]. Thus in order to bring the sample to its associated state, the right temperature must carefully be chosen. As shown in Figure 3.6, for the iron measurements done in this work the sample was heated to 80°C for 10 minutes, in order to prepare the associated state.

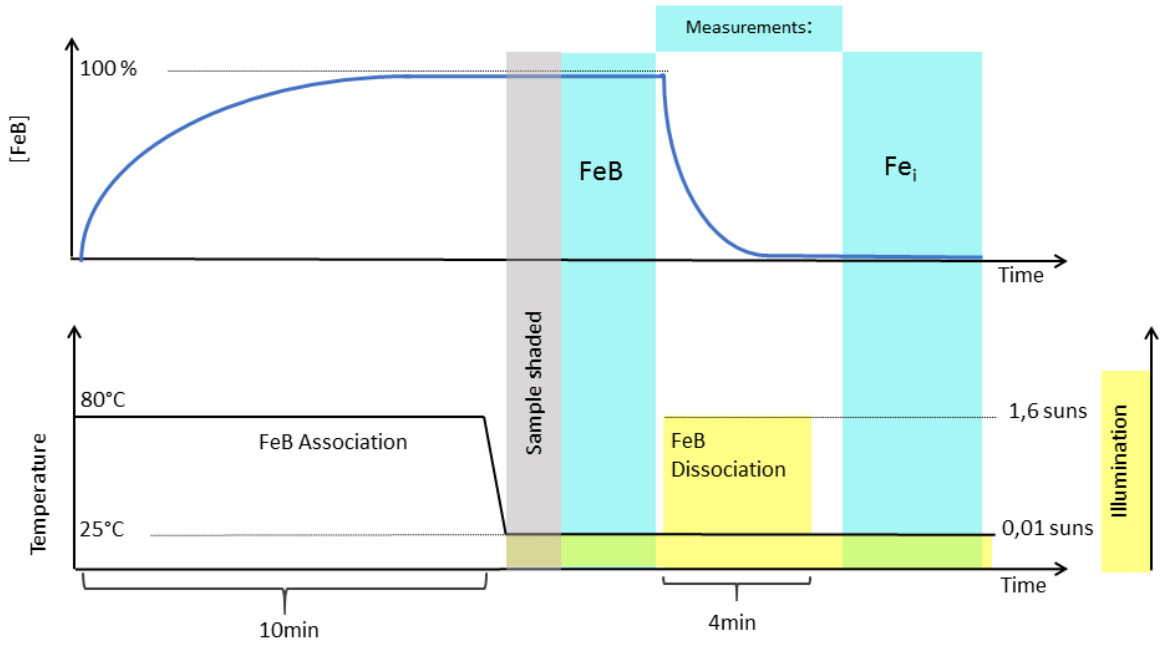


Figure 3.6: *Measurement procedure for iron imaging. The sample is brought to its associated state by thermal association of FeB pairs at 80˚ for 10 minutes. After the FeB measurement, the sample is illuminated in order to prepare the dissociated state. The upper graph shows the percentage of iron atoms in the sample being present in the FeB state, whereas the lower graph shows the temperature and illumination conditions to which the sample is subjected.*

Secondly, *dissociation due to illumination* before and during the FeB measurement must be avoided. After preparation of the associated state, the wafer is therefore handled in darkness until the FeB measurement is started. Moreover, the laser illumination intensity during the FeB measurement should be kept low. Ideally, the integration time should also be kept short.

However, in order to achieve clear images with a favourable signal-to-noise ratio, a high generation rate and/or a long integration time is needed. Since the splitting of FeB increases in proportion to the *square* of the generation rate, whereas the PL signal is directly proportional to it, a low generation rate (i.e. laser intensity) and a long integration time provide the best overall result [16]. The measurements for this work were mostly done by illumination intensities of 1/100 sun and with integration times between one and three minutes.

Thirdly, *complete dissociation* of the FeB pairs must be ensured when the dissociated state is prepared. For this purpose, as shown in Figure 3.6, the sample was illuminated with an intensity of approximately 1.6 suns for 4 minutes before the Fe_i measurement.

Below the cross-over point, the lifetime – and thus the PL signal – is lower in the dissociated state than in the associated state. Thus, a wafer in its dissociated state will give a weaker PL signal than the same wafer in its associated state, when subjected to the same illumination. For these reasons, longer integration times must be chosen in the dissociated state than in the associated state. In this work, integration times up to five minutes have been used.

3.5 HYPERSPECTRAL IMAGING

3.5.1 Measurement setup

The hyperspectral imaging technique used in this work is a scanning procedure for which a camera moving across the sample is applied. The excitation laser is connected to the same movable rig as the camera, thus moving in parallel with it. The sample is placed on a cryogenic cooler below the camera. A drawing of the measurement setup for the hyperspectral measurements is shown in Figure 3.7

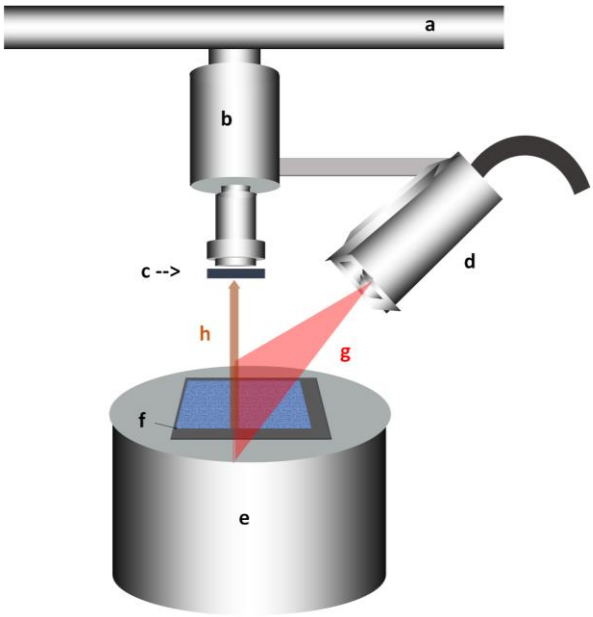


Figure 3.7: *Setup for the hyperspectral measurements. A movable rig (a), to which both the camera (b) with long-pas filter (c) and excitation laser (d) are connected, moves across the sample. A cryogenic cooler (e) filled with liquid nitrogen and equipped with a sample holder (f) ensures that the sample is cooled to approximately -190ºC. After [8].*

Camera

The hyperspectral images were recorded with a HgCdTe SWIR camera (Specim, Spectral Imaging, Ltd, Oulu Finland), capable of detecting light with wavelengths in the range 926-2527 nm (corresponding to 0.49-1.34 eV). While measuring a sample, the camera moves slowly, recording the signal from one row of pixels simultaneously, before moving to the next row of pixels. For each pixel, the incoming light falls onto a diffraction grating, causing a light deflection, for which the diffraction angle depends on the wavelength of the incoming light. For any wavelength λ , the deflection angle θ_n for diffraction maximum n is given by [55]

$$sin\theta_n = \frac{n\lambda}{d} \tag{3.15}$$

By normal operation of the camera, only the 1st order maximum of the diffracted light will be detected. Thus, for each row of spatial pixels, a two-dimensional image is made, in which the first dimension corresponds to the spatial dimension of the wafer, and the second dimension

corresponds to the wavelength. When the camera is scanned over the sample, recording one two-dimensional image for each row of pixels, a three-dimensional tensor of data is produced. This tensor is denoted a *hypercube.* Figure 3.8 explains the data acquisition of the hyperspectral camera.

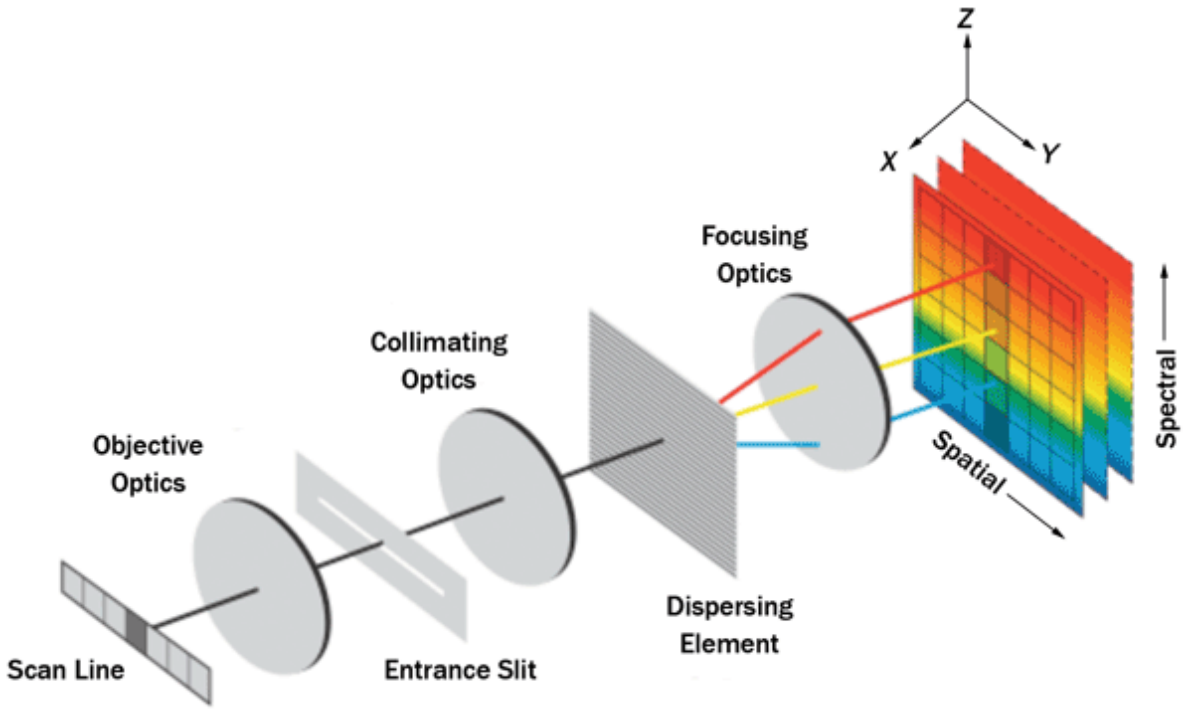


Figure 3.8: *Working principle for the hyperspectral camera. Photoluminescence signal from one line of pixels at the sample is recorded simultaneously. The light is collimated before hitting a diffraction grating, thus resulting in a two-dimensional image with one spatial and one spectral dimension. By letting the camera scan the sample, a tensor with two spatial directions and one spectral dimension is created. Taken from [9].*

In the hypercube, the x- and y-directions (as denoted in Figure 3.8) represent the two spatial directions of the wafer, while the z-direction represents the wavelength dimension. For clarity, in this work, the spatial direction parallel to the row of pixels recorded simultaneously is denoted *x*, while the spatial direction in which the camera is moved is denoted *t.* The wavelength dimension is denoted λ. Figure 3.9 illustrates how the dimensions of the hypercube are denoted in the following. The two-dimensional images recorded by the chip (shown in Figure 3.9 b), having the dimensions x and λ, are called *frames.*

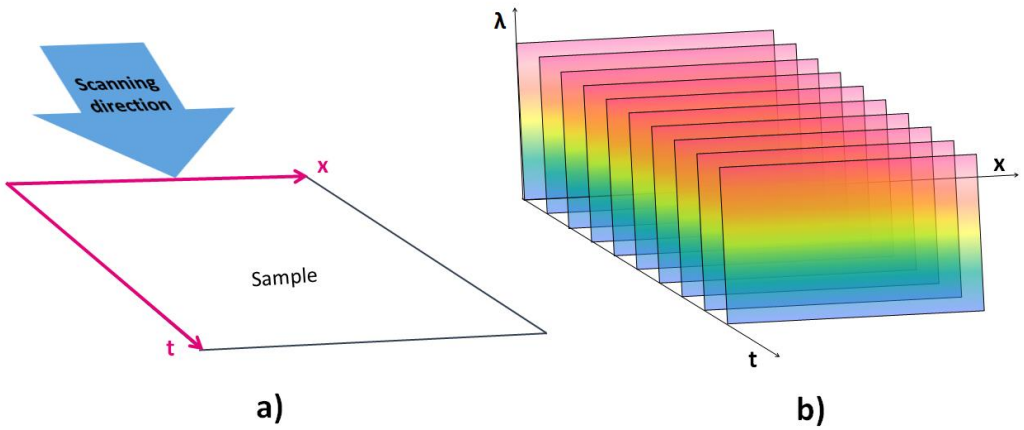


Figure 3.9: *The dimensions of the hypercube as used in this work. Part a) shows the wafer sample and the directions in which it is being scanned. Part b) shows the resulting hypercube, and its tree directions, x, t and* ***λ****. The two-dimensional images (with dimensions x and* ***λ)*** *of which the hypercube consist are denoted frames. The dimensions x and t are the two spatial dimensions of the wafer, being orthogonal and parallel to the scanning direction respectively.*

In the λ-direction, the camera chip has 256 pixels, giving a spectral resolution of approximately 0.033eV. The energy intervals recorded per pixel (in this work denoted as *wavelength bands*) are, however, not constant throughout the camera chip.

Excitation source

As excitation source for the hyperspectral imaging, a 808 nm line-shaped laser from Coherent (Lasiris Magnum II, Gamatada Instrument AB, Uppsala, Sweden) was employed [8, 56]. The laser is mounted to the same movable rig as the camera, meaning that when the camera scans the sample, the laser will always illuminate the line momentarily photographed by the camera.

In order to prevent reflected laser light from disturbing the measurement, the camera was equipped with a (850nm) long-pass filter. Despite this filter, however, some light originating from the laser enters the camera optics. Because this light has shorter wavelengths than light within the normal detection range of the camera, its diffraction due to the diffraction grating in the camera will be less than that of the light within the detectable range. The 1^{st} order maximum of this light does therefore not hit the camera chip, and is not detected. The 2^{nd} order maximum, on the other hand, is deflected exactly twice as much as the 1^{st} order maximum (cf. equation (3.15)). It thus hits the camera chip exactly at the point at which light with twice its wavelength normally is being detected. As a result, a strong signal at 1616 nm, corresponding to 0.77 eV is seen in the measured data (see section 5.1).

Cryogenic cooler

In order to minimize the impact of phonons (see section 2.6.1) thus making the spectral resolution of the images clearer, it is necessary to cool the samples before the measurements are conducted. For this purpose, a cryogenic cooler, consisting of an aluminium vessel covered with Styrofoam isolation was applied. When the vessel is filled with liquid nitrogen, the entire aluminium body cools down to temperatures of about -190ºC. The wafer lying on it, will virtually reach a state of thermal equilibrium with the nitrogen in the vessel. In order to achieve a cooling also of the air above the sample, and thus avoiding frost, the plate on the top of the

sample holder is surrounded by small holes through which evaporated nitrogen from the vessel may blow. The sample holder is equipped with a thermometer, allowing the sample temperature to be measured.

3.5.2 Laser intensity and distribution

Spatial distribution of the laser light

According to its product sheet [57], the laser is supposed to give an evenly illuminated line, with only a slight increase in intensity at the ends of the line. Its beam power is user adjustable, having a maximum value of 5600 mW. Adjusting the laser power is done manually by regulating the bias voltage with a potentiometer. For bias voltage in the range 0.5-4.5 V, a linear increase in the laser power is expected [57].

As a part of this work, the spatial distribution of the laser power was measured, using an optical power meter (model 1918-R) and a UV silicon photodetector (model 918D-UV-ODR3), both from Newport Corporation. The measurement was conducted by moving the power meter spatially, in steps of 1 cm along an axis perpendicular to the scanning direction of the laser, and measuring the laser output for each position. The measurement was done for bias voltages of 5V and 2V. The results, obtained as a mean of three separate measurement series for each of the voltages, and with accordingly calculated standard deviations, are shown in Figure 3.10.

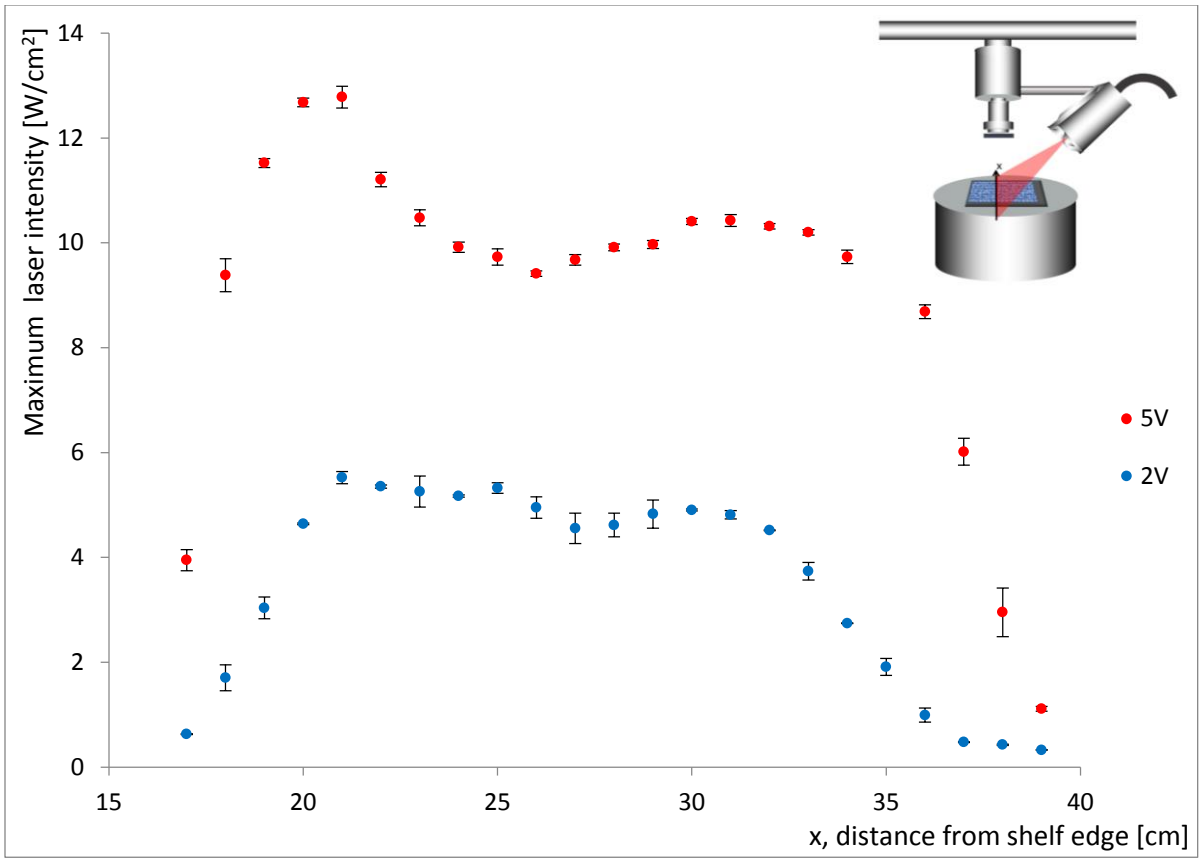


Figure 3.10: *Measured intensity of the excitation laser, as function of distance from measurement shelf edge, for laser bias voltages of 5V (red) and 2V (blue). The inset in the upper right corner illustrates how the x-axis was chosen relative to the measurement setup. Wafers are normally placed with their left edge about 19cm from the shelf border.*

It should be mentioned that the absolute values of the laser intensity shown in Figure 3.10 are based on the assumption that the width of the laser line equals the minimum value given in the product sheet (0.2 mm) [57]. Thus, the values may be regarded as a maximum possible value for the laser intensity rather than its true value. Nevertheless, the relative variation of the laser intensity throughout the width of the sample is independent of this assumption.

Adjustment of laser line to region photographed by camera

The measurement described above was done independently of the camera, meaning that the measured laser profiles describe the *total* laser power as function of position, irrespective of the adjustment of the laser to the line photographed by the camera.

Adjusting the laser line to the region of the wafer recorded by the camera is done manually by inspecting the camera signal on the computer. As illustrated in Figure 3.11, the position of the laser line can be adjusted in the direction perpendicular to the recorded line, as well as angularly. An inaccurate adjustment will mean that some parts of the line recorded by the camera is illuminated by the total available laser power for that spatial position, whereas other parts of the photographed line are illuminated with only parts of the laser power at their position. The remaining part of the laser light illuminates neighbouring parts of the sample (i.e. parts that are

not momentarily being photographed).

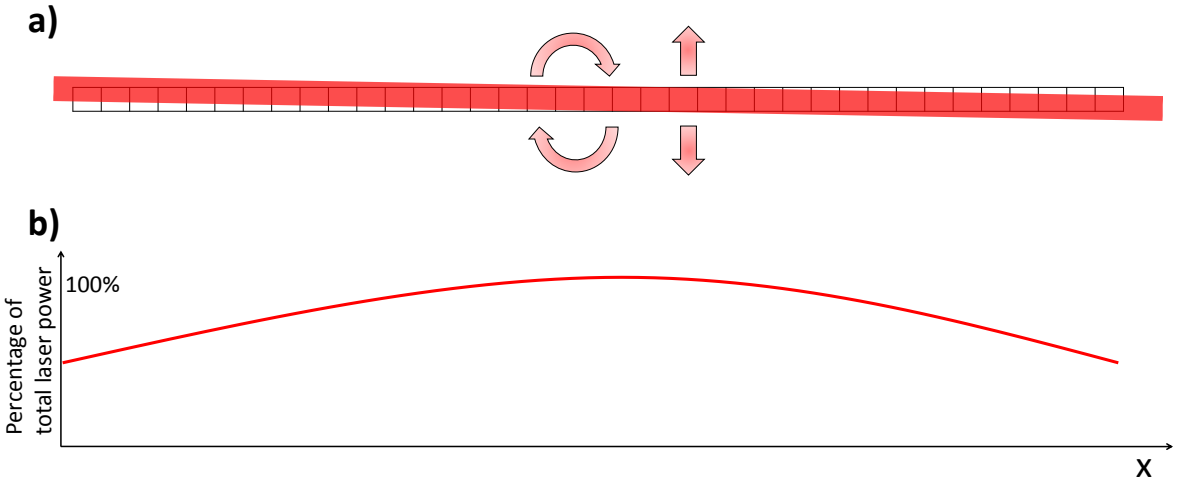


Figure 3.11: *a) Illustration of possible angular deviation between area recorded by the camera (grid) and area illuminated by the excitation laser (red line). The arrows indicate how the laser can be adjusted as compared to the area recorded by the laser. b) Resulting variation in percentage of total laser power for position x illuminating the area recorded by the camera.*

Thus, in addition to the spatial variation of the *laser power as such* (shown Figure 3.10), a spatial variation of the *percentage of the laser light illuminating the recorded line of the sample* (illustrated in figure Figure 3.11) exists. In an ideal case the effects of the two types of spatial variation may partly cancel each other. This happens if, for example, the distinct peak at the left wing of the measured laser power profile illuminates an area not coinciding exactly with the line momentarily being photographed. In other cases, however, the two effects will reinforce, making the actual spatial variation in illumination intensity of the line being recorded by the laser even larger than the variations of the laser power as such.

Figure 3.12 shows two images of the band-to-band emission line of wafer p-twin-central, made from measurements between which the wafer is rotated 90°. The more intense part of the laser therefore illuminates different parts of the wafer. The differences between the two images are remarkable.

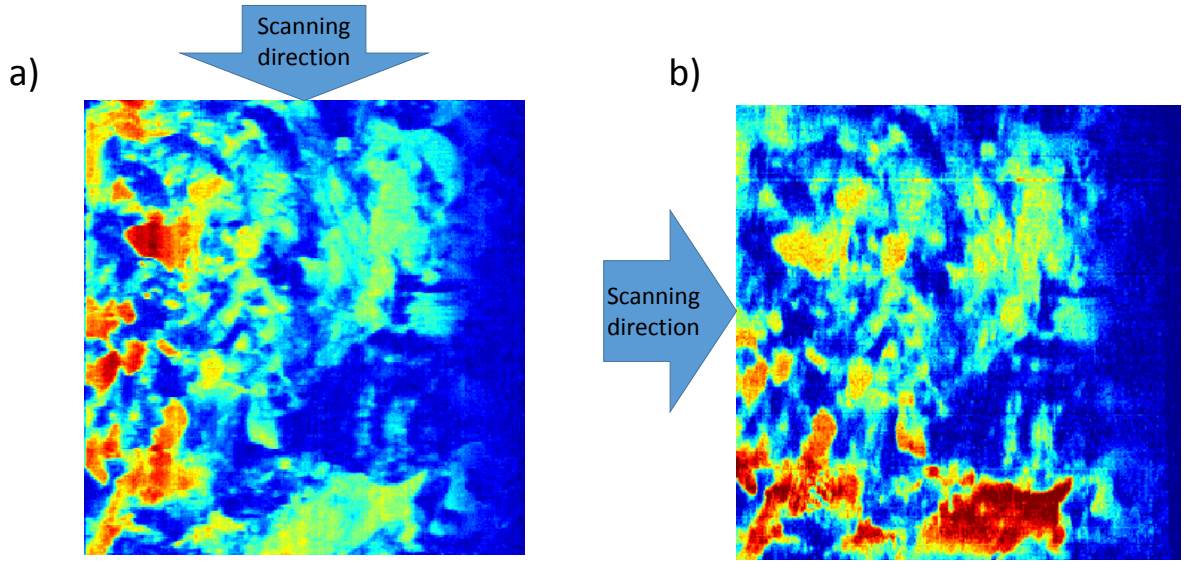


Figure 3.12: *Comparison of spatial images of the band-to-band emission line from wafer p-twin-central-as-cut scanned a) vertically and b) horizontally.*

Within the frames of this work, no correction for the inhomogeneous laser illumination has been done. The effects of this inhomogeneity must be kept in mind when drawing conclusions based on hyperspectral images recorded with this setup. In section 4.5, the question of reliability of the hyperspectral measurements is further discussed, also taking the inhomogeneous laser profile into account.

Temporal stability of the laser

The temporal stability of the laser was also measured. This was done by placing a multicrystalline wafer on the sample holder, and letting the camera and the laser scan it – exactly as it would do by a normal measurement. The temporal variation of the laser intensity was found as a mean of 20 centrally placed pixels in the spatial direction x of the wafer, for the wavelength band corresponding to the 2nd order diffraction signal of the laser. Since this signal originates from the laser itself, and do not depend on the sample[5], it can be assumed that its intensity is proportional to the laser intensity. The evolution of the laser signal through time (shown in Figure 3.13) corresponds to the evolution in the spatial direction t.

[5] In principle, changes in the reflection of the sample surface, could impact the measured signal. Such an effect is, however, assumed to be negligible compared to the variations of the laser itself.

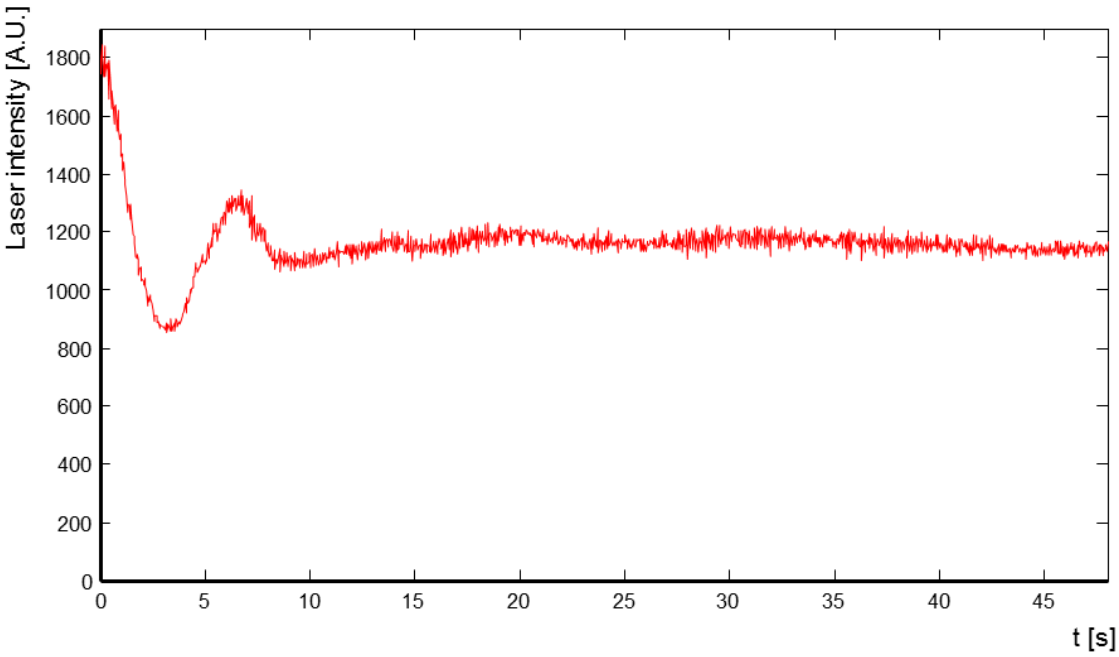


Figure 3.13: *Example of temporal variation of laser intensity. The laser was turned on at time t =0. The laser intensity is found as mean of 20 pixels centrally placed in the spatial direction x of the image, for the* $\boldsymbol{\lambda}$ *value corresponding to the 2nd order diffraction maximum of the laser light. The evolution through time corresponds to the evolution in the spatial direction t. The measurement was done for a laser bias voltage of 3V.*

As seen from the figure, during the first 10 seconds after the laser is turned on, its intensity varies dramatically. In order to prevent the laser – which potentially has very high irradiation intensity per area – from causing damage to the measurement setup, it has normally been turned on just before starting the measurement. Taking into account that scanning across a wafer lasts between 10 and 15 seconds, it is clear that the instability of the laser may influence the measurement results. In order to minimize this effect, the laser should be turned on at least 10 seconds before starting the measurement. If the scanning process is started at the same time as the laser is turned on, meaning that the laser does not illuminate the same part of the setup over a long time, this can be done without damage risk.

