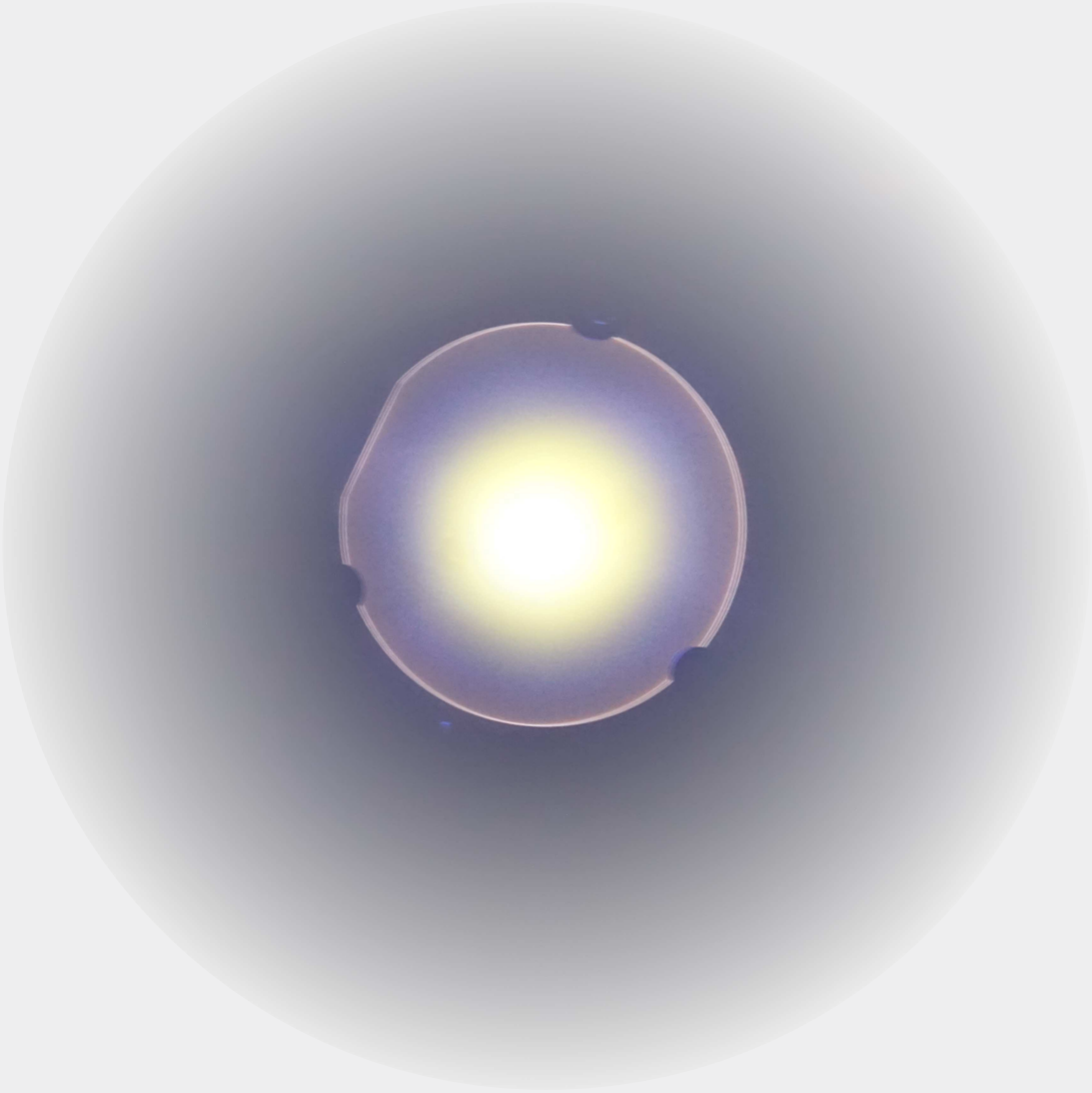
Characterization of GaN thin films and growth by plasma-assisted molecular beam epitaxy



Thore Aunsborg and Rasmus Hjelmgart
Master's thesis - June 2016
Supervised by Kjeld Pedersen
Department of Physics and Nanotechnology
Aalborg University

AALBORG UNIVERSITY
DENMARK

Characterization of GaN thin films and growth by plasma-assisted molecular beam epitaxy

Master's thesis by

Thore Aunsborg and Rasmus Hjelmgart
Department of Physics and Nanotechnology
Skjernvej 4A, DK-9220, Aalborg Øst

Supervised by

Kjeld Pedersen
Department of Physics and Nanotechnology
Skjernvej 4A, DK-9220, Aalborg Øst

Abstract

This thesis is concerned with the investigation of the properties of GaN thin films. The basic theory of PL spectroscopy, Raman spectroscopy, and Hall measurements is accounted for. This is then applied to GaN, which also includes the calculation of the wurtzite band structure. General theory behind PAMBE growth of GaN is described, including different GaN growth schemes, and how these can be used to optimize the growth.

Several GaN samples were grown by PAMBE, and were along with MOCVD samples characterized with respect to surface, electrical, and optical properties. The capabilities of different characterization methods are discussed with respect to GaN. Here it is argued that the interplay between several methods is necessary to give an accurate and reliable sample description.

Danish abstract

Denne afhandling undersøger GaN tyndfilms egenskaber. Den grundlæggende teori bag PL spektroskopi, Raman spektroskopi og Hall målinger er forklaret. Dette betragtes i forhold til GaN, hvilket ogsåindeholder en beregning af båndstrukturen for wurtzite. Generel teori bag PAMBE GaN syntese er beskrevet inklusiv forskellige GaN syntesemåder, og hvordan disse kan anvendes til at optimere syntesen.

Flere GaN prøver er blevet groet med PAMBE, som sammen med MOCVD prøver er blevet karakteriseret med hensyn til overflademæssige, elektriske, og optiske egenskaber. De forskellige karakteriseringsmetoders anvendelighed er diskuteret med hensyn til GaN. Der bliver her argumenteret for at sammenspillet mellem forskellige metoder er nødvendigt for at give en præcis og pålidelig prøvebeskrivelse.

Concluded June 15, 2016
Number of pages: 144 (152)
Number of appendices: 4

Preface

This Master's thesis was written by the authors during the period from September 1, 2015 - June 15, 2016, as a partial fulfilment of the MSc in Engineering in Nanomaterials and Nanophysics at the Department of Physics and Nanotechnology at Aalborg University.

Sources in this thesis are denoted by numbers in square brackets, i.e. [#], with the references numbered in order of appearance in the thesis and listed in this order in the bibliography. If the reference is placed before a full stop, it refers to that sentence only. When placed after a full stop, it refers to the entire paragraph. Figures and formulas are numbered according to the chapter in which they are found, with the first figure in chapter 2 denoted by 2.1, the second by 2.2, etc. All graphs and images are referred to as figures. Figures without citation are made by the authors.

The authors wish to thank those who supported the work in this thesis, including the following people from the Department of Physics and Nanotechnology at Aalborg University:

Kim H. Jensen, for his help with the cleaning of sapphire substrates and passivation of GaN samples.

Hans Nilsson, for his help with the making of an aluminium table for Hall measurements.

Thomas G. Pedersen, for teaching us the empirical pseudopotential method for the modelling of the band structure of GaN.

Mads Brincker, for helping with the XRD measurements.

Mariia M. Rozhavskaia, for providing the MR MOCVD GaN samples.

Peter K. Kristensen, with whom we worked closely in all parts of the MBE growth process, and to whom we are grateful for his assistance with lab work in general.

Finally, thank you to **Piotr Caban** of the Institute of Electronic Materials Technology, Warsaw, for providing the PC MOCVD samples.

The front page image is a photograph of the yellow luminescence from a 2-inch MOCVD-grown Si-doped GaN-on-sapphire sample and the blue luminescence from a piece of paper, excited by the 351 nm line of an Ar-ion laser.

Thore Aunsborg
t.aunsborg@gmail.com

Rasmus Hjelmgart
rhjelmgart@gmail.com

Table of contents

1 Introduction

Gallium nitide (GaN) is perhaps the most promising material for future electronics, and it could likely become the most important semiconductor next to Si. GaN has numerous applications within the semicondcutor industry, where its fascinating application as a high electron mobility transistor (HEMT) as well as its optoelectronic applications has paved the way for much recent research into GaN thin films. The high mobility is achieved from a 2D electron gas formed in the AlGaN/GaN heterostructure, where the modulation doping limits impurity scattering due to dopants. The wide band gap of GaN is another desired property. In high-power electronic, this allows for the more economically viable solution of GaN grown on sapphire compared to the more expensive SiC devices. A comparison of GaN HEMT device- and material properties with respect to other common semiconductors is shown in figure 1.1, presenting very desirable electronic characteristics. The figure shows that GaAs-HEMT structures can achieve much higher electron mobilities than GaN-HEMTs. However, for high-power electronic applications, GaAs-HEMTs suffer from low thermal conductivities and low breakdown voltage, which makes GaN-HEMTs a very promising alternative for these applications. [1, 2, 3]

In addition to this, for optoelectronic applications, GaN laser diodes could optimize devices for communication in the ultraviolet regime and provide more sustainable solutions for UV laser diodes. Furthermore, GaN allows for a wide range of tune-ability of its band gap when doped e.g. with Al or In (III-V materials are already the leading materials in todays optoelectronics market). In either case, these applications require low dark currents and a low density of recombination centres located within the semiconductor band gap. [1, 3]

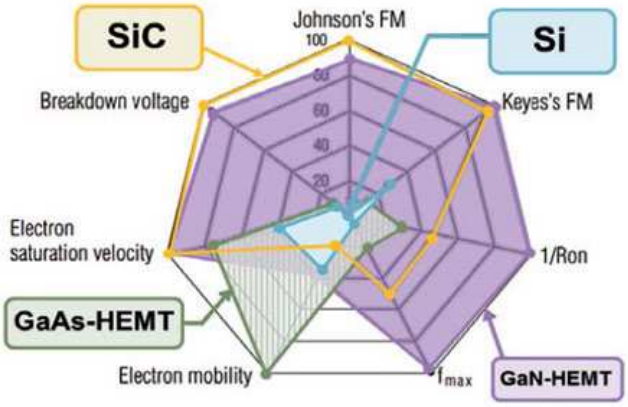


FIGURE 1.1: *Comparison of figure of merit (FM) and other material performance characteristics for some high power transistors. Taken from [4].*

However, the as-grown GaN thin films often suffer from both structural and electronic defects, which reduce their efficiency in the desired applications, where the defects generally have a negative influence on the properties as well as the lifetime of the device. As of now, there are indeed high quality GaN based devices spreading on the market, but their cost must be reduced before GaN electronics will have the potential to invade more market sectors. One extremely interesting solution is the integration of GaN on silicon, for which the current solutions require complicated growth schemes utilizing several epitaxial nucleation layers to buffer for the growth of a fairly good GaN film. Furthermore, solutions for GaN/Si growth would significantly reduce the complications involved in the integration of GaN devices in current microelectronic applications. Nevertheless, GaN/Si growth has not yet matured into a viable technological solution, because of the considerable difficulties associated with the growth. Therefore, sapphire substrates are still the most cost-efficient alternatives to native GaN substrates. [1, 5]

The structural and electronic defects in GaN have been intensively studied during the past 20 years. However, for either case some occurrences remain a puzzle and more research is still needed to fully understand the characteristics and formation mechanisms of these defects. This naturally calls for more thorough investigation of the properties of produced GaN thin films, where the comprehension between different characterization techniques could be the missing piece. It is well known that epitaxial growth of GaN on a lattice mismatched substrate results in strain and the formation of structural defects as a relaxation mechanism in order to remove the strain induced by the lattice mismatch. Furthermore, it is essential to consider the thermal expansion coefficients for all epitaxial layers, since the growth of high quality crystalline GaN is achieved at elevated temperatures. A large mismatch in thermal expansion coefficients will induce more material strain upon cooling, and thus more structural defects that reduce the crystalline quality of the film. [6, 7]

Other crystal imperfections include electronic defects through the inclusion of impurities in the crystal lattice, and also the vacant position of e.g. a N or Ga atom. Both may serve as donors or acceptors and can therefore attract electrons or holes excited with a different purpose, thus reducing both optical and electrical properties of the film. Additionally, it has been widely suggested that vacancies and impurities in the crystal lattice may cause local strain fields, which in turn can result in more structural defects. Native defects and contaminants cannot be entirely avoided, and, in the absence of doping, the background impurities present in any growth cause the films to grow as n-type. Some of these impurities are also responsible for the reduced properties of the GaN films, and thus it is necessary to understand which native defects are embedded during film growth, and how they affect the electrical and optical properties of the film. [1]

In this thesis, the primary concern is three main scattering mechanisms that reduce the performance of GaN based devices.

- **Interface scattering** due to the roughness of the GaN surface layer.
- **Dislocation scattering** due to structural scattering sites.
- **Background impurity scattering** as a result of impurities embedded in the GaN layer.

These scattering mechanisms are mainly produced during GaN growth. In order to identify to what extent each one of these scattering mechanisms affect the as-grown films, it is necessary to apply proper characterization methods.

1.1 Thesis approach

The purpose of this thesis will be to apply and optimize several characterization schemes in order to gain an understanding of the most important properties of GaN as a material. The characterization is performed on several GaN samples grown by various methods, and the results obtained from this will be outlined and discussed with the purpose of giving a description of the sample properties. Having obtained a methodological approach to GaN characterization, the scene is set for the development of optimal growth parameters. In this perspective, molecular beam epitaxy (MBE) is considered as a possible candidate for GaN growth. An MBE system is used for GaN growth on sapphire substrates, and the original purpose of this thesis was to use the characterization methods to find optimal growth parameters for GaN on sapphire and silicon with the MBE system. However, this was never achieved, because the gallium containing crucible cracked, resulting in gallium contamination of the evaporation source. Several attempts were made to remove the contamination before it was deduced that the evaporation source would not be functioning properly within the time frame of this work. Therefore only a few samples were grown with the MBE system. Instead the purpose was turned towards a more general approach to the characterization of GaN, with additional samples generously provided from other sources.

The thesis begins with a theoretical part, where it is of interest to first consider the general theory of the main characterization methods. This is followed by a consideration of the fundamental properties of GaN and how these can be investigated experimentally. Both the zinc blende and wurtzite crystal structures of GaN are modelled with the purpose of outlining their optical properties through calculations of the band structure, density of states, and optical dielectric functions. This is followed by a description of both luminescent and Raman active properties of GaN. These sections will outline how structural defects and impurities in GaN can be studied. Finally, a chapter is devoted to the description of GaN growth by MBE. This description is fundamental in order to achieve a proper comprehension between growth and characterization. Here it is also essential to consider the prominent and unavoidable effects of heteroepitaxial growth.

The experimental part will include an outline of the characterization of GaN films, along with the work with the MBE system. The characterization methods include AFM for surface analysis, Hall and van der Pauw measurements for electrical characterization, Raman and XRD studies for identification and evaluation of the crystalline properties of GaN, and photoluminescence measurements for identifying the properties of common radiative defects in GaN.

Characterization techniques

2

This chapter is devoted to giving an overview of the theory of some important characterization techniques used in this thesis. First, photoluminescence is considered, including recombination processes in semiconductors and their spectral characteristics. Next, the theoretical basis of Raman scattering is described from both a classical and a quantum mechanical point of view. Finally, the Hall and van der Pauw methods for studying the electrical properties of semiconductors are outlined.

2.1 Photoluminescence processes in semiconductors

Photoluminescence (PL) is the process in which a recombination between an electron and a hole releases energy in the form of photons with an energy corresponding to that of the electronic transition to a lower energy level. It is called photoluminescence because, in this method, the semiconductor is excited by light. Photoluminescence is a powerful, versatile tool for determining many optical properties of semiconductors, and it may thus provide very accurate information about electronic transitions in semiconductors. This section begins with an overview of several radiative recombination processes that may take place in semiconductors and how they can be identified through photoluminescence spectroscopy. [8, 9, 10]

Luminescence from an imperfect crystalline semiconductor may be observed from several different electronic transitions within the semiconductor band gap. This is illustrated in figure 2.1, where examples are given of the most common transitions that may be identified in a PL spectrum. Starting from left to right of the figure, the different general transitions will be denoted as follows:

- **(e-h)**: radiative recombination of a free electron-hole pair
- **(h-D^0)**: radiative recombination of a neutral donor with a free hole in the valence band
- **DAP**: radiative recombination of a donor-acceptor pair
- **(e-A^0)**: radiative recombination of a free electron with a neutral acceptor
- **X(n=1,2,3,. . .)**: radiative recombination of a free exciton in the n'th energy state
- **(D^0, X)**: radiative recombination of an exciton bound to a neutral donor
- **(A^0, X)**: radiative recombination of an exciton bound to a neutral acceptor

It should be mentioned that an exciton can also be bound to an ionized donor (D^+, X) or acceptor (A^-, X), however, these scenarios have not been mentioned in the list, because these transitions, typically, are not observed in the same material simultaneously (this will be further elaborated in the section about bound excitons). Also, in figure 2.1, only one donor and acceptor level is shown. However, several impurities may be present in a semiconductor, which means that several donor and acceptor levels may exist within the band gap. These are typically divided into deep and shallow levels. Shallow levels are those that lie in the vicinity of the conduction and valence band and thus become thermally ionized at temperatures up to room temperature. Deep levels lie farther away from the band edges, meaning they may not be thermally ionized at room temperature. A thorough description of recombinations between charge carriers and impurity levels will be given later on. [8, 9, 10]

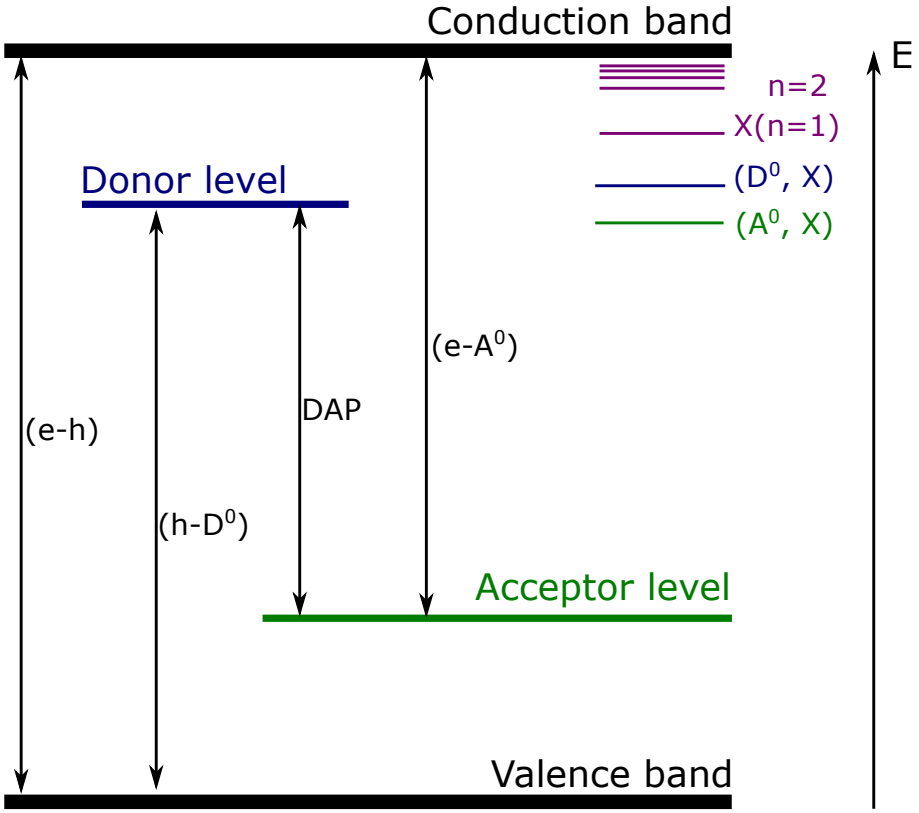


Figure 2.1: *Band diagram showing some common excitation processes in a semiconductor crystal. Inspired by [11].*

2.1.1 Recombination of free electron-hole pairs

This section will be devoted to describe the (e-h) recombination scheme, and what a typical PL emission spectrum will look like, if only band-to-band recombination is considered. The scheme is illustrated in figure 2.2, where the decay of an electron from an energy state E_2 to an energy state E_1 is illustrated. At E_1 the electron spontaneously recombines with a hole, resulting in emission of a photon with an energy $h\nu = E_2 - E_1$. In this scheme the recombination process may seem to occur away from the extrema of the energy bands. While this process is in theory possible, it is much more likely that the charge carriers will undergo thermalization and relax towards the extrema of the bands ($E_2 - E_1 \approx E_g$), where they may reach quasi-thermal equilibrium distributions. This is because the recombination time of an (e-h) pair in a semiconductor is much larger than the time it takes to reach this quasi-thermal equilibrium.

This also explains why the photons emitted in the recombination process may have significantly different energy compared to the photons used to excite the (e-h) pairs. [8, 10]

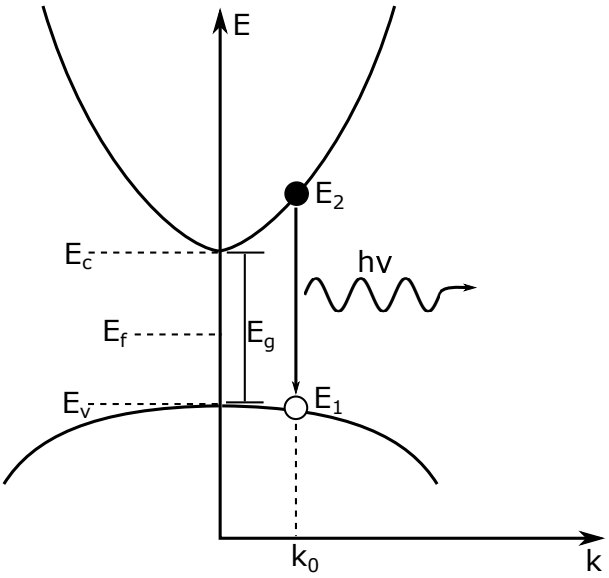


FIGURE 2.2: *(e-h) recombination scheme in a direct band gap semiconductor. E_v is the valence band edge, E_c is conduction band edge, E_g is the band gap, and E_f is the Fermi level.*

Continuing with the scheme from figure 2.2, the probability density of spontaneous emission (I_{sp}) of a photon due to the recombination of an (e-h) pair can, in the dipole approximation where the (e-h) pair is assumed not to interact with other charge carriers, be written as

$$I_{sp} \approx \frac{1}{\tau_r} f_e(\nu)\rho(\nu) \tag{2.1}$$

where τ_r is the recombination time of the electron and a hole (meaning that τ_r^{-1} is approximated as the transition probability). $f_e(\nu)$ is the Fermi-Dirac function representing the probability of a state E_2 being occupied in the conduction band and a state E_1 being unoccupied in the valence band. $\rho(\nu)$ is thus the joint density of states for free electrons and holes in each band. Considering the Fermi-Dirac function given by

$$f(E) = \frac{1}{\mathrm{e}^{(E-E_f)/k_BT} + 1} \tag{2.2}$$

where k_B is Boltzmann constant, and T is the temperature, $f_e(\nu)$ may then be written as the product of occupancies in either band, that is $f_c(E_2)$ and $f_v(E_1)$. Here $f_v(E_1)$ represents the probability of finding a hole in the valence band, which is similar to a missing electron in the valence band. $f_e(\nu)$ is thus

$$f_e(\nu) = f_c(E_2)f_v(E_1) = f(E_2)[1 - f(E_1)] \tag{2.3}$$

This can be further simplified if the excitation power is assumed to be low enough for the Fermi level to stay in the middle of the gap, meaning that the semiconductor stays in thermal equilibrium during the excitation process. If the semiconductor is also assumed to be intrinsic, having a wide band gap, and at low temperature, it is found that $k_BT << E_g/2 \approx E_2 - E_f \approx$

$E_f - E_1$. Thus the following approximations may be formulated

$$f(E_2) = \frac{1}{e^{(E_2-E_f)/k_BT}+1} \approx e^{-(E_2-E_f)/k_BT}$$
$$[1-f(E_1)] = 1 - \frac{1}{e^{(E_1-E_f)/k_BT}+1} = \frac{e^{(E_1-E_f)/k_BT}}{e^{(E_1-E_f)/k_BT}+1} = \frac{1}{1+e^{-(E_1-E_f)/k_BT}} \approx e^{-(E_f-E_1)/k_BT}$$

which when inserted in eq. 2.3 yields

$$f_e(\nu) \approx e^{-h\nu/k_BT} \tag{2.4}$$

The density of states from eq. 2.1 can be found by considering that the number of states in an energy interval dE_2 should be equal to the number of photons emitted with an energy $E_2 - E_1 = h\nu$ in a frequency interval $d(\nu)$, and thus the density of states $\rho(\nu)$ can be written as

$$\rho(\nu)d\nu = \rho_c(E_2)dE_2$$
$$\rho(\nu) = \rho_c(E_2)\frac{dE_2}{d\nu} \tag{2.5}$$

So in order to obtain $\rho(\nu)$, the density of states in the conduction band at E_2, and the energy, E_2, must be determined. From the parabolic band shown in figure 2.1, the energies E_2 and E_1 can be found from the parabolic approximation

$$E_1 = E_v - \frac{\hbar^2 k_0^2}{2m_h} \tag{2.6}$$
$$E_2 = E_c + \frac{\hbar^2 k_0^2}{2m_e} \tag{2.7}$$

where k_0 is the displacement along the k-axis from the band edges, m_h and m_e is the hole and electron effective masses, respectively. By using that $E_2 - E_1 = h\nu$, k_0^2 may then be determined

$$E_2 - E_1 = E_c - E_v + \frac{\hbar^2 k_0^2}{2}\left(\frac{1}{m_e}+\frac{1}{m_h}\right) = h\nu$$
$$k_0^2 = \frac{(h\nu - E_g)2m_r}{\hbar^2} \tag{2.8}$$

where m_r is the reduced mass of the electron and hole, defined by $1/m_r = 1/m_e + 1/m_h$. Inserting eq. 2.8 in eq. 2.7 then yields

$$E_2 = E_c + \frac{m_r}{m_e}(h\nu - E_g) \tag{2.9}$$

which must be used along with the density of states for the conduction band. $\rho_c(E)$ can for a simple parabolic band be found from a simple particle in a box model. The well know result is

$$\rho_c(E) = \frac{(2m_e)^{3/2}}{2\pi^2\hbar^3}(E - E_c)^{1/2} \;; \qquad E \geq E_c \tag{2.10}$$

Now $\rho(\nu)$ can be found from eq. 2.5 by insertion of eqs. 2.10 and 2.9, which gives

$$\rho(\nu) = \frac{(2m_r)^{3/2}}{\pi\hbar^2}(h\nu - E_g)^{1/2} \;; \qquad h\nu \geq E_g \tag{2.11}$$

which is the density of states per unit frequency. [8]

Thus an expression for I_{sp} may be formulated by inserting eq. 2.4 and eq. 2.11 in eq. 2.1, yielding

$$I_{sp}(h\nu) \approx D_0(h\nu - E_g)^{1/2} \mathrm{e}^{-(h\nu - E_g)/k_B T} \tag{2.12}$$
$$D_0 \equiv [(2m_r)^{2/3}/\pi\hbar^2\tau_r] \exp(-E_g/k_B T)$$

where everything that is independent of the photon energy is put into D_0. From D_0 it is seen that the expected luminescence intensity from a band-to-band recombination depends only on the reduced mass, m_r, meaning that m_e and m_h cannot be determined solely from photoluminescence measurements. The luminescence intensity may then be plotted with respect to the difference between the photon energy and the band gap, $h\nu - E_g$. The plot is shown in figure 2.3 for a number of different temperatures. [8, 10]

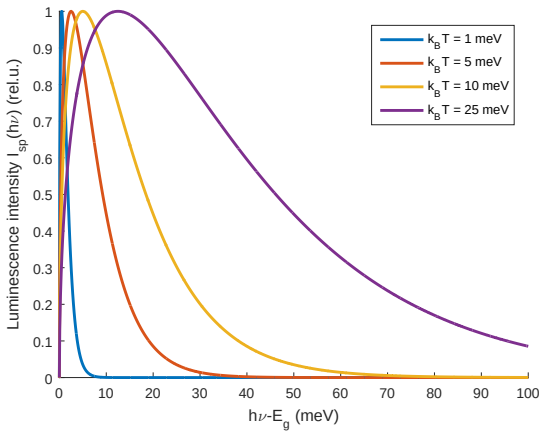


FIGURE 2.3: *The intensity is plotted in relative units with respect to the difference between the photon energy and the band gap. The spectra show a Maxwell-Boltzmann distribution, which broadens with increasing temperature.*

The shape of the curves all follow Maxwell-Boltzmann distributions and the peak is right shifted with an increase in temperature. Furthermore, the widths of the curves also increase with temperature, and for the full-width half maximum (FWHM), it is seen to increase with temperature as

$$\mathrm{FWHM} \approx 1.8 k_B T$$

However, despite this broadening factor, the curves in figure 2.3 are quite narrow. The FWHM broadening in a temperature-range from 4 K to 100K is only on the scale of 1 meV to 20 meV, corresponding to a broadening of less than 10 nm in the wavelength spectrum. The physical reason for the broadening with temperature may be noted from the Fermi-Dirac statistics. As the temperature is increased, excited electrons and holes may occupy states further away from the extrema of the energy bands. This means that photons of higher energy are released during the recombination process, which is seen from the much more prominent broadening of the Boltzmann tail in the spectrum compared to the peak shift. It should be noted, however, that (e-h)-pairs are rarely observed at very low temperatures, with the exception of very narrow band gap semiconductors. At these low temperatures the spectrum will typically be dominated by exciton recombinations, the reason for this will be elaborated later in this chapter. However, the shape of the spectra still follows a Maxwell-Boltzmann distribution at elevated temperatures, where transitions due to (e-h)-pairs are readily expected. [8, 10]