3.5.3 Measurement procedure

Before measuring, liquid nitrogen is frequently poured into the cryogenic cooler until the sample holder reaches its desired temperature of about -190°C. The cooling process lasts approximately two hours. Also between the measurements, small amounts of fluent nitrogen must be added to the cooler striving to keep the temperature constant. Before each new measurement, frost must be removed manually from the sample holder, a process causing the temperature to increase slightly [58]. However, due to the large heat capacity of the cooler, the sample holder (and thus the wafer to be measured) quickly reaches a state of thermal equilibrium with the cooler, allowing the measurement to be started.

The scanning speed of the movable rig must be adjusted such that the spatial distance Δt the movable rig moves between two subsequent frames equals the spatial width Δx of a pixel. Only then, the proportions of the image will equal the proportions of the imaged object. When the

setup is adjusted for 165mm·165mm wafers, the pixel width Δx equals 0.504mm. The frame rate is set to 25 Hz, meaning that the time between two subsequent frames equals $\frac{1}{25\ s^{-1}} = 0.04\ s$. Thus, in order to get quadratic pixels, the scanning speed was set to $0.504\ \mathrm{mm}/0.04\ \mathrm{s} = 12.6\ \mathrm{mm/s}$. At this speed, one measurement takes about 18 seconds, not including preparations [58]. For measurements with higher resolution, meaning that the pixel width Δx is smaller, also the scanning speed has to be lower, in order to achieve quadratic pixels. If the length of the sample is unchanged, the entire measurement therefore takes longer.

In this work, all wafers were firstly measured at standard illumination conditions (5V laser bias voltage) and integration time (20 ms). Under those conditions, many wafers caused oversaturation of the camera chip, especially for the wavelengths close to the band gap. Therefore, also measurement series with lower illumination intensity and with shorter integration time were taken.

As in an earlier work based on the same setup[59], all measurements were repeated three times. This was originally done in order to gain the possibility of taking the median of the three measurements, thus obtaining images with an improved signal-to-noise ratio. However, as will be explained in sections 4.3 and 4.4, in this work the repeated measurements were rather used for an investigation of the background noise level and the development of a new method to correct for its variations.

4 DATA PROCESSING

An important part of this work is the development of a new data processing regime correcting for the oscillation of the background noise level of the utilized hyperspectral camera. This chapter is dedicated to description of this new method, as well as other data processing and analysis techniques used in this work. First, the theory and working principle of the useful analysis tool *multivariate curve resolution* (MCR) is described (section 4.1). Further – in section 4.2 – the standard data processing regime, routinely used for the analysis of hyperspectral images from the utilised setup, is presented. In section 4.3, through a comparison of subsequent measurements of the same wafer, the weakness of this method is identified. A temporal oscillation of the background is recognized as the reason for the difference between subsequent measurements. Finally – in section 4.4 – the problem related to the temporal background oscillation is solved through the development and application of the new background correction regime. The last subchapter – section 4.5 – describes and evaluates various methods for analysis of hyperspectral data, taking the uncertainties related to background noise level and laser distribution into account.

4.1 MULTIVARIATE CURVE RESOLUTION (MCR)

Data from hyperspectral measurements mostly consists of various signals – caused by different physical mechanisms – that are partly mixed and overlaid, spectrally as well as spatially. Thus, there is need for a mathematical method that can extract the main data components. Multivariate curve resolution, which mathematically builds on the following bilinear model

$$\boldsymbol{D} = \boldsymbol{C}\boldsymbol{S}^{\boldsymbol{t}} + \boldsymbol{E} \qquad (4.1)$$

is one such method. In equation (4.1) **D** may represent any data matrix. In the case of a hyperspectral measurement, **D** will contain the spectra recorded for each pixel in the hyperspectral image. $\mathbf{S}^{\mathbf{t}}$ represents the main spectra (also called the signature or the pure spectra) in **D**, practically meaning the "most frequently occurring" spectra in **D**. For each pixel in the original data matrix **D**, the matrix **C** gives the weights that each spectrum in $\mathbf{S}^{\mathbf{t}}$ must be given in order to reconstruct the original spectrum for that pixel. These weights are commonly denoted the *concentrations* or of the spectra in $\mathbf{S}^{\mathbf{t}}$. The multivariate curve resolution technique is, in other words, based on the assumption that the observed spectra **D** are a linear combination of the pure spectra in $\mathbf{S}^{\mathbf{t}}$. **E** represents the matrix of residuals not explained by **C** and $\mathbf{S}^{\mathbf{t}}$ [60, 61].

Principal component analysis (PCA), which is another way of reducing the dimension of multivariate data, has as goal to find a new basis for the data matrix D with the property that when D is described in this basis, all its values are uncorrelated and arranged in order of decreasing variance. For datasets in which most of the variation is explained by changes in only a few of the components in the new basis, other components may be omitted with only a very small loss of information [62]. Thus, the dimensions of the dataset may be reduced accordingly. However, due to the orthogonality constraint and the constraint of maximum variance, PCA will not necessarily find the true pure spectra and their concentrations in a data material [61].

As opposed to PCA, the MCR algorithm strives to find the true spectra of which a mixed signal consists, regardless of orthogonality of the spectra. This is achieved through the application of an alternating least square algorithm (ALS) that iteratively solves equation (4.1) [60]. The iterative method starts with an initial estimation for the concentration matrix **C.** As a first step, an estimation for the pure spectra $\mathbf{S^t}$ may then be calculated by

$$\boldsymbol{S^t = C^+ D} \qquad \textit{(4.2)}$$

for which $\mathbf{C^+}$ is the pseudoinverse[6] of **C**. In the second step, a new estimate for **C** is calculated by the use of the pseudoinverse of $\mathbf{S^t}$, denoted $\mathbf{(S^t)^+}$, according to

$$\boldsymbol{C = D(S^t)^+} \qquad \textit{(4.3)}$$

For each step, $\mathbf{S^t}$ and **C** may be modified according to certain constraints such as non-negativity (meaning that neither the pure spectra nor the concentration matrix **C** can contain negative values) or unimodality (meaning that the sought pure spectra have to be without double peaks). For hyperspectral imaging applications, it could mostly be assumed that neither the pure spectra nor their concentrations contain negative values. Thus, a non-negativity constraint may be applied to both $\mathbf{S^t}$ and **C** [61, 63].

The two steps, combined with the application of the constraints, are repeated. For any two consecutive iteration cycles A and B, the relative difference of the standard deviation of the residual matrices $\mathbf{E_A}$ and $\mathbf{E_B}$ is calculated. When this value is less than a previously selected value (usually set to 0.1%), convergence is said to be achieved [60], and it is then assumed that the matrix **D** is well explained by $\mathbf{S^t}$ and **C** [63].

As any mathematical method for linear decomposition of matrices, MCR-ALS is subject to *ambiguities*, meaning possible variations in the matrices that solve the bilinear model in equation (4.1). Variations in the shape of the spectra represented by the matrices are denoted *rotational ambiguity*, and variations in their magnitude is denoted *intensity ambiguity* [64].

Theoretically, an infinite number of the matrix pair $\mathbf{S^t}$ and **C** may be found that reproduce the data **D** within the uncertainty of the preselected convergence limit. This may be explained by replacing **C** and $\mathbf{S^t}$ in equation (4.1) with $\boldsymbol{C' = CT}$ and $\boldsymbol{S'^t = T^{-1}S^t}$ respectively. Even though **C'** and $\mathbf{S'^t}$ are not the sought solutions of the bilinear model given in equation (4.1), their product will be the same as the product $\mathbf{CS^t}$ [64]. Therefore, they may be falsely interpreted as a "true" solution of the system. The ambiguities will to a large extent diminish through the use of suitable constraints, forcing **C** and $\mathbf{S^t}$ to obey certain conditions. However, according to Zhang and Tauler [61], for hyperspectral imaging, the amount of rotational ambiguity present in a solution may not be negligible.

[6] The inverse of a matrix is only defined for square matrices. For non-square matrices however, a *psedoinverse* (also called the *More-Penrose inverse*) might be defined through a singular value decomposition. The psedoinverse of a non-square matrix has similar properties as the inverse for square matrix [62].

4.2 STANDARD BACKGROUND SUBTRACTION PROCEDURE

The images recorded with the hyperspectral camera used in this work have a very high background noise level. As shown in Figure 4.1, the background noise level is more than one order of magnitude larger than the signal. Moreover, the background level $b_{x1,\lambda1}$ for one certain pixel corresponding to the spatial line $x = x_1$ and the wavelength band $\lambda = \lambda_1$ does not equal the level $b_{x2,\lambda2}$ of another pixel corresponding to another spatial position in the x-direction $x = x_2$ and another wavelength band $\lambda = \lambda_2$. Thus, at the end of each measurement, the camera shutter was kept closed for some seconds. This results in a *dark area* at the bottom of the picture in which no luminescence signal is present (marked by the dashed rectangle in the right image of Figure 4.1). The dark area can therefore be utilized for finding the background level of the other parts of the image. Normally, the background value for each pixel of the camera chip is found as the median value for that pixel in 50 successive frames in the dark area. In the t, x, λ-coordinate system, this means that the background for pixel (x_j, λ_k) on the two-dimensional camera chip is found by taking the median of pixel values from (t_i, x_j, λ_k) to $(t_{i+50}, x_j, \lambda_k)$ in the three-dimensional hypercube. An example of an image before and after background subtraction is shown in Figure 4.1.

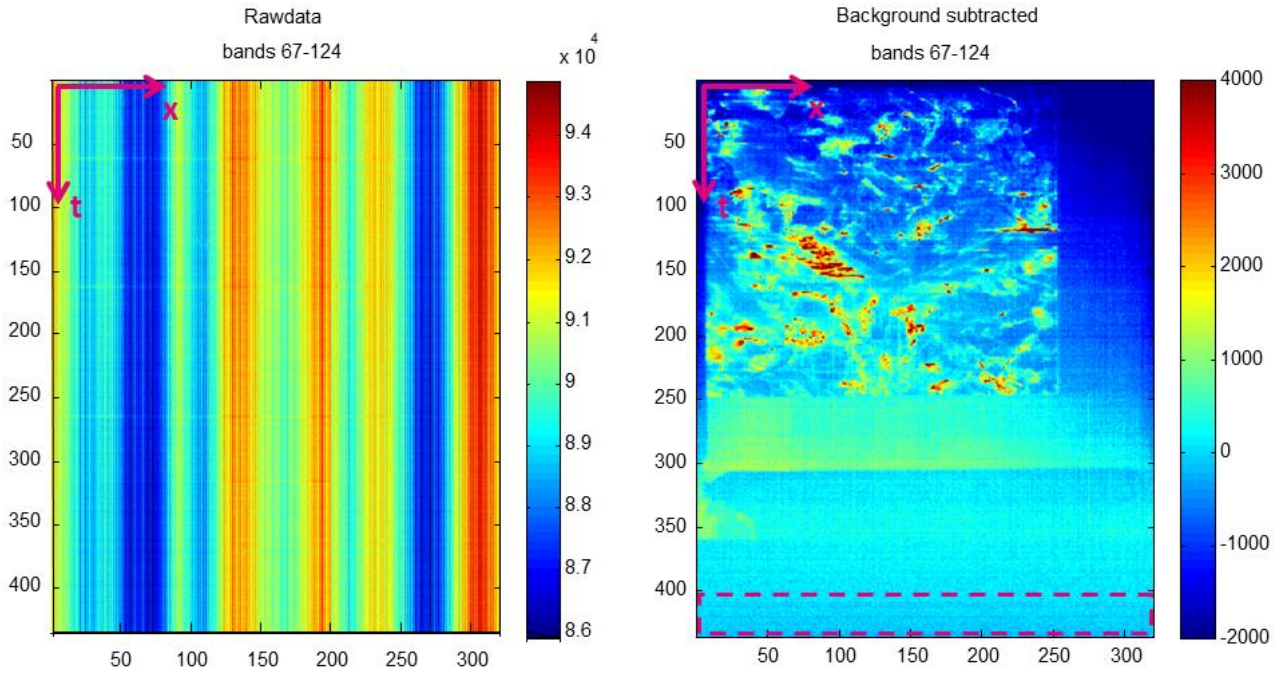


Figure 4.1: *Illustration of the background noise level compared to measurement signal. The left image shows an integral over wavelength bands 67-124 before the background subtraction is performed, whereas the right image shows the same wavelengths after background subtraction. At the right image, the wafer, as well as the part of the sample holder not covered by the sample (yellow area below wafer and blue area at the right side of the wafer) are visible. The dashed rectangle at the right image shows the "dark area" from which the background level for each pixel is found.*

4.3 DIFFERENCES BETWEEN SUBSEQUENT MEASUREMENTS

As mentioned in section 3.5.3, for each combination of measurement parameters (illumination, integration time and temperature), three measurements were done. In prior experiments, this procedure was generated in order to create a median of the three images [65]. Accordingly, it has previously been closely examined whether spatial points in subsequent images correspond. In this work, however, the spectral intensities of subsequent images of the same wafer are for the first time thoroughly compared.

4.3.1 Spectral comparison through MCR of single measurements

First, the spectral information of subsequent images was compared through the use of MCR. The theoretical background for this method has been explained in sections 4.1 and a closer description of how it is applied to hyperspectral measurements can be found in section 4.5.4. Figure 4.2 shows a comparison of the MCR loading plots from three measurements of wafer p-twin-central with equal measurement parameters.

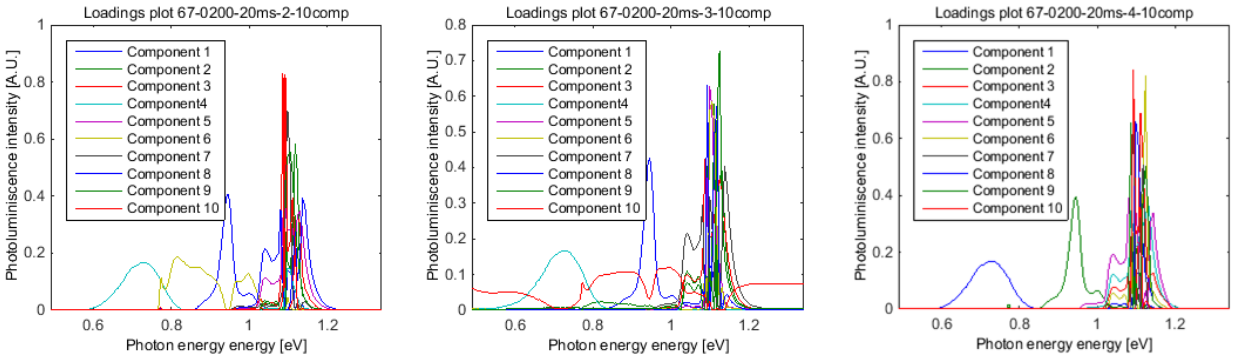


Figure 4.2: *Comparison of MCR loading plots from three subsequent measurements of the same wafer with equal measurement parameters.*

Differences between the three plots are easily seen. As explained in section 4.1, due to rotational and intensity ambiguity, the results of an MCR analysis may differ, even if the raw data matrices **D** which is input for the analysis is unchanged. Mehl [59] has compared several MCR analyses of the same hyperspectral measurement done with the same setup as has been used for this work. He found that the differences between subsequent MCR analyses of the same measurement are, indeed, negligible. Based on this finding, it seems unlikely that the large variations between the plots in Figure 4.2 are caused solely by artefacts related to the MCR analysis as such.

4.3.2 Spatial comparison through integrated images

Secondly, the spatial information in the images was compared. In order to avoid possible confusions due to ambiguities of the MCR analyses, this was done directly by integrating over certain wavelength bands. An example of three such images, made by integration over the interval 0.73 – 0.92 eV, is shown in Figure 4.3. All the images are from wafer p-twin-central with equal measurement conditions.

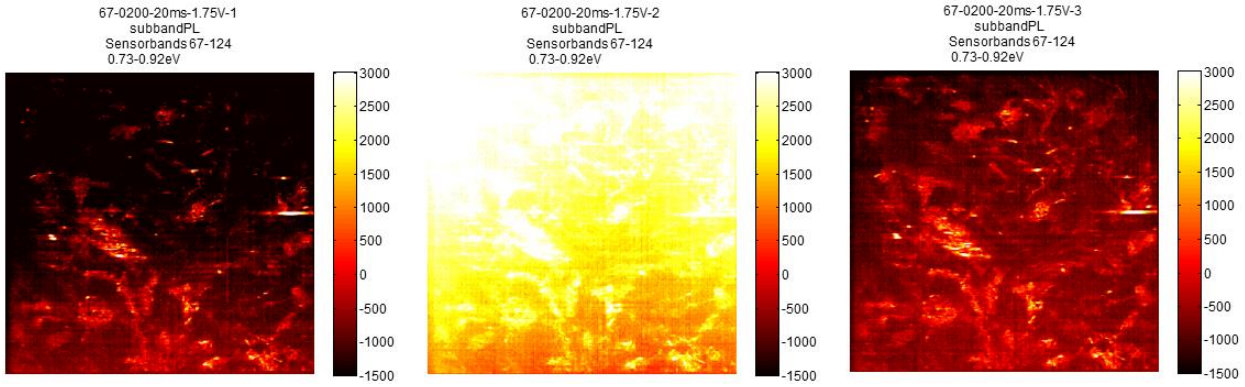


Figure 4.3: *Comparison of the Sub-band PL signal from three subsequent measurements of the same wafer with equal measurement parameters.*

Whereas Figure 4.2 suggests that the relative intensity of the signal from different wavelength intervals varies, an effect that partly could be attributed to a combination of rotational ambiguity and intensity ambiguity, Figure 4.3 clearly shows that the number of counts varies from one measurement to the next. In addition to the large variation in mean number of counts between the three images, the count rate seems to *increase* from top to bottom during the measurements 1 and 3 (left and right), whereas it *decreases* from top to bottom during measurement 2 (central). This feature suggests some variable to fluctuate periodically.

4.3.3 Comparison of subsequent images through combined MCR

In order to investigate the idea of a periodically fluctuating variable further, a combined MCR analysis was done for three subsequent measurements of the same wafer. This was obtained by horizontally joining the hypercube representing each of the three measurements, and thereafter performing an MCR analysis of the combined hypercube. The resulting loading plot is shown in Figure 4.4.

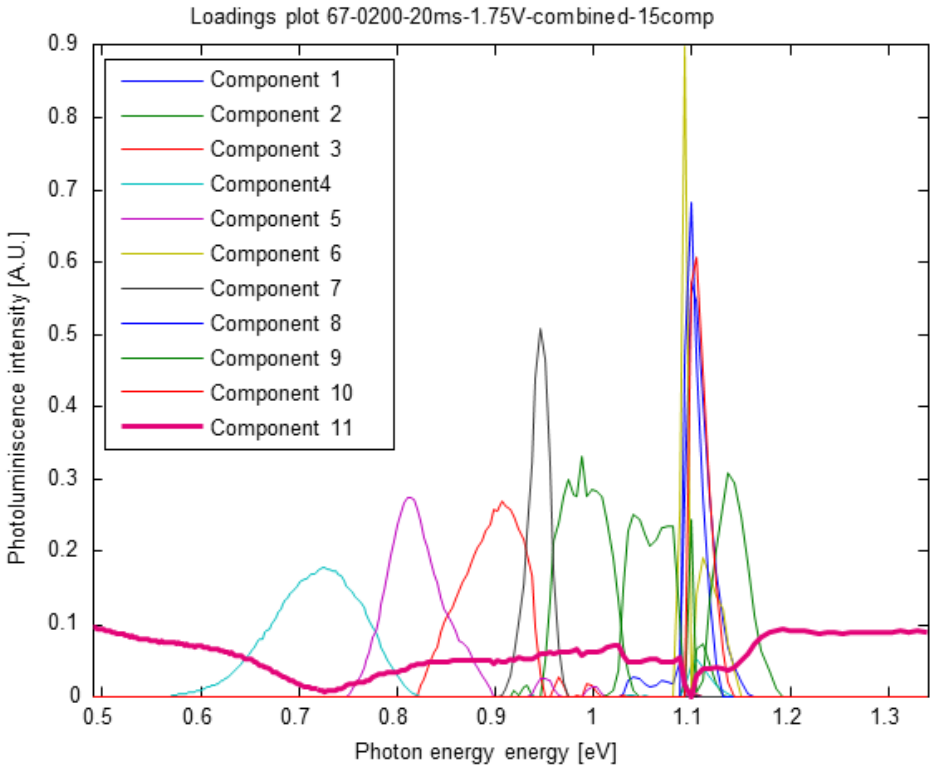


Figure 4.4: *Loadings plot from an MCR analysis of a combination of three subsequent measurements of the same wafer. The three images were joined horizontally before the analysis was done.*

Component 11 in this plot suggests the presence of a signal ranging over the entire wavelength area. The score plot of this component (Figure 4.5) reveals that this component is present only at measurement number 2 of 3, and that its intensity varies slightly across the wafer.

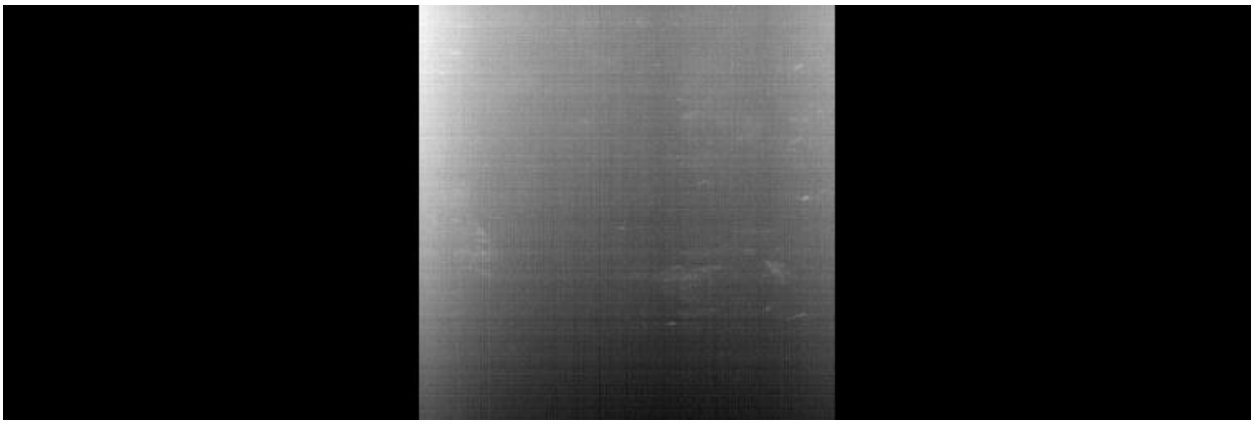


Figure 4.5: *Score plot of component 11 from the above shown MCR-analysis. Since the MCR analysis is done for three hypercubes (with spatial dimensions t×x) that are horizontally joined, the score plot have dimensions t×3x. The left part of the score plot correspond to the first measurements, the central part to the second measurement and the right part to the third measurement. Note that the measurements are of the same wafer, with identical measurement parameters.*

For other wafers, similar components where found that where present in measurement 1 and 3, but not in measurement 2. This, again, suggests a periodic behaviour of some measurement parameter.

4.3.4 Identification of thermal background oscillation as a probable reason for the difference between measurements

Having identified a periodical behaviour of the measurement data, possible reasons for this behaviour were sought.

First, it was hypothesized that the laser may, under some conditions, need longer time to stabilize than what is visible in Figure 3.13. Therefore, measurements similar to that shown in Figure 3.13, for time intervals corresponding to five wafer measurements were done. However, except from a strong oscillation during the first 10 seconds, the laser intensity seemed not to oscillate notably. To further rule out the possibility of such an effect, wafer measurements were performed after the laser had been allowed to stabilize for five minutes. The same periodical behaviour of the count rate was still seen. The hypothesis of the laser causing the periodicity seems therefore very unlikely.

Secondly, it was suspected that the difference in count rate was caused by fluctuations in the background noise level of the camera detector during the measurements. In order to investigate this possibility, measurements over a long time interval (3 min) with a constantly closed shutter, allowing only the background noise level to be recorded, were done. A plot of the count rate for each pixel (x_1, λ_i) along the line $x = x_1$ of the camera chip as a function of time, shown in Figure 4.6, confirmed the suspicion of a periodically oscillating behaviour of the background noise.

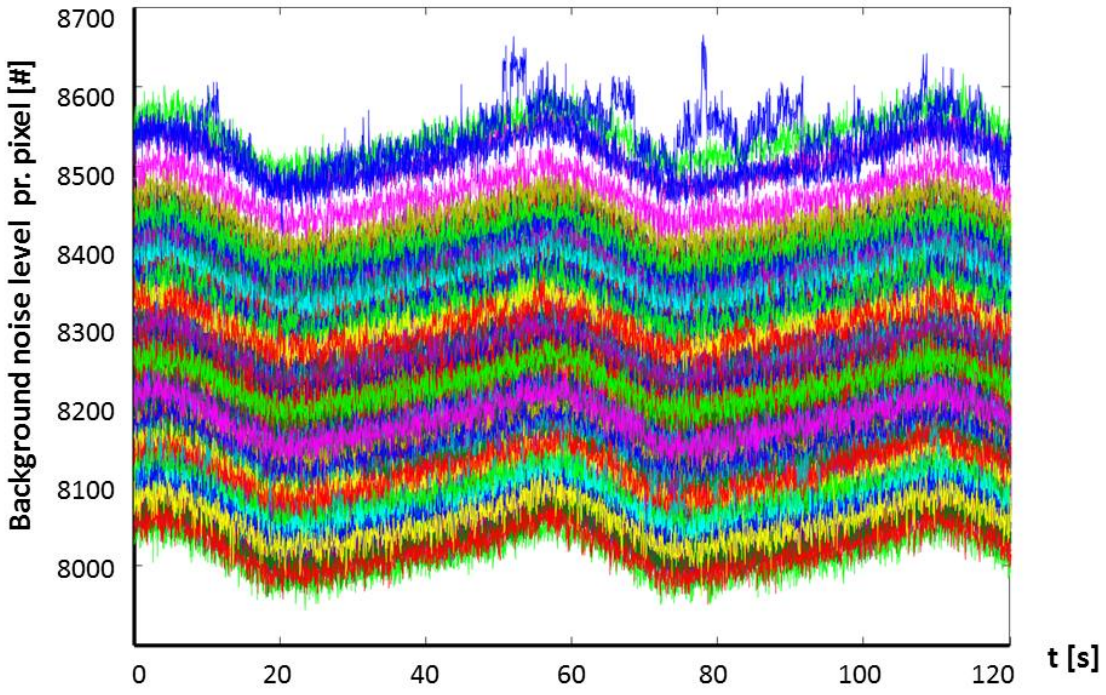


Figure 4.6: *Plot of the background noise level for a line of pixels* $\boldsymbol{x = x_1}$*, along the central part of the camera chip. Each plot represents the background noise level for one of the 256 wavelength bands.*

The plot illustrates both the difference in background noise level from pixel to pixel, and the variation of the background level for all pixels as a function of time. Both variations are believed

to be caused by the camera cooling system. If the cooling were inactive, the temperature would increase, causing the background noise to increase (close to) linearly with time. This isfor example seen between $t = 25s$ and $t = 55s$ in the figure. When the temperature reaches a certain level, the cooling would be activated, causing the temperature, and thus the background noise level to decrease quickly. This is seen for example between $t = 60s$ and $t = 70s$. After reaching a certain minimum temperature, the cooling system would again be deactivated, causing a new temperature increase.

Figure 4.7 shows the spatial variation of the background noise on the camera chip (shutter closed) for one certain time t. Based on the observed pattern, one can speculate that the differences are caused by flowing liquid or gas cooling the camera chip.

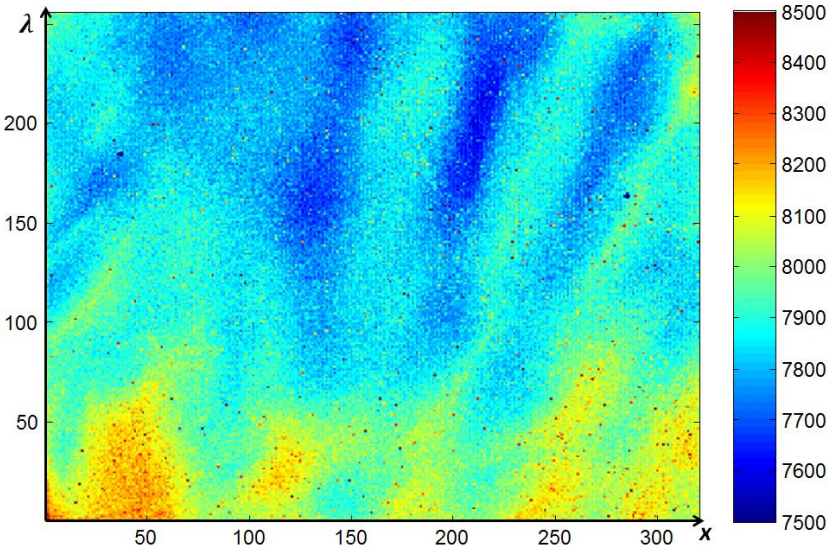


Figure 4.7: *Visualisation of the background noise and its spatial variation on the camera chip. The colorbar gives the number of counts per 20ms.*

4.3.5 Discussion of the effect of the background oscillation on the measurement result

An approximate picture of the effect of the background oscillation on the measurement result could be achieved by a closer investigation of Figure 4.6. When studying the figure, one should take into account that scanning one wafer takes about 18 seconds, and that the defect PL signal per wavelength interval often contributes by less than 50 counts (see for example Figure 5.15). The oscillation of the background level however has an amplitude of almost 100 counts. This means that if the old scheme for the background subtraction is used, and if the dark area of the image accidentally is recorded in a period with a high background noise level (e.g. at t = 55 s in Figure 4.6), most of the pixels in the resulting image will attain negative values. This explains, for example, the large areas with negative count values on measurement 1 and 3 in Figure 4.3.