2.1.2 (h-D^0), (e-A^0) and DAP

When impurities are present in a semiconductor, they may participate in optical transitions. The most common types of impurities are the so called *shallow impurities*, the energy states of which lie localized near the extrema of the energy bands as depicted in figure 2.4, where the two schemes show the (h-D^0) and (e-A^0) recombinations, respectively. [8, 10]

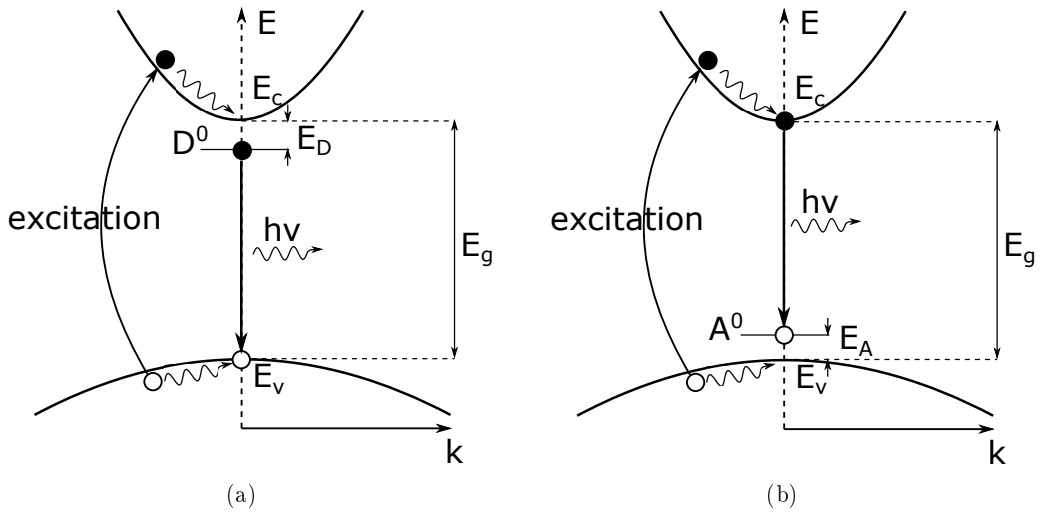


Figure 2.4: *Radiative recombination process between* **a***: a hole and a neutral donor (h-D^0).* **b***: an electron and a neutral acceptor (e-A^0).*

In figure 2.4 it is shown how an electron may first undergo excitation from the valence band to the conduction band. This process generates a free hole and a free electron in either band, thus allowing for the donor bound electron to undergo radiative recombination with the hole in the valence band. This further leaves an unoccupied state at D^0, to which any free electron in the conduction band may relax to, returning the system its original state. The radiative release in this recombination process is then given by $h\nu \approx E_g - E_D$. A similar process takes place in the case of an acceptor impurity, where the released photon energy is given by $h\nu \approx E_g - E_A$. In either case the energy span between the impurity state and the nearest energy band must be larger than the thermal energy, i.e. $k_B T < E_{D,A}$, since else the impurity states will be ionized and the donor or acceptor will donate its electron or hole to the conduction band or valence band, respectively. This means that PL may be used to determine the activation energy of impurities and thus even the impurity specimen e.g. by varying the sample temperature while measuring the luminescence spectrum. [8, 10]

Another possible recombination process is that of a donor-acceptor pair (DAP), which is possible when the donor and acceptor states both are present in a semiconductor. Thus the DAP recombination process may simply be described by a combination of figures 2.4(a) and 2.4(b), since the recombination process follow the same steps. [8, 10]

However, determination of the energy of the emitted photons in the DAP recombination process may not be straight forward. Starting by considering the reaction of the recombination process

$$D^0 + A^0 \longrightarrow D^+ + A^- + h\nu \tag{2.13}$$

from which it is seen that an ionized DAP is also generated in the process. Thus, this term must naturally yield a positive contribution to the emitted photon energy in the form of a Coulomb attraction between the two ionized impurity atoms, and the influence of this term is then determined by the distance, R, between the donor and acceptor atoms. Thus the photon energy emitted in a DAP recombination process may be written as

$$h\nu = E_g - (E_D + E_A) + \frac{e^2}{4\pi\epsilon_r\epsilon_0 R} \tag{2.14}$$

where e is the electron charge, ϵ_r is the relative permittivity, and ϵ_0 is the vacuum permittivity. However, the distance between the ionized donor and acceptor may be a somewhat complex quantity, since many factors may influence it. First of all, R can only take a discrete value, which depends on the crystal structure and the lattice constant of the semiconductor crystal the impurity atoms are embedded in. Thus several DAP lines may be observed in a PL spectrum from a semiconductor with a sufficient concentration of dopants, where each line represents different contributions from the Coulomb term in eq. 2.14, until a point where $R \to \infty$ and $h\nu = E_g - (E_D + E_A)$. Another peculiarity arises if the impurity concentration in the semiconductor is low, where no DAP transitions may be seen at all, because the distance between donors and acceptors becomes too large and thus the probability of a DAP recombination will be small compared to other recombination schemes such as those mentioned earlier in this section. Other parameters that may influence the probability of DAP transitions are e.g. the excitation power of the excitation source. At low excitation intensities there are fewer photons per area, so the probability of exciting D^0 and A^0 states in close proximity to each other may be rather low, meaning that it may be less likely for a DAP transition to occur between donor and acceptor atoms that are positioned in the vicinity of each other, compared to those that lie at a more intermediate distance from each other. Thus a characteristic blue shift of the DAP intensity maximum can be observed with increased excitation intensity, in accordance with eq. 2.14. [8, 10]

2.1.3 Free and bound excitons

In this section excitons and their possible energy states will be treated. A free exciton consists of an electron and a hole bound to each other by an attractive force. This means that the exciton is a free quasi particle, where its constituents have to be treated as a two-particle system, conversed to that of the (e-h) pair, where the electron and hole were treated as two free uncorrelated particles. Typically, a few different subclasses of excitons are mentioned in literature, depending on the strength of the interaction between the electron and hole. Since most semiconductors have large dielectric constants, this results in a larger screening of the Coulomb interaction, meaning that the electron and hole are weakly bound. Thus the electron and hole may be separated by many lattice constants of the semiconductor crystal, and the resulting exciton is often called a *Wannier exciton*. The aim of this section will thus be to elucidate the properties of the Wannier excitons and see how these properties may reveal themselves in a PL spectrum. [6, 8, 10]

In order to consider a 3-dimensional two-particle system consisting of a hole in the valence band and an electron in the conduction band, the exciton motion is considered as two particles, each with a relative mass orbiting a mutual center of mass (CM). This means that from the point of the CM, the position of the electron and hole have translational invariance. Furthermore, the CM of the exciton can be treated as a free particle with a mass $M = m_e + m_h$ and thus

its wavevector is given by $\mathbf{K} = \mathbf{k}_e + \mathbf{k}_h$, where $\mathbf{k}_{e,h}$ are the wavevectors describing the positions of the electron and hole in k-space, respectively. In the vicinity of the extrema of the energy bands, the energy bands can often be considered spherically symmetric within a crystal, and the following is thus described under this approximation. The aim will now be set on solving the Schrödinger equation for this two particle system in order to find the discrete energy levels of the exciton. Starting with the exciton wavefunction, $\Psi(\mathbf{r}_e, \mathbf{r}_h)$, which may be written as a linear combination of Bloch functions, namely $\psi_{k_e}(\mathbf{r}_e)$ and $\psi_{k_h}(\mathbf{r}_h)$ for the electron and hole, respectively, such that

$$\Psi(\mathbf{r}_e, \mathbf{r}_h) = \sum_{\mathbf{k}_e, \mathbf{k}_h} C(\mathbf{k}_e, \mathbf{k}_h) \psi_{\mathbf{k}_e}(\mathbf{r}_e) \psi_{\mathbf{k}_h}(\mathbf{r}_h) \tag{2.15}$$

where $\mathbf{r}_{e,h}$ is the real space coordinates of the electron and hole, and $C(\mathbf{k}_e, \mathbf{k}_h)$ are the eigenvectors. However, it may be simpler to rewrite eq. 2.15 in terms of Wannier functions, which are functions of the electron and hole positions relative to the CM, i.e. $a_n(\mathbf{r}; \mathbf{R})$ is a function of $\mathbf{r} - \mathbf{R}$, where $\mathbf{r}$ is the relative coordinate and $\mathbf{R}$ is the coordinate of the CM. The coordinate system is shown in figure 2.5.

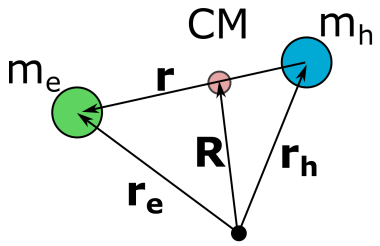


Figure 2.5: *Coordinate system for an exciton with the center of mass (CM).*

In accordance with the figure, the new coordinates may thus be written as

$$\mathbf{R} = \frac{m_e \mathbf{R}_e + m_h \mathbf{R}_h}{m_e + m_h} \tag{2.16}$$

$$\mathbf{r} = \mathbf{R}_e - \mathbf{R}_h \tag{2.17}$$

The Wannier functions are obtained by Fourier transforming the Bloch functions, thus writing eq. 2.15 in terms of the Wannier functions gives [10]

$$\Psi(\mathbf{r}_e, \mathbf{r}_h) = \frac{1}{N^{1/2}} \sum_{\mathbf{R}_e, \mathbf{R}_h} \Phi(\mathbf{R}_e, \mathbf{R}_h) a_{\mathbf{R}_e}(\mathbf{r}_e) a_{\mathbf{R}_h}(\mathbf{r}_h) \tag{2.18}$$

where N is the number of unit cells, and $\Phi(\mathbf{R}_e, \mathbf{R}_h)$ is the exciton envelope wavefunction. The exciton Schrödinger equation can then be written only in terms of $\Phi(\mathbf{R}_e, \mathbf{R}_h)$, since it is the only term given in coordinates relative to the center of mass

$$\left[-\left(\frac{\hbar^2}{2m_e} \nabla^2_{\mathbf{R}_e} \right) - \left(\frac{\hbar^2}{2m_h} \nabla^2_{\mathbf{R}_h} \right) - \frac{e^2}{4\pi\epsilon_0 \epsilon_r |\mathbf{R}_e - \mathbf{R}_h|} \right] \Phi(\mathbf{R}_e, \mathbf{R}_h) = E_{exc} \Phi(\mathbf{R}_e, \mathbf{R}_h) \tag{2.19}$$

now by introducing the coordinates of the CM given in eqs. 2.16 and 2.17, along with the reduced exciton mass $1/\mu = 1/m_e + 1/m_h$, the exciton wave equation may then be written in two uncorrelated terms by separation of variables, i.e. $\Phi(\mathbf{R}_e, \mathbf{R}_h) = \psi(\mathbf{R})\phi(\mathbf{r})$, and thus eq. 2.19 is separated in two equations as

$$\left(-\frac{\hbar^2}{2M} \right) \nabla^2_{\mathbf{R}} \psi(\mathbf{R}) = E_R \psi(\mathbf{R}) \tag{2.20}$$

$$\left(-\frac{\hbar^2}{2\mu} \nabla^2_{\mathbf{r}} - \frac{e^2}{4\pi\epsilon_0 \epsilon_r r} \right) \phi(\mathbf{r}) = E_r \phi(\mathbf{r}) \tag{2.21}$$

where it is quickly realized that the eigenvalue to eq. 2.20, E_R, is simply that of a free particle, solved by using the eigenfunction $\psi(\mathbf{R}) = 1/N^{1/2}\exp(i\mathbf{K}\cdot\mathbf{R})$, so that the energy becomes

$$E_R = \frac{\hbar^2 K^2}{2M} \tag{2.22}$$

Furthermore, the form of eq. 2.21 is similar to the well known solution to the Schrödinger equation for the hydrogen atom, where the eigenfunction is typically separated in two functions, one dependent on the radial coordinates, $R_{nl}(r)$ and the other dependent on the angular coordinates, $Y_{lm}(\theta,\phi)$. $R_{nl}(r)$ is typically expressed as the associated Laguerre polynomial, and $Y_{lm}(\theta,\phi)$ is the spherical harmonics. n, l, and m are the quantum number, angular momentum quantum number, and the magnetic quantum number, respectively. If only the s-type eigenstates are considered, only the quantum number, n, is of relevance and thus the eigenvalue of eq. 2.21 is given by

$$E_r(n) = E_r(\infty) - \frac{R^*}{n^2} \qquad\qquad n = 1, 2, 3, ... \tag{2.23}$$

where $E_r(\infty)$ is the energy before the energy states form a continuum of states, meaning that $E_r(\infty) = E_g$. R^* is the constant prefactor that comes from solving eq. 2.21. It is usually called the exciton Rydberg constant and is thus defined in terms of the Rydberg energy as

$$R^* = \frac{\mu e^4}{32\hbar^2\pi^2\epsilon_0^2\epsilon_r^2} = \left(\frac{\mu}{m_0\epsilon_r^2}\right) \times 13.6\,\text{eV} \tag{2.24}$$

where it was used that $13.6\,\text{eV}/m_0 = e^4/(32\hbar\pi^2\epsilon_0^2)$, where m_0 is the electron rest mass. It should be mentioned here, that the use of ϵ_r, means that the volume occupied by the exciton is considered large enough to be a homogeneous dielectric. This is obviously not always the case, and in more accurate theoretical calculations the dielectric constant of the exciton medium is thus often calculated with an interpolation between ϵ_r and ϵ_∞ [8]. Nonetheless, the eigenvalues from from eq. 2.19 may now be evaluated, thus describing the discrete energy states, which the exciton may occupy

$$E_X(\mathbf{K}) = E_g + \frac{\hbar^2 K^2}{2M} - \frac{R^*}{n^2} \tag{2.25}$$

which means that the exciton energy states lie as parabolic bands with an energy difference of $R^*(1/n^2 - 1/(n+1)^2)$ between the n'th and $n+1$ band. Thus the ground state lies farthest away from the band edge, as was also depicted in figure 2.1. R^* is thus also the binding energy of the free exciton. Finally, it should be mentioned that this simple parabolic model is not adequate for most semiconductors, especially for wide band gap semiconductors such as GaN. Nonetheless, the parabolic model derived here gives an impression of how exciton recombinations reveal themselves in a PL spectrum, given that the thermal energy is less than the binding energy of the exciton. [6, 8, 10, 12]

Bound excitons

So far only free excitons have been considered. However, when impurities or lattice defects are present in a semiconductor, these serve as potential wells that may trap free excitons, hence they become *bound excitons*. When the exciton is bound to e.g. a donor impurity, it is usually denoted by (D^0, X), or (D^+, X) if the donor is ionized. The bound exciton state (D^0, X) must thus consist of two electrons, a hole, and a donor ion, and the (D^+, X) state thus consists of an electron, a hole, and a donor ion. [8]

When an exciton has been bound to an impurity its kinetic energy dissipates, after which the exciton stays localized at the impurity until it recombines and emits a photon. The ratio of the semiconductor effective masses $\sigma = m_e/m_h$ is a crucial parameter, when considering excitons bound to ionized impurities. It was mentioned earlier that the (D^+, X) and the (A^-, X) states might not exist simultaneously in a semiconductor. The reason for this may actually be explained solely by the relative electron and hole masses. Starting by considering the luminescence process in the case of excitons bound to e.g. an ionized donor

$$D^+ + X \rightarrow (D^+, X) + D_1 \tag{2.26}$$
$$(D^+, X) \rightarrow D^+ + h\nu_{BE} \tag{2.27}$$

where D_1 is the binding energy of the exciton bound to the ionized donor. The emitted photon would then have an energy $h\nu_{BE} = E_X - D_1$, just like earlier. However, in order for this process to take place, it turns out that σ must take on a specific value, the same statement being valid for ionized acceptors. The reason for this is that e.g. an ionized donor needs to be able to overcome the kinetic energy of the hole of the exciton. It can only do this if the effective mass of the hole is sufficiently small. It can be considered in a sense, where the ionized donor attracts the electron of the exciton. On the other hand the hole of the exciton will keep pulling in its electron partner, and it may even rip it loose from the ionized impurity. In other words, the energy that the ionized donor can hold the electron with must also be able to cope with the kinetic energy of the hole. This may only be possible beneath a certain value of $\sigma = m_e/m_h$ ratio. The opposite situation is then related to an ionized acceptor, where it is the kinetic energy of the electron that needs to be considered. Here σ has to be larger than some value, and hence no semiconductor exists, where it is possible to have both of these conditions fulfilled at once, and there may even be semiconductors that exhibit neither (D^+, X) or (A^-, X) PL lines (this is e.g. the case in silicon, where $\sigma \approx 0.61$. An example where the (D^+, X) PL line may be observed is GaAs with $\sigma \approx 0.11$). It can be shown theoretically (something that lies outside the scope of this report), that a critical ratio of $\sigma_c \approx 0.43$ exists for ionized donors [8]. Excitons bound to ionized donors may thus only be observed in semiconductors with a σ value lying below this value, whereas ionized acceptors may only be found in semiconductors, where σ is larger than $1/\sigma_c \approx 2.33$. [6, 8]

Luminescence may also be observed from recombination of excitons bound to neutral impurities. In the case of e.g. a neutral donor atom, the following processes take place

$$D^0 + X \rightarrow (D^0, X) + E_{BX} \tag{2.28}$$
$$(D^0, X) \rightarrow D^0 + h\nu_{BE} \tag{2.29}$$

where the emitted photon should then have an energy $h\nu_{BE} = E_X - E_{BX}$, with E_{BX} being the binding energy of the bound exciton. It is generally a very complicated task to determine the binding energies of excitons bound to neutral impurities theoretically, which is also why most methods rely on experimental observations. Among these *Haynes' rule* should be mentioned [13, 14]. Haynes found a quantitative expression stating that E_{BX} increases linearly with respect to the binding energy of e.g. a shallow donor E_D or acceptor E_A as

$$E_{BX} = a + bE_{D,A} \tag{2.30}$$

where a and b are constants that depends on the semiconductor material as well as if it is a donor or acceptor impurity. An example could be for a donor in silicon, where $a = 0$ and

$b \approx 0.1$, meaning that the binding energy of the bound exciton equals approximately one tenth of the binding energy of the donor the exciton is bound to (excitons in silicon may be misleading example, however, it was for silicon that Haynes first did his experimental findings of excitons). Thus in order to observe photoluminescence from bound excitons in e.g. silicon, liquid helium temperatures are generally required. Eq. 2.30 is also valid for acceptors, however, the constants will differ from those used for the donors. E_{BX} for acceptors is generally larger than E_{BX} for donors. The reason for this is similar to that for ionized donors and acceptors, except that for neutral impurities there is always a stable σ ratio, where the bound exciton has a non-zero binding energy to the impurity for both acceptors and donors. This may be explained by considering the two limits of a heavy hole ($m_h \to \infty$) and a light hole ($m_h \to 0$). In the limit of a heavy hole, where the electron of the exciton is bound to an impurity, the hole will also be bound there since it has a very low kinetic energy. The complex made of the exciton and the impurity is similar to a H_2 molecule, which is a very stable molecule. In the opposite limit, the hole will have a very large kinetic energy and it cannot be bound to the impurity. Thus only the electron of the exciton is bound, and the complex is now similar to a H^- ion, which is also stable. From this consideration it is also evident, why the exciton binding energy to a neutral donor is generally smaller than the binding energy to a neutral acceptor, given that $m_h > m_e$ (as is the case for most semiconductors). Qualitatively explained, a donor atom generally attracts the hole of the exciton in order to remove its unbound valence electron (in case of shallow donors), whereas an acceptor atom attracts the electron of the exciton. This means that for the donor it is the electron that decreases the binding energy of the complex, whereas for the acceptor it is the hole. [8]

2.1.4 Participation of phonons in optical phenomena

So far this chapter focused mainly on photo-excitation of electronic states and subsequent recombination of charge carriers by emission of light. However, this picture does not give the full story about PL. Another major participant in many optical process, namely the *phonon*, has not yet been considered. The phonon is a quasi-particle representing lattice vibrations of the semiconductor crystal. To understand the more general properties of phonons, the vibration of a 1-dimensional diatomic chain of atoms is typically considered, because the general trends from the 1D model are the same as in a 3D model. From this model two phonon branches are discovered, an acoustic branch where neighbour atoms oscillate in phase, and an optical branch where neighbour atoms oscillate in counter-phase. Furthermore, both of these oscillations are subdivided in two modes called longitudinal and transverse modes, which for optical and acoustic phonons are thus denoted as LO, TO, LA, and TA, respectively. The abbreviation is always with respect to the direction of the crystal wavevector $\mathbf{k}$. [8, 10]

Phonons participate in optical processes through interactions with the semiconductor electrons. With respect to PL, the most interesting example of this interaction is when a phonon participates in the transition of an electron from an excited state to its ground state and still emits a photon. The interaction from the electron-phonon coupling may lead to the emission of one or more phonons, where a single phonon has an energy $\hbar\omega$. This means that the electron effectively lowers its energy, which can then be seen in the PL spectrum as what is a called a *phonon replica* and it will typically lie at a specific photon energy depending on the number of phonons, n, emitted in the interaction and thus lie at $h\nu - n\hbar\omega$. The easiest way to identify these replicas is probably just by looking at the curve shape in the PL spectrum, since it should be identical to the curve it is a replica of. [8, 10]

There are three main ways in which excitons in polar semiconductors can interact with the lattice. The first is the exciton-phonon interaction through the deformation potential interaction, the second is through the piezoelectric mechanism, and the third is the interaction through the *Frölich mechanism*. The latter only occurs in polar semiconductors, and can be explained by polar lattice vibrations causing an internal electric field. Because the effective mass of the electron is usually much lower than that of the hole, a charge distribution is created within the crystal, which can then interact with the lattice through the Frölich mechanism. This is illustrated in a very simple manner in figure 2.6, which also illustrates that the coupling to the LO phonon is much stronger than the coupling to the TO phonon, because the LO phonon creates a varying electric field **E** along the **k** direction. Since III-V semiconductors are highly ionic, the Frölich mechanims is usually the dominating electron-phonon interaction, and thus the replica produced by the coupling to a LO phonon is the most commonly observed phonon replica in a PL spectrum. [8, 10, 15]

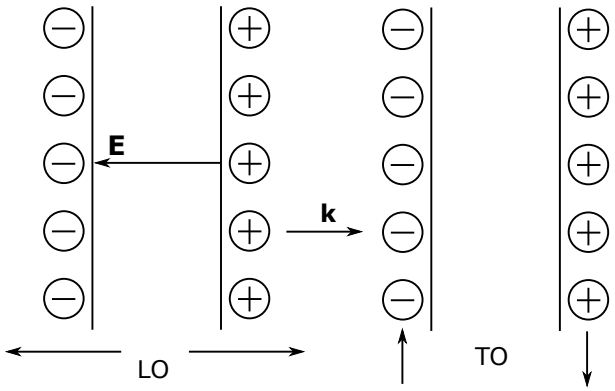


Figure 2.6: *LO and TO phonons in a polar crystal.*

The intensity of phonon replicas in a PL spectrum decreases as the number of phonons emitted in the process increases, because the probability for the process to occur lowers. It is also possible that an electron and a hole recombines entirely through phonon recombinations, and thus this is called a non-radiative recombination. In order for this to occur, the total amount of phonons involved is then found from $E_g = n\hbar\omega$. Thus in a rough approximation, this non-radiative transition probability $1/\tau_{nr} \approx \exp(-E_g/\hbar\omega) \approx \exp(-n)$ strongly depends on the size of the band gap. In a typical semiconductor, let E_g =2 eV and $\hbar\omega$ =25 meV and hence n =80, this means that the non-radiative recombination rarely ever occurs in semiconductors with a wide band gap.

Electron-phonon interaction and the configurational coordinate model

While the argument given above for the probability of non-radiative recombination processes is valid for pure crystals, this ideal view falls apart when impurities are considered as luminescence centers. The photon emission from localised centers, such as deep defects, may play a dominant role in the luminescence spectrum, and can be used to characterize the defect types responsible for the emission.

In order to illustrate this phenomenon, a deep acceptor, which can be raised to an excited state by capturing a hole, is considered. This situation is chosen because it is of pivotal relevance to the luminescence in n-type GaN. If this impurity is excited by hole capture, the hole wave function is highly localised to one bond of the defect. This causes atoms in the vicinity to shift

in position from their unperturbed lattice sites, and find a new equilibrium position. Once the hole recombines with either a free electron or one from a donor atom, the atoms shift back to their original position. Thus, some of the energy that would otherwise be radiatively emitted is lost through phonons in the re-arranging of lattice atoms. This description of excitation and relaxation is called the *configuration coordinate model*, and is shown in figure 2.7, where the x-axis represents some generalized 1D real-space coordinate. The y-axis shows the total potential energy, both nuclear and electronic, for the ground and excited states of the impurity atom.

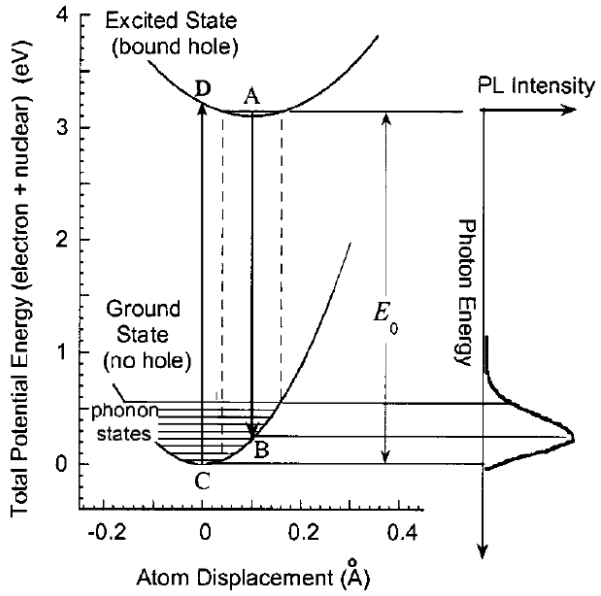


FIGURE 2.7: *Illustrative configuration coordinate diagram showing the ground and excited states of a luminescent defect center, and the PL spectrum which results from its relaxation. Modified from [6].*

In the figure, it is shown how the excited state energy minimum is displaced in space relative to that of the ground state. If it is assumed that the atomic nuclei are stationary in the time frame of an excitation event, a reasonable assumption since photon absorption occurs within $\sim 10^{-15}$s while the frequencies of phonons are $\sim 10^{-12}-10^{-13}$s, all optical transitions are vertical in this diagram. This assumption is called the adiabatic approximation. At low temperatures, where the phonon energy $\hbar\omega >> k_B T$, the absorption and emission of a photon occurs from the zero vibrational levels of the ground state and the excited state, respectively. Once an electron is photo-excited from the valence band to the conduction band, this causes the acceptor to bind a hole and enter the excited state (C $\rightarrow$ D), which again causes the acceptor to be displaced to a new minimum energy position (D $\rightarrow$ A). After some time in the conduction band, the photo-excited electron might recombine with the acceptor-bound hole with the release of a photon, with the most probable emission energy being that of transition A $\rightarrow$ B in the figure. From point B, the system again reconfigures to the original minimum energy position in the ground state, reaching point C. [6, 8]

In the simplest case, only one phonon state is considered in the interaction with the electron/hole-phonon coupling, and it is assumed here that these have equal frequencies (a simplifying, but not accurate assumption for GaN). Under this assumption, it is obvious that the potential energy wells in the ground and excited states have the same shape, and followingly that the energy difference between D and A is the same as that between B and C, which is defined here as the *relaxation energy*, E_R. This is the mean energy transferred to phonons as a

result of the excitation or relaxation, and can be written as

$$E_R = \mathcal{S}\hbar\omega \tag{2.31}$$

where the dimensionless factor $\mathcal{S}$ is the *Huang-Rhys* factor, which expresses the mean number of phonons emitted during the relaxation from B $\rightarrow$ C. The strength of the electron-phonon coupling is thus represented by this parameter, as a larger $\mathcal{S}$ results in a larger nuclear displacement as a result of excitation. [6, 8]

Using the CC model and the Huang-Rhys factor, it is now possible to explain the spectral lineshape and width of the luminescent transitions of the defect. Assuming low temperature, so that the excited state is at its energy minimum, the most likely recombination is naturally between point A and B. However, with a probability determined by $\mathcal{S}$, it is also possible that the emission may terminate at an energy slightly higher or lower by $\hbar\omega$, or at any level $\pm n\hbar\omega$ with a likelyhood decreasing for higher n. The highest emission energy possible is the one corresponding to emission the difference between the minima of the ground and excited states, E_0 in figure 2.7, which is denoted the *zero-phonon line* (ZPL) for obvious reasons. Naturally, the probability of emission at the ZPL energy is inversely proportional to the atomic displacement between the ground and excited states, and completely dominates the spectrum for $\mathcal{S} \rightarrow 0$. Ignoring line broadening, the emission spectrum thus consists of a series of $n = 0, 1, 2...$ phonon replica lines spaced by $\hbar\omega$, with the intensity of each line modulated by the probability of relaxation from the electronic excited state minimum to each vibrational level of the electronic ground state. By taking the overlap of the nuclear wave functions of the ground and excited states, this probability is found to have the form of a Poisson distribution [6, 8]

$$I(n) = e^{-\mathcal{S}}\frac{\mathcal{S}^n}{n!}, \qquad n = 0, 1, 2, ..., \tag{2.32}$$

Naturally, even at low temperatures, these lines are broadened, both due to variations in the local phonon energy, and because of e.g. broadening due to the lifetime of the excited state. Thus, in the case of a weak electron phonon coupling, a sharp, intense ZPL with a few lower intensity phonon replicas is expected, whereas for stronger coupling, a broader spectrum with weaker ZPL is observed. For even larger $\mathcal{S}$, the Poisson distribution approximates a Gaussian distribution, and no ZPL is observed.