As explained in section 4.1, non-negativity constraints are applied in the MCR-ALS algorithm, meaning that neither spectra nor concentrations are allowed to attain negative values. Analysing a dataset in which many of the values are negative, which will often be the case after applying the old background subtraction scheme to a measurement, may therefore cause incorrect results. This may explain the differences between the MCR loading plots in Figure 4.2.

An important purpose of this work was to *compare* defect luminescence from various wafers (e.g. after different processing steps) or from various wafer regions (e.g. with different content of metal contamination). Based on the considerations above, it should be clear that conclusions regarding such comparisons cannot be drawn without firstly solving the problems related to the thermal background oscillation. Therefore, an improved version of the background subtraction scheme which corrects for this oscillation had to be developed.

4.4 Development of a new regime for background subtraction

By the development of a new background subtraction scheme, mainly two aspects of the background oscillation had to be considered. These were

i) the dependence of the background noise level on time, i.e. the change of the background noise level along the t-dimension of the sample, and
ii) the spatial variation of the background noise level on the camera chip, i.e. the variation of the background noise level along the x-dimension of the sample and along the λ-dimension.

These aspects are considered in the following two subsections, and methods are found to correct for their influence. Thereafter, the working principle of the new background subtraction scheme is explained. Lastly, as a quality control for the scheme, measurements on which it is applied, are compared.

4.4.1 Variation of the background noise level with time

For any frame in a hyperspectral measurement, it is necessary to determine at which point in the thermal oscillation cycle (Figure 4.6) the frame is recorded, i.e. to determine the magnitude of the "mean" background noise level for that frame. For this purpose, it was desired to find an area of the camera chip that was not at all influenced by incoming light for the duration of the entire measurement. Because the variation of counts with time in such an area would be caused only by the variation in the background level, this area could function as a reference background level for the entire camera chip.

Practically, there are two possibilities for finding such a reference area; either it can be taken from the part of the chip detecting light with energy higher than the bandgap E_G of silicon, or it can be taken from the part of the chip detecting light with energies even lower than the detected signals. A visual comparison of the signal from these two parts of a typical measurement is provided in Figure 4.8. Note that standard background subtraction has been performed on the measurement, causing the negative count values.

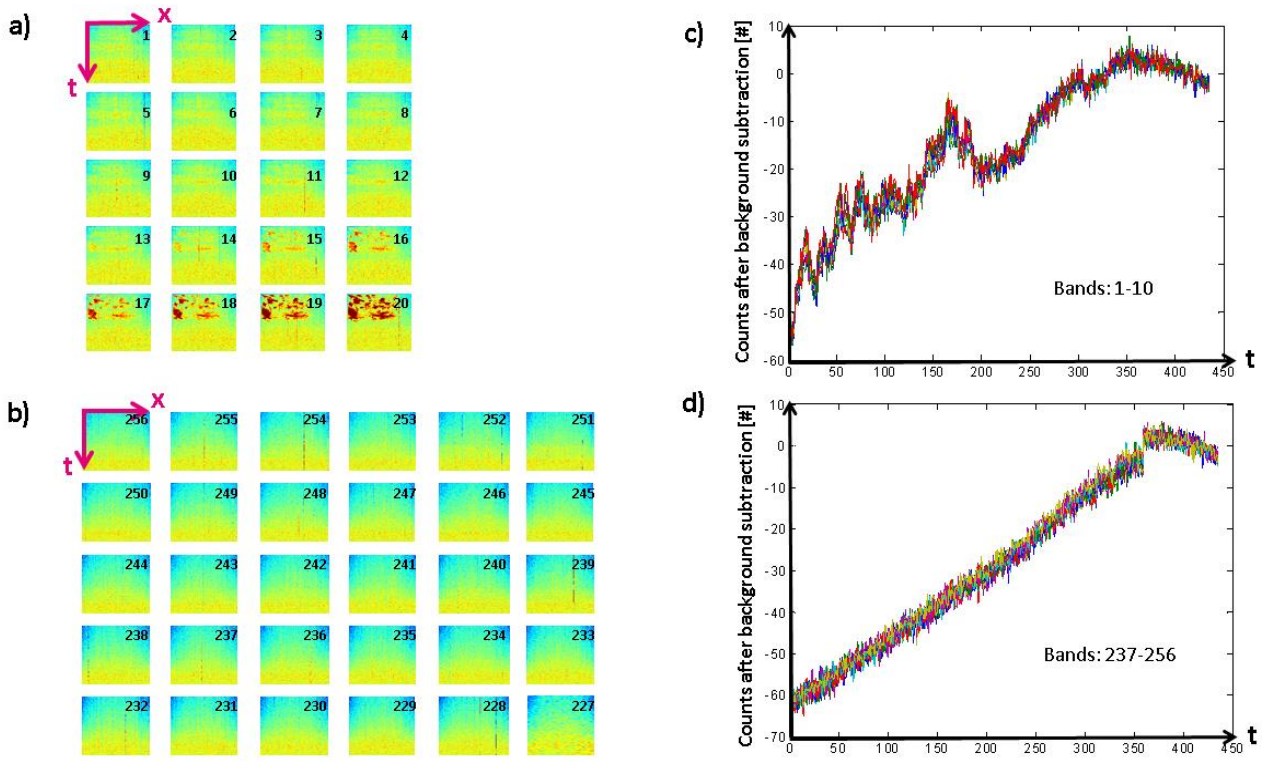


Figure 4.8: *a) Spatially resolved images of the 20 wavelength bands corresponding to the shortest wavelengths, from a typical hyperspectral measurement of a multicrystalline sample. The spatial dimensions are shown in pink, and the black numbers indicate the wavelength band number. b) Corresponding images for the 30 wavelength bands corresponding to the longest wavelengths from the same measurement. c) Median number of counts for all x-values in wavelength bands 1-10 as function of time. d) Corresponding plot for wavelength bands 237-256. Note that t is given in unit of pixels along t-dimension of the image and not in seconds.*

From part a) of the figure, which shows images of the 20 shortest wavelength bands, it is clear that the camera detects signals that are not related to background noise at least from the 10th wavelength band and upwards. Part c) reveals that also the 10 first bands receive relatively strong signals, making them unsuited as reference area. Parts b) and c) in the figure indicates, however, that the longest wavelength band are more or less uninfluenced by incoming light. The observed gradient in the number of counts along the t-dimension is probably caused by the changing background noise level. Therefore, the last wavelength bands can be used as a reference area for the background noise in the t-dimension. For further analyses, it was decided to use the 20 longest wavelength bands (Bands 237-256), and the x-values from 70 to 250 as reference area. As will be explained below (see section 4.4.2), the edge regions of the measurement ($x < 70$ and $x > 250$) have other characteristics which make them unsuited as reference area. The reference area is illustrated in Figure 4.9.

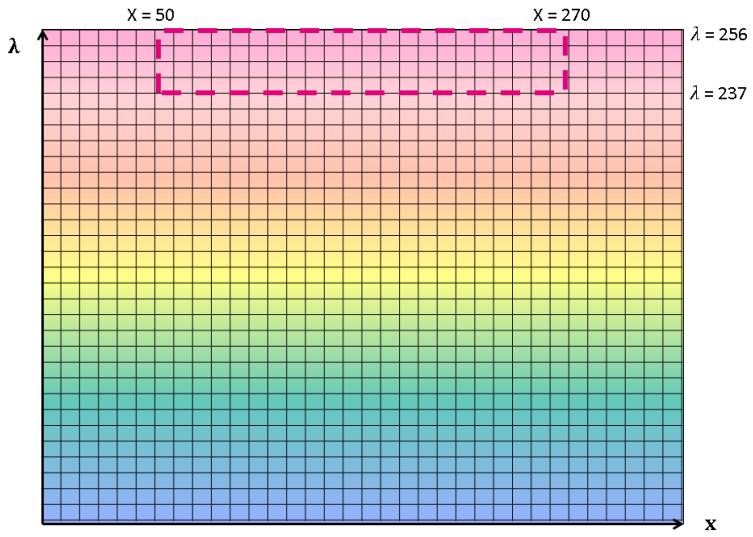


Figure 4.9: *Illustration of the reference area (marked with the dashed rectangle) of the camera chip that was used to determine the background noise level of the entire chip.*

4.4.2 Spatial variation of the background noise level on the camera chip

It needs to be addressed whether the variations in background noise level are the same for every location on the camera chip. For this purpose, more spatially resolved images of the background noise on the chip, similar to Figure 4.7, were analysed. The images were selected from frames originating from different points in background oscillation cycle (i.e. different values of *t* in Figure 4.6). It was found, that along the edges of the chip and in its corners, the amplitude of the background oscillation attained its maximum value of about 120 counts. In the central area, however, the amplitude of the background change was about 60 counts.

Based on the difference between the minimum and the maximum possible value for the background noise level (corresponding to e.g. $t = 20\ s$ and $t = 55\ s$ in Figure 4.6), a matrix with the same dimensions as the camera chip was constructed. For each camera pixel (x_i, λ_j), the corresponding pixel (i, j) in the matrix gives the difference of the mean number of counts for that pixel at the highest and lowest possible background noise level. In the following, this matrix is referred to as the *background correction matrix (**BCM**)*. The median of the numbers in the area of **BCM** corresponding to the reference area of the frames is referred to as the *background correction matrix reference (BCMR)*. Note that both the matrix **BCM** and the number BCMR are assumed to be constant; they depend neither on the frame t nor on the measurement to be studied. A visualisation of the background correction matrix is shown in Figure 4.10.

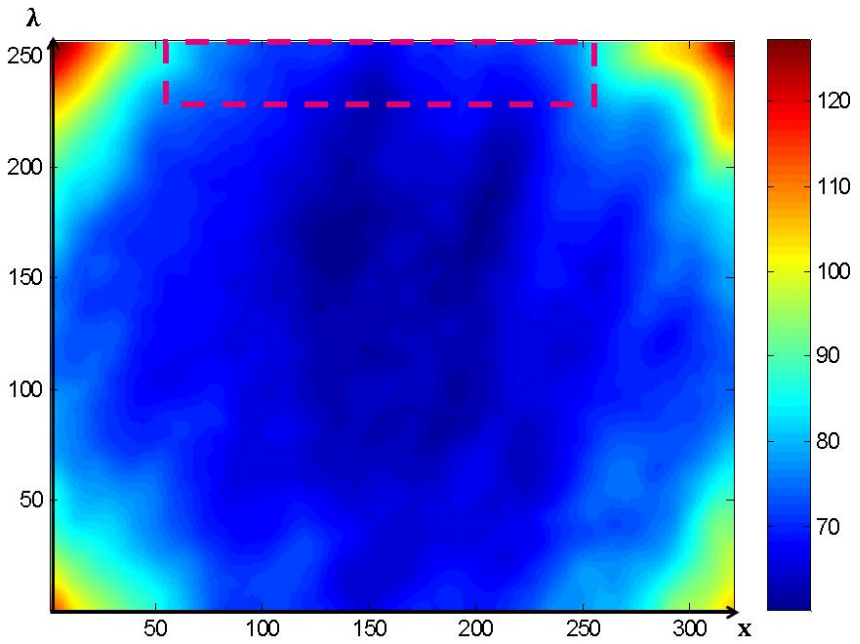


Figure 4.10: *Visualisation of the background correction matrix (**BCM**). Each pixel gives the difference between the highest and the lowest background noise level for the corresponding point on the camera chip. The dashed rectangle indicates the reference area of the matrix. The median of the numbers in this area are referred to as background correction matrix reference (BCMR).*

4.4.3 Working principle of the background correction

Based on the reference area and the background correction matrix, a much more accurate background subtraction than the one explained in section 4, can be performed. The working principle of the improved background correction scheme can be summarised in the following steps:

I) For each frame t, the median number of counts in the reference area of the frame is calculated.
II) The *dark reference* (DR) of the measurement is found as the median number of counts in the corresponding area of the dark part of the image (the part that originally was used to find the background for the entire image. See section 4.). Note that this number is a constant, and does not depend on t. It will, however, vary from one measurement to another.
III) For each frame, the DR is subtracted from the median number of counts in the reference area. The resulting number is denoted $Ref(t)$.
IV) The normal background subtraction (as explained in section 4) is performed.
V) We now, for each pixel (x, λ) in each frame t, want to find the "*thermal background*" $TB(t, x, \lambda)$ that additionally must be subtracted from the raw count value for that pixel due to the oscillating background level. By assuming that the ratio of the thermal background TB for one value of t to the BCM (which, loosely spoken, is the maximum value of the thermal background) is constant for the entire camera chip, we can claim the following equation to be true:

$$\frac{Ref(t)}{BMCR} = \frac{TB(t,x,\lambda)}{BMC(x,\lambda)} \qquad \textbf{(4.4)}$$

VI) Since $Ref(t)$, $BMCR$ and $BMC(x,\lambda)$ are known, the thermal background value $TB(t,x,\lambda)$ for each pixel can thus easily be calculated as

$$TB(t,x,\lambda) = \frac{Ref(t)}{BMCR} \cdot BMC(x,\lambda) \qquad \textbf{(4.5)}$$

and be subtracted from the corresponding pixel value.

The Matlab script performing this background correction can be found in Appendix A.

4.4.4 Comparison of images after the background correction

After correcting the background, measurements of the same wafer with equal measurement parameters (temperature, laser excitation and integration time) were again compared. From Figure 4.11, it is clear that the reproducibility of the measurements has increased drastically after implementing the new background subtraction method. The fact that the images after the background subtraction seem to be similar, also indicates that the assumption done under point V in the description of the background correction procedure above, seems to hold true for these measurements.

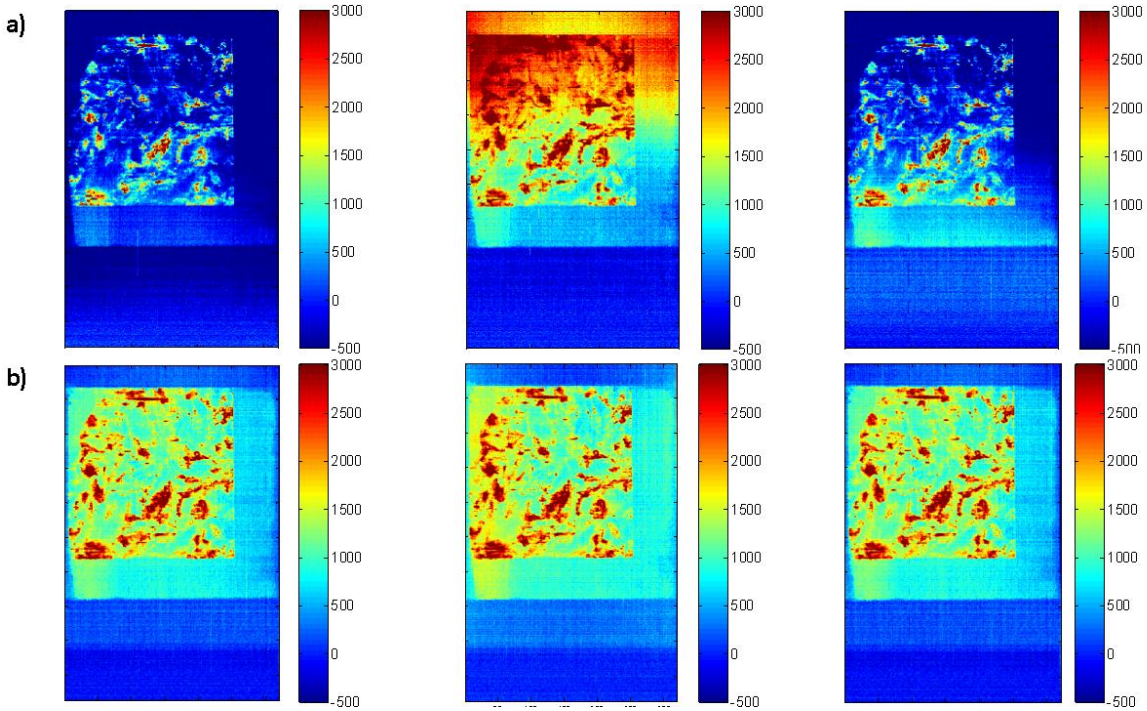


Figure 4.11: *Subband PL images (0,73 – 0,92 eV) from three subsequent measurements of the same wafer. In a) the standard background subtraction is used, and in b)the new background subtraction scheme developed as a part of this work is applied. Note that, also in b), some minor differences between the images are seen.*

When comparing the three images in Figure 4.11 b, one observes that some minor differences between the three images are still present. Most visible is the fact that both edges of the central

measurement shows more counts than the two other measurements do at their edges. Taking this into account, one still has to take care when drawing conclusions based on comparisons of the number of counts for measurements of different wafers. However, when comparing the images in Figure 4.11a (without the new background correction regime), to those in Figure 4.11b (with the new background correction regime) it is clear that the possibilities for comparing DRL from various wafers, which is a necessity for the rest of this work, is dramatically enhanced.

4.5 VARIOUS METHODS OF ANALYSING MULTISPECTRAL DATA AND THEIR RELIABILITY

The information in a hypercube representing a multispectral measurement from which the background is subtracted can be analysed and presented in different ways. In this subchapter, the data analysis methods used in this work are shortly presented, and their reliabilities are discussed taking into account the influence of the inhomogeneous laser illumination.

4.5.1 Mean spectrum of entire wafer

Description

One of the simplest ways of extracting information from a hypercube is by plotting and studying its *mean spectrum* (see e.g. Figure 5.1). By mean spectrum it is here meant that for each wavelength band of the hyperspectral image, a mean of the pixel values in the *xt*-plane is found. These mean values are plotted against an axis representing the energy of each wavelength band.

Reliability and disadvantages

Due to the inhomogeneous laser illumination (see Figure 3.12), some parts of a measured wafer will be more intensely illuminated than other parts. When taking a mean *spatially* across the entire wafer, the contribution of the strongly illuminated regions to the mean spectrum will be larger than the contribution of the less illuminated regions. Therefore, conclusions should not be drawn from such spectra alone. For neighbouring wafers, the spatial distributions of crystal imperfections such as dislocation clusters, grain boundaries and contaminations, are very similar. This means that the regions being strongly illuminated in one sample, exactly corresponds to those being strongly illuminated in another sample. Mean spectra originating from such measurements are therefore comparable.

For wafers from different ingot positions or from different ingots, however, the crystal structure differ. This means that in one sample a region affected e.g. by dislocation clusters or high contamination concentration is strongly illuminated, whereas, in another sample the strongly illuminated region is of high material quality. Such differences may affect the mean spectra of the wafers. Care must therefore be taken when comparing mean spectra from wafers that do not have the same crystal structure.

Saturation and bleeding effects cause a severe challenge when studying mean spectra of entire wafers. In many of the measurements done with high excitation laser voltages, saturation of the camera chip pixels corresponding to the band-to-band emission peak has occurred. Saturation of a pixel (or a pixel region) of the camera chip probably causes lateral diffusion of charge carriers

(called bleeding) to neighbouring camera pixels, thus causing artificially high count values in the neighbouring pixels. In the case of a hyperspectral camera, the two spatial dimensions of the camera chip (x *and* λ) correspond to one spatial dimension of the sample and the energy dimension respectively. Therefore, a saturated camera pixel will "bleed" not only to spatially neighbouring areas on the wafer, but also to neighbouring energy values. For the measurements done in this work, bleeding from the band-to-band emission line to its "spectrally neighbouring" emission line, D4, is most often observed.

When studying a mean spectrum for a large area (e.g. an entire wafer), bleeding effects between neighbouring peaks can occur *also when a local minimum between the peaks is seen*. Despite this local minimum, which emerges from the part of the image in which bleeding has not occurred, bleeding in other regions of the image may increase the intensity of the D4 peak in the mean spectrum. The challenges related to bleeding can, to a large extent, be omitted by keeping the injection at a low level, for example by reducing the excitation laser intensity, or by analysing as-cut wafers instead of passivated wafers. Because of the high surface recombination rates of as-cut wafers, their injection level, and thus the intensity of the band-to-band emission line will be low. Since the injection dependence of the D line emissions are lower than that of the band-to-band emission (see section 5.2.3), they will still be detectable.

The *strong tails* of some of the emission signals, causing signals from various emission processes to partly overlap spectrally, constitutes another challenge when studying the mean spectra of wafers. Due to such strong tails, it is in some cases hard to recognise one signal from another. Moreover, signals at close energies may appear more intense than they should, due to spectral overlap with other signals. Such problems are by many authors (e.g. [7, 38]) solved by using Gaussian fit algorithms to resolve individual peaks. Such techniques could probably have been used for the data analysed here, but are not included in this work.

4.5.2 Spectra from wafer regions

Description

In exactly the same way as described for entire wafers in the previous section, a mean spectrum for a *wafer region* can be found.

Reliability and disadvantages

As opposed to spectra from entire wafers, spectra from parts of wafers do not have the advantage that errors due to laser inhomogeneity are eliminated by averaging. If spectra from very differently illuminated regions (Figure 4.12a) are compared, this may cause false conclusions to be drawn. If, however, the regions over which a mean is taken are chosen from regions orthogonal to the scanning direction and extending over the entire width of the wafer (Figure 4.12 b), the effects of the inhomogeneous laser profile will be eliminated, allowing for reliable comparison between regions.

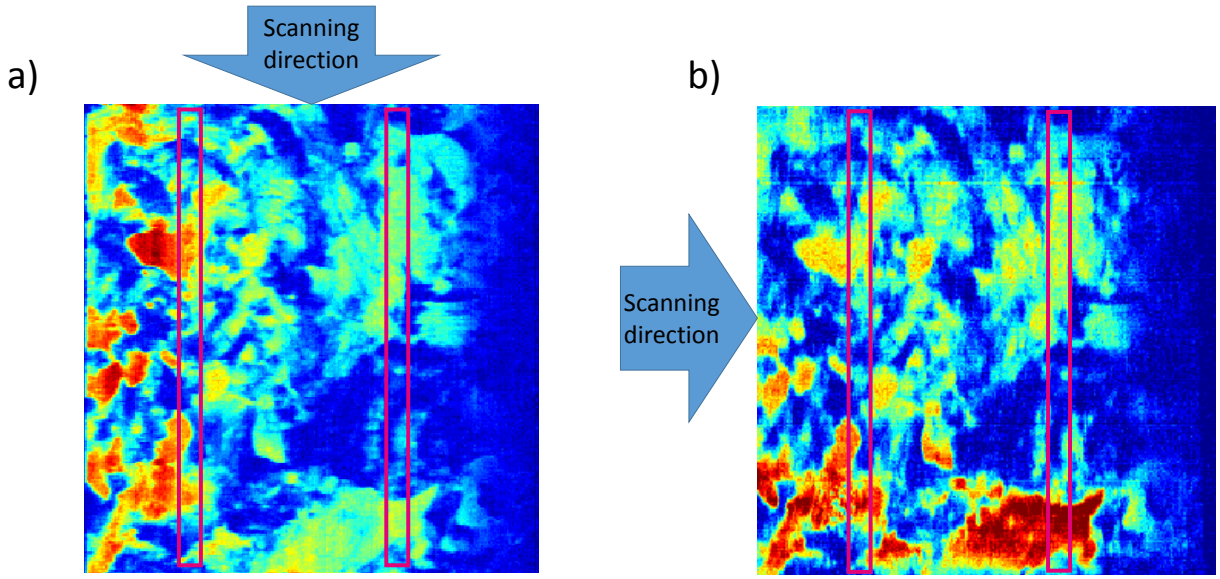


Figure 4.12: *Illustration of regions from which a) mean spectra are not comparable due to the inhomogeneous laser profile and b) mean spectra are comparable despite of the inhomogeneous laser profile.*

The challenges related to saturation and bleeding, that occur for spectra from entire wafers, can to a large extent be avoided when spectra from smaller regions are compared. This can be done by choosing to study the spectra for spectral regions in which no saturation has taken place. Moreover, from wafer with different crystal structures, regions with similar density of defects and similar laser illumination intensity can be localized and compared.

4.5.3 Spatially resolved images integrated over certain wavelength bands

Description

Based on theory, and on features observed in wafer spectra, certain interesting wavelength bands can be defined, e.g. corresponding to the band-to-band emission and the four D line emissions. Further, the measurement signal can be integrated over these wavelength bands, thus creating spatially resolved images of the emission lines on the wafer. Figure 4.13 illustrates the wavelength bands over which spatially resolved images of the D lines as well as the band-to-band emission line (BB) in this work standardly are integrated. The line "D0" will be further discussed in section 5.2.1. Examples of integrated images can be seen for example in Figure 5.3.

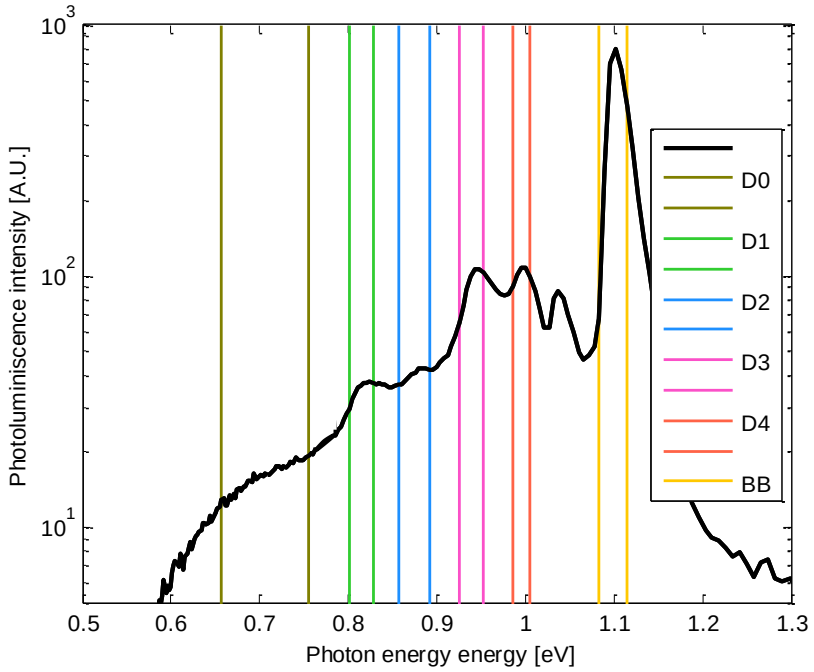


Figure 4.13: *Illustration of the wavelength bands over which the D lines and the band-to-band emission (BB) standardly are integrated.*

Table 4.1 gives an overview of the limits between which the measured signal are integrated in order to obtain spatially resolved images of the D lines, as well as the energy corresponding to the peak maximum. In the following, when the terms D0, D1, D2, D3, D4 and BB are used, it refers to the signal found between the integration limits given here.

Table 4.1: *Overview of the D lines as well as the BB emission line and the energy intervals over which it in this work standardly is integrated to obtain spatially resolved images of the emission lines.*

Name of emission line *Used in this work*	Lower integration limit *Used in this work* [eV]	Upper integration limit *Used in this work* [eV]	Peak maximum *Based on literature* [eV]	Ref.
D0	0.66	0.75	0.75	[35],[47]
D1	0.80	0.83	0.812	[36]
D2	0.86	0.89	0.875	[36]
D3	0.93	0.95	0.934	[36]
D4	0.99	1.01	1.000	[36]
BB	1.08	1.11	1.15	[14]

Reliability and disadvantages

Integrated images give valuable information on the spatial distribution of various emission lines on wafers. Due to the inhomogeneous laser distribution, two locations within one integrated image cannot necessarily be directly compared. Comparisons can, however, reliably be

conducted between corresponding spatial regions at images of different wafers, as long as their scanning directions do not differ (cf. Figure 4.12).

As in the case of mean spectra of entire wafers, *saturation and bleeding* cause challenges also when comparing images integrated over certain wavelengths. Also in this case, the band-to-band emission line tends to cause saturation, and thus “bleed” into the D4 peak, causing an improper increase of the D4 signal in regions with high BB signal. If it is desired to make an exact analysis of the intensity of e.g. D4, one should therefore take care to use only images (or parts of images) without saturation of the BB emission line. Also in this case, analysing as-cut wafers helps omitting the problems related to saturation.

4.5.4 Analysis by MCR

Description

Multivariate Curve Resolution (MCR) is a useful way of extracting information from a multispectral measurement. The theoretical background and working principle of this method are explained in section 4.1. Since the MCR algorithm needs a matrix (two-dimensional) as input, and the hyperspectral data is organized as a three-dimensional tensor, the hypercube must firstly be reshaped to a matrix. After the analysis, the resulting concentration matrix **C** (c.f. equation (4.1)) is reshaped back to the original spatial dimensions of the analysed hypercube.

The results of an MCR analysis can by visually inspected by looking at its *loading plot* and its *score plots*, meaning a plot of the vectors of S^t and images showing the values of each "plane" in C respectively (c.f. equation(4.1)). Each score plot corresponds to one of the components in the loading plot (one of the vectors of S^t), and shows from which areas of the wafer emission at the corresponding energy level was detected. Figure 4.14 shows an example of a loading plot and some corresponding score plots from an MCR-analysis.

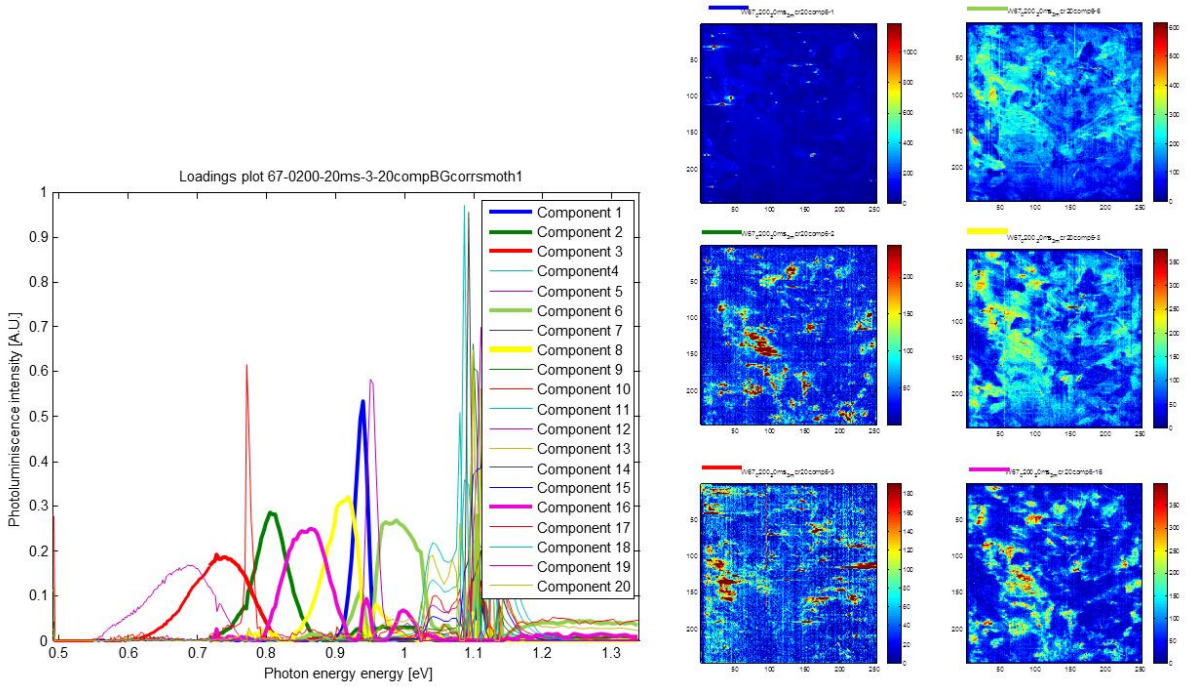


Figure 4.14: *Example of an MCR-analysis. To the left, the loading plot is shown. Some of the components are highlighted. To the right, the score plots corresponding to the highlighted components are shown. The coloured bars in the upper left corners of the score plots indicate to which component in the loading plot they correspond.*

Reliability and disadvantages

The problems related to saturation and bleeding and to spectrally overlapping signals that are described for the other analyse methods, will to a large extent be avoided by the use of MCR. This is possible, since the MCR algorithm takes the spatial and the spectral dimension into account simultaneously, rather than only the spectral dimension (such as a mean spectrum), or only the spatial dimensions (such as integrated images).

Comparing various hyperspectral measurements through the use of MCR has, however, proven to be challenging. This is, among other reasons, due to a confusion regarding the number of components that are to be extracted by the MCR algorithm. This number can be user defined to any integer from 1 to 20. For measurements in which the pixels corresponding to the BB peak are saturated, the MCR algorithm tends to split the band-to-band emission line in a variety of components, rather than extracting the DRL signals. Thus, it is necessary to preselect a high component number in order to find the desired DRL signals. On the other hand, when a high number of components are preselected, some of the DRL signals are – in some cases – split in more components. It is not clear whether the splitting in more signals always is an analysis artefact caused by the high sensitivity of MCR, or if it relates to real physical effects.

Due to these uncertainties, it is in this work decided *not* to use MCR for the comparison of different wafers. MCR may however be applied in order to find the spectral and spatial distribution of signals within one wafer. For such purposes MCR is a useful supplement to the extraction of images by integration. The results of the two methods should however be critically compared.

When comparing results from the two methods, one should keep in mind that MCR strives to extract components causing *variation* in the analysed data material. Thus, if there is a certain background level stretching spatially over the entire wafer or spectrally over all wavelengths, MCR will quantify components regarding to their deviation from this background level. When integrating over a certain wavelength interval, however, also the background level for each wavelength bond will contribute to the signal. This will not only cause a value different from zero in areas without a real signal in the certain wavelength interval, but also an increase of the total number of counts in signal-rich regions. Since the issue of subtracting the correct background level for all spatial positions and all wavelengths is challenging, the integrated images of the D lines may suffer from small errors regarding the background level. This, in turn, may cause differences between the images found by MCR and those found by integration.

5 RESULTS AND DISCUSSION

The motivation of this work was to investigate the difference between DRL in n- and p-type silicon, the change of DLR along the length of an ingot, the impact of metal impurities on DLR and the change of DLR due to solar cell processing steps. Accordingly, in this chapter, one subchapter is dedicated to each of these topics. At the end of each subchapter, the main findings related to the topic in question are shortly summarized. Thereafter, these results are discussed in relation to literature, and attempts to explain the observed effects physically are made.

Before the results are presented, however, some general features of the measured luminescence spectra will be described.

5.1 GENERAL FEATURES OF THE MEASURED LUMINESCENCE SPECTRA

Figure 5.1 shows a typical luminescence spectrum from one of the multicrystalline wafers studied, expressed as spatial mean of the entire wafer, as explained in section 4.5.1. In the figure, some of the most characteristic features of the recorded luminescence spectra are highlighted.

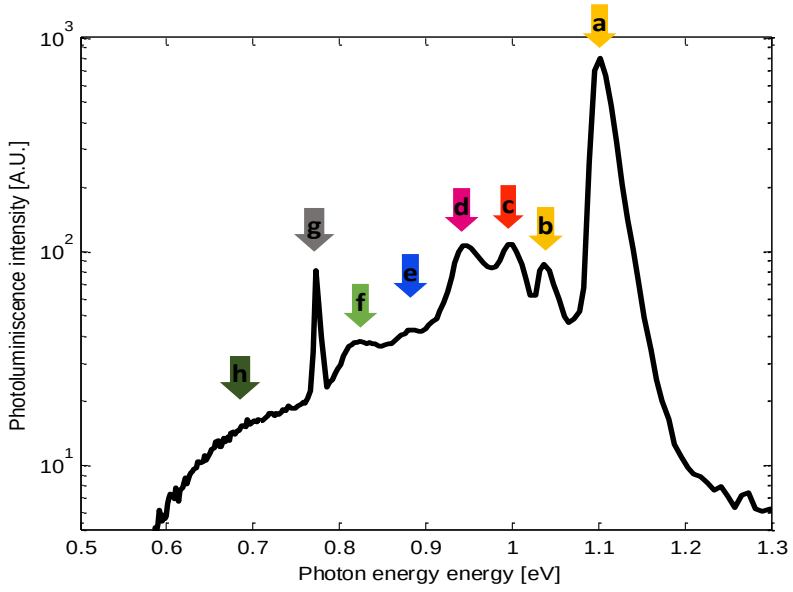


Figure 5.1: *Typical luminescence spectrum with highlighted emission lines and main features: a) band-to-band (BB)emission line, b) Phonon replica of BB, c) D4 emission line, d) D3 emission line, e) D2 emission line, f) D1 emission line, g) signal caused by the 2nd order diffraction maximum of the laser reflection, h) emission line at approximately 0.7 eV, in this work denoted D0.*

At the right side of the plot a very clear band-to-band (BB) emission line (a) is seen. On its low-energy side, a less intense peak (b) follows. This is a phonon replica (see section 2.6.1) of the BB-transition. Further, D4 (c) and D3 (d) are clearly visible. D2 (e) appears here only as a weak signal, whereas D1 (f) appear as a somewhat more distinct peak. At its low energy side, a distinct signal (g) appears. This very sharp peak is due to the 2nd order diffraction maximum of the laser

light, as explained in section 3.5.1, and is therefore not a part of the photoluminescence spectrum of the sample. Further, a broad peak (h) with its maximum close to 0.7eV is seen. This signal, which in this plot barely is visible, but in other plots is a clear peak (see e.g. Figure 5.4), is in this work denoted D0. A similar signal has been described also by Flø et al. [35] and Burud et al.[47].

Properties that may influence the luminescence signal of a sample

Various physical properties may influence the DLR as well as the BB luminescence of a wafer. In section 4.5, effects that *falsely* cause apparent changes to a measured spectrum or to spatially resolved images were explained. In the following, some properties that may cause actual changes to the emission intensities of the D lines as well as the BB emission line are mentioned.

As explained in section 2.3.1, under LLI conditions, the radiative band-to-band emission of a sample is proportional to the excess carrier density Δn as well as to the net doping concentration N. Accordingly, when comparing the luminescence spectra of two wafers with different doping concentrations, if all other parameters are equal, one expects a more intense BB emission line in the sample with higher doping concentration. Further, if the lifetime, and accordingly the injection level for a sample is high, one expects a strong BB emission line. For wafers with equal doping concentrations, one can assume that the height of the BB peak is an indication for the charge carrier density. This statement, however, holds true only when no saturation has occurred in the pixels corresponding to this peak.

Based on the equations in section 2.3.3, it is clear that also the rates of recombination due to defects depend on the injection level. If we assume that the D line emissions are caused by defects that can be described by the SRH-theory, we must therefore take into account that also the intensities of these emissions are impacted by the injection level, and accordingly by the lifetime of a sample. It is not known whether all the D lines have the same injection dependence and if this injection dependence equal that of the BB emission line. Therefore, the shape of the luminescence spectra of a sample may vary due to varying injection levels.

The net doping concentration of a sample influences its total concentration of free charge carriers (c.f. equations (2.8)), which, in turn, influences the recombination rate of a certain defect level (c.f. equation (2.36)). Therefore, one cannot exclude that also emission intensities of the D lines, such as the BB emission intensity, depend on the net doping concentration.

5.2 Comparison of defect luminescence in MC p- and n-type silicon

The study in which the D lines firstly were mentioned [36], was based on n-type silicon. Except from that, however, DRL is until now mostly studied in p-type silicon (e.g.[2, 3, 33, 35, 38]). To the best of our knowledge, no thorough comparison of defect luminescence in n- and p-type material has been done. Such a comparison is therefore of interest.

This subchapter starts with a comparison of the defect luminescence signal of an n-type and a p-type wafer from the npn-ingot and a comparison with lifetime images of the same samples. The fact that these are *neighbouring* wafers from the *same* crystal ingot, allows for particularly good possibilities for a comparison of the two material types, spatially as well as spectrally. Because of the special characteristics of this ingot, due to its compensation (see sections 2.1.2 and 3.1.2),

samples from the n-twin and the p-twin ingots are also compared. Moreover, the dependence of the luminescence spectra on injection level is investigated.

5.2.1 Spatial comparison of DRL of n- and p- wafers from the npn-ingot and its relation to lifetime

Since defects, such as dislocation clusters and grain boundaries, from which DRL may originate, can be detrimental for charge carrier lifetime, a spatial comparison of DRL with lifetime is of worth, not only for the comparison between DRL in n- and p-type material, but also for the understanding of DRL and is origins as such.

Figure 5.2 shows images of lifetime and corresponding injection level in the investigated n-type (a, c) and p-type (b, d) sample from the npn-ingot. The images are made by photoluminescence imaging calibrated by modulated PL (see section 3.3). Dislocation clusters and grain boundaries appear as darker regions and lines. These features can mostly be found at the same location in both samples, illustrating that they have the same crystal grain distribution. A region of reduced lifetime due to solid-state in-diffusion of impurities from the crucible into the crystal is clearly seen as a dark area at the right edge of both wafers. The three dark, circular spots observed on both wafers (most clearly on the p-type wafer) are due to additional contamination with iron during the solidification process.

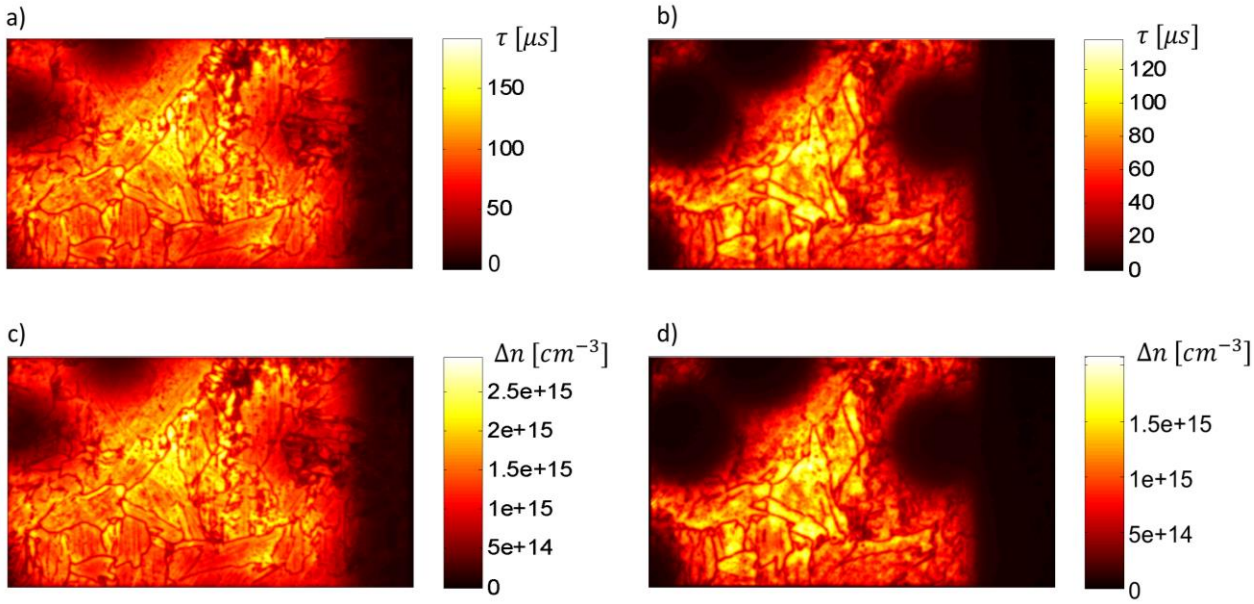


Figure 5.2: *Lifetime and* $\boldsymbol{\Delta n}$ *maps of wafers from the npn ingot based on photoluminescence imaging at 1sun illumination, calibrated by modulated PL. a) Lifetime map of n-type wafer, b) Lifetime map of p-type wafer, c) Map of* $\boldsymbol{\Delta n}$ *in n-type wafer d) Map of* $\boldsymbol{\Delta n}$ *in p-type wafer. Note that the maps are differently scaled. Figure reprinted from[49].*

In the high-lifetime regions of the samples, both lifetime and injection level is 25-50% higher in the n-type than in the p-type wafer. In the in-diffusion region, however, the difference between the two samples is even larger, illustrating the large difference in impact of metal contamination on n-type and p-type material (section 2.3.3). A more accurate discussion of the measurement results related to these wafers, can be found in [49].

In Figure 5.3, spatially resolved images of each of the defect emission lines, as well as the band-to-band emission line of these samples are shown. The images are made by integrating the luminescence signal over the wavelengths given in Table 4.1.

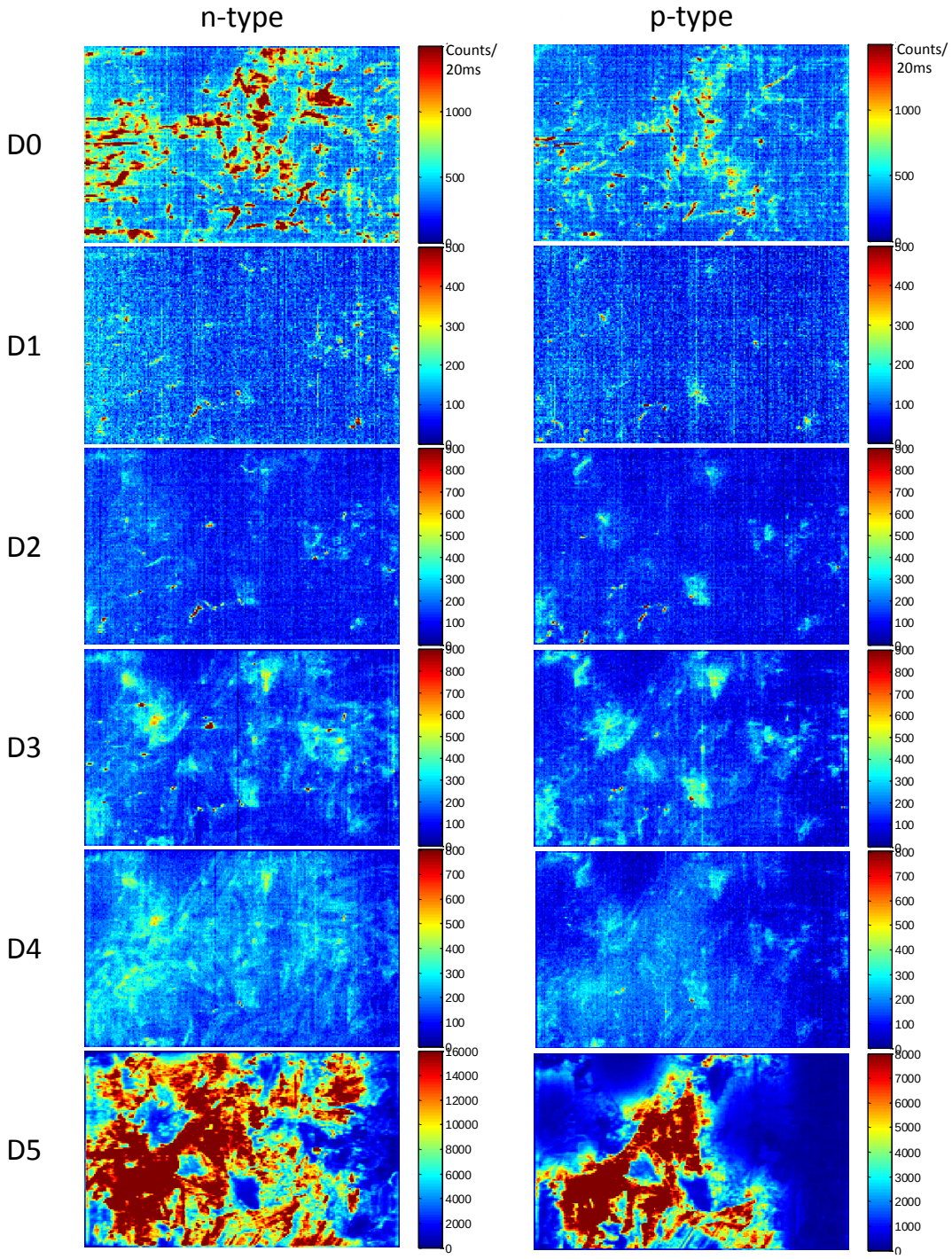


Figure 5.3: *Comparison of the spatial distribution of the D lines in an n-type and a p-type sample from the npn ingot. The wavelengths over which it is integrated are given in Table 4.1. The colorbars give the numbers of counts per 20ms integrated over the given wavelength bands. Note that for the images of the BB emission, the colormaps are adjusted slightly differently, in order to enhance the contrasts in the image.*

When comparing these images to the lifetime maps in Figure 5.2, one should keep in mind that the hyperspectral measurements are done at cryogenic temperatures, whereas the lifetime images are obtained at room temperature. Due to the dependence of recombination activity on temperature, explained in section 2.4.4, this cause differences in the distribution of the PL signal. Some other features found from a visual comparison of the D line emissions originating from the p-type and the n-type samples in Figure 5.3, and the lifetime maps in Figure 5.2, are mentioned below.

All the D line emissions seem to have similar distributions in the n-type and the p-type sample. Further, one notes that within one wafer, D1 and D2 mostly occur in the same regions. D3 and D4 also partly seem to originate from the same regions. In the images of D4, however, bleeding effects (see section 4.5.1) from the BB emission line are seen, causing apparent differences between the D3 and the D4 emissions. Except from the regions in which the D4 signal attains artificially high values due to such effects, there is a tendency for the D-line emissions D1-D4 to be anticorrelated to the BB emission. Comparison with Figure 5.2 suggests that the D line emissions mostly originate from regions that have low lifetime due to dislocation clusters and grain boundaries. As opposed to the other D line emissions, the D0 emission seems to originate mostly from regions with high lifetime. All emission lines – and especially D0 and D4 – seem to be somewhat more intense in the n-type than in the p-type sample. This can probably be attributed to the higher injection level in the n-type sample compared to the p-type sample (see Figure 5.2).

5.2.2 Spectral comparison of DRL of wafers from the npn-ingot

A further comparison of the DRL in the two samples can be done by investigating their spectra. Since the crystal structure of the two wafers are very similar (see Figure 5.2), such a comparison can be done without large impacts of the laser inhomogeneity.

Figure 5.4 shows a comparison of mean luminescence spectra from the p-type and n-type wafers from the npn-ingot. The coloured lines in part b) of the figure highlight the wavelength intervals in which the D lines, according to literature, should be found (see Table 4.1).

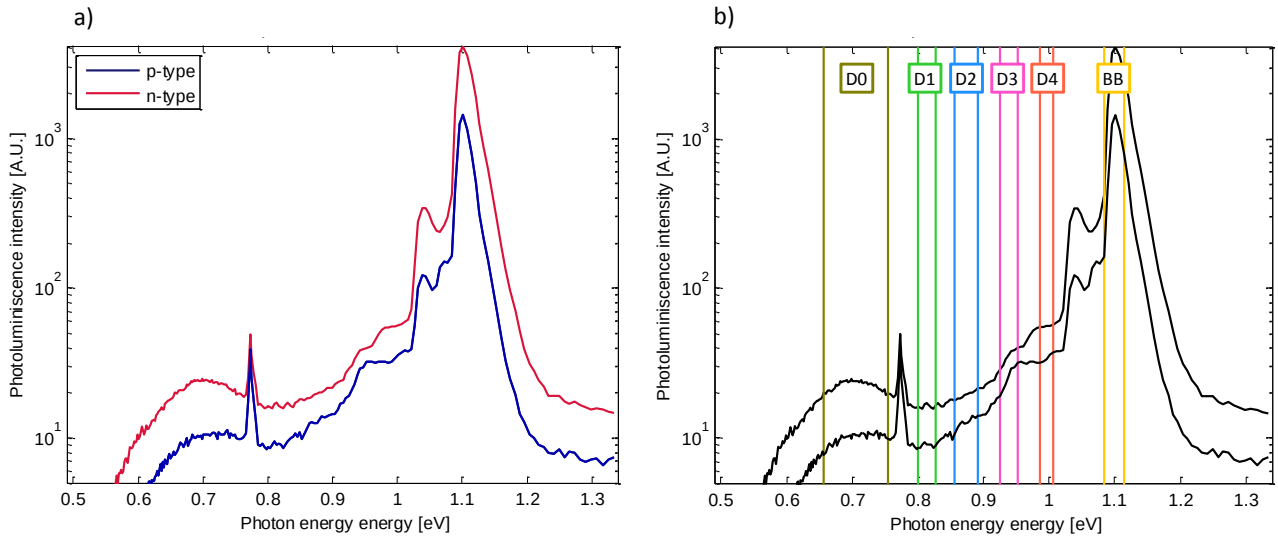


Figure 5.4: *a) Comparison of luminescence spectra from n-type (red) and p-type (blue) wafer. b) Visualisation of the wavelength intervals in which – according to literature – the D lines emission should be found overlaid on the two spectra in a.*

In both plots, the BB peak as well as phonon replicas are seen. The n-type spectrum has only one phonon replica, whereas two replicas are visible in the p-type spectrum. D3 and D4 are seen as shoulders of the wide band-to-band peak, whereas D1 and D2 for these plots hardly are visible. The D0 signal is seen as broad peak in both spectra.

When comparing the two spectra, one observes two important differences. Firstly, the n-type spectrum is for all wavelengths more intense than the p-type spectrum. Secondly, close to the value of D3, the difference in intensity between the two spectra is smaller than for other wavelengths. For D4 however, the difference between the two spectra seems to be larger than for other intensities. These two features will be discussed in the following.

Higher PL intensity in the n-type than in the p-type sample explained by lifetime differences

As explained in section 2.3.3, iron, as well as most other transition metals, has larger capture cross sections for electrons than for holes, meaning that their effect is more detrimental in p-type than in n-type silicon. Thus, when other material characteristics, such as crystalline structure and contamination level are similar, the charge carrier lifetime is mostly higher in n-type than in p-type material. The images of lifetime and Δn in Figure 5.2 show that the lifetime and accordingly the excess carrier concentration of the investigated n-type sample is significantly higher in that of the p-type sample. If we assume that the recombination rates related to the defects that cause the observed spectrum are dependent on injection level, the difference in lifetime and Δn seen in Figure 5.2 may explain the intensity difference between the n-type and the p-type spectra in Figure 5.4 as well as the differences between the D line images in Figure 5.3.

Further investigation of regions with high intensity of D3 and D4

The special features found close to the value of D3and D4 in the spectral comparison of the n- and p-type samples (Figure 5.4), and the large difference in the spatial images of D4 (Figure 5.3)

both suggest a difference between the DRL of the n-type and the p-type sample in the wavelength interval of D3 and D4.

In order to investigate this possible difference further, regions with high intensity of D3 and D4 were chosen, and their n- and p-type spectra compared. Spectra from some chosen regions are shown in Figure 5.5.

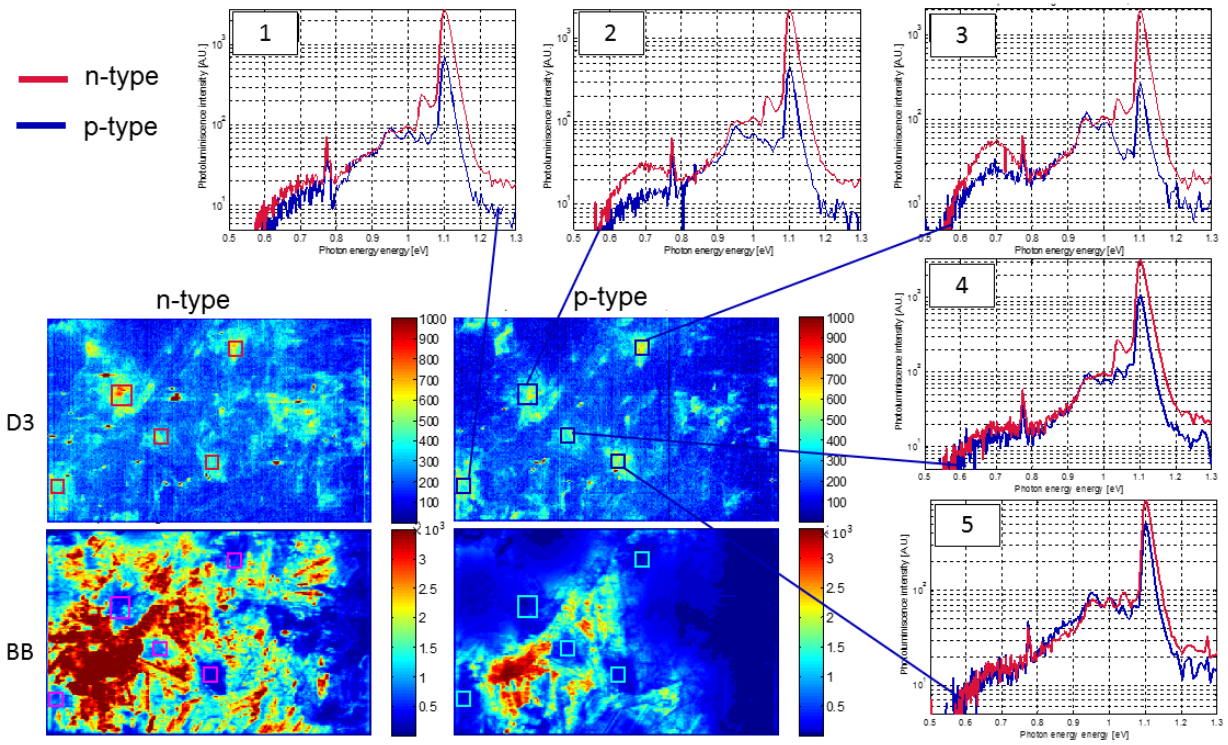


Figure 5.5: *Photoluminescence spectra of D3/D4 rich regions on p-type and n-type wafers from the npn-ingot. The blue quadrats highlight the regions at the p-type samples from which the spectra are collected, and the red quadrats highlight the corresponding regions on n-type sample. The regions from which the spectra are collected are also shown on images off the BB emission from both wafers. Note that the chosen D3/D4 rich regions always correspond to regions with low BB emission intensity. The colorbars give the numbers of counts in the images integrated over the energy intervals given in Table 4.1.*

In the chosen regions, both D3 (~0.93 eV) and D4 (~1.0 eV) can be recognized for both the n-type sample (red plots) and the p-type sample (blue plots). However, by inspecting the two peaks in all the spectra, one notes that in the n-type sample D3 and D4 seem to have approximately the same intensity, whereas D3 is more intense than D4 in the p-type sample. This corresponds to the findings in Figure 5.4 and Figure 5.3. One may therefore suspect that this is a general feature for these samples.

As explained in section 4.5.1, bleeding effects and effects of overlap from the strong tail of the BB emission line could possibly influence the shape of the mean spectra. Since the BB peak is stronger in the n-type than in the p-type sample, the intensity of the D4 emission line would – when affected by such effects – increase more in the n-type sample than in the p-type sample. This could give spectral shapes similar to those shown in Figure 5.5.