In general, the electron-phonon interaction is particularly important in III-nitride semiconductors, due to their ionic nature giving rise to a large Fröhlich interaction. It is obvious, then, that if the electron-phonon coupling is very strong, a situation might arise where the ground state potential energy "band" might overlap with that of the excited state at a relatively small atomic displacement. In most cases, a sharp drop in defect luminescence with temperature corresponds to the thermal escape of carriers bound to the defect to the valence or conduction bands (this is dicussed further with respect to GaN in section 3.2). However, if the ground and excited states overlap in the CC model, luminescence might be thermally quenched even at very low temperatures due to the possibility of multi-phonon recombination, so that these defects act predominantly as non-radiative recombination centers.

2.2 Raman spectroscopy

In the study of the crystalline properties of solid materials, photoluminescence spectroscopy is beautifully complimented by Raman spectroscopy. This technique is widely used both for the investigation of rotational and vibrational states in molecules, and for investigating the crystallinicity and vibrational states in solids, the latter being the focus of this section. Raman spectroscopy utilizes Raman scattering, which is fundamentally different from absorption and emission from electronic levels in photoluminescence. For non-resonant incident light on a material, some fraction of it will be scattered by the material. Generally, scattering of light in a crystal may be either elastic or inelastic. Elastic scattering, or *Rayleigh scattering* (R in figure 2.8), can arise due to inhomogeneities in the crystalline structure (such as point defects or grains), or it can be caused by the interaction of the photons with the lattice itself. This type of scattering, where the energy of the incident photons remain unchanged, is the dominant mechanism of scattering.

Inelastic scattering, or *Raman scattering*, is where the energy of the emitted light is modified slightly with respect to that of the incident light through interaction with phonons. This increase or decrease of the energy relative to the incident photon corresponds to the energy of a certain phonon mode and is called a *Raman shift*. The Raman spectrum of a material thus consists of a number of discrete frequencies of both higher and lower energy than the exciting radiation. Referring to figure 2.8, the lines of lower frequency than the incident photon are called *Stokes lines* (S) and those of higher frequency are called *anti-Stokes lines* (A). The intensity of Raman scattering is around five orders of magnitude lower than for Rayleigh scattering, and thus a laser is typically needed for Raman spectroscopy to achieve sufficient intensities. [16, 17]

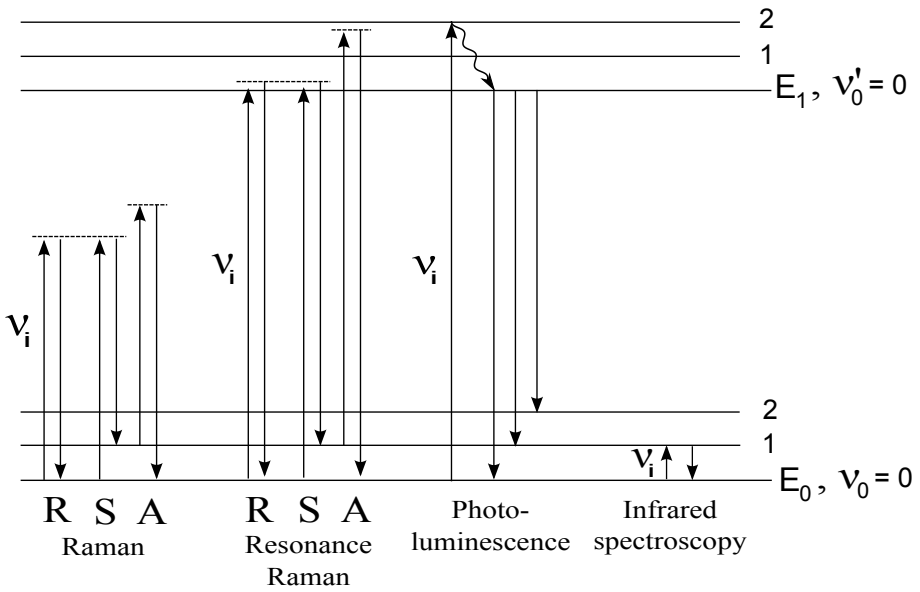


FIGURE 2.8: *Illustration of the energy levels involved with different spectroscopy techniques. E_0 and E_1 denote the electronic ground state and first excited state, respectively, and ν_0 and ν_0' denote vibrational energy states. The dotted lines involved in Raman scattering are "virtual" energy states.*

In general, there are several ways in which the Raman scattering process can be described. For a meticulous mathematical derivation, both classical and quantum mechanical approaches require third-order perturbation theory, which is beyond the scope of this report. However, more

simplistic approaches can give qualitative insight into the parameters and variables involved in the Raman process, and allow the understanding of experimentally observed Raman spectra. This will be the aim of the next sections. [10, 16].

2.2.1 Simplified quantum-mechanical description

In order to describe the Raman scattering process in a semiconductor, it is necessary to denote the initial states of the three systems participating in the scattering: incident and scattered photons of frequencies ω_i and ω_s, respectively, the electrons in the semiconductor, and the phonons which take part in the scattering. Before scattering, there is some number of photons, $N(\omega_i)$ and $N(\omega_s)$, with frequencies ω_i and ω_s, respectively, there are N_q phonons, and the system is in its electronic ground state. If, for example, a Stokes scattering event occurs, there will be $N(\omega_i) - 1$ photons of frequency ω_i, while $N(\omega_s)$ and N_q have increased by one each. It thus seems reasonable that the scattering can be described solely by the interaction Hamiltonian for photons and phonons. However, this interaction is very weak if the photons and phonons have a large energy difference, and can often be neglected [18]. On the other hand, for light at optical frequencies, the photons couple strongly to electrons, creating electron-hole pairs which can then interact with phonons. The process takes place as three distinct events [10, 19]:

- **1**: The incident photon is annihilated, creating an electron-hole pair
- **2**: The excited electron is scattered by a phonon, losing some energy and exciting the lattice to a higher vibrational level
- **3**: The scattered photon is created by annihilation of the electron-hole pair

As such, it is seen that the electrons in a semiconductor mediate the Raman scattering process. The energy of the intermediate excited state (and thus the incident radiation) generally has no effect on the Raman shift itself, but it can have a large influence on the intensity of different scattering processes. Referring to figure 2.8, standard Raman scattering experiments are carried out with an excitation wavelength far below the band gap, and thus there are no nearby electronic states for the electron to occupy. However, Heisenberg's uncertainty principle allows the existence of virtual states with a lifetime inversely proportional to the energy difference between the virtual level and the closest real level, $\Delta E \Delta t \geq \hbar/2 \Rightarrow \hbar/(2\Delta E) \geq \tau$, where $\tau = 1/\Delta t$. This means that, in principle, any optical frequency can be used for Raman scattering. [10, 16, 19]

The ratio between the intensities in the Stokes- and anti-Stokes bands is easily explained by Bose-Einstein statistics. Here, the ratio between the population of two states is given as $\exp(\hbar\omega/kT)$, which means that at low temperatures, the probability that an incident photon encounters the semiconductor in a vibrationally excited state is much lower than the probability that the photon encounters it in the ground state, leading to the Stokes intensity being higher than the anti-Stokes one. [17]

2.2.2 Classical description

While the quantum description outlined above provides some insight into the Raman process, this is advantageously supplemented by a classical description, which shows how material parameters

influence the scattering process. In this description, a planar electromagnetic field incident on the crystal is considered, which induces the polarization

$$\mathbf{P} = \epsilon_0 \left(\chi^{(1)} \mathbf{E} + \chi^{(2)} \mathbf{E}^2 ... \right) \tag{2.33}$$

where $\overline{\chi}^{(1)}$ the first order electric susceptibility of the crystal, $\overline{\chi}^{(2)}$ is second order, and $\mathbf{E}$ is the incoming electric field. The second term on the right hand side is responsible for nonlinear optical interactions, which can be also be utilised for Raman scattering, but as the effects are negligible at low excitation intensities, it is not considered in this derivation.

If it is assumed that the crystal is isotropic or that $\mathbf{E}$ is parallel to a high symmetry axis, $\chi^{(1)}$ (subsequently denoted χ) is a scalar. If the incident radiation (subscript i) has a cosine variation, $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \cos(\mathbf{k}_i \mathbf{r} - \omega_i t)$, the induced polarisation can then be written as

$$\mathbf{P}(\mathbf{r}, t) = \epsilon_0 \chi \mathbf{E}_0 \cos(\mathbf{k}_i \cdot \mathbf{r} - \omega_i t) \tag{2.34}$$

The thermally excited atomic vibrations present in the crystal, for which the normal modes are quantized into phonons, are plane waves on the form $\mathbf{Q}(\mathbf{r}, t) = \mathbf{Q}_0 \cos(\mathbf{q} \cdot \mathbf{r} - \omega_0 t)$, where q is the phonon wavevector and ω_0 is the frequency. If the adiabatic theorem is invoked, such that it is assumed that $\omega_i >> \omega_0$, and if the amplitude of the lattice vibrations are small, χ can be expanded in a Taylor series [10]

$$\chi = \chi_0 + \left(\frac{\partial \chi}{\partial \mathbf{Q}} \right)_0 \mathbf{Q}(\mathbf{r}, t) + \left(\frac{\partial^2 \chi}{\partial \mathbf{Q}^2} \right)_0 \mathbf{Q}^2(\mathbf{r}, t) + ... \tag{2.35}$$

where χ_0 is the susceptibility in the absence of phonons. From this, it is obvious that the response of the system to the electromagnetic field is dependent on fluctuations in atomic positions. Neglecting the quadratic $\mathbf{Q}$-term for the time being (which is responsible for second-order Raman scattering), the term linear in $\mathbf{Q}$ denotes a susceptibility fluctuation induced by the lattice vibration. Introducing a polarization in-phase with the electric field $\mathbf{P}_0$ and a polarization $\mathbf{P}_{ind}$ induced by the phonon, we can write

$$\mathbf{P}(\mathbf{r}, t, \mathbf{Q}) = \mathbf{P}_0(\mathbf{r}, t) + \mathbf{P}_{ind}(\mathbf{r}, t, \mathbf{Q}) \tag{2.36}$$

with

$$\mathbf{P}_0(\mathbf{r}, t) = \epsilon_0 \chi_0 \mathbf{E}_0 \cos(\mathbf{k}_i \mathbf{r} - \omega_i t) \quad ; \quad \mathbf{P}_{ind}(\mathbf{r}, t, \mathbf{Q}) = \epsilon_0 \left(\frac{\partial \chi}{\partial \mathbf{Q}} \right)_0 \mathbf{Q}(\mathbf{r}, t) \mathbf{E}_0 \cos(\mathbf{k}_i \cdot \mathbf{r} - \omega_i t)$$

The former is the term responsible for Rayleigh scattering. Rewriting the latter, it is obtained that

$$\mathbf{P}_{ind}(\mathbf{r}, t, \mathbf{Q}) = \frac{\epsilon_0}{2} \left(\frac{\partial \chi}{\partial \mathbf{Q}} \right)_0 \mathbf{Q}_0 \mathbf{E}_0 \left\{ \cos\left[(\mathbf{k}_i + \mathbf{q}) \cdot \mathbf{r} - (\omega_i + \omega_0) t \right] + \cos\left[(\mathbf{k}_i - \mathbf{q}) \cdot \mathbf{r} - (\omega_i - \omega_0) t \right] \right\} \tag{2.37}$$

This expression consists of Stokes and anti-Stokes waves with frequencies and wavevectors $\mathbf{k}_S = \mathbf{k}_i + \mathbf{q}$, $\omega_S = \omega_i + \omega_0$ and $\mathbf{k}_{AS} = \mathbf{k}_i - \mathbf{q}$, $\omega_{AS} = \omega_i - \omega_0$, respectively. This is a result of wavevector and energy conservation, namely that the initial and final state of the total system must conserve both properties. Since the energy of the incident photon, for excitations with visible light, is much larger than that of the phonon, it follows that $\omega_i \approx \omega_S$ (and similarly for ω_{AS}). It follows, then, as there is virtually no dispersion in the refractive

index between these frequencies, that the largest possible magnitude of the phonon wave vector is $|\mathbf{q}| = (n(\omega_i)\omega_i - n(\omega_S)\omega_S)/c$. This value, corresponding to $10^6 \mathrm{cm}^{-1}$, is very small compared to the size of a semiconductor Brillouin zone (BZ). Consequently, first-order Raman scattering only probes the very center of the first BZ, $\mathbf{q} \approx 0$. From this, it is again seen how the Raman spectrum is a map of the energies of long-wavelength phonons. [10, 18]

The parameter that is at the centre of this investigation is the intensity of the light scattering from a sample, I_s, that is, the ratio of the average power scattered into a solid angle by the induced polarization, P_{ind}. This is proportional to $|P_{ind}|^2$, and can be written as (lifting the restriction that the crystal is isotropic) [19]

$$I_s \propto |\mathbf{e}_i \cdot \left(\frac{\partial \overline{\chi}}{\partial \mathbf{Q}}\right) \mathbf{Q}(\omega_0) \cdot \mathbf{e}_s|^2 = |\mathbf{e}_i \cdot \overline{R} \cdot \mathbf{e}_s|^2 \tag{2.38}$$

where $\mathbf{e}_i$ and $\mathbf{e}_s$ are unit polarization vectors of the incident and scattered light, respectively, and $\overline{R}$ is the *Raman tensor*. It was assumed here that the phonon wavevector is zero, as was reasoned for one-phonon scattering. As the intensity is proportional to the square of the lattice displacement, it is seen that there is no Stokes scattering if there are no lattice vibrations present. This is, however, never the case, as even at zero temperature, there will always be zero-point motion. [10]

The definition in 2.38 is convenient because the Raman tensor is usually the property that is given in the literature for the symmetry modes in crystals. From this equation, it is also seen that in order for any phonon mode to participate in Raman scattering, at least one component of the Raman tensor must be non-zero. This very important point forms the basis of the Raman spectroscopy *selection rules* which, returning to quantum mechanical terms, describe whether a transition between quantum states is possible (allowed) or very unlikely (forbidden). The probability of a transition can be given through the overlap of two states with some transition matrix, which for optical transitions in the dipole approximation is merely the momentum matrix, as seen in appendix A. As reasoned in section 2.2.1, a Raman scattering event requires the simultaneous participation of three transitions, and it is thus a generally unlikely process. Fortunately, while the calculation of the Raman tensor can be a time-consuming process, the non-zero components of the Raman tensor can be determined solely from symmetry arguments, as a transition can only be allowed if the parity of the overlap integral of the transition is non-zero. By consulting group character tables, which are produced through the results of group theory, it can be found whether a given phonon mode is Raman allowed (active) or forbidden (inactive) for a given crystallographic point group. An in-depth discussion of the symmetries of the Raman tensor is, however, beyond the scope of this thesis. [10, 16, 18]

2.2.3 Raman scattering methods and variations

There are several variations of the Raman scattering scheme which can be tuned to extract different information about a sample. Some of the most common variations are outlined here, but the following is by no means an exhaustive overview of the utilization of the Raman process.

Resonant Raman scattering

Typically, the property of interest in Raman measurements is the phonon energy. However, it is also possible to extract information about a wide range of other parameters, including the deformation potential and the electronic band structure. Because there are usually several

intermediate states involved in Raman scattering, the Raman transition probability matrices contain far too many terms to accurately extract this information. By using an excitation laser which has an energy close to an electronic interband transition, all other intermediate states usually has negligible contribution in comparison. This allows for much better isolation of the variables involved, which makes it possible to extract additional information from the scattered intensity. [10]

Another possible advantage of resonant Raman scattering is that, while standard Raman scattering has little chance of occurring and thus probes the whole sample specimen, the resonant enhancement allows for a much higher scattering rate, such that only the top layers of a sample are probed. This can be advantageous for e.g. the study of thin film morphology.

Infrared spectroscopy

While infrared spectroscopy is fundamentally different from Raman scattering, in that it involves a "real" absorption process, it is often used in conjunction with Raman spectroscopy. Infrared spectroscopy is a variation of common spectroscopy where the absorption is greatly enhanced if the energy of an incoming photon is equal to the difference between two vibrational levels. Thus, IR spectroscopy, just as Raman spectroscopy, studies phonon modes. The difference lies in which modes are active for each technique. In centrosymmetric crystals, for example, the Raman scattering is forbidden for odd-parity phonons but is allowed for even-parity phonons, while the opposite is true for IR spectroscopy. Thus, in centrosymmetric crystals, Raman scattering and IR spectroscopy are complimentary in that they yield information about the phonon modes which are forbidden for the other technique. Additionally, in some crystals, some phonon modes can be found which are neither Raman nor IR active, and are thus said to be *silent*. In crystals without a centre of inversion, phonon modes may be both Raman and IR active. [10, 17]

Brillouin scattering

In any crystal with two atoms per primitive cell, it is well known that there exists both optical and acoustic phonon branches. Raman scattering is typically only referred to as scattering off of optical phonons, while scattering of acoustic phonons is called *Brillouin scattering*. In essence, the physics of these processes is very similar, but for one-phonon Raman spectroscopy, where only the BZ center is probed, the acoustic phonon energy is much lower than that of the optical phonons. The primary usage of Brillouin scattering is that the information that it gives about long-wavelength acoustic phonons (sound waves) can be analysed to find the elastic constants of a material [19]. [10]

Two-phonon Raman scattering

So far, only phonons with $\mathbf{q} \approx 0$ have been considered as a result of the wave vector conservation requirement. However, if two phonons are involved in the scattering process, corresponding to the second term on the right hand side of eq. 2.35, wave vector conservation requires only that the sum of the wave vectors of the two phonons is approximately zero. This means that a phonon with a wave vector of any magnitude can contribute to two-phonon Raman scattering, provided that the sum of the wave vectors is zero. Usually, the two-phonon Raman spectrum is dominated by phonons with equal but opposite wave vectors, i.e. $\mathbf{q}_a = -\mathbf{q}_b$, which are called *overtones*. Thus, the two-phonon Raman spectrum can often be interpreted as the one-phonon density of states multiplied by two, although modified by the scattering efficiency. [10, 19]

2.3 Characterization of electrical properties of semiconductors

This section starts by considering the Hall effect, which can be used to characterize the electrical properties of a semiconductor. It is quickly realized that to do this, certain geometric and electrical considerations must be taken into account. This leads to the van der Pauw method, which describes the most ideal geometric configurations for accurate Hall measurements. Furthermore, some practical aspects concerning the preparation of a semiconductor sample for Hall measurements must be considered. Herein special concern is given to the choice of metal-semiconductor contacts.

2.3.1 The Hall effect

The Hall effect can be observed in a semiconductor material, when it is subject to both an electric and a magnetic field. To understand how the Hall effect can be utilized figure 2.9 can be considered.

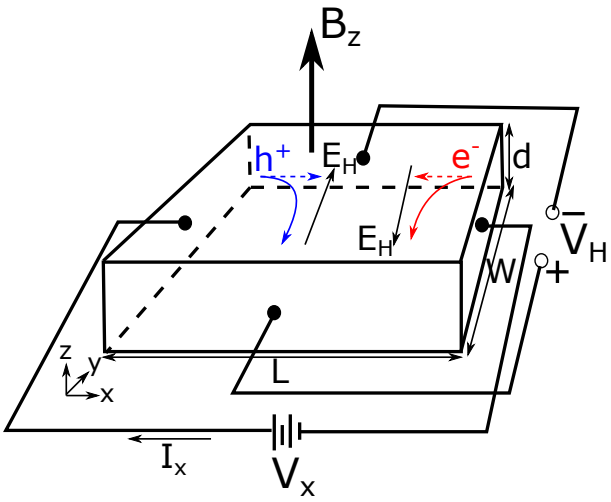


Figure 2.9: *The Hall scheme is shown here, where the square slab is the semiconductor sample and a current is applied perpendicular to the measured Hall field.*

In the figure, a voltage is applied between two ends of a square slab of non-magnetic semiconductor material with a geometry such that this produces an electric field along the x-direction. The figure shows how the charge carriers move in response to this with a dotted line. However, as a magnetic field is applied along the z-direction ($\mathbf{B} = \hat{z}B_z$), this changes the movement of the carriers, in accordance with the Lorentz force

$$\mathbf{F} = q[\mathbf{E} + \mathbf{v} \times \mathbf{B}] \tag{2.39}$$

where q is the charge, $\mathbf{E}$ is the electric field, $\mathbf{v}$ is the velocity of the carrier, and $\mathbf{B}$ is the magnetic field. The force on the carriers due to the magnetic field is easily realized by considering the cross product in eq. 2.39 and the sign of q. Electrons and holes thus move in the negative y-direction. This builds up a charge difference between either end of the semiconductor, resulting in an electric field in the y-direction, and its sign thus depends on the type of charges, that accumulate at the end of the semiconductor. At some point the electric field in the y-direction will balance the force from the magnetic field, which points in the opposite direction. Thus a

steady state is reached, where

$$\begin{aligned} \hat{y}F_y &= q[\hat{y}E_y + \hat{x} \times \hat{z}v_x B_z] = 0 \\ qE_y &= qv_x B_z \end{aligned} \tag{2.40}$$

and in this steady state, the equilibrating electric field is called the *Hall field*, and it is shown as E_H in the figure. The Hall field can then be measured through the Hall voltage, $V_H = E_H W$. In the figure, the polarity of the Hall voltage depends on the majority carriers, which determine the direction of the Hall field. Thus if the Hall voltage is positive, this means that holes are the majority carriers, and thus the semiconductor is of p-type. Thus for an n-type semiconductor, the Hall voltage will have a negative polarity. [20, 21, 22]

At equilibrium $E_H = E_y$, and $E_H = V_H/W$ can thus be inserted into eq. 2.40. If a p-type semiconductor is considered, V_H is positive. When the holes are subjected to an electric field in the x-direction, they are subjected to a drift force, and their velocity v_x is thus a drift velocity given by $v_{dx} = J_x/(ep)$, where e is the elementary charge, J_x is the current density, and p is the concentration of holes. Eq. 2.40 may thus be rewritten as

$$\begin{aligned} \frac{V_H}{W} &= \frac{J_x}{ep} B_z \\ V_H &= \frac{I_x B_z}{epd} \end{aligned} \tag{2.41}$$

where it was used that $J_x = I_x/A$ with $A = Wd$ in the x-direction (W, d, and L are the geometric parameters of the slab, also shown in the figure). Since the current can be controlled, and the Hall voltage can be measured, the concentration of holes, p, can thus be obtained by rearranging eq. 2.41 to

$$p = \frac{I_x B_z}{edV_H} \tag{2.42}$$

and similarly for an n-type semiconductor

$$n = -\frac{I_x B_z}{edV_H} \tag{2.43}$$

which will yield a positive quantity, because the Hall voltage will also be of negative polarity for an n-type semiconductor. [20, 21, 22]

The mobility of the main charge carrier may also be calculated. Assuming a p-type semiconductor and using the relation between current density and an electric field $J_x = \sigma E_x$ with σ being the conductivity, an expression for the mobility of holes μ_p is obtained

$$J_x = \sigma E_x \qquad => \qquad ep\mu_p \frac{V_x}{L} = \frac{I_x}{Wd} \qquad => \qquad \mu_p = \frac{1}{ep}\frac{I_x}{V_x}\frac{L}{Wd} \tag{2.44}$$

where it was used that $E_x = V_x/L$, $J_x = I_x/(Wd)$, and $\sigma \approx ep\mu_p$ by assuming that $p >> n$. A similar equation is found for the electron mobility in case of an n-type semiconductor with $n >> p$ as

$$J_x = \sigma E_x \qquad => \qquad en\mu_n \frac{V_x}{L} = \frac{I_x}{Wd} \qquad => \qquad \mu_n = \frac{1}{en}\frac{I_x}{V_x}\frac{L}{Wd} \tag{2.45}$$

Thus the mobility of a hole in a p-type semiconductor and an electron in an n-type semiconductor has been determined. The mobility is a quantity that describes the response of a charge carrier,

when it is subjected to an electric field. It is thus an important property of a semiconductor in order to better understand its electrical properties. [20, 21, 22]

When doing an actual experiment on a semiconductor, the semiconductor material is often produced in the form of a thin film on an underlying substrate. An example is GaN on top of Al_2O_3, where the GaN layer thickness is typically in the range of a few hundreds of nanometres and up to a few microns. Thus a few different quantities are often introduced to simplify the calculations. First of all, I_x and V_x can be converted to the sheet resistance by means of Ohms law such that $R_s = V_x/I_x$. This quantity will be addressed later in this section, but it is an experimentally determinable quantity. Also the sheet density is introduced as $n_s = nd$ and is thus in units of cm^{-2}. Furthermore, the sample can be geometrically designed such that $W = L$. Thus in case of e.g. an n-type semiconductor, eq. 2.42 and eq. 2.44 become

$$n_s = \frac{I_x B_z}{e|V_H|} \qquad \& \qquad \mu_n = \frac{1}{e n_s R_s} \tag{2.46}$$

where the popular Hall coefficient is defined as $R_H = 1/(en_s)$. Finally, it should be mentioned that the simple derivation presented here is approximative, because carrier statistics have not been considered. This can be done e.g. through the Boltzmann transport equation, which would introduce a term describing scattering mechanisms. [22]

2.3.2 The van der Pauw method

Attention is now turned towards the sheet resistance R_s. L. J. van der Pauw [23] showed a method by which the sheet resistance could be obtained experimentally with minimum measurement error, and subsequently the method has been widely applied in semiconductor physics. In order to apply the method, a number of idealized conditions have to be met

- The sample thickness must be uniform and of negligible thickness compared to its area
- The composition of the sample must be homogeneous
- The sample must be symmetrical
- The contacts should be placed at the perimeter of the sample
- The area of the contacts should be less than 1/10th of the area of the sample

Thus any violation towards these conditions may lead to an increased error rate, naturally depending on the degree of the error. In order to reduce the errors in measurements, suitable van der Pauw geometries can be applied, which are shown in figure 2.10. [21, 23, 24, 25]

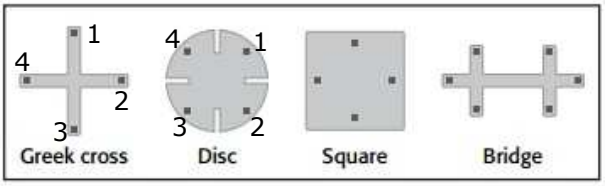


Figure 2.10: *Examples of van der Pauw geometries. The disc shape is often referred to as the van der Pauw clover. Taken from [24].*

If it is assumed that a suitable geometry is chosen (take e.g. the disc shape from the figure), and that the conditions listed above are met, van der Pauw showed that the sheet resistance could be determined as

$$e^{-\pi R_{vertical}/R_s} + e^{-\pi R_{horizontal}/R_s} = 1 \tag{2.47}$$

where $R_{vertical}$ and $R_{horizontal}$ are the resistances measured along the vertical and horizontal directions, respectively. These resistances are defined through the reciprocal theorem, which states that the resistance measured in one direction, e.g. from 4 to 1 in disc scheme of the figure should equal the resistance measured from 1 to 4. Thus in order to increase the accuracy of the resistance measurements used in eq. 2.47, the resistances are often defined as the average values

$$R_{vertical} = \frac{R_{12,34} + R_{34,12} + R_{21,43} + R_{43,21}}{4} \tag{2.48}$$

$$R_{horizontal} = \frac{R_{41,32} + R_{32,41} + R_{14,23} + R_{14,32}}{4} \tag{2.49}$$

where the notation is such that $R_{12,34}$ means that this resistance is determined by letting a current flow between contact 1 and 2 while the voltage is measured between contact 3 and 4. [21, 22, 23, 24]

2.3.3 Measurement of the Hall voltage

The determination of the Hall voltage may in theory seem like a simple task by considering the geometry of the Hall setup scheme in figure 2.9. However, a number of different voltages may "hide" in the voltage measurements

- V_M: voltage that arises from small misalignment of the contacts parallel to the Hall field. It will be proportional to the current flow, but not the magnetic field.
- V_S: voltage due to the Seebeck effect. It arises if there is a thermal gradient between the contacts, where V_H is measured. This effect is independent of applied current and magnetic field.
- V_E: voltage from the Ettingshausen effect. Due to electrons moving with different velocities (often termed hot or cold electrons), the Lorentz force on these electrons is different, and thus they will accumulate at different positions inside the semiconductor. This results in an internal thermal gradient in the semiconductor between the hot and cold part, and thus a diffusion current towards the colder part of the semiconductor. This effect is most prominent in poor thermal conductors. This effect will thus be proportional to both the applied current and magnetic field.
- V_N: voltage due to the Nernst effect. If the semiconductor is subjected to an external thermal gradient, this will cause a diffusion current of electrons towards the colder part of the semiconductor. These electrons will also be affected by the magnetic field, and thus this may generate an electric field normal to direction of the diffusion current and the magnetic field. It can occur e.g. if there is a resistance across the contacts, which thus results in heating. It will thus be proportional to the magnetic field.
- V_R: voltage arising from the Righi-Leduc effect. The temperature gradient creating the Nernst effect will also lead to a new temperature gradient perpendicular to the magnetic field and the temperature gradient responsible for the Nernst effect. This effect will thus be proportional to the magnetic field.