Further, as explained in section 5.1, the injection level may influence the spectral shape. Because the lifetime is lower in the p-type than in the n-type sample, one may suspect that the weaker D4 emission from the p-type than from the n-type wafer is a pure effect of injection differences, rather than having a relation to the different material type. If, for example, the injection dependence of the D4 emission is larger than that of D3 emission, it would cause spectral shapes similar to those shown in Figure 5.5. Thus, it should be studied whether the same difference between the intensity of the emission lines can be found for other injection levels.

The npn-ingot is, due to its compensation effects (see sections 2.1.2 and 3.1.2), a very special ingot, whose characteristics do not necessarily equal those of other ingots. Therefore, it needs to be addressed whether the differences between n- and p-type materials found in the npn-ingot also can be observed in ingots with uncompensated material.

5.2.3 Generalisation to other ingots and injection levels

In order to investigate whether the characteristics observed for samples from the npn-ingot hold true also for other samples and injection conditions, the DLR of wafers from central ingot positions of the n-type and the p-type twin ingots can be compared. Since the crystal structure of these wafers, as opposed to the wafers from the npn-ingot, differ, this structure should be investigated first. Figure 5.6 therefore displays lifetime images (made by photoluminescence imaging calibrated by modulated PL) of the wafer p-twin-central and n-twin-central.

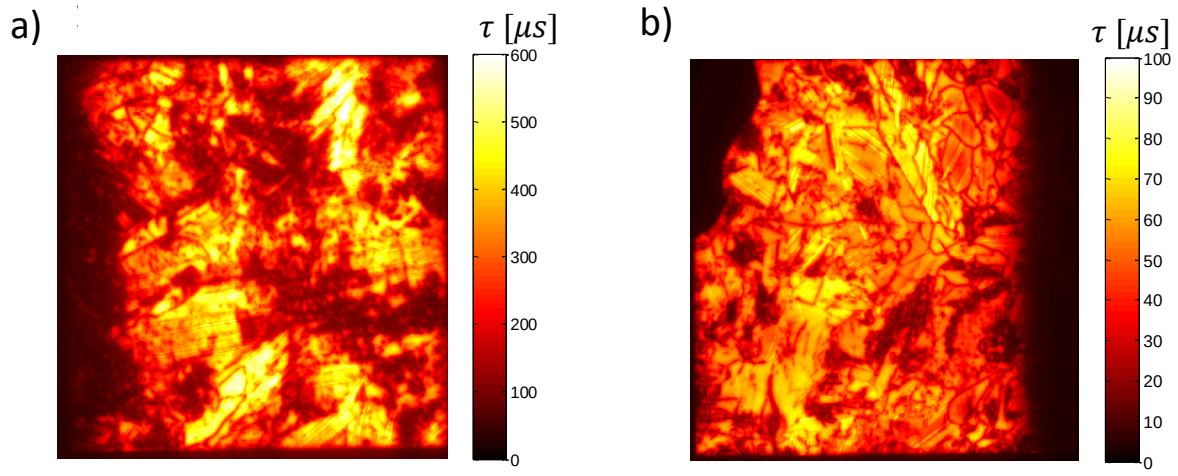


Figure 5.6**:** *Comparison of lifetime images of wafers n-twin central and p-twin central based on photoluminescence imaging at 1sun illumination, calibrated by modulated PL.*

From the images one can note a strong influence of solid-state inn-diffusion of impurities on the right side of the p-type sample, as well as similar but much less evident effect on the left edge of the n-type sample (see section 2.4.1 and 0)[7]. Moreover, one can see that of the n-type sample,

[7] As explained in section 4.5.1, it should be questioned whether the inhomogeneous laser spectrum influences the result when wafers with different crystal structure are imaged. The asymmetric distribution of iron between wafers from the n- and p- twin ingots is a feature that possibly could have caused such differences. However, as we will see later, if such an effect was present here, it would cause an effect which is opposite to the one we observe, regarding the spectral shapes of the two samples. Therefore, such effects are here considered negligible.

relatively large areas seem to be affected by dislocation clusters (seen as darker cloud-like areas and spots), whereas in the p-type sample, only smaller areas seem to be affected by such defects. These differences will be further discussed in section 5.2.5.

Figure 5.7 displays a spectral comparison of the wafers n-twin-central and p-twin-central. Since measurements done at various laser illumination intensities are included, also the injection dependence of the emission lines can be investigated.

In the figure, the difference in the intensity of D3 and D4, already observed for wafers from the npn-ingot, is recognised. Additionally, two peaks corresponding to the energies of D1 and D2 are clearly visible for the n-type wafer, whereas they hardly can be seen for the p-type wafer. As before, two phonon replicas are recognised in the p-type spectrum and only one in the n-type spectrum. The height difference of the BB peaks between the n-type and the p-type sample, as well as the change of the spectra at varying excitation voltages (i.e. at various injection levels), will be discussed in the following.

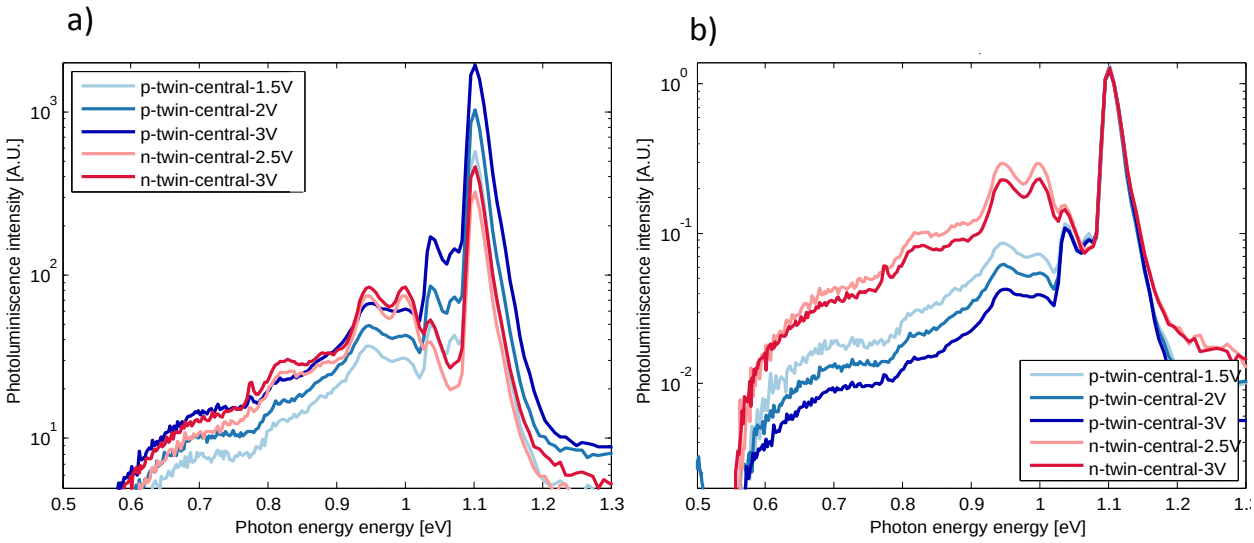


Figure 5.7: *Comparison of luminescence spectra for wafers n-twin-central and p-twin-central. The numbers following the sample name indicate the laser excitation voltages at which the measurements were done. Part a) shows the absolute intensities of the spectra, whereas part b) shows the spectra normalized to the height of the BB peak.*

Higher intensity of the BB peak in the n-type than the p-type sample explained by difference in doping concentration

Ackowledging that the lifetime of n-type material normally is higher than that of p-type material, if other characteristics are similar, one would expect the BB peak of wafer n-twin-central to be higher than that of wafer p-twin-central. That would correspond to the observed difference between the samples from the npn- ingot in Figure 5.4.

However, among the samples from the twin ingots, the doping concentration differ (see Table 3.1). As explained in section 5.1, the low doping concentration of the sample n-twin-central

compared to the sample p-twin-central explains that the BB peak is less intense in the n-type sample.

The fact that the BB peak now is less intense for the n-type than the p-type sample, weakens the suspicion that the differences between D3 and D4 in the n- and p-wafers are caused by bleeding or by overlapping signals. If the strong D4 signal in the p- type wafer of the npn-ingot (observed in Figure 5.5) were caused only by bleeding or overlapping from the strong BB peak, we would expect the same effect to take place for the twin samples, causing the wafer p-twin-central to get a more intense D4 peak than n-twin-central. Since this is not the case, the observed difference between D3 and D4 in the investigated n- and p-type wafers is probably related to material differences, rather than to these measurement artefacts.

Dependence on injection level

In Figure 5.7, when the excitation voltage (and consequently the injection level) is increased, the intensities of D3 and D4 seem to increase with the same rate. This suggests that the difference between the intensity of D3 and D4 in the p-type compared to the n-type sample is *not* caused only by differences in the injection dependence of the two peaks and differences in injection level between the two samples.

From part b) of Figure 5.7, in which the spectra are normalized to the height of the BB peak, one notes that for both material types (n and p), the intensity of all the D line emissions seem to increase less than the BB emission when the excitation intensity is increased. Therefore, the figure suggests that for the investigated injection levels, the D line emission rates are less injection dependent than the band-to-band emission rate.

5.2.4 Summary of section results

A spectral and spatial comparison of DRL from n- and p-type wafers gives the following results:

- The four D lines (D1-D4) are observed for the same wavelengths in n-type and p-type emission spectra. The emissions seem to originate from the same regions.
- For the investigated n-type wafers, the intensities of D3 and D4 are similar, whereas D3 is stronger than D4 in the investigated p-type wafers. The difference is probably not caused only by injection differences and difference in injection dependence.
- The intensities of D1 and D2 are slightly higher in the n-type than in the p-type samples.
- The injection dependence of the BB emission line seems to be higher than that of the D lines.

5.2.5 Discussion of section results

Since until now, no thorough comparison of the D-lines in n- ad p-type material has been done, the observation of the D lines at the same wafer regions for n- and p-samples with comparable crystal structure brings new information to the discussion of the D line emission, suggesting that their origins are independent of material type.

Discussion on the difference between D3 and D4 intensities based on SRH-statistics

The observed difference in the emission intensities of D3 and D4, between n-type and p-type material, suggests that the recombination mechanism related to D4 is more efficient in n-type

than in p-type material. Considering that holes are the minority charge carriers in n-type material, one therefore may hypothesise that the defect level related to this recombination mechanism is close to the valence band. Assuming that the probability of charge carrier transition decreases with increasing energy difference, for this defect level the probability for hole capture will be larger than the probability for electron capture. Since the number of majority carriers is much larger than the number of minority carriers, the capture rate for minority carriers will be limiting for the recombination process (cf. equation (2.26)). A trap level close to the valence band edge will therefore, theoretically, lead to a higher recombination rate in n-type than in p-type material when the injection level is unchanged.

Other factors that may cause differences between the investigated n- and p-type samples

One must note that between the investigated samples, not only the material types (n or p) differ. Other material characteristics such as doping level (see Table 3.1) and crystal structure are also different – at least among the samples from the twin ingots. Therefore, the observed differences are not necessarily related to material type only.

Even if the twin ingots are solidified in identical crucibles and under very similar process conditions (see section 3.1.1) some material differences between the two ingots exist. As seen in Figure 5.6, the sample n-twin- central is – to a larger extent than p-twin-central – influenced by dislocation clusters. These differences – which is not directly related to material type (n or p) – may cause differences in the DRL spectra of the two samples.

Therefore, before conclusions are drawn based on the observed differences between the DRL spectra of the samples n-twin-central and p-twin-central, it should be questioned whether the observed differences might be caused by other material characteristics than the doping type and the doping concentration as such.

5.3 VARIATION OF DEFECT LUMINESCENCE OVER THE INGOT HEIGHT

Due to dislocation multiplication during the solidification process (see section 2.4.2), the influence of dislocation clusters mostly increases along the growth direction of an ingot. Moreover, the concentration of metal impurities from the crucible, as well as the doping concentration is different at different ingot positions. Comparing the DRL spectra of wafers from different ingot positions may therefore reveal differences due to influence from these factors.

In this section, lifetime images of wafers from different ingot positions are compared in order to investigate their differences regarding crystal structure. Further, luminescence spectra of the wafers are compared and analyses regarding their dependence on injection level.

5.3.1 Comparison of crystal structure and lifetime at different ingot positions

To investigate differences in crystal structure over the height of an ingot, Figure 5.8 displays lifetime images from samples at various ingot positions of the p-twin ingot.

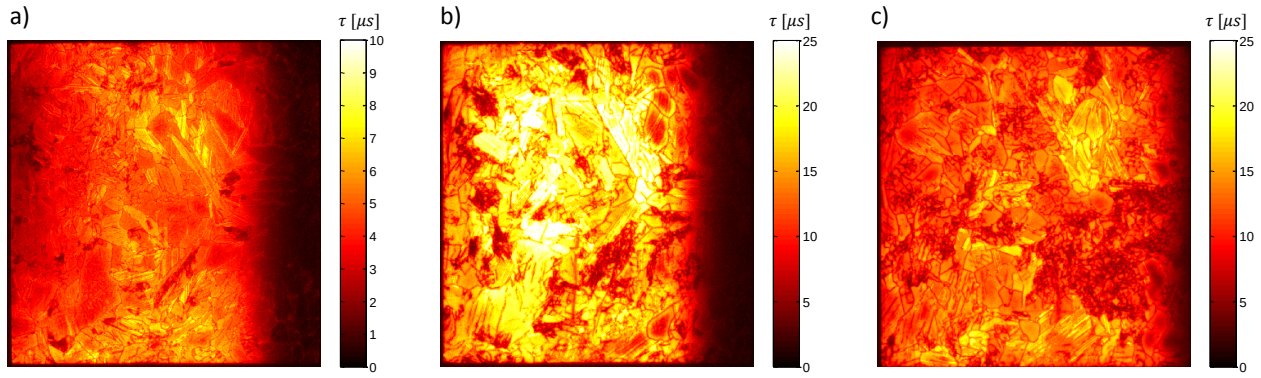


Figure 5.8: *Lifetime images, made by photoluminescence imaging calibrated by modulated PL, of wafers from a) bottom b) central and c) upper parts ingot positions of the p-twin ingot. Note that the image in a) is scaled differently compared to the other images. Note further that these measurements are obtained at illuminations of 1/100 sun, whereas the lifetime images in Figure 5.6 are obtained at illuminations of 1 sun. Reprinted from [Kleiber, 2014 #75].*

One notes, as expected, that the influence of dislocation clusters (seen as darker cloud-like spots) seem to increase towards the top of the ingot. Moreover, the inn-diffusion region at the right edge of the wafers decreases towards the top of the ingot. The charge carrier lifetime increase form the bottom to the central part of the ingot. From the central to the topmost ingot position, a very clear reduction in the influence of solid-state in-diffusion is seen. The mean lifetime however, seem not to change drastically.

5.3.2 Spectral comparison of DRL from wafers at different ingot positions

In order to display the development of DRL along the ingot height, Figure 5.9 shows spectra of wafers from the three different ingot positions of the p-twin-ingot. Large spectral differences are seen, especially regarding the intensity of D4. For the topmost wafer, the intensity of the D4 emission line is the same as that for the D3 emission line. For the wafer from the central ingot position, D4 is clearly less intense than for the topmost wafer, and for the wafer from the lower ingot position, D4 is not at all present, or is so weak that it cannot be recognised from the neighbouring peaks. The D4 signal seems, in other words, to increase throughout the length of the ingot, from the bottom to the top. The wafer from the ingot bottom also shows a less intense D1 signal than the other two wafers.

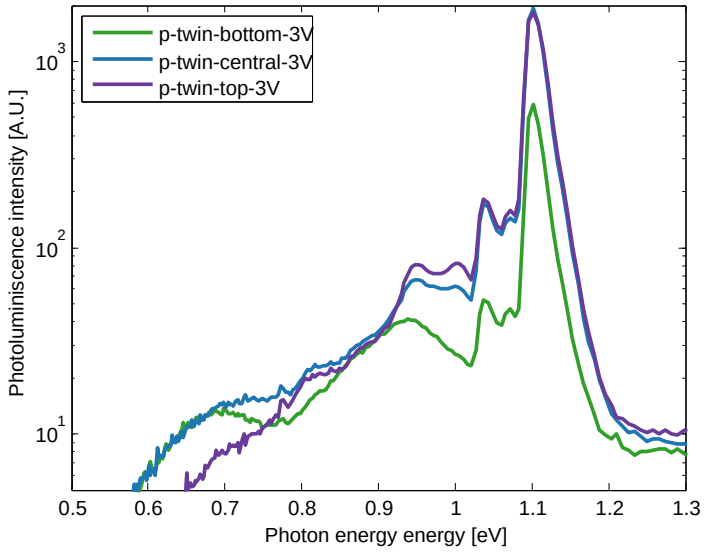


Figure 5.9: *Comparison of luminescence spectra from the bottom, central part and upper part of the p-type twin block. The numbers following the sample name indicate the laser excitation voltages at which the measurement was done.*

The difference in the height of the BB emission peaks in Figure 5.9 is probably caused by two different effects. Firstly, the doping concentration N increases from the bottom to the top of the ingot (see Table 3.1). Secondly, as observed in Figure 5.8, the lifetime, and accordingly the injection level, increase from the bottom to the central part of the ingot. Its mean value does, however, not change much from the central to the topmost ingot positions.

Due to the large increase of the BB emission peak, and probably of the injection level, along the ingot height, one may suspect that the differences between the defect luminescence spectra from wafers at different ingot height are caused by injection differences rather than differences related to the ingot height itself. The injection dependence of the DRL spectra therefore needs to be investigated.

Injection dependence

For the purpose of displaying the injection dependence of the DRL signals from the investigated wafers, Figure 5.10 shows a comparison of photoluminescence spectra from various ingot heights at different laser excitation voltages, and accordingly different injection levels.

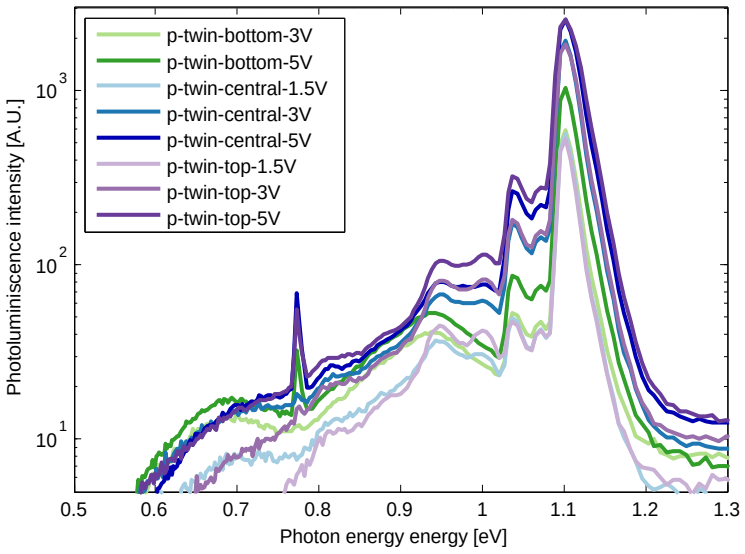


Figure 5.10: *Comparison of luminescence spectra from wafers at different ingot heights at various laser excitation voltages. The numbers following the sample name indicate the laser excitation voltages at which the measurements are done.*

The figure indicates that when the excitation laser voltage is increased, the luminescence spectrum for one ingot height is simply shifted upwards without changing its shape significantly. When comparing wafers at different block heights, however, substantial differences are seen. This indicates that the observed differences of the defect luminescence spectra shapes are not – or at least not exclusively – caused by differences in charge carrier injection, but rather in differences related to the crystal structure, independently of the related lifetime.

Comparison with spectra from n-type wafer

In Figure 5.11, the spectra from *various ingot positions* of the p-twin-ingot shown in Figure 5.10 are plotted together with spectra from a *central ingot position* in the n-twin-ingot (also shown in Figure 5.7). The similarity between the spectral shape of the topmost wafer in the p-type ingot and the central wafer in the n-type ingot is remarkable. Wafers from other ingot position of n-type ingot were not included in this study.

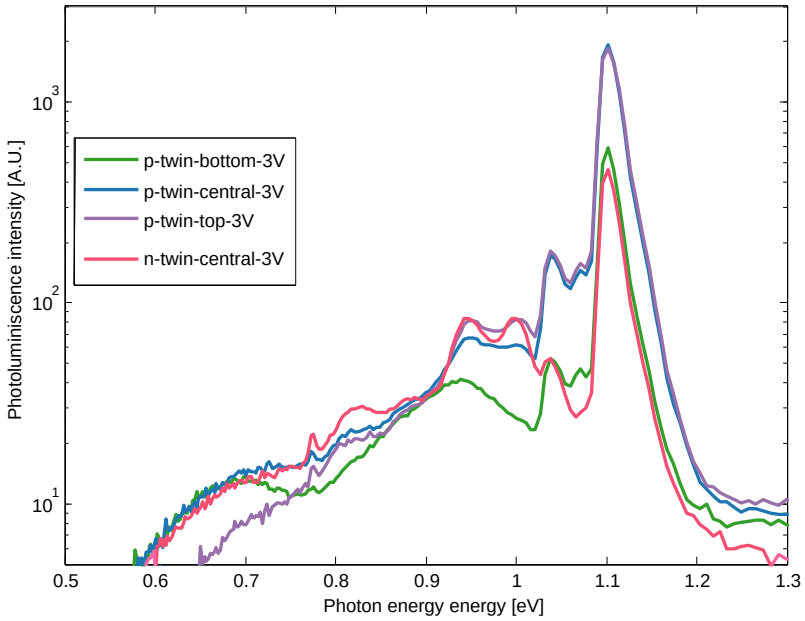


Figure 5.11: *Spectra from the central position in the n-type ingot compared with spectra from various block positions in the p-type ingot. The numbers following the sample name indicate the laser excitation voltages at which the measurement are done.*

5.3.3 Summary of section results

Through a comparison of lifetime images, we have observed an increase in the influence of dislocation clusters along the growth direction of the p-twin-ingot

A comparison of the luminescence spectra from wafers at high, central and low ingot positions of the p-twin-ingot indicates that the intensity of D4 seems to increase towards the top of the ingot. A weak decrease of D1 is also seen. This change in the spectral shape along the ingot height is not caused solely by injection differences.

When approaching the top of the p-type ingot, the shape of the luminescence spectra gradually approaches that of the central ingot position of the n-type ingot.

5.3.4 Discussion of section results

It is generally accepted [2, 33, 39] (see section 2.6.1) that the D3 and D4 emissions are due to intrinsic properties of dislocations in silicon. Thus, one may hypothesize that the defect from which the D4 emission originates is present only to a very low extent at the ingot bottom, and that its concentration, due to dislocation multiplication increases towards the ingot top. That would correspond to the observed increase in the D4 emission intensity along the growth direction of the ingot.

However, in subchapter 5.2, we observed differences in the intensity ratio between D3 and D4 related to material type (p/n). These differences are very similar to those observed here related to ingot positions. Thus, it should be questioned if there is a relation between the observations.

In Figure 5.6, we noted that the wafer n-twin-central seem, to a larger extent than its p-type counterpart p-twin-central, to be influenced by dislocation clusters. Further, since iron and other metallic impurities are less detrimental in p-type material, this wafer is only to a low extent impacted by recombination due to impurities.

Figure 5.8 illustrates that the influence of dislocation clusters increases towards the top of the p-twin ingot. Moreover, the figure shows that, in our samples, the impact of metal in-diffusion from the crucible decreases towards the ingot top. This means that when approaching the top of the p-twin ingot the material is getting *gradually more similar to that of the n-twin ingot*. This observation corresponds to the spectral comparison displayed in Figure 5.11. Based on these observations, we suggest that the spectral differences observed between n-type and p-type material in section 5.2 are related to other material differences, rather than the doping type.

We speculate a) that wafers for which recombination due to dislocation clusters limits the carrier lifetime have DRL spectra characterised by comparable intensities of D3 and D4, and b) that samples in which the lifetime is limited by recombination due to impurities or their precipitates have a reduced emission intensity of D4 compared to D3.

The observations in section 5.2.2 related to the npn-samples does also match this speculation. The crystal structure and the distribution of dislocations of the two samples from this ingot are comparable. However, the *recombination activity* due to these dislocations is more important in the n-type than in the p-type sample, because the lifetime of the p-type sample to a larger extent is limited by recombination due to impurities.

The observed variation in the ratio of the intensities of D3 and D4, for the different ingot positions as well as between the n- and p-type samples, contradicts the idea of D3 being the phonon replica of D4, suggested by some authors [33]. The observation is, however, in line with findings by Burud et al.[47] and Nguyen et al. [3] both reporting large variations in the ratio of the intensities of the D3 and D4 emissions from different locations on multicrystalline wafers. Knowing that the concentration of impurities, such as iron, as well as the impact of dislocations can vary spatially across a wafer, one may suggest that findings of Burud et al. as well as that of Ngyen et al. also can be related to our speculation regarding the relative importance of recombination due to dislocations and recombination due to impurities.

5.4 DEPENDENCE OF DEFECT LUMINESCENCE ON IRON

Iron is an important contaminant in solar cells, whose presence is detrimental to the minority carrier lifetime. Thus, the dependence of DRL on iron and other metal impurities has been extensively discussed (e.g. in [4],[3] and [35]), and is an important topic in this work.

This subsection starts with a spatial comparison of the D line emissions in a p-type sample with images of interstitial iron concentration, followed by a similar comparison for an n-type wafer. Further, spectral comparisons of regions with varying iron content are compared.

5.4.1 Relation between defect luminescence and iron in p-type Si

Figure 5.12 shows an image of the local lifetime in wafer p-twin-central (a) and an image of interstitial iron concentration of the same wafer (b). On the lifetime image, dislocation clusters and grain boundaries causing regions with decreased lifetime are recognized. Additionally, a region of very low lifetime along the right edge of the sample is seen. The iron image clearly suggests that the reduced lifetime in this region is due to the in-diffusion of iron from the crucible during the solidification process (see section 2.4.1). Because of a measurement artefact related to the iron measurements close to grain boundaries (further explained in [16]), only this very clear increase in iron concentration in the right part of the iron map can be considered as reliable. The slight increase in iron concentration in small regions and spots spread across the wafer area cannot necessarily be trusted. One should further note that, in addition to iron, also other metal impurities, such as chromium, diffuse from the crucible during the solidification process, causing similar concentration profiles to that of interstitial iron.

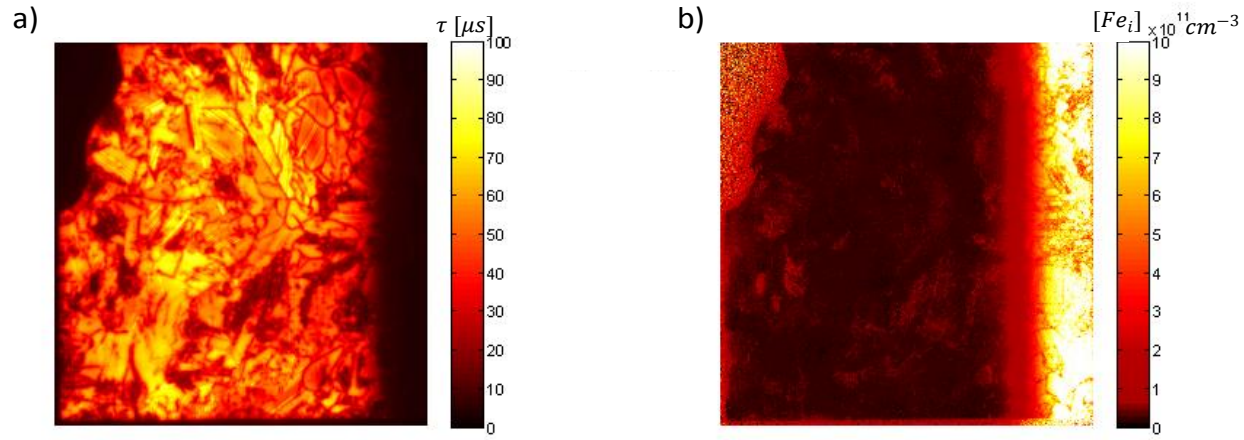


Figure 5.12: *a) Spatially resolved image of lifetime at 1sun illumination and b) image of interstitial iron concentration for wafer p-twin-central. The images are based on photoluminescence imaging calibrated by modulated PL, combined with FeB pair splitting (see section 3.4).*

In order to compare the spatial distribution of the D lines with the distribution of interstitial iron, MCR analysis was applied to extract the various components of a hyperspectral image corresponding to the various D lines. Figure 5.13 presents the results of an MCR analysis done for a hyperspectral measurement of wafer p-twin-central. The result of this MCR analysis is carefully compared to integrated images of measurements of the same wafer, and no severe discrepancies are found. Since the MCR component images appear clearer than the integrated

images, the MCR images are chosen for presentation. Among the 20 components found by the MCR analysis, six components that seem to correspond to D0-D4 and BB are picked out and displayed in the figure. In addition, the loading plot is shown. Note that the scanning direction of this measurement is *orthogonal* to the iron in-diffusion region, avoiding false conclusions to be due to the inhomogeneity of the laser profile.