Considering all of these voltages as part of the measured voltage along with the Hall voltage, the measured voltage will be

$$V_{HM} = V_H + V_M + V_S + V_E + V_N + V_R \tag{2.50}$$

However, most of these additional voltages that will be measured may actually be cancelled out by doing a number of different measurements. If a sample e.g. with the disc geometry from figure 2.10 is considered along with the dependencies of the undesirable voltages in eq. 2.50 on the magnetic field and the current, it is possible to remove these voltages by doing an average of the measurements in an appropriate manner, i.e.

$$\frac{1}{4}[V_{24,P} + V_{24,N} + V_{42,P} + V_{42,N}] = V_H + V_E \tag{2.51}$$

where $V_{24,P}$ denotes the voltage measured from contact 2 to 4 with a positive contribution from the magnetic field. $V_{24,N}$ is then the same but with a magnetic field in the opposite direction. Bear in mind that e.g. $V_{42,P}$ is then the potential measured in the opposite direction, and its sign will then be opposite to that of $V_{24,P}$. Thus the effects only dependent on the magnetic field are cancelled by changing the polarity of the magnetic field, and those dependent on the current are cancelled from different measurement directions. Only V_E cannot be cancelled out in this manner, since it is dependent on both the magnetic field and the current. However, it can often be ignored due to the thermal conductivity of the semiconductor ensuring that $V_H >> V_E$. [21, 22]

Working with eq. 2.51 thus requires several measurements, especially if the sample is not completely symmetric (which should always be checked in the Hall measurements). Therefore a method described by C. Kirkpatrick [22] is followed instead, which in principle allows for the removal of all voltages except for V_E with a single measurement. First, splitting $V_H M$ in two terms such that

$$V_{HM} = V_0 + V_H \tag{2.52}$$

where $V_0 = V_M + V_S + V_E + V_N + V_R$. If a current linearly varying with time is applied, i.e.

$$I(t) = m_I t + b \tag{2.53}$$

where t is the time, and the other two constant can be manipulated by the user. The measured Hall voltage from eq. 2.52 then becomes

$$V_{HM} = V_0 + \frac{B(m_I t + b)}{q n_s} \tag{2.54}$$

It should be underlined that m_I should be small, such that the current does not vary too rapidly. This is required in order for the system to maintain an equilibrium state during the measurements. Now, by taking the derivative of eq. 2.54 with respect to time and magnetic field, the terms in V_0 that only depend on either current or magnetic field are all removed. This is done as following

$$\frac{\mathrm{d}V_{HM}}{\mathrm{d}t} = \frac{\mathrm{d}V_0}{\mathrm{d}t} + \frac{B m_I}{q n_s} \tag{2.55}$$

and then differentiating with respect to magnetic field

$$\frac{\mathrm{d}^2 V_{HM}}{\mathrm{d}B\mathrm{d}t} = \frac{\mathrm{d}^2 V_0}{\mathrm{d}B\mathrm{d}t} + \frac{m_I}{qn_s} = \frac{\mathrm{d}^2 V_E}{\mathrm{d}B\mathrm{d}t} + \frac{m_I}{qn_s}$$
$$\frac{\mathrm{d}^2 V_{HM}}{\mathrm{d}B\mathrm{d}t} \approx \frac{m_I}{qn_s} \tag{2.56}$$

since the $\frac{\mathrm{d}^2 V_E}{\mathrm{d}B\mathrm{d}t}$ term can often be neglected in comparison to the Hall voltage [22]. The term on the left hand side of eq. 2.56 can be measured as follows: first the Hall voltage is measured with respect to the current and time. The slope of this measurement is then $\mathrm{d}V_{HM}/\mathrm{d}t$. This measurement should then be acquired at three or more different magnetic field strengths. This allows for a plot of $\mathrm{d}V_{HM}/\mathrm{d}t$ vs magnetic field strength. The slope of the curve made by these points is then $\mathrm{d}^2 V_{HM}/\mathrm{d}B\mathrm{d}t$, and the sheet carrier density along with the mobility can then be extracted. [22]

2.3.4 Metal-semiconductor contacts

When a metal-semiconductor contact is formed, the interface is typically illustrated by aligning the band diagrams of each material and then letting the respective Fermi levels equilibrate, which happens due to electrons flowing from the lower work function material to that of the higher work function material. This will as usual result in band bending in the semiconductor material, by which the degree of bending mainly depends on the difference between the work function of the metal ϕ_m and the semiconductor ϕ_s. This difference in work function may lead to different types of contacts. One typically distinguish between Schottky and ohmic contacts, wherein the Schottky contact is better known as a *Schottky diode* meaning it is of rectifying nature. This type of contact is obviously undesirable with respect to the van der Pauw method. The ohmic contact, however, is a non-rectifying contact and is thus the desired contact type. [20, 21]

In order to obtain an ohmic contact a scheme as the one illustrated in figure 2.11 can be considered. Here a metal and an n-type semiconductor are first considered before contact (a of the figure), and the scheme is then illustrated after contact (b of the figure). This is an ohmic contact, because the conduction band of the semiconductor bends downwards crossing the equilibrated Fermi level of the heterostructure. This is possible because their work functions $\phi_s > \phi_m$ and because the difference between these work functions is small enough. In the opposite case where $\phi_s < \phi_m$ the heterostructure would have formed a Schottky diode, where the conduction band would bend upwards, thus resulting in a diode-characteristic built-in potential (often called a Schottky barrier). In the ohmic contact shown in the figure, the difference between $\phi_s - \phi_m = \phi_{Bn} \approx \phi_n$ gives the barrier height, and it should be small enough that the application of a positive potential to the semiconductor will easily overcome it and thus enable the flow of electrons from the semiconductor and into the metal. Conversely, there is no barrier from the metal to the semiconductor, since the equilibrated Fermi level overlaps the conduction band of the semiconductor. [20, 21]

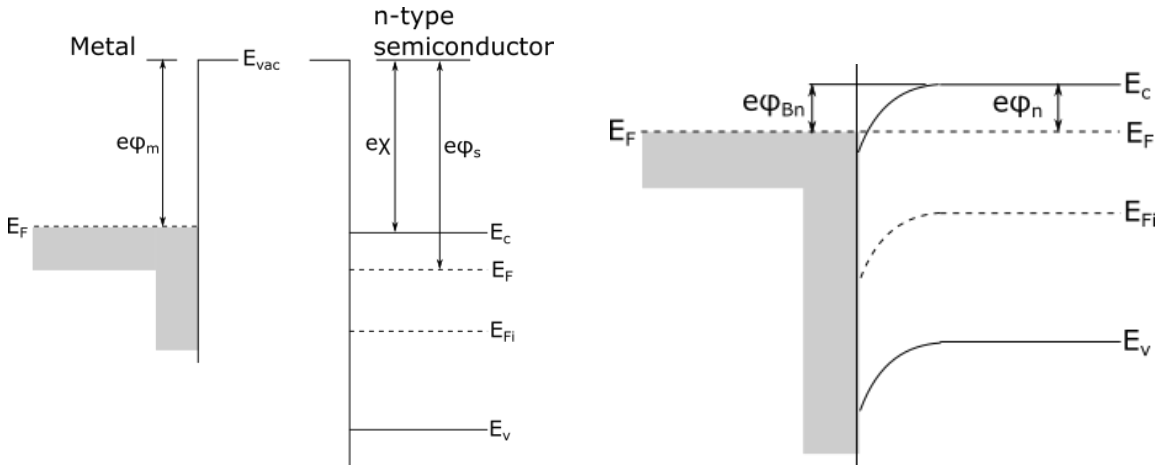


Figure 2.11: *Band diagram of an ohmic contact. ϕ_m and ϕ_s are the metal and semiconductor work functions, respectively. χ is the electron affinity, E_F and E_{Fi} refer to the extrinsic and intrinsic Fermi levels, respectively.* ***a****: shows the scheme before contact.* ***b****: shows the scheme when contact is made, and the Fermi levels have equilibrated.*

It is thus of importance to know the electrical properties of the semiconductor in order to choose a suitable metal in order to form ohmic contacts. Furthermore, it is important to know, whether the semiconductor is of n-type or p-type nature, since if it was p-type it would be required that $\phi_m > \phi_s$ in order to form an ohmic contact. Also the barrier height ϕ_n and the width of the barrier region (the region in the figure, where it bends) can be significantly reduced by increasing the doping concentration of the semiconductor. This thus makes the non-rectifying nature of the contact more ideal. Another possibility is to form a Schottky diode-type contact, but with a semiconductor that contains a heavily doped region between the metal contact and the semiconductor itself. This may result in a very narrow depletion region thus increasing the probability of tunnelling through the barrier, which may ideally have the same effect as the ohmic contact. [20, 21]

General properties of GaN

3

The properties of a solid stem from its crystal structure, and it is thus of essence to begin this chapter with a description of the crystal structure of GaN. Here, both zinc blende and wurtzite structures are considered and modelled in a simple empirical manner. Furthermore, exciton effects and the effects of strain on the band structure of GaN are discussed as well. It is then integral to give a thorough discussion of the luminescence of GaN, and through this gain an idea of the reasons for the very complex luminescent behaviour seen in the literature. While the theory in this chapter is discussed with respect to GaN, most of it is also valid for the remaining compounds in the III-nitride group.

3.1 Crystal properties

GaN exists in two crystal forms, zinc blende and wurtzite. The zinc blende structure can be visualized by considering two interpenetrating fcc lattices separated by a vector $\vec{\tau} = a(1/4, 1/4, 1/4)$, where a is the lattice constant. The two fcc lattices then consists of Ga and N atoms, respectively. This is illustrated in figure 3.1(a), where the zinc blende unit cell is shown. Similarly, the wurtzite crystal structure consists of two interpenetrating hcp lattices with Ga in one hcp lattice and N in the other. In wurtzite the two hcp lattices are separated by $\vec{\tau} = (0, 0, 5c/8)$, where c is the vertical lattice constant. The unit cell of wurtzite is shown in figure 3.1(b). It is the lattice vectors $\vec{a}_1$, $\vec{a}_2$, and $\vec{a}_3$ that span the primitive unit cell for either crystal structure. From the volume they expand, it can be seen that the primitive unit cell of zinc blende contains 2 atoms (1 Ga and 1 N), because the 4 corner atoms are each shared by 4 adjacent cells. The volume expanded by the primitive unit cell of wurtzite contains 4 atoms (2 Ga and 2 N), which is quickly realized by noting that there are 8 corner atoms each shared by 8 cells, whereas the 4 nitrogen atoms at the edges are each shared by 4 cells. It may seem inconvenient that the planar lattice vectors here are chosen to form a 60° angle, and not a 120° angle, however, if the $\vec{a}_2$ vector is placed to point along the y-axis, the two vectors will indeed form the 120° angle, and thus both options of planar lattice vectors are equivalent and expand an equivalent volume. [1, 2, 26, 27]

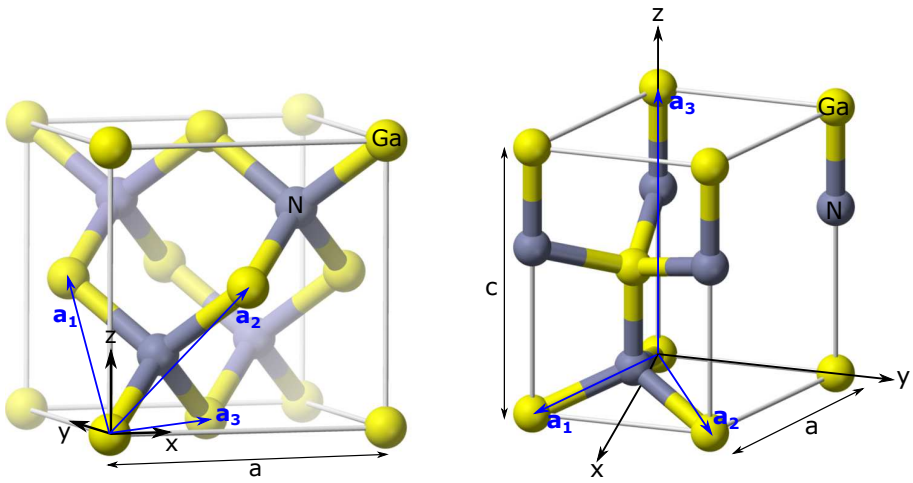


Figure 3.1: **a**: *The 8-atom cubic unit cell for zinc blende GaN, and* **b**: *the primitive unit cell for wurtzite GaN. Both images are from Wikipedia with a modified coordinate system. The lattice constants a in each figure are not the same, but are related by* $a_{zb} = \sqrt{2}a_{wz}$, *i.e. the lattice constant for zinc blende equals the length of* $\vec{a}_2$ *in wurtzite.*

From the definition of the lattice vectors in the two figures, their vectorial descriptions for zinc blende become

$$\vec{a}_1 = (0, \frac{a}{2}, \frac{a}{2}) \qquad \vec{a}_2 = (\frac{a}{2}, 0, \frac{a}{2}) \qquad \vec{a}_3 = (\frac{a}{2}, \frac{a}{2}, 0) \tag{3.1}$$

and for wurtzite

$$\vec{a}_1 = (\frac{a\sqrt{3}}{2}, \frac{-a}{2}, 0) \qquad \vec{a}_2 = (\frac{a\sqrt{3}}{2}, \frac{a}{2}, 0) \qquad \vec{a}_3 = (0, 0, c) \tag{3.2}$$

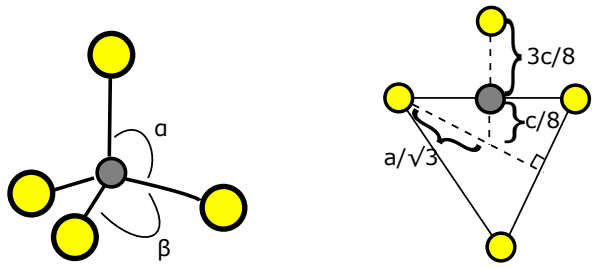
Also, from figure 3.1 it may be seen that the bonding between the atoms form tetrahedrons, where 4 Ga atoms are located at the corners of the tetrahedron and thus with equivalent distance to the N atom, which is located at the center of the tetrahedron, and vice versa. This type of bonding ensures that the atoms are very densely packed, and thus makes it difficult to excite an atom in either of these crystal structures. This type of bonding for these atoms is possible only through sp^3 hybridization, where one s orbital and three p orbitals (p_x, p_y, and p_z) are mixed in order to form 4 equivalent bonding orbitals. In an ideal structure, the two bond angles shown in figure 3.2(a) are equal. For wurtzite an ideal relation between its two lattice constants may thus be obtained. The positions of the atoms in the 4-atom primitive unit cell of wurtzite are