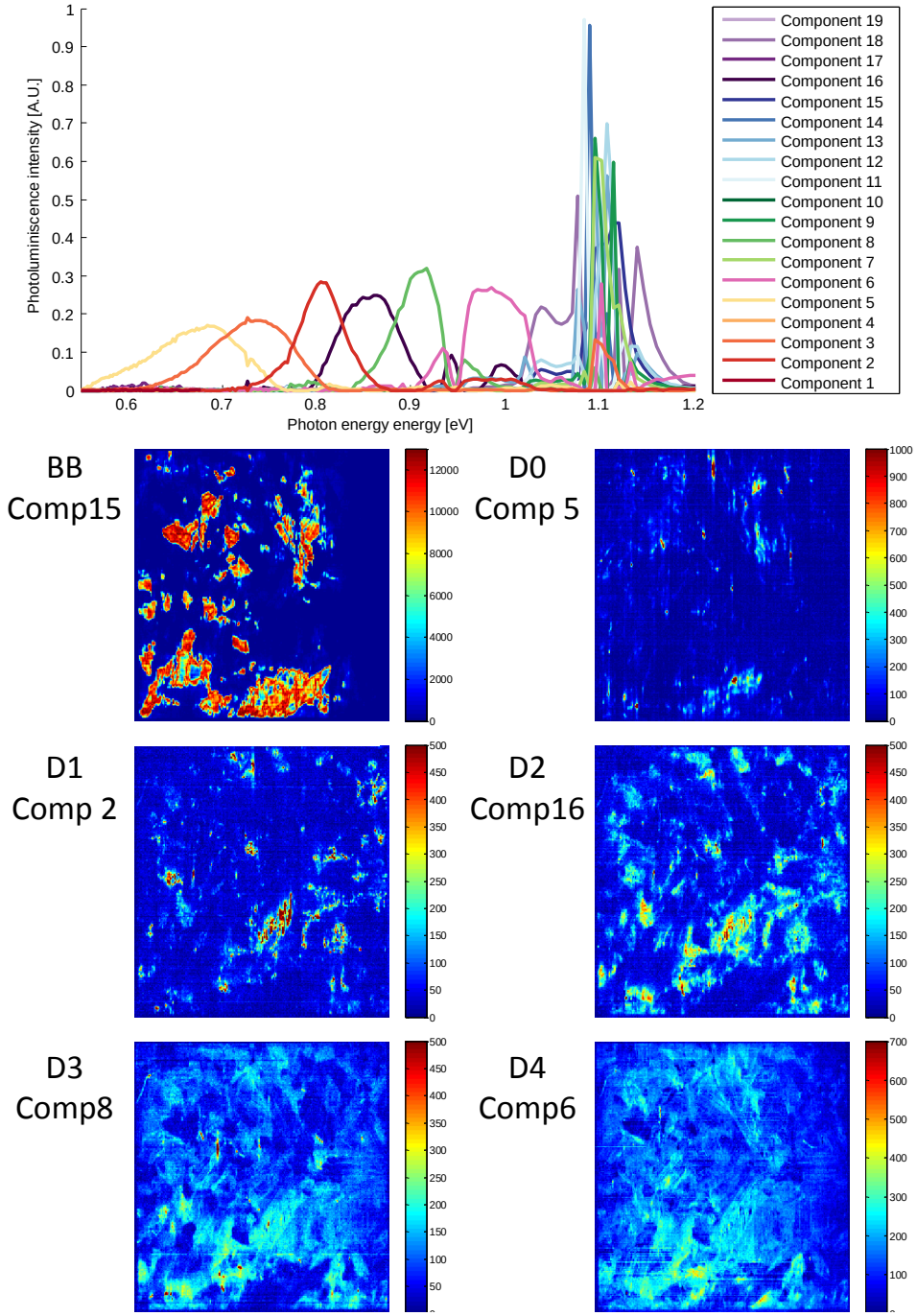


Figure 5.13: *MCR analysis for a hyperspectral measurement of the wafer p-twin-central done at an excitation laser voltage of 5V. Loadings plot for each of the 20 extracted components and score plots representing six components that may correspond to D0-D4 and BB are shown. For clarity, only some of the MCR components are shown in the loadings plot .*

Comparing the spatial distribution of the D line emissions and the BB emission in Figure 5.13, with the lifetime and the iron image in Figure 5.12 one notes that the BB emission – as expected – seem to be supressed in areas with high iron concentration. Further, one notes that the distributions of D1 and D2 clearly corresponds to regions with dislocation clusters, and accordingly reduced lifetime. This corresponds well to the accepted idea that D1 and D2 originates from impurities or secondary defects in the vicinity of dislocations [2]. However, contrary to the BB emission, the intensity of the D1 and D2 signals seem not to be influenced by the increase in iron content. Regions with signals in the energy intervals of D1 and D2 are evenly distributed throughout the entire extent of the wafer, regardless of the local iron concentration. This observation will be further discussed in section 5.4.5.

D3 and D4 also seem to correlate negatively with the BB emission on most of the wafer. In the iron rich part, however, the intensity of both D3 and D4 seems to decrease, despite the decreasing lifetime and BB intensity. This observation suggests that iron or other metal impurities may – at least partly – supress both the D3 and the D4 emissions. However, the observed reduction of D3 and D4 in the iron rich region may also be caused by the extreme reduction of lifetime and injection level caused by the iron, thus being only an indirect consequence of the high iron content.

5.4.2 Relation between defect luminescence and iron in n-type Si

The influence of interstitial iron and other transition metals on lifetime and thus on the PL intensity is less in n-type than in p-type material. Therefore, n-type mc-Si can – to a much larger extent than p-type mc-Si – be used to investigate the impact of iron and other metal contaminations *directly* on DRL intensity, with less impact from other effects, such as decreased lifetime and injection level.

Because the technique for making spatially resolved iron measurements of silicon wafers are based on the association of Fe atoms with boron atoms, and their dissociation, the method is only applicable for p-type wafers (in which boron is the doping material). Exact measurements of the iron content in the n-type wafers can therefore not be made with the iron imaging method utilized in this work. However, due to the fact that the p-type and n-type twin blocks are crystallized in identical crucibles and under very similar temperature conditions, one can assume that the in-diffusion of iron from the crucible walls in the two blocks has taken place in a similar fashion and to a similar extent[8]. Consequently, for the wafer n-twin-central, we can assume an in-diffusion area similar to the one shown in Figure 5.12b for the wafer p-twin-central. This allows for a study of the impact of iron on DRL also in n-type material.

As mentioned during the process of labelling the wafers, the n-type wafers were oriented differently from the p-type wafers (see 0). Thus, as opposed to the p-type wafers, the in-diffusion zone occurs at the left edge of the n-type wafers.

[8] One may ask why the samples from the npn-ingot were not used for this analysis. This is because the high lifetime (and accordingly high injection level and BB emission peak) combined with low DRL signal (see e.g. Figure 5.4) make them unsuited to extract sharp images of the D lines.

Figure 5.14 shows an overview of D lines images (made from integration over the wavelengths given in Table 4.1) as well as a lifetime image for the wafer n-twin-central. In order to visualize the distribution of iron in the sample, an image of the interstitial iron concentration in the wafer p-twin-central (rotated for better comparison) is also displayed.

A comparison of the D line images with the iron image clearly suggests a reduction of the D3 and D4 emission intensities in the in-diffusion region. This is similar to the finding for the wafer p-twin-central. Moreover, when comparing those images with the lifetime image, one notes that the reduction of the D3 and D4 emission intensities, *does not* correspond directly to the region of reduced lifetime at the left wafer edge. The region of reduced lifetime is uneven and cornered, probably due to crystal imperfections such as grain boundaries and dislocation clusters that coincidentally are placed in or close to the region of iron in-diffusion. The region of reduced D3 and D4, however seem to run parallel to the wafer edge, independently of the varying lifetime. This observation clearly suggests that the reduction of the D3 and D4 emission intensities is not related to reduced lifetime as such, but rather to the presence of iron or other in-diffusing contamination materials.

A similar pattern is also very clear for the sample n-twin-central-as-cut regarding the distribution of the D4 emission line with respect to the iron concentration (see Figure 5.20 a). This finding will be further discussed in section 5.4.5.

When studying the images D1 and D2 in Figure 5.14, it is clearly seen that the emissions seem to run undisturbed throughout the wafer, independently of the local iron concentration. The spatial distribution of these emission lines seems to correspond to regions with dislocation clusters. This is similar to the pattern found for the wafer p-twin central in Figure 5.13.

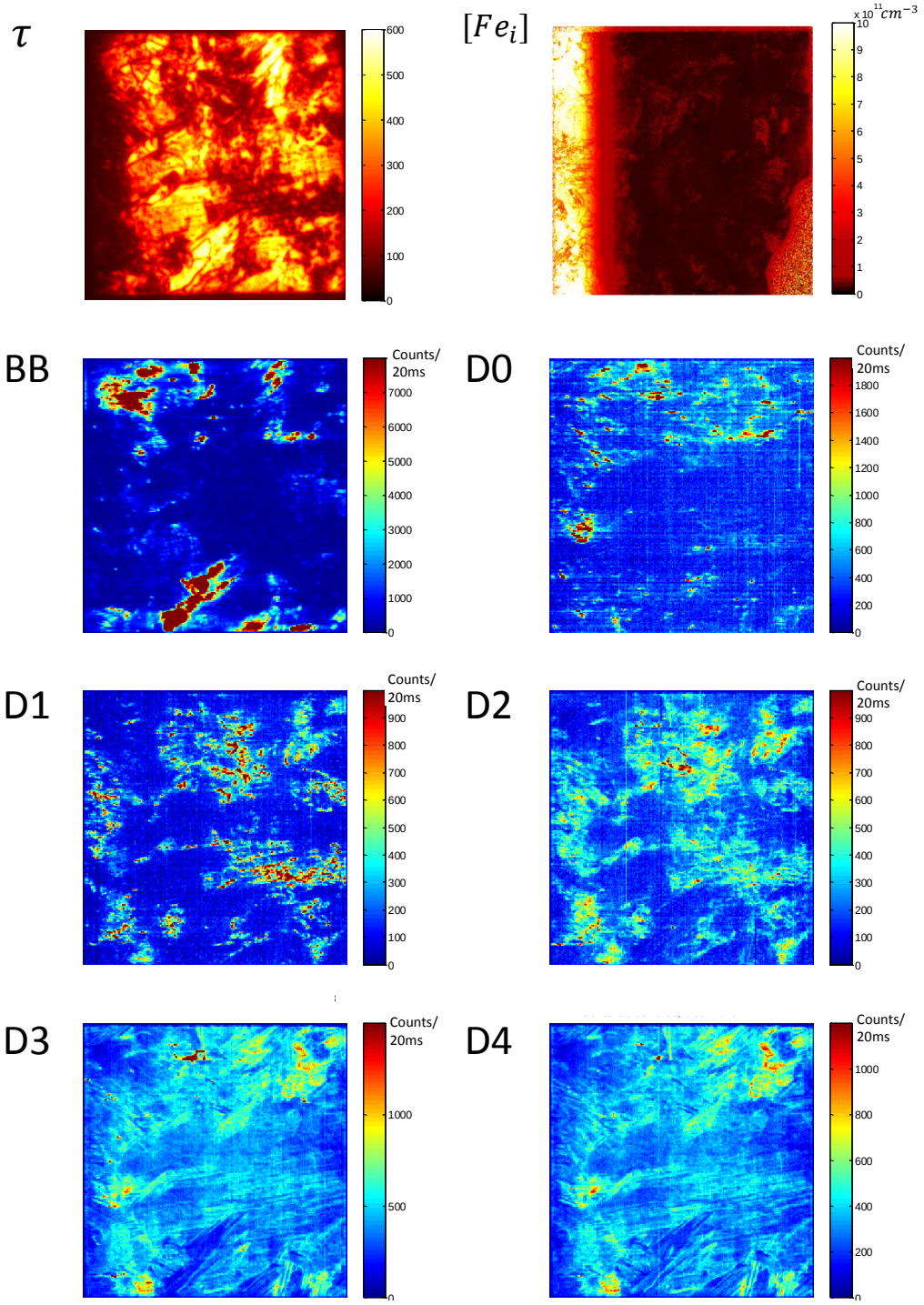


Figure 5.14: *Lifetime image as well as images of the spatial distribution of the D lines on the wafer n-twin-central. In addition, an image of the assumed iron concentration, based on a measurement of the sister wafer p-twin-central, is shown. The colorbar scales of the emission line images give the numbers of counts found by integration over the energy intervals given in Table 4.1. The mean spectrum from this hyperspectral measurement can be seen in Figure 5.7.*

5.4.3 Comparison of luminescence spectra in regions with varying iron content

We wanted to investigate whether the apparent negative influence of iron or other metal impurities on D3 and D4 emission intensities observed in *spatially* resolved images, also can be recognised by *spectral* comparison of regions with various iron content. In order to avoid challenges related to saturation and bleeding, spectra based on as-cut wafers are chosen for such a comparison.

Because of the low charge carrier lifetime of as-cut wafers, their iron concentration cannot be measured with the method applied for this work. The approximate iron content can, however, be found from passivated neighbouring wafers. Simulations of the interstitial iron concentration in wafers from the central part of ingot p-twin [66] show that the change of iron concentration after passivation is less than 10%. This value is considered to be within the frames of uncertainty of the hyperspectral measurements. Thus, for the sake of comparison of iron content with intensity of D line emissions, the iron profile of wafer p-twin-central-*as-cut* can be approximated by that of wafer p-twin-central. Likewise, the iron profile of wafer n-twin-central-*as-cut* can be approximated to that of n-twin-central, which in turn is a mirroring of the iron profile of *p*-twin-central. The iron concentration profiles in Figure 5.15 and Figure 5.16 are based on these assumptions.

Spectra of n-type wafer

Figure 5.15 a) shows the assumed concentration of interstitial iron as function of distance from the wafer edge for wafer n-twin-central-as-cut. The values are found as an arithmetic mean of the corresponding row of pixels in the iron map of wafer p-twin-central. In part b), a spatially resolved image of the D4 emission line is shown. Regions from which the spectra in part c) are taken are indicated. The regions are chosen parallel to the wafer edge. Since the in-diffusion of iron is supposed to be the same for all points at an equal distance from the crucible wall, the interstitial iron concentration within one region can be assumed to be constant. Part c) shows spectra from each of the regions highlighted in b). The colours indicate the region from which each spectrum originates. The scanning direction of this measurement is perpendicular to the chosen regions, meaning that the inhomogeneity of the laser profile has the least impact on the differences between the mean spectra from the regions (cf. section 4.5.2).

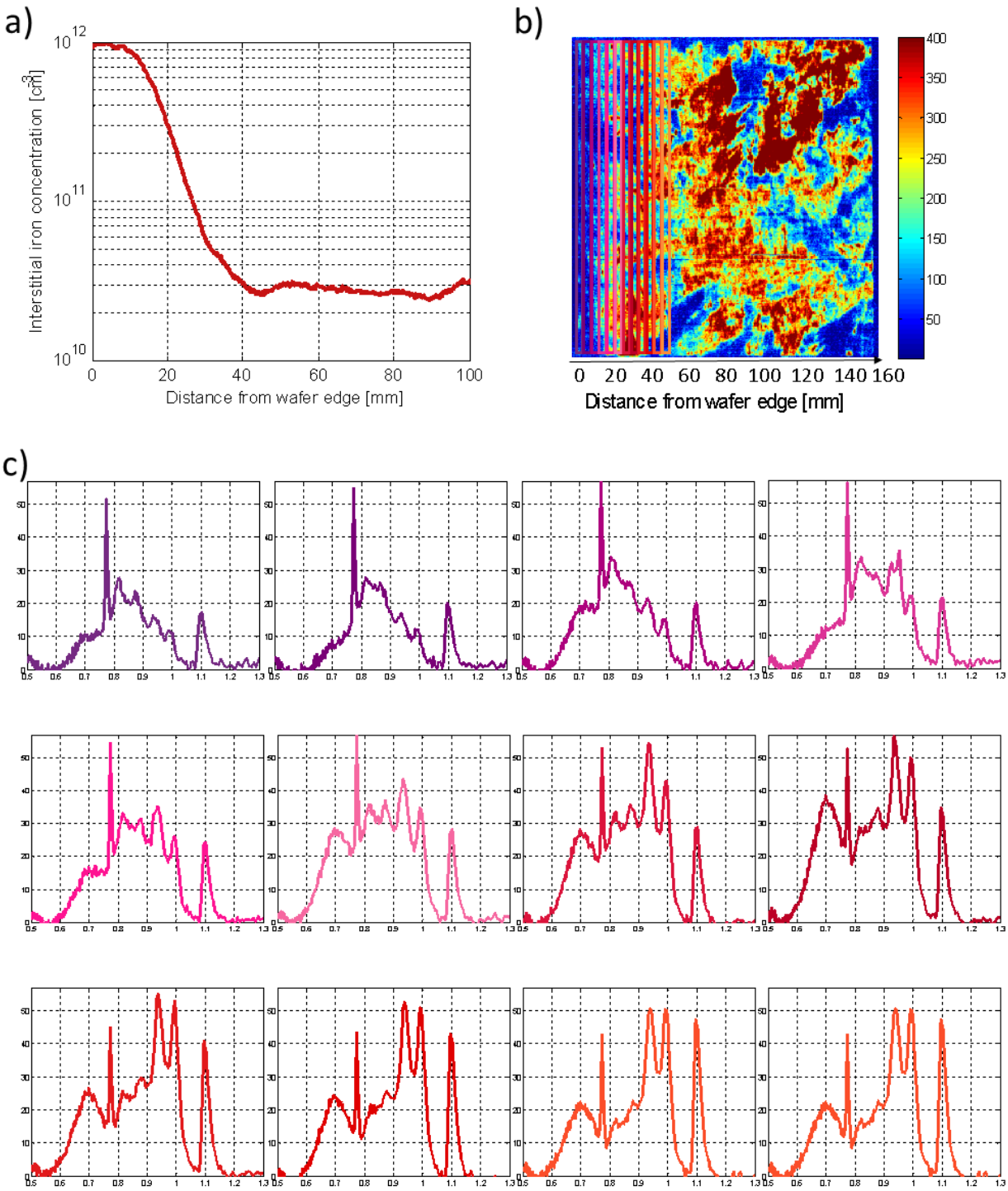


Figure 5.15: *a) Measured concentration of interstitial iron in wafer n-twin-central-as-cut as function of distance from wafer edge, based on the assumption that the inn-diffusion of iron in the n- and p- twin ingots are equal. b) Spatially resolved image of the D4 emission of the same wafer, with rectangles highlighting the regions from which spectra in c) are chosen. c) Defect luminescence spectra of regions with increasing distance to the left wafer edge, and accordingly decreasing concentration of interstitial iron. The colours indicate the regions from which each spectrum is taken. The x-axis gives the photon energy in units of electron volt, and the luminescence intensity in arbitrary units (counts).*

Figure 5.15 clearly shows that the intensities of D3 and D4 decrease towards the wafer edge. The measured increase of the interstitial iron content in the same wafer region (which indicates an increase also of other metal impurities), underpins the theory that iron, or other in-diffusing metals have a negative impact on the D3 and D4 emission intensities.

Moreover, another interesting effect is observed: Far from the wafer edge, and thus with lower impurity concentration, the intensity of D3 and D4 seem to be similar. When approaching the wafer edge, however, the intensity of D4 seems to decrease more than that of D3. This observation suggests that the presence of iron or other impurities supresses the emission process related to D4 more efficiently than the one related to D3. This observation will be further discussed in section 5.4.5.

Spectra of p-type wafer

In order to investigate whether a similar effect is observable in the p-type sample, Figure 5.16 displays an analysis of wafer p-twin-central-as-cut, similar to the one of n-twin-central-as-cut in Figure 5.15. Also in this sample the intensities of both D3 and D4 seem to decrease when approaching the wafer edge (i.e. when the concentration of iron and other metal impurities increases). The effect is, however, not as evident as for the n-type sample. Moreover, in this sample, as opposed to the n-type sample, the emission intensities of D3 and D4 seem here to undergo a comparable absolute decrease when the interstitial iron concentration increases.

In the fourth DRL spectrum in Figure 5.16 (upper right plot) an abrupt increase of the D3 intensity is seen. This is probably due to the presence of a defect emitting signals of the category *very intense D3* (VID3), described by Flø et al. [35]. These are very intense signals, distributed in small points throughout the wafer. Such a point accidentally being included in an area from which a mean spectrum is taken, can probably alter the spectrum significantly. The apparent increase of the mean D3 signal in this region can, in other words, not necessarily be trusted.

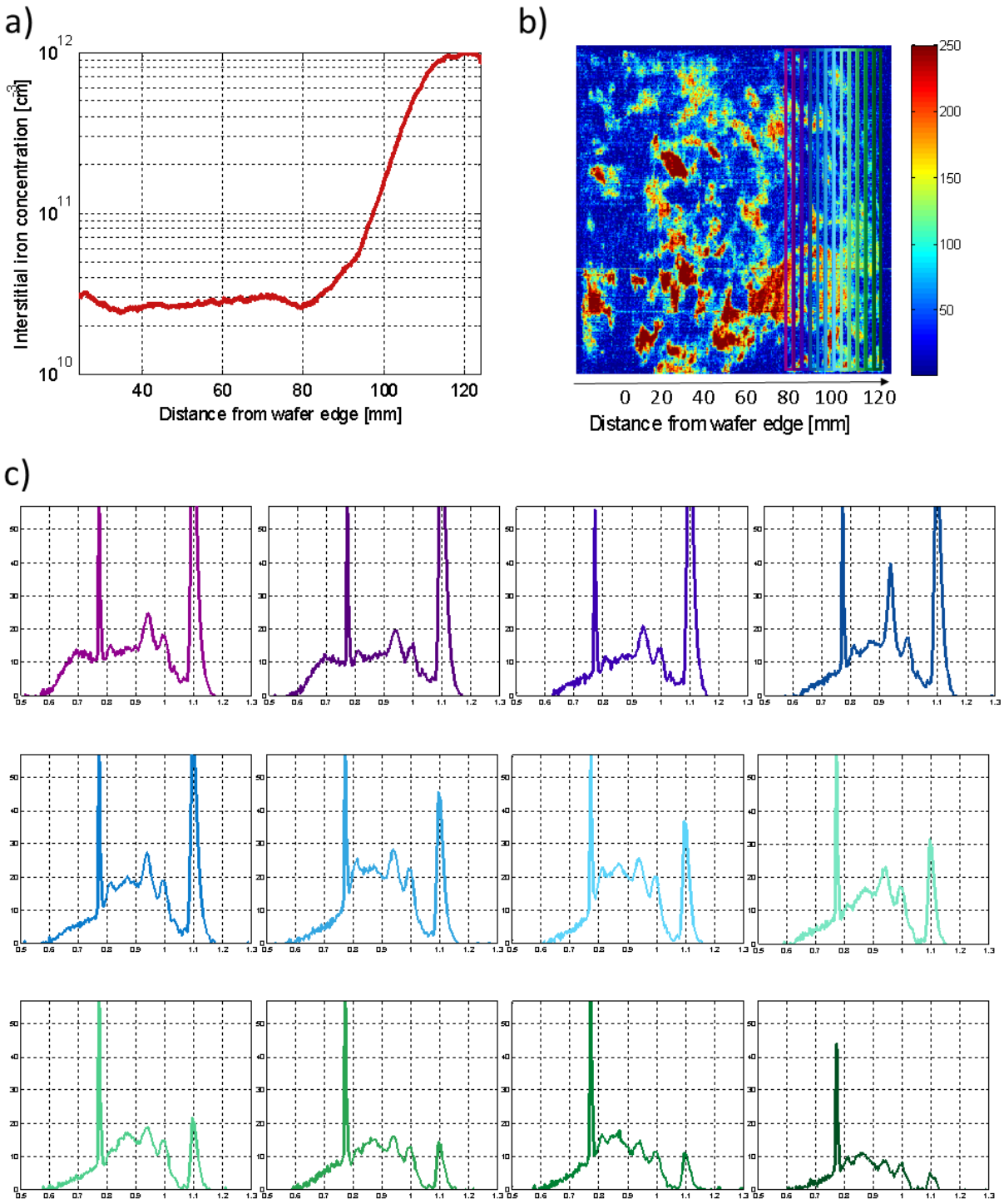


Figure 5.16: *a) Mesured concentration of iron in wafer p-twin-central-as-cut as function of distance from wafer edge, based on iron mapping on a passivated neighbouring wafer. b) Spatially resolved image of the D4 emission of wafer p-twin-central-as-cut, with rectangles highlighting the regions from which spectra in c) were chosen. Note that the distance scale is defined to have its origin at the line along which the wafer was later cut to the dimensions 125x125 cmm². When the measurement was done, the wafer had its original 156x156mm² dimensions. c) Defect luminescence spectra of regions with increasing distance to the right wafer edge, and accordingly decreasing concentration of interstitial iron. The colours indicate the regions from which each spectrum is taken. The x-axis gives the photon energy in units of electron volt, and the luminescence intensity in arbitrary units (counts).*

5.4.4 Summary of section results

A comparison of D line emissions in p- and n-type wafers with lifetime images and images of interstitial iron concentration suggests that both in n-type and p-type wafers, the emission intensities of D3 and D4 decreases in regions with high iron concentration. The decrease is most evident in the n-type sample.

In the investigated n-type sample, the decrease of the D4 emission intensity is higher than the decrease of the D3 emission intensity. In the investigated p-type sample, however, the reduction of D3 and D4 emission intensities seems to be comparable. The decrease in the intensities of D3 and D4 in iron rich regions seems to not only be a consequence of the reduced lifetime and injection level. It may therefore be a consequence of iron or other metal impurities.

5.4.5 Discussion of section results

As explained in section 2.6.1, D3 and D4 are assumed to be due to intrinsic properties of dislocations, whereas D1 and D2 are due to secondary defects or impurities trapped in the strain fields around dislocations. Therefore, it seems, in the first place, surprising that the D3 and D4 emission intensities decrease in regions with high concentrations of iron and other metals, and that D1 and D2 seem not (or only to a low extent) to be altered by the contamination concentration.

When studying the images in Figure 5.14, however, one can see that the regions of reduced emission intensities of D3 and D4 seem to be somewhat narrower than the region of high interstitial iron concentration due to solid-state in-diffusion. The narrow region along the wafer edge in which the D3 and D4 emission intensities decreases, corresponds very well to the region in which one expects the presence of *iron precipitates*. Solid state inn-diffusion of iron typically causes an concentration profile of interstitial iron similar to that shown in Figure 5.16 [12]. The profile is characterized by a steep increase followed by a “plateau” at which the concentration stays relatively constant. This plateau arise when the concentration is so high that no more iron can exist in its interstitial form. As explained in section 2.4.3, precipitates will then build.

Therefore, based on the analyses presented here, we suggest that the observed decrease of D3 and D4 emission intensities in regions with high iron concentration is caused by iron precipitates or other metallic precipitates, rather than by interstitial iron. The following paragraph, accompanied by Figure 5.17, explains how we suggest that precipitates may supress the emission intensity of D3 and D4 without altering the intensities of D1 and D2.

A dislocation or slip line (Figure 5.17 a, indicated by a black line) in a crystal region without dissolved impurities or precipitates, emits DRL signals in the wavelength regions of D3 and D4 [2, 33, 39] (indicated by orange arrow). When secondary defects or impurities are trapped in the strain field around the dislocation (Figure 5.17 b), additional *discrete energy* levels in the band gap will be introduced and signals in the wavelength region of D1 and D2 will be emitted (indicated by red arrow) [33, 39, 46]. These signals are shown to originate from regions distributed *around* the grain boundaries or slip lines [2, 39, 46].

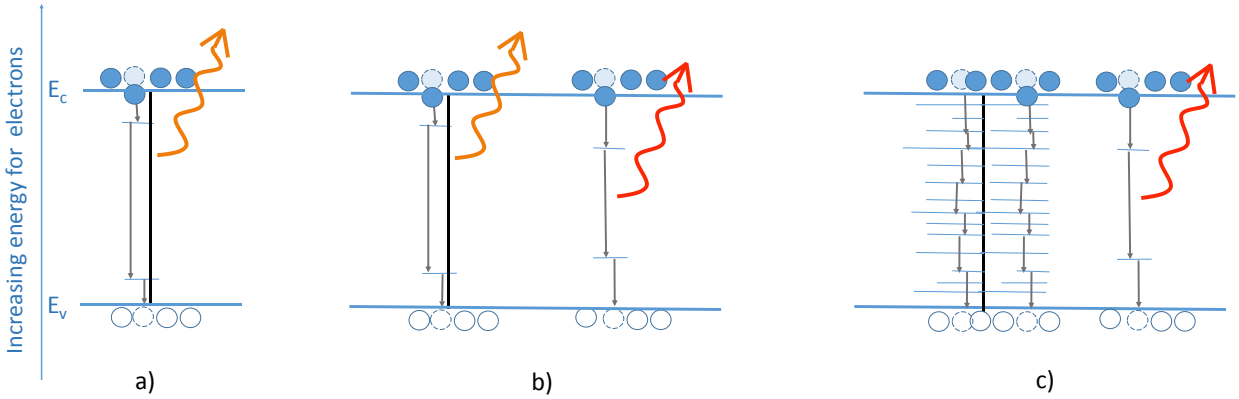


Figure 5.17: *Suggested model for the impact of interstitial iron and iron precipitates on DRL. Three different regions are shown: a) region with intrinsic dislocation, emitting D3 and D4 signals, b) region in which secondary interstitial impurities (e.g.iron) are caught in the strain field of the dislocation, resulting in additional emission of D1 and D2 and c) region with metallic precipitates on dislocation line acting as non-radiative recombination centre supressing the D3/D4 emission.*

Precipitates of iron or other impurities may lead to the introduction of *a variety of* energy levels, creating a *defect band* in the band gap close to the dislocation line onto which the precipitates are built (Figure 5.17c). The precipitates may therefore function as non-radiative recombination centres with high recombination activity, partly supressing the intrinsic radiative nature of the dislocation (i.e. suppressing the intensities of the D3 and D4 emissions). The secondary point defects, however, which are distributed around the dislocation, are not altered by the precipitates. The D1 and D2 emission intensities may therefore stay unchanged irrespective of the local concentration of impurity precipitates.

A similar model is for the recombination activity of dislocations is described by Seibt et al. [67]. Moreover, a similar explanation for the reduction of D3 and D4 emission intensities in regions with high concentrations of oxygen precipitates is suggested by Tajima et al. [2].

As mentioned in section 2.6.3, Gundel et al. [4] reports a shift of the subPL band to longer wavelengths (i.e. lower energies) in points with iron precipitates compared to iron contaminated regions without precipitates. Their measurements are done at room temperature. According to Tajima et al. [2], D3 and D4 do at increasing temperature merge into a broadband at approximately 0.94 eV, whereas D1 and D2 merge to a broadband of approximately 0.79 eV. A similar temperature behaviour was also reported by Inoue et al.[68]. Assuming that the subPL band observed by Gundel et al. is a blend of the two broad bands described by Tajima et al., a suppression of the upper band (the band corresponding to D3 and D4) could, in principle, be equivalent with a shift of the subPL to lower energies. Gundels findings may, in other words, match the idea that D3 and D4 are supressed by metal precipitates.

The findings of Flø et al. [35], suggesting correlation between interstitial iron concentration and the emission intensity of D4, seem in the first place to contradict the hypothesis of D3 and D4 being supressed by iron or other metallic precipitates. Their measured interstitial iron concentrations are, however, all below 10^{11} cm^{-3}. For interstitial iron concentrations at this order

of magnitude, one may assume that iron is present almost exclusively in interstitial form. Studying Figure 5.15 and Figure 5.16 one finds that an evident decrease of the D3 and D4 signals in our samples are seen only at interstitial iron concentrations above 10^{11} cm^{-3}. As explained above, this high iron concentration points in direction of the presence of iron precipitates.

Based on the results presented in section 5.2 and 5.3, we speculated that the observed differences in spectral shape among the investigated wafers might be explained by the recombination mechanisms by which the charge carrier lifetimes of the samples mainly is limited. For samples whose lifetime probably is limited by recombination due to dislocation clusters, we observed a symmetrical shape regarding the intensities of D3 and D4 emissions, and for samples whose lifetime probably is limited by iron or other impurities we observed reduced intensities for D4 compared to D3.

In Figure 5.15 and Figure 5.16 we observed that the emission intensities of *both* D3 and D4 were reduced in regions with high concentration of iron and other impurities. For the n-type sample, however, we observed that the D4 intensity was reduced *more* than the D3 emission intensity in regions with increasing iron (see Figure 5.15). The large reduction of D4 compared to D3 causes the DRL spectrum to get the asymmetrical shape which – according to our speculation – characterizes regions in which the lifetime is limited by impurities rather than by dislocations. This observation perfectly fits the assumption that the lifetime in this region gradually changes from being limited by recombination active dislocations to rather being limited by impurities. It therefore underpins our theory.

5.5 Change of defect luminescence during processing steps

Studying changes of defect luminescence during solar cell processing is important, not only as a tool for understanding the nature of DRL, but also in order to find ways of improving the efficiency of finished solar cells through defect engineering. Recently, such changes have been studied among others by Johnston et al. [6], finding large differences between the defect luminescence in various regions of a solar cell after different processing steps (see section 2.6.4).

In this subsection the defect luminescence of mc-Si wafers at different processing steps are compared, spatially as well as spectrally.

5.5.1 Spatial comparison of DRL from wafers after various solar cell processing steps

In order to determine whether effects similar to those found by Johnston et al. are present in our samples, spatially resolved images of the D lines (obtained by integration) from differently processed wafers are compared. Figure 5.18 displays the D lines as well as the BB emission line for four neighbouring wafers. These are one as-cut wafer (p-twin-central-as-cut), one passivated wafer (p-twin-central), one wafer additionally subjected to P-gettering (p-twin-central-P) and one wafer additionally subjected to P-gettering and H-passivation (p-twin-central-P-H). The exact wafer processing procedures are further described in section 3.1.1. For a better comparison with the findings by Johnston et al. [6], in addition to the D line images, images integrated over the

wavelengths denoted as subPL (0.73-0.92 eV) by Johnston et al. are also included in the figure. This wavelength interval includes D1 and D2, but not D3 and D4. Figure 5.19 shows a similar comparison for differently processed n-type samples[9].

[9]) Since Johnston's investigations are of p-type wafers, and the images of „subPL" for our p-type samples show now notable changes, this images is for the n-type sample omitted

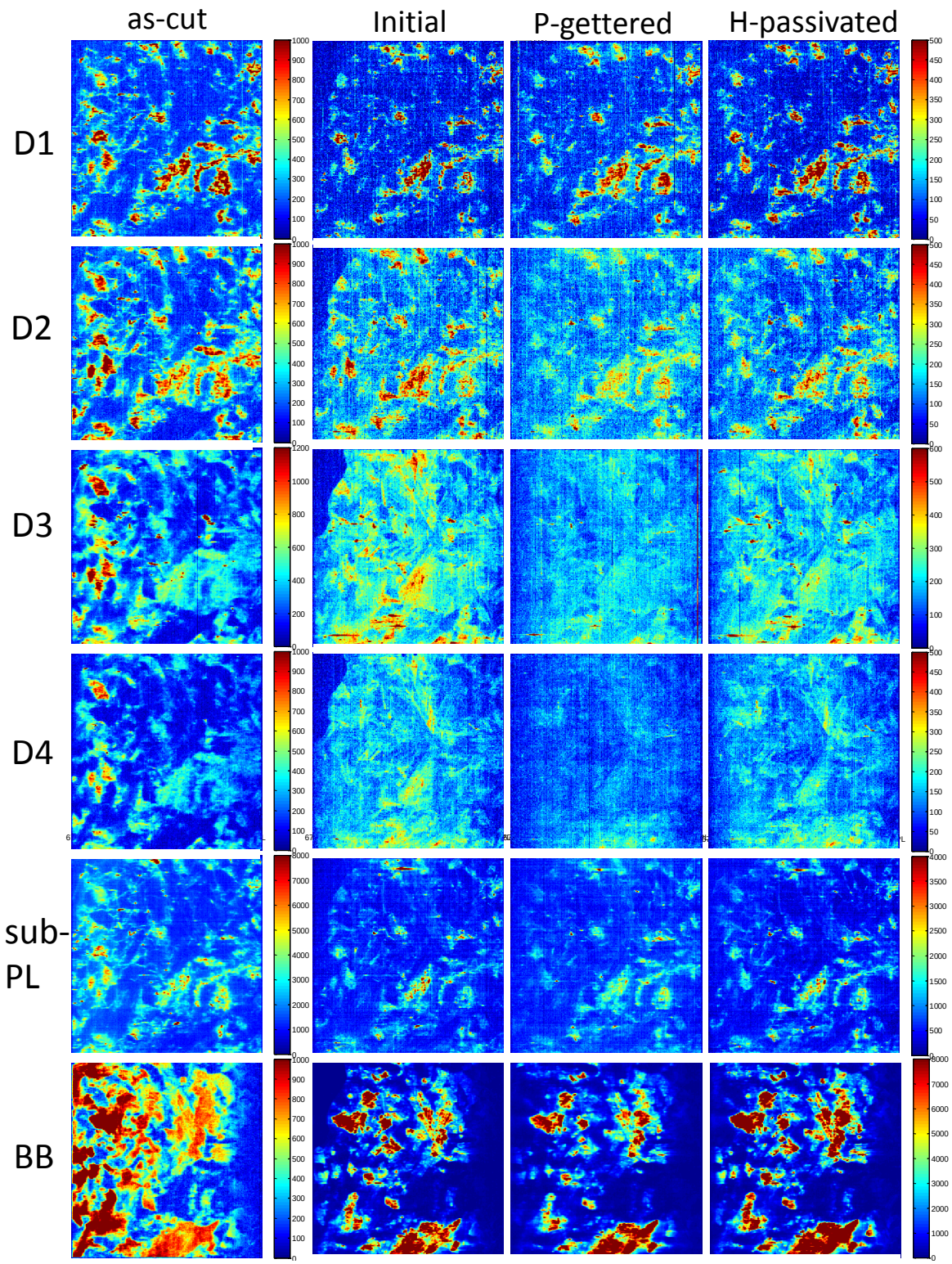


***Figure 5.18**: Comparison of DRL from neighbouring wafers from the central part of ingot p-twin in the as-cut state, in the initial state, after P-diffusion and after H-passivation (see Table 3.1). The colorbar scales give the numbers of counts found by integration over the energy intervals given in Table 4.1. Note that the scales of the as-cut wafer images differ from those of the other images.*

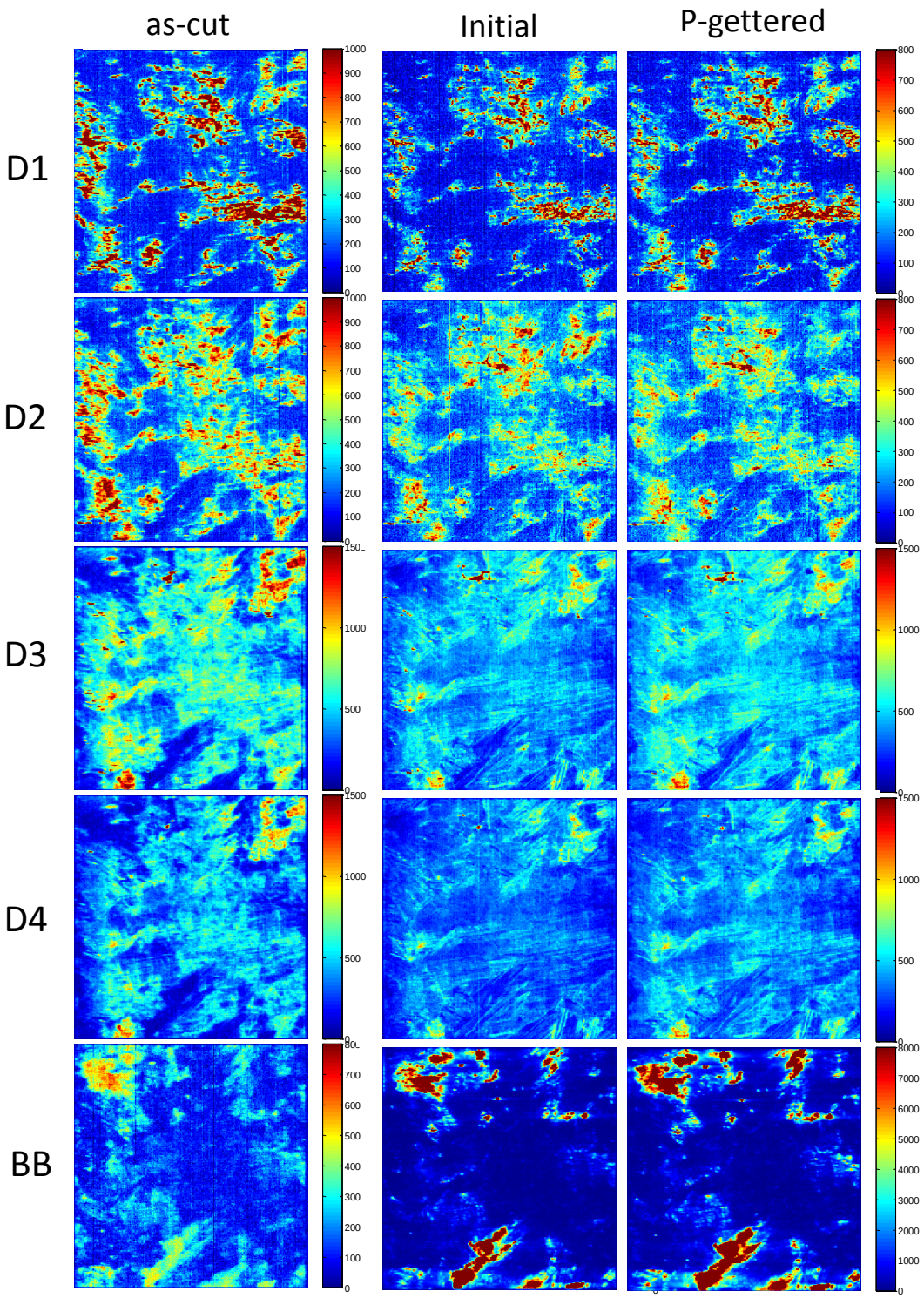


Figure 5.19: *Comparison of DRL from neighbouring wafers from the central part of ingot n-twin in the as-cut state, after initial treatment, after P-diffusion (see Table 3.1). The colorbar scales give the numbers of counts found by integration over the energy intervals given in Table 4.1. Note that the scales of the as-cut wafer images differ from those of the other images.*

Based on these images, no changes in the subPL signal (only shown in Figure 5.18 for the p-type samples) comparable to those described by Johnston et al. are seen. The distribution and intensities of the D1 and D2 emission lines also seem not to be remarkably impacted by the processes investigated here. Similar comparisons have been conducted for wafers from the upper and lower part of ingot p-twin. Also among those wafers no large differences were found. This will be further discussed in section 5.5.4.

Regarding the spatial distribution and the intensity of the D3 and D4 emission lines, however, two interesting effects are seen. Firstly, both for the n- and the p-type sample, the spatial distribution of the D3 and D4 emission lines seem to change from the as-cut to the initial (surface-passivated) state. Secondly, for the p-type samples these emission lines seem to decrease in intensity after the P-gettering step. These two features will be further investigated later in this subsection and in the following subsection, respectively. First, however, a note on the laser intensities applied for the measurements shown in Figure 5.18 and Figure 5.19 is necessary.

Note on laser intensity

In Figure 5.18 and Figure 5.19, the samples n/p-twin-central, n/p-twin-central-P and p-twin-central-P-H are all illuminated at low laser intensities in order to avoid saturation and bleeding. The as-cut samples, however, are due to their large surface recombination rate illuminated at maximum laser intensity (i.e. 5V laser bias voltage). From Figure 3.10 we know that at a laser bias voltage of 5V, the laser distribution curve has distinct increase at its left edge. This increase is not present by low excitation laser voltages. On the contrary, at low excitation laser voltages, the laser intensity is significantly reduced at both wafer edges.

This difference in laser intensity explains e.g. why some areas at left part of wafer p-twin-as-cut in Figure 5.18 appear to have a more intense PL signal than its neighbouring passivated wafers. The effect is most evident for the BB images, but can be seen also for all the D line images as well as the image of subPL. The apparent differences between the samples in this area must not be falsely interpreted as a change in DRL due to wafer processing.

Change of the spatial distributions of D3 and D4 after surface passivation

In order to elucidate the change of the spatial distribution of D3 and D4 after surface passivation, enlarged images of the D4 emission originating from samples in the as-cut state and after surface passivation, for n-type as well as p-type samples, are shown in Figure 5.20. The images of the D3 emission lines are similar to those of the D4 emission lines, and are here not included.

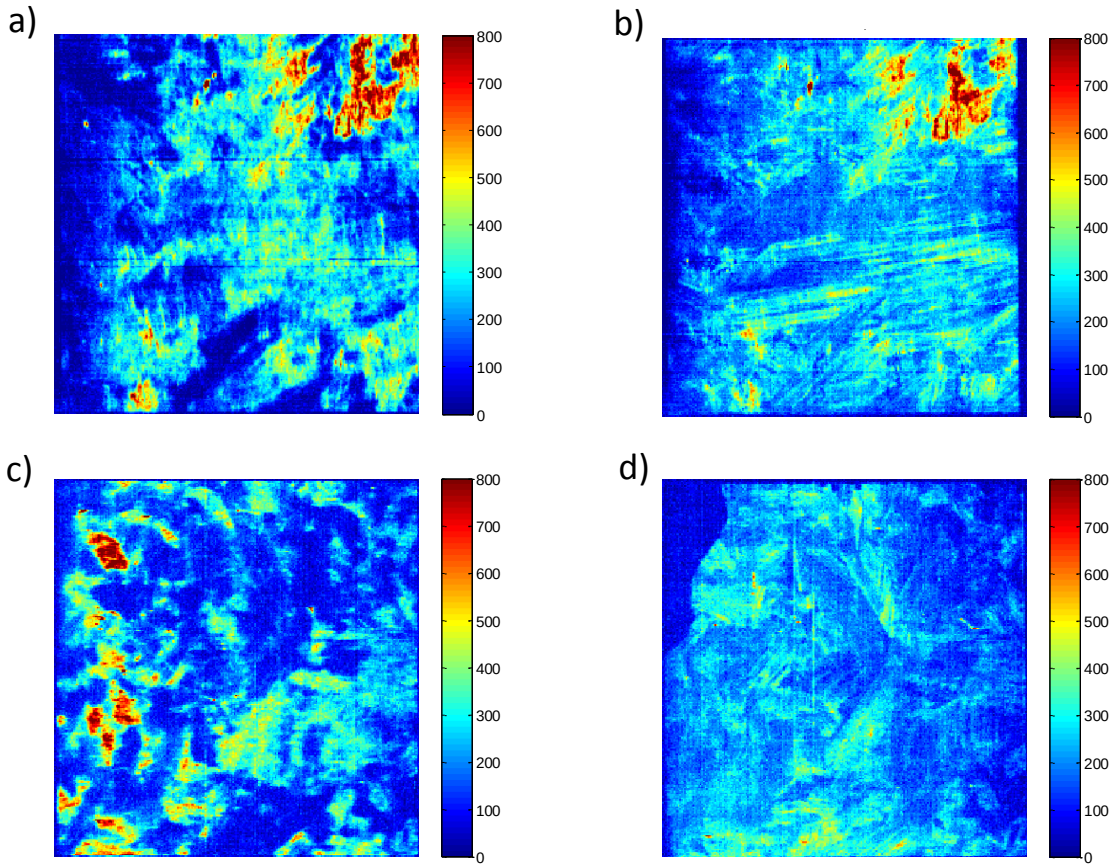


Figure 5.20: *Comparison of the spatial distribution of the D4 emission line in the as-cut state (a, c) and after surface passivation (b, d), for n-type (a, c) as well as p-type (c, d) samples. The colorbar scales give the numbers of counts found by integration over the energy intervals given in Table 4.1. Note that the apparent decrease in PL intensity from the as-cut to the passivated state at the left part of the p-type sample probably is due to the changes in of the laser distribution.*

In the as-cut state, the spatial distributions of these D line emissions are very similar to those of D1 and D2 (cf. Figure 5.18 and Figure 5.19). Moreover, the emissions seem to be distributed in cloud-like formations. After the surface passivation however, the D3 and D4 emission signals seem, to a larger extent, to originate from narrow lines rather than from cloud-like areas. Note that the apparent decrease in PL intensity from the as-cut to the passivated state at the left part of the p-type sample, as explained above, probably is due to the changes in illumination intensity.

Comparing the images in Figure 5.20 to the images of the BB emission line in Figure 5.18 and Figure 5.19, one notes that some regions that are bright in the BB images also appear to get a brighter D4 signal after the surface passivation. This effect is probably caused by the strong tail of the BB signal overlapping with the D4 signal, causing it to appear stronger in regions with strong BB emissions. It must thus not be falsely interpreted as an increase of the D4 signal in BB rich regions. However, effects of overlapping emission lines cannot explain the apparent change in the D4 distribution from cloud-like to needle-like patterns.

This change is probably related to the CP-etching process that the wafers were subjected to before the surface passivation (see section 3.1.1). During this process, the surface structure is changed causing crystal structures, such as twins (see section 2.4.2), to be visible at the sample

surface. The surface passivation, which strongly alters the beam path of emitted light, may also contribute to the observed change.

5.5.2 Spectral comparison of DRL from wafers after various solar cell processing steps

In order to further investigate the apparent reduction of the emission intensities of D3 and D4 after the P-gettering step, the spectrum of the P-gettered sample (p-twin-central-P) was compared to spectra from samples p-twin-central and p-twin-central-P-H. Figure 5.21 displays the absolute spectra from the three samples, as well as the spectra normalised to the height of the BB emission line in each of the samples.

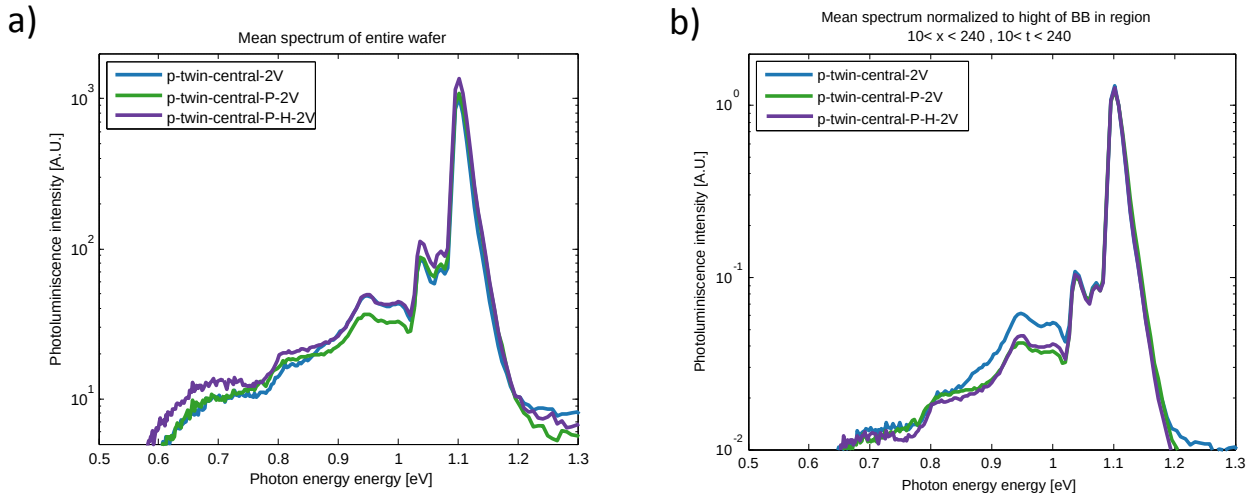


Figure 5.21: *Comparison of absolute spectra (a) and spectra normalized to the height of the BB peak (b) of wafers being subjected to surface passivation only (p-twin-central), additional P-gettering (p-twin-central-P) and additional P-gettering and H-passivation (p-twin-central-P-H). The numbers following the sample name indicates the laser excitation voltages used for the measurements.*

Figure 5.21 a) clearly shows that the emission intensities of D3 and D4 seem to decrease after the P-gettering process. After the H-passivation, the intensities seem to revert to their original levels (i.e. to the level they have in the sample which is only surface passivated). Note that the BB emission line is higher for the P-gettered sample than for the passivated sample, indicating that the decrease of the D3 and D4 emission intensities is not caused by a general decrease of charge carrier injection. Figure 5.21 b) further clarifies that the ratio of the D3 and D4 emission lines to the BB emission decreases after the P-gettering step. After the H-passivation, however, it seems to be more or less unchanged.

Similar results to those shown here are found also for measurements at other laser intensities from the central part of the p-twin ingot. For samples from other ingot positions, the results are not clear. These findings will be discussed in section 5.5.4.

5.5.3 Summary of section results

A spatial comparison of the D line emission from wafers in its as-cut state, after surface passivation, after additional P-gettering and after additional P-gettering and H-passivation reveal

no notable changes of the intensities or distributions of D1 and D2, or of emissions in the wavelength interval 0.73-0.92eV (subPL).

The total intensity of the D3 and D4 emission lines seems to decrease after the P-gettering process. A spatial comparison of the D line emission from wafers in its as-cut state, after surface passivation, after additional P-gettering and after additional P-gettering and H-passivation reveal no notable changes of the intensities or distributions of D1 and D2, or of emissions in the wavelength interval 0.73-0.92eV (subPL).

The total intensity of the D3 and D4 emission lines seems to decrease after the P-gettering process.

5.5.4 Discussion of section results

Abscence of subPL changes

Considering the substantial changes in the subPL signal observed by Johnston et al. [6] (see Figure 2.6), it is in the first place surprising that no similar changes are found in the samples investigated here. As explained in section 2.6.4, Johnston et al. observed the largest changes in the subPL signal after the antireflection coating (ARC) step, which consists of a deposition of SiN_x at elevated temperatures. This is similar to the H-passivation step included in this study (see section 3.1.1). Thus, one would primarily expect to observe changes from the P-diffused to the H-passivated sample shown in Figure 5.18.

However, comparing the details of the process through which the wafers have been, one finds that the ARC process employed by Johnston et al. includes a temperature step at 450˚C for 30 minutes, whereas our H-passivated wafers are heated to 815 ˚C for only 15 seconds. If the changes in the subPL signal observed by Johnston et al. are due to the temperature step, rather than to the deposition of SiN_x as such, the difference in process temperature can explain why no large changes are observed in our samples. This hypothesis is further strengthened by another publication from Johnston et al. [41], stating that the change of subPL signal may revert back to its initial intensities at temperatures above 500 ˚C (see section 2.6.4).

On the other hand, if the change of the subPL intensities observed by Johnston et al. are caused exclusively by the 450˚C temperature step, one would expect that the 400˚C temperature step through which all our samples – except from the as-cut wafers – have been as a part of the surface passivation (see section 3.1.1), would give similar effects. Accordingly, one would expect to observe differences between the subPL signals from the as-cut to the initially treated samples.

It must also be mentioned that the measurements done by Johnston et al. are conducted in room temperature, whereas our measurements are done at cryogenic temperatures. As explained in section 2.4.4, the effect of some defects on lifetime depends on temperature. Thus, we cannot exclude that differences between our samples could have been detected at room temperature.

Change of D3 and D4 due to solar cell processing

The intensities of D3 and D4 seem to decrease after the P-gettering process. As explained in sections 2.5.1 and 3.1.1, the gettering process includes a subsequent removal of the PSG formed due to due to the phosphorous diffusion. This process is therefore an *external gettering* and is supposed to decrease the total amount of iron and other metal contaminations in the sample. However, during the temperature step included in the P-diffusion process, besides the intended external gettering, some internal gettering might take place [3]. This means that impurities do not only diffuse from the bulk material to the PSG layer. They do also, to some extent, tend to decorate grain boundaries and sub-grain boundaries in the sample. Consequently, the concentration of interstitial iron and other impurities, as well as their precipitates, in the vicinity of grain boundaries may actually be higher after the P-gettering step than before. One may also speculate that a redistribution of iron and other impurities to smaller precipitates takes place during the gettering process. This would mean that even if the impurity concentration decreases (also in the vicinity of dislocations), a larger percentage of dislocations might be influenced by precipitates of iron or other impurities.

Based on the suggested model from the foregoing subchapter (section 5.4.5), we therefore speculate that the decrease of the D3 and D4 intensities after P-gettering is due to a formation or a redistribution of iron precipitates or precipitates of other impurities decorating the grain boundaries and sub-grain boundaries from which the emissions originate.

A similar explanation is mentioned by Nguyen et al.[3] regarding the intensity of the D1 emission line. As explained in section 2.6.3, he observed an increase of the D1 intensity after P-gettering of a mulicrystalline sample. He speculates the change to be due to increased concentration of impurities close to sub-grain boundaries after the P-gettering step, but finally concludes that "D1 is enhanced when the sub-grain boundaries are cleaned of metal impurities". He assumes, in other words, that in his samples the concentration of impurities close to the sub-grain boundaries is reduced as a result of the P-gettering.

However, during the P-gettering process, the amount of impurities diffusing to the emitter region (external gettering) compared to the amount of impurities precipitating on nearby sub-grain boundaries and dislocations (internal gettering) probably depends both on defect density and on the density of various contamination species. Accordingly, in our samples from the central ingot position, as opposed to the samples investigated by Nguyen et al., an internal gettering effect may have taken place, causing increasing concentration of impurities close to the grain boundaries, which, in turn, causes a decrease of the D3 and D4 emission intensities. The fact that this effect is not clearly observable for other ingot positions could be attributed to differences in contamination level and defect density between various areas in the ingot.

After the H-passivation, the emission intensities of D3 and D4 seem to increase to its original level. This might, in principle, be a consequence of hydrogen atoms saturating unsaturated bondings related to the non-radiative recombination centers introduced by the impurity precipitates (cf. Figure 5.17). However, after the H-passivation step also the BB emission peak increases, indicating that the total injection level of the sample increases as well. The increase of

the D3 and D4 emission lines from the P-gettered to the H-passivated sample may thus be an effect of the increased injection level only.

5.6 General discussion

Related to all subtopics in this work (i.e. differences in DRL between n- and p-type material, variations of DRL over the ingot height, dependence of DLR on metal contaminations and changes of DRL due to solar cell processing), features related to the intensities of D3 and D4 are found. No clear trends related to D1 and D2 are identified. Thus, it should be questioned why no such trends were observed in our samples, even if other authors report changes in the D1 and D2 intensities due to e.g. metal contaminations and solar cell processing [3, 5, 7] (see sections 2.6.3 and 2.6.4).

The samples investigated in this work are all from ingots solidified for research purposes, rather than from industrially produced ingots. As mentioned in section 3.1.1, these ingots are crystalized in G2 crucibles (i.e. crucibles that are smaller than those used for industrial solidification). Moreover, during the solidification process, the temperature ramps differed from those normally used in industrial crystallization [12]. As explained in section 2.4.2, during the solidification of an ingot, dislocations are formed due to strains in the crystal material. One may assume that due to quicker temperature variations and smaller ingot area (i.e. smaller are over which stress can be distributed), an enhanced amount of dislocations is present in these ingots compared to industrially produced mc-Si ingots. Considering that D3 and D4 are due to intrinsic dislocations [3, 39, 46], this may lead to an extraordinarily high intensity of D3 and D4 emissions in our samples, compared to industrially produced wafers.

The setup used for the multispectral measurements in this work has an unfavorable signal-to-noise ratio, causing challenges in detection of weak signals. Further, due to the reasons mentioned above, the emission intensities of D3 and D4 in our samples generally seem to be high compared to those of D2 and D1. One may therefore speculate that the lack of obvious trends related to the D1 and D2 emission lines simply is due to a low measurement sensitivity.

Further, it should be noticed that in most recent papers treating defect luminescence in relation to contaminations and wafer processing (e.g. [3, 5, 7]), the luminescence has been investigated at a microscopic scale. Spectra are mostly found for single points or very small regions rather than for entire wafers or wafer regions. Considering that D1 and D2 often originate from relatively small parts of the wafer, this may explain why our observed mean signals are too low to extract reliable information, whereas other authors report strong signals. Due to the uncertainties of our measurements related to laser inhomogeneity and background subtraction, studying and comparing spectra from small regions or points is not considered to be reliable.

6 CONCLUSIONS AND OUTLOOK

A new background subtraction method has been developed for the hyperspectral imaging setup at NMBU. This method significantly enhances the possibilities for extracting spectral and spatial information from the recorded hyperspectral images, and thus makes the hyperspectral camera a promising tool for investigating radiative defects in solar cell material. Limitations regarding the reliability of the measurements related to inhomogeneous laser excitation are discussed.

Defect luminescence of multicrystalline n- and p-type sister wafers with the same crystal structure were compared. The detected emission lines were found for the same energies for both material types, and the emissions originate from corresponding wafer regions, indicating that the origin of the D lines are independent of material type.

Differences in the intensity ratio between the D3 and D4 emission were found between wafers of different material types (n/p) as well as between wafers from different ingot positions. This contradicts the idea that D3 constitutes the phonon replica of D4, suggested by some authors [33], but confirms experimental observations by Burud et al.[47] and Nguyen et al. [3]. We suggest that the observed differences in spectral shape among the investigated wafers may be explained by the recombination mechanisms mainly limiting the charge carrier lifetime of the sample. The DRL spectra of samples or regions whose lifetime probably is limited by recombination active dislocation clusters, seem to be characterized by a symmetrical shape regarding the intensities of D3 and D4 emissions, whereas the DRL spectra for wafers or regions whose lifetime probably is limited by iron or other impurities seem to have a reduced emission intensities for D4 compared to D3.

Based on a comparison of defect luminescence images with images of interstitial iron concentration, a reduction of the D3 and D4 in regions with high iron concentration is observed. We suggest that the D3 and D4 emission intensities partly are suppressed by precipitates of iron or of other contamination species. This is in line with findings by Tajima et al. [2]. Our observation of decreasing intensities of the D3 and D4 emissions after P-gettering matches this hypothesis if we assume that some internal gettering or redistribution of precipitates takes place along with desired external gettering.

We observe no changes in the subPL signal due to solar cell processing steps that are comparable to those observed by Johnston et al. [6]. We suggest that the deviation from their results is due to differences in the wafer processing procedure or differences in the sample temperature during measurement.

Table 6.1 provides an overview of our main findings and suggested explanations.

Table 6.1: **Summary of observations and suggested explanations.**

Observation	**Conclusion/Suggested explanation**
D line emissions found in the same spatial regions and the same wavelength intervals in n-type as well as p-type material	The origins of the D line emissions are independent of doping type.
Lower intensity of D4 in p-type than in n-type wafers	Wafer regions being limited by recombination active dislocation clusters are characterized by comparable intensities of D3 and D4. Wafer regions being limited by impurities are characterized by reduced D4 emission intensity compared to D3 emission intensity.
Increasing intensity of D4 along ingot growth direction	
Negative correlation of D3 and D4 with iron or other impurities	D4 and D3 emissions are suppressed due to precipitates decorating the dislocations from which the emissions originate.
Decrease in the emission intensities of D3 and D4 after P-gettering	

Further investigations are needed in order to strengthen the hypotheses suggested here. For example, by combining hyperspectral imaging at higher spatial resolution with µ-PL (micro photoluminescence) and iron imaging, the dependence of the spectral shape on lifetime limitation as well as the apparent negative correlation between D3/D4 and iron or other impurities might be more closely studied. By means of such experiments, relations to the D1 and D2 emission intensities might also be found. In order to omit challenges related to bleeding and overlapping peaks, the hyperspectral imaging being a part of the investigations should preferably be conducted at unpassivated wafers.

Hyperspectral imaging is a promising method for the detection of DRL. We have utilized the unique combination of hyperspectral imaging with advanced PL based techniques for lifetime and iron imaging in order to broaden our understanding of DRL in mc-Si. Hopefully, this knowledge can contribute to the development of more efficient mc-Si solar cells.

7 REFERENCES

1. *International Technology Roadmap for Photovoltaic (ITPRV), 2014 Results* ITPRV, Editor. 2015.
2. Tajima, M., et al., *Deep-level photoluminescence due to dislocations and oxygen precipitates in multicrystalline Si.* Journal of Applied Physics, 2012. **111**(11): p. 113523.
3. Nguyen, H.T., et al., *Micrometer-Scale Deep-Level Spectral Photoluminescence From Dislocations in Multicrystalline Silicon.* Photovoltaics, IEEE Journal of, 2015. **PP**(99): p. 1-6.
4. Gundel, P., et al., *Micro-photoluminescence spectroscopy on metal precipitates in silicon.* physica status solidi (RRL) - Rapid Research Letters, 2009. **3**(7-8): p. 230-2.
5. Tajima, M., et al., *Photoluminescence analysis of iron contamination effect in multicrystalline silicon wafers for solar cells.* Journal of Electronic Materials, 2010. **39**(6): p. 747-50.
6. Johnston, S., et al., *Correlating Multicrystalline Silicon Defect Types Using Photoluminescence, Defect-band Emission, and Lock-in Thermography Imaging Techniques.* Photovoltaics, IEEE Journal of, 2014. **4**(1): p. 348-354.
7. Nguyen, H.T., *Effects of Solar Cell Processing Steps on Dislocation Luminescence in Multicrystalline Silicon* Energy Procedia 2015 (preprint), 2015.
8. Flø, A., *Hyperspectral Imaging as a tool for Characterization of Multicrystalline silicon Wafers* in *Faculty of Environmental science and Technology*. 2014, NMBU: Ås.
9. Giesecke, J.A., et al., *Minority carrier lifetime imaging of silicon wafers calibrated by quasi-steady-state photoluminescence*. Solar Energy Materials and Solar Cells, 2011. **95**(3): p. 1011-1018.
10. Schubert, M.C., H. Habenicht, and W. Warta, *Imaging of metastable defects in silicon.* IEEE Journal of Photovoltaics, 2011. **1**(2): p. 168-73.
11. Schindler, F., et al., *Potential Gain in Multicrystalline Silicon Solar Cell Efficiency by n-Type Doping.* Photovoltaics, IEEE Journal of, 2015. **5**(2): p. 499-506.
12. Schubert, M.C., et al., *Impact of Impurities From Crucible and Coating on mc-Silicon Quality—the Example of Iron and Cobalt.* Photovoltaics, IEEE Journal of, 2013. **3**(4): p. 1250-1258.
13. Schubert, M.C., et al., *Quantitative iron concentration imaging*. Solid State Phenomena, 2010. **156-8**: p. 407-12.
14. Sze, S.M., *Semiconductor devices: physics and technology*. Vol. 2nd. 2002, Singapore: John Wiley & Sons 523.
15. Rein, S., *Lifetime Spectroscopy - A Method of Defect Characterization in Silicon for Photovoltaic Applications*. Vol. 1st. 2005, Heidelberg: Springer-Verlag Berlin Heidelberg. 489.
16. Habenicht, H., *Charakterisierung leistungsmindernder Defekte und deren Umverteilung während der Herstellung von multikristallinen Silizium-Solarzellen*. 2011, Alber-Ludwigs-Universität,: Freiburg im Breisgau. p. 255.
17. Chen, J., *Physics of Solar Energy*. 1st. ed. 2011, New Jersey: John Wiley & Som´ns. 326.
18. Dziewior, J. and W. Schmid, *Auger Coefficients for Highly Doped and Highly Excited Silicon.* Applied Physics Letters, 1977. **31**(5): p. 346-348.
19. Shockley, W. and W.T. Read, *Statistics of the Recombinations of Holes and Electrons.* Physical Review, 1952. **87**(5): p. 835-842.
20. Hall, R.N., *Electron-hole recombination in germanium.* Physical Review, 1952. **87**(2): p. 387.
21. Istratov, A.A., H. Hieslmair, and E.R. Weber, *Iron and its complexes in silicon.* Applied Physics A (Materials Science & Processing), 1999. **A69**(1): p. 13-44.

22. Niewelt, T., *Analyse von Defekten in Kristalinem Silicium*, in *Fachbereich 11 Material- und Geowissenschaften, Bereich Materialwissenschaften*. 2012, Technische Universität Darmstadt (TUD): Darmstadt
23. Ashcroft, N. and D. Mermin, *Solid State Physics*. 1.st ed. 1976, Philadelphia: Saunders College. 826.
24. Schön, J., *Modellierung von Prozessschritten zur umgelagerung rekombinasjonsaktiver Defekte in kristallinem Silizium*. 2011, Universität Konstanz. p. 169.
25. Schindler, F., et al. *The potential of multicrystalline n-type silicon for high efficiency solar cells*. in *Proceedings of the 29th European Photovoltaic Solar Energy Conference and Exhibition*. 2014. Amsterdam, The Netherlands.
26. Kwapil, W., *Alternative materials for crystalline silicon solar cells - Risks and implications*, in *Universität Konstanz Fakultät für Physik*. 2010.
27. Green, M., *Solar Cells - Operating Principle, Technology and System Application* 1992, Kensington: University of new south Wales
28. Geerligs, L. and M. D., *Dynamics of light-induced FeB pair dissociation in crystalline silicon.* Applied Physics Letters, 2004. **85**(22).
29. Lauer, K., et al., *Iron gettering and slip dislocations in Czocralski silicon.* Solid State Phenomena, 2011. **178-179**: p. 211-216.
30. Chen, J. and T. Sekiguchi, *Carrier Recombination Activity and Structural Properties of Small-Angle grain boundaies in Multicrstalline silicon* Japanese journal of Applied Physics, 2007. **46**.
31. *Impurity Diffusion and Gettering in Silicon*. Materials Research Society, ed. B. Fair, C. Pearce, and J. Washburn. Vol. 36. 1984: Pittsburg, Pensylvania.
32. Kveder, V., M. Kittler, and W. Schröter, *Recombination activity of contaminated dislocations in silicon:\quad{}A model describing electron-beam-induced current contrast behavior.* Physical Review B, 2001. **63**(11): p. 115208.
33. Arguirov, T., et al., *Temperature behaviour of extended defects in solar grade silicon investigated by photoluminescence and ebic.* Materials Science and Engineering B, 2003. **102**(1-3): p. 251-6.
34. Karzel, P., et al., *Influence of hydrogen on interstitial iron concentration in multicrystalline silicon during annealing steps.* Journal of Applied Physics, 2013. **113**(11): p. 114903.
35. Flø, A., et al., *Distribution of radiative crystal imperfections through a silicon ingot.* AIP Advances, 2013. **3**(11): p. 112120.
36. Drozdov, N.A., A.A. Patrin, and V.D. Tkachev, *Recombination Radiation on Dislocations in Silicon.* Jetp Letters, 1976. **23**(11): p. 597-599.
37. Arguirov, T., *Electro-optical properties of dislocations in silicon and their possible application for light emitters*. 2007, Brandenburgischen Technischen Universität Cottbus.
38. Ostapenko, S., et al., *Defect monitoring using scanning photoluminescence spectroscopy in multicrystalline silicon wafers.* Semiconductor Science and Technology, 2000. **15**(8): p. 840-8.
39. Tajima, M., *Spectroscopy and Topography of Deep-Level Luminescence in Photovoltaic Silicon.* Photovoltaics, IEEE Journal of, 2014. **4**(6): p. 1452-1458.
40. Lausch, D., et al., *Classification of recombination active defect structures in multicrystalline silicon solar cells.* Energy Procedia, 2011. **8**(0): p. 28-34.
41. Johnston, P.e.a., *Comparison of Photoluminescence Imaging on Starting Multi-Crystalline Silicon Wafers to Finished Cell* in *IEEE Photovoltaic Specialists COnference*. 2012, NREL: Austin, Texas.
42. Sauer, R., et al., *Dislocation-Related Photoluminescence in Silicon.* Applied Physics a-Materials Science & Processing, 1985. **36**(1): p. 1-13.

43. Tajima, M. and Y. Matsushita, *Photoluminescence related to dislocations in annealed Czochralski-grown Si crystals.* Japanese Journal of Applied Physics, 1983. **22**(9): p. L589-91.
44. Kveder, V.V., et al., *Dislocation-related electroluminescence at room temperature in plastically deformed silicon.* Physical Review B, 1995. **51**(16): p. 10520-6.
45. Lausch, D., et al., *Classification of recombination-active defects in multicrystalline solar cells made from upgraded metallurgical grade (UMG) silicon.* Solid State Phenomena, 2011. **178**: p. 88-93.
46. Lightowlers, E.C. and V. Higgs, *Luminescence associated with the presence of dislocations in silicon.* Physica Status Solidi A, 1993. **138**(2): p. 665-72.
47. Burud, I., A.S. Flø, and E. Olsen, *On the origin of inter band gap radiative emission in crystalline silicon.* AIP Advances, 2012. **2**(4): p. 042135.
48. Kleiber, A., *Materiallimitierungen in multikristallinem n-Typ Siliziumnach verschiedenen solarzelltypischen Hochtemperaturprozessen*, in *Technische Fakultät*. 2014, Albert-Ludwigs-Universität Freiburg im Breisgau. p. 74.
49. Knörlien, M., *Rekombinationseigenschaften von Kristalldefekten in multikristallinem n-Typ Silizium*, in *Fakultät für Maschinenbau, Verfahrens- und Energietechnik*. 2014, Technische Universität Bergakademie Freibeg. p. 61.
50. Giesecke, J., *Messung von Minoritätsladungsträger-Diffusionslängen in Silicium-Solarzellen mit Lumineszenzmethoden*, in *Fachbereich Physik*. 2008, Universität Freiburg: Freiburg, Germany. p. 120.
51. Giesecke, J., et al. *Spatially resolved carrier lifetime calibrated via quasi-steady-state photoluminescence*. in *Proceedings of the 1st International Conference on Silicon Photovoltaics*. 2011. Freiburg, Germany: Elsevier Energy Procedia.
52. Zoth, G. and W. Bergholz, *A fast, preparation-free method to detect iron in silicon.* Journal of Applied Physics, 1990. **67**(11): p. 6764-71.
53. Macdonald, D., J. Tan, and T. Trupke, *Imaging interstitial iron concentrations in boron-doped crystalline silicon using photoluminescence.* Journal of Applied Physics, 2008. **103**(7): p. 073710.
54. Macdonald, D., et al., *Doping dependence of the carrier lifetime crossover point upon dissociation of iron-boron pairs in cystalline silicon.* Applied Physics Letters, 2006. **89**(14): p. 142107.
55. Tipler, P. and G. Mosca, *Physics for Scientists and Engineers* 6th ed. 2008: W. H. Freem,an and Company. 1412.
56. Coherent, *Operator's Manual Coherent Magnum Laser II*. Wilsonville.
57. Coherent *Lasiris Magnum II - Product sheet*
58. Fremme, A., *Hyperspectral photoluminescence scans of mc- Si wafers analysed through Multivariate Curve Resolution (MCR)*, in *Institutt for matematiske raealfag og teknologi*. 2014, NMBU: Aas
59. Mehl, T., *Karakterisering av mulikrystallinske solceller ved bruk av hyperspektral avbildning*, in *Institut for Matematiske realfag og teknologi*. 2014, NMBU: Ås. p. 77.
60. Jaumot, J., et al., *A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB.* Chemometrics and Intelligent Laboratory Systems, 2005. **76**(1): p. 101-110.
61. Zhang, X. and R. Tauler, *Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging.* Analytica Chimica Acta, 2013. **762**(0): p. 25-38.
62. Lay, D., *Linear Algebra and its Applications*. 3rd ed. 2006.
63. Tauler, R. and A. de Juan *Multivariate Curve Resolution Tasks*.
64. AS, C.S., *The Unscrambler Methods*, in *The unscrabler User Manual*, C.S. AS, Editor.

65. Mehl, T., *Karaktrisering av multikrystallinske solceller ved bruk av hyperspektral avbildning*, in *Fakultet for miljøvitenskap og teknologi*. 2014, NMBU: Ås.
66. Schön, J., et al., *Understanding the distribution of iron in multicrystalline silicon after emitter formation: theoretical model and experiments.* Journal of Applied Physics, 2011. **109**(6): p. 063717.
67. Seibt, M., et al., *Electronic states at dislocations and metal silicide precipitates in crystalline silicon and their role in solar cell materials.* Applied Physics A, 2009. **96**(1): p. 235-253.
68. Inoue, M., et al., *Microscopic and spectroscopic mapping of dislocation-related photoluminescence in multicrystalline silicon wafers.* Journal of Materials Science: Materials in Electronics, 2008. **19**(1): p. 132-134.

APPENDIXES

A MATLAB FUNCTION FOR BACKGROUND CORRECTION

The following function performs the background correction which was developed during this work. This script is developed in cooperation with Torbjørn Mehl .

```
%----------------------------
% load_subtract_correct.m
%----------------------------

% Input arguments:
% fn              - name of measurement
% BGcorrversion   - desired version of the background subtraction/correction
scheme
    % Version 0 => old backgroundsubtractionscheme (no correction)
    % Version 3 => newest version of new bacggrouncorrectioscheme

% Output arguments:
% a1              - hypercube with corrected background
% BGindex         - index indicating the utilized version of background
% correctioncurve - vector indicating the development of the background
%                   along the spatial dimension "t" of the image
% maxcorr         - number giving the maximal value of the background
%                   correction for the actual hypercube


 function
[a1,BGindex,correctioncurve,maxcorr]=load_subtract_correct(fn,BGcorrversion)
  % Reading from header file
  thdr = readenvihdr([fn,'.hdr']);
  % Finding the correct dimensions for the hypercube
  nl = str2num(thdr.lines);
  ns = str2num(thdr.samples);
  nb = str2num(thdr.bands);

  %Reading image from file
  a1 = multibandread([fn,
'.raw'],[nl,ns,nb],'uint16',0,thdr.interleave,'ieee-le');

  % Finding the background from the dark part at the end of the image:
  dx_BGend = 49;
  BGend=squeeze(median(a1(nl-dx_BGend:nl,:,:),1));

  % Defining the part of the image that should be used as a reference value
  % to scale the background:
  lambda_min=237;
  lambda_max=256;
  y_min=70;
  y_max=250;
```

```
  % Finding the reference value for the end background
  BGend_ref = median(median(BGend(y_min:y_max,lambda_min:lambda_max),1),2);

  % Finding the reference value for each frame
  frame_ref = median(median(a1(:,y_min:y_max,lambda_min:lambda_max),2),3);
  frame_ref = smoothn(frame_ref);

  % Subtracting the end background from each frame in the entire image
  for i=1:nl
    a1(i,:,:)=squeeze(a1(i,:,:))-BGend;
  end

  % Loading the general background correction matrix
  load('BGdifcorrmatrTM');

  % Finding its reference value
  BGcorrmatr_ref =
median(median(BGcorrmatrTM(y_min:y_max,lambda_min:lambda_max),1),2);

  % Creating the "extra backgroundsubtraction tensor" and subtrcting it
  % from the images
  %-----------------------------------------------------
  correctioncurve = frame_ref-BGend_ref;
  frame_ref = (frame_ref - BGend_ref)/BGcorrmatr_ref;

  % Version 0 => old background subtraction scheme (no correction)
  % Version 3 => newest version of new background correction scheme

  if BGcorrversion == 0;
      for i=1:nl
       a1(i,:,:)=squeeze(a1(i,:,:)) - ones(size(BGend))*correctioncurve(i) ;
      end
      BGindex = 'BGcorrsmooth0'

  elseif BGcorrversion == 3;

     for i=1:nl
       a1(i,:,:)=squeeze(a1(i,:,:)) - BGcorrmatrTM(:,:)*frame_ref(i) ;
     end
     BGindex = 'BGcorrsmooth3'
  end

  % Finding the maximal value of the background correction for reference
  maxcorr = -(max(abs(correctioncurve.')) .* (2*(abs(min(correctioncurve.'))
< max(correctioncurve.'))-1));
end
```

B FUNCTION FOR EXTRACTING D LINE IMAGES

This function creates integrated images of the D-lines, based on the integration limits given in Table 4.1.

```
%----------------
% D_lines.m
% ---------------
% Input argument:
% www             - Hypercube from which the background is subtracted

% Output arguments:
% Dlines          - Structure array containing one field for each of the six D
                    line images, as well as the names of the peaks.

function[Dlines]= D_lines(www)
% Creating a structure array, with "peaks" as one field
% Defining the peak names in this field:

Dlines. peaks =     {'D0'; 'D1'; 'D2';'D3';'D4';'BB'};

% Defining integration limits
Dlines.minlim =    [ 114, 91, 74, 60, 49, 30 ];
Dlines.maxlim =    [ 153, 99, 83, 66, 53, 35 ];


% creating defect images
dim = size(www) % Finding the dimensions of the hypercube

defect_images = zeros(dim(1),dim(2),length(Dlines.peaks));

for k = 1:length(Dlines.peaks);
    for i = 1:dim(1);
        for j = 1:dim(2);

           defect_images(i,j,k) =
           sum(www(i,j,Dlines.minlim(k):Dlines.maxlim(k)),3);

        end
     end
 end

% Assigning the D line images to correct filed in structure array
Dlines.D0 = defect_images(:,:,1);
Dlines.D1 = defect_images(:,:,2);
Dlines.D2 = defect_images(:,:,3);
Dlines.D3 = defect_images(:,:,4);
Dlines.D4 = defect_images(:,:,5);
Dlines.BB = defect_images(:,:,6);

end
```

C SCRIPT FOR COMPARING SPECTRA FROM DIFFER WAFERS OR WAFER REGIONS

The following script was developed in order to easily compare the same region(s) at many samples. The script requires that the hypercubes to be investigated already are treated by the function “load_subtract_correct” , i.e. that the correct background already is subtracted. The hypercubes must be placed in one folder (inputdir), with no other files.

```
%% User inputs
%-----------------------------------------------------------------
inputdir= 'd:\Defect_imaging\Analysed data\Spectra comparison\Change
due to processing steps\n-typ\inputs\'
outputdir= 'd:\Defect_imaging\Analysed data\Spectra comparison\Change
due to processing steps\n-typ\'


%% Loading data
%-----------------------------------------------------------------
addpath (inputdir);
flist = dir(fullfile(inputdir)); % get structure array containing all
filenames in inputdir

% Placing names in list
for i=3:length(flist);
  filename{i-2} = flist(i).name;
end

% creating a structure array with the measurement names as one and the
hypercubes as a second field
results= struct;
  for i=1:length(filename);
      results(i).name=filename(i)
      results(i).www = load([cell2str(filename(i))])
  end
rmpath (inputdir);


% Showing a BB image of each of the wafers with imtool
 for i =1 :length(filename)
        imtool(results(i).www.www(:,:,31),[],'Colormap',[jet],...
        'InitialMagnification','fit');
 end


% Making image for each of the D lines and saving them to athird field
in the structure arrray
for i = 1:length(results);
```

```
    results(i).Dlines = D_lines(results(i).www.www,'TM_wide__')
end
```

```
%% Defining color

%---------------------------------------
color = zeros(length(results),3);
color(1,:) = [220 20 60];
color(2,:) = [225 20 147];
color(3,:) = [199 21 133];
color(4,:) = [148 0 211];
color(5,:) = [65 105 225];
color(6,:) = [0 0 205];
color(7,:) = [200 0 200];

color = color./256



%% Defining regions manually
% ---------------------------------------
regions =struct;

% Central region
regions(1).xstart = 100;
regions(1).xstop = 130;
regions(1).tstart = 10;
regions(1).tstop = 240;

% Right edge region
regions(2).xstart =200;
regions(2).xstop = 230;
regions(2).tstart = 10;
regions(2).tstop = 240;

% Left edge region
regions(3).xstart =10;
regions(3).xstop = 40;
regions(3).tstart = 10;
regions(3).tstop = 240;


% Entire wafer
regions(8).xstart = 10;
regions(8).xstop = 240;
regions(8).tstart = 10;
regions(8).tstop = 240;



for i=1:length(regions);
    regions(i).deltat = regions(i).tstop-regions(i).tstart;
    regions(i).deltax = regions(i).xstop-regions(i).xstart;
end
```

```
%% Defining regions "automatically"
%---------------------------------------

% Creating structure array for the regions
regions =struct;

% Manual input for the maximal size of the regions and the region
whidt:
Xstart = 210;
Xstop =  300;
deltax =  7;
xjump = 8 ;


 regions(1).xstart=Xstart;
 regions(1).tstart= Tstart;

% Creating regions with the given whidt
k = 1;
xstop = 1;

while xstop < Xstop
    regions(k).xstart= Xstart+(k-1)*xjump;
    regions(k).xstop = regions(k).xstart+ deltax;

    regions(k).tstart= Tstart+(k-1)*tjump;
    regions(k).tstop = regions(k).tstart+ deltat;

     xstop_mm= (regions(k).xstop-xedge)*delta_mm
     k=k+1;
end


for i=1:length(regions);
    regions(i).deltat = regions(i).tstop-regions(i).tstart;
    regions(i).deltax = regions(i).xstop-regions(i).xstart;
end



%% Showing the regions on a BB image of each wafer in folder
%-----------------------------------------------------------
for i = 1:length(results);
     % Creating figure
     figure;
     clims = [1 8000];
     imagesc(results(i).Dlines.BB, clims);
     colorbar;
     axis image;
     axis off;
     colormap(jet);
```

```
      % Making correct name
      name = regexprep(results(i).name,'_','-')
      name = cell2str(name)
            name = regexprep(name,'-BG(\S*)','')  % removing the part of
      the name from BG and onwards
      title(name);

      % Highlighting all the regions
            for j = 1 :length(regions);

                  rectangle('Position',[regions(j).xstart,regions(j).tst
                  art,regions(j).deltax,regions(j).deltat],'LineWidth',2
                  .5,'LineStyle','-','edgecolor',color(2,:))
            end

      %  saving figures as fig and png
      saveas(gcf,[outputdir,'_spektrum_region_',regexprep(name,'\.','p'
      ),'.fig']);
      saveas(gcf,[outputdir,'_spektrum_region_',regexprep(name,'\.','p'
      ),'.png']);
 end



%% Calculating mean spectra for each region
%-----------------------------------------
for i = 1:length(results);
    results(i).areaspectre = zeros (256,length (regions));
    results(i).normspectre = zeros (256,length (regions));
end

for i =1:length(results);
     for j=1:length(regions);

           results(i).areaspectre(:,j)=squeeze(sum(sum(results(i).www.w
           ww(regions(j).tstart:regions(j).tstop,regions(j).xstart:regi
           ons(j).xstop,:),1),2))/((regions(j).tstop-
           regions(j).tstart)*(regions(j).xstop-regions(j).xstart));
     end
end

% calculating spectra normalized to the height of the BB peak
%------------------------------------------------------------
 for i =1:length(results);
     for j=1:length(regions);

     results(i).normspectre(:,j)=
     results(i).areaspectre(:,j)./(median(results(i).areaspectre(30:32
     ,j)));

    end
```

```
end
%% Plotting mean spectrum of the chosen regions
%-------------------------------------------------
%(Here one figure per region and one plotted line per wafer.
% With small changes, other combinations can be made)

load ('Energi_specimOct2014')


shortnames='';
colorindex = 0;

for j = 1:length(regions);
  figure;
     for i = 1:length(results)

           % Plotting
           colorindex=colorindex+1;
           semilogy(Ev,
           results(i).areaspectre(:,j),'Color',color(colorindex,:),'Lin
           eWidth',3);
           hold on

           % Ensuring correct name
           name = regexprep(results(i).name,'_','-')
           name = regexprep(name,'-BG(\S*)','')  % removing the part of
           the name from BG and onwards
           shortnames  =[shortnames, name];
           grid off;
           ylim([5 1e3
           xlim([0.5 1.3]);
           % Saving figure
           saveas(gcf,[outputdir,'meanspectrum_withname',int2str(j),
           int2str(i),'.fig']);
           saveas(gcf,[outputdir,'meanspectrum_withname',int2str(j),
           int2str(i),'.png']);



           i =i+1
     end
   j = j+1
end
```

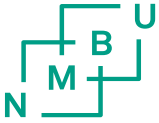
```
%% Plotting spectra normalized to the height of BB
%-----------------------------------------------------
shortnames='';
colorindex = 0;

for j = 1:length(regions);
  figure;
     for i = 1:length(results)

           % Plotting
           colorindex=colorindex+1;
           semilogy(Ev,
           results(i).normspectre(:,j),'Color',color(colorindex,:),'Lin
           eWidth',3);
           hold on

           % Ensuring correct name
           name = regexprep(results(i).name,'_','-')
           name = regexprep(name,'-BG(\S*)','')  % removing the part of
           the name from BG and onwards
           shortnames  =[shortnames, name];
           grid off;
           ylim([0.001 2])
           xlim([0.5 1.3]);
           % Saving figure
           saveas(gcf,[outputdir,'normspectrum_withname',int2str(j),
           int2str(i),'.fig']);
           saveas(gcf,[outputdir,'normspectrum_withname',int2str(j),
           int2str(i),'.png']);

           i =i+1
     end
   j = j+1
end
```



Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no