1. Ga-atom (0, 0, 0)
2. Ga-atom $(\frac{a}{\sqrt{3}}\ 0\ \frac{c}{2})$
3. N-atom $(\frac{a}{\sqrt{3}}\ 0\ \frac{c}{8})$
4. N-atom $(0\ 0\ \frac{5c}{8})$

where the positions of the N-atoms are realized from the displacement vector $\vec{\tau}$. From these positions the geometrical relation between the lattice constants a and c may be obtained. This is illustrated in figure 3.2(b) and by the Pythagorean relation

$$\frac{c^2}{64} + \frac{a^2}{3} = \frac{9c^2}{64} \qquad => \qquad \frac{a}{c} = \sqrt{\frac{3}{8}} \tag{3.3}$$

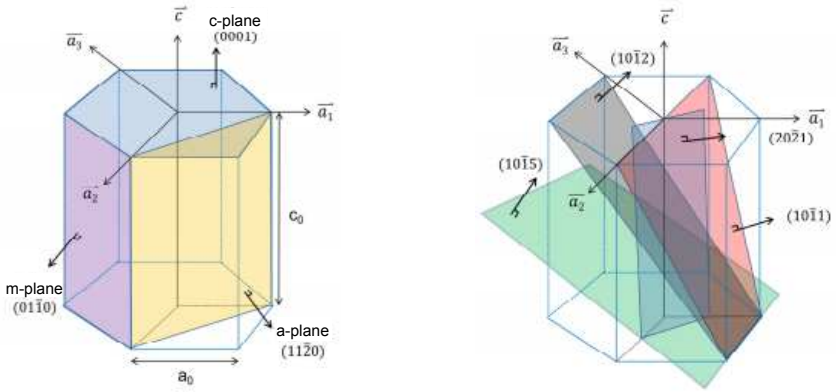
However, the ideal wurtzite structure does not exist for GaN. The non-ideality is a result of the two bond angles of the tetrahedron not being equal, i.e. $\alpha < \beta$. This further means that $a/c > \sqrt{3/8}$, and this non-ideality is often described by a parameter u (if it was ideal $u = (a/c)^2$), which thus gives the displacement of the crystal along the c-axis. The value of u is thus often defined from experimental data, since the reason for the non-ideal wurtzite structure must stem from the difference in size between its two constituent atomic species. [1, 2, 26, 27]



(a) (b)

FIGURE 3.2: *Sketches of the tetrahedral bonding scheme in GaN.* **a***: shows the most relevant bond angles, and* **b***: shows a geometric description for wurtzite, visualizing the ideal relationship between its two lattice constants, i.e.* $a/c = \sqrt{3/8}$*, which is true for* $\alpha = \beta = 109.5°$.

Finally, another consequence of the wurtzite structure is the lack of inversion symmetry along the (0001) plane. In figure 3.3 some common crystal planes of the hexagonal wurtzite structure are shown along with the definition of these planes with a four-index Miller notation with respect to the given coordinate system. The lack of inversion symmetry may be seen from the wurtzite unit cell in figure 3.1(b) by considering the direction of the N-bonds with respect to the Ga-bonds. If the crystal is inverted, these bonding directions are different, and thus the polarity of the crystal changes. A (0001) wurtzite GaN layer will thus be Ga-polar with Ga-atoms terminated at the surface, and the inverted $(000\bar{1})$ direction must then have N-polarity with N-atoms terminated at the surface. [1, 2, 26, 27]



(a) (b)

FIGURE 3.3: *Both figures show some of the crystal planes in the wurtzite GaN. The surfaces are here defined by the given coordinate system as* (a_1, a_2, a_3, c)*. The figures are taken from [2].*

3.1.1 Band structure

It is now of interest to consider the electronic properties of GaN through its band structure. In order to do this in a numerically feasible manner, the Empirical Pseudopotential Method (EPM) is deployed. A description of the EPM is given in appendix B, followed by the construction of a standard matrix eigenvalue problem (eq. B.15) from the most general evaluations of the Schrödinger equation in a periodic potential. The resulting eigenvalue problem is then

$$\begin{aligned} H_{\vec{G},\vec{G}'} &= \langle \psi_{\vec{G}} | \hat{H} | \psi_{\vec{G}'} \rangle \\ H_{\vec{G},\vec{G}'} &= \frac{\hbar^2}{2m}(\vec{k}+\vec{G})^2 \delta_{\vec{G},\vec{G}'} + S_S(\vec{G}-\vec{G}')V_S(|\vec{G}-\vec{G}'|^2) + iS_A(\vec{G}-\vec{G}')V_A(|\vec{G}-\vec{G}'|^2) \end{aligned} \quad (3.4)$$

which consists of a kinetic and a potential part. In order to evaluate this equation, the reciprocal lattice vectors must be used and expanded such that a sufficient amount of plane waves will be included in the calculations. Considering eq. 3.1, the reciprocal lattice vectors for zinc blende are then

$$\vec{G}_1 = \frac{2\pi}{a}(-1,\,1,\,1) \qquad \vec{G}_2 = \frac{2\pi}{a}(1,\,-1,\,1) \qquad \vec{G}_3 = \frac{2\pi}{a}(1,\,1,\,-1) \quad (3.5)$$

and for wurtzite (see eq. 3.2)

$$\vec{G}_1 = \frac{2\pi}{a}\left(\frac{1}{\sqrt{3}},\,-1,\,0\right) \qquad \vec{G}_2 = \frac{2\pi}{a}\left(\frac{1}{\sqrt{3}},\,1,\,0\right) \qquad \vec{G}_3 = \frac{2\pi}{a}\left(0,\,0,\,\frac{a}{c}\right) \quad (3.6)$$

where everything is written with a similar front factor $2\pi/a$, which is henceforth omitted, such that most expressions can be written in units of this front factor. Inclusion of more reciprocal vectors within the neighbouring unit cells is compiled by

$$\vec{G}_{klm} = k\vec{G}_1 + l\vec{G}_2 + m\vec{G}_3 \quad (3.7)$$

where k, l, and m are integers e.g. in the range $-5 : +5$. The amount of vectors included in $\vec{G}_{klm}$ is then restricted by limiting its length. This is necessary since the inclusion of an infinite amount of plane waves is computationally impossible, however, for e.g. zinc blende it was found that the limitation $|\vec{G}_{klm}| = 5$ was sufficient to achieve accurate results. This is feasible, since larger $\vec{G}_{klm}$'s would correspond to something that oscillates very fast, and thus their contribution is expected to be small. [26, 28, 29]

The Brillouin zone must be constructed and the high symmetry points are found in the irreducible part of the zone. Both of the Brillouin zones are shown in figure 3.4 with their irreducible part highlighted. For zinc blende (**a** of the figure) the volume of the irreducible zone is 48 times smaller than the volume of the Brillouin zone, whereas for wurtzite (**b** of the figure) the ratio is 24.

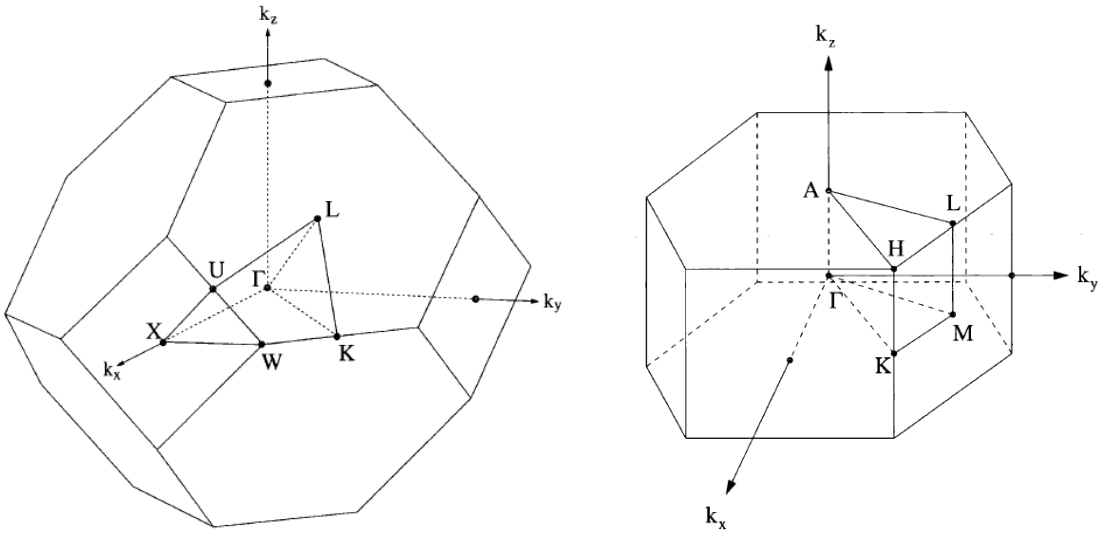


FIGURE 3.4: *Brillouin zone of* **a***: zinc blende and* **b***: wurtzite. The irreducible part of the Brillouin zone is highlighted along with its high symmetry points. The figures are taken from [26].*

Before continuing with the modelling of the properties of wurtzite GaN, the interested reader is referred to appendix C, where the modelled band structure and dielectric functions for the GaN zinc blende structure are presented, along with additional details about the computation method and theory behind it.

For wurtzite the chosen path through a set of critical points in the irreducible Brilloin zone start in A=(0, 0, $\sqrt{a/4c}$), then moves to L=(0, $1/\sqrt{(3)}$, $\sqrt{(a)/2c}$), then down to M=(0, $1/\sqrt{(3)}$, 0), then to Γ =(0, 0, 0), up to A, then to H=(1/3, $1/\sqrt{(3)}$, $\sqrt{a/4c}$), down to K=(1/3, $1/\sqrt{(3)}$, 0), and back to Γ. Since the primitive unit cell of wurtzite contains 4 atoms, the structure factors are slightly different from those of zinc blende. The wurtzite structure factors and corresponding form factors are derived in appendix B.2. These can be readily inserted in eq. 3.4 and the eigenvalue problem can be solved just like for zinc blende using the form factors listed in table B.2 (also taken from Wang et al. [28]). The calculated band diagram is shown in figure 3.5, and it agrees well with the ones calculated by both Wang et al. and Dugdale [26]. The structure calculated here may succumb to minor deviations e.g. due to the choice of lattice parameters. Furthermore, Dugdale used non-ideal pseudopotentials in order to better model the non-ideal wurtzite structure where $u \neq 3/8$. This not only means that the form factors are slightly different, it also means that the symmetric form factors for $|\vec{G}_{klm}|^2 = 6$ and $|\vec{G}_{klm}|^2 = 7 + 1/3$ are non-zero [26]. Nonetheless, these form factors are still small, and thus only cause slight differences for the calculated band structure. [26, 27, 28]

Other features such as spin-orbit coupling is often included in the literature, when considering semiconductors composed of heavy elements such as gallium. However, for GaN (and the remaining III-V nitride semiconductors) the effect from spin-orbit coupling is rather small [26]. The reason for this was shown by Dugdale [26], who calculated the influence of both Ga and N on the density of states (DoS) for wurtzite GaN. Here he found that for the top most valence bands the DoS is dominated by nitrogen, and since it is a light element the effect of spin-orbit coupling at the topmost valence bands is small. Dugdale further notes that the inclusion of spin-orbit

coupling doubles the size of the Hamiltonian matrix element used to calculate the eigenvalues, ultimately doubling the computational time for EPM scheme, which is as always an important consideration in computational methods. Dugdale did, however, employ the spin-orbit effects in his calculations, which resulted in a splitting of 11 meV at the Γ-point compared to calculations he did without spin-orbit coupling. [26, 27, 28]

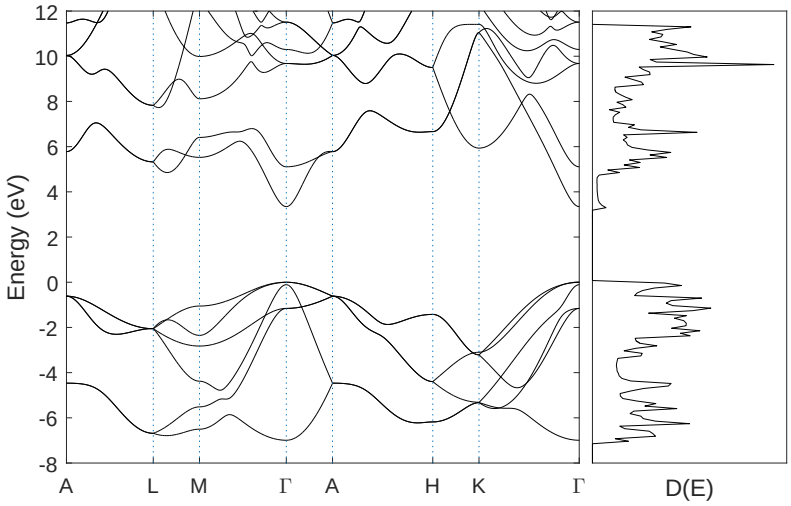


Figure 3.5: *Calculated band structure for wurtzite GaN. The band diagram shows the six topmost valence bands and some of the lowest conduction bands. The corresponding density of states have been included as well up to the energy of 11 eV. The lattice constants used are a =3.189 Å and c =5.185 Å, and the non-ideallity factor was chosen to be u =0.3768.*

In the calculations a non-ideality factor u =0.3768 was chosen from an experimental value given by Dugdale [26]. This value was only used in the calculations of the symmetric and anti-symmetric structure factors (see eq. 2.40) in order to include crystal field splitting in spite of the ideal parameters used everywhere else in the model. The effect of this is a slight splitting between the two top-most valence bands and the third top-most valence band of 110 meV, which is similar to what Wang et al. [28] obtained. The resulting energy differences in the critical points of the band diagram are summarized in table 3.1. The results show only a slight deviation compared to the results by Wang et al. (deviating at the second decimal value), and e.g. their band gap value was E_g =3.36 eV. Comparison with Ziade et al. [30], who used different form factor values than those used here, generally shows deviation of the first decimal value and their results are always slightly higher compared to the results obtained here.

Critical points	Γ	A	L	K
$E_c - E_v$ [eV]	3.347	6.396	7.373	9.033

Effective mass	band 9	band 8	band 7	band 6
Γ to A	$m_e^{\parallel} = 0.139$	$m_{hh}^{\parallel} = 1.723$	$m_{lh}^{\parallel} = 1.723$	$m_{ch}^{\parallel} = 0.124$
Γ to M	$m_e^{\perp} = 0.156$	$m_{hh}^{\perp} = 2.106$	$m_{lh}^{\perp} = 1.321$	$m_{ch}^{\perp} = 0.166$

Table 3.1: *Summarizing the energy differences between the lowest conduction band and the topmost valence band for a few critical points in the wurtzite band diagram. The calculated effective masses are given in units of the free electron mass m_0.*

The effective masses calculated here are listed in table 3.1. The effective masses have been calculated for the three topmost valence bands and the lowest conduction band, all at the Γ-point. The denotions $\parallel$ and $\perp$ refer to directions along the c-axis and along the in-plane axis, respectively, and the only values used are those that lie within the parabolic approximation. The effective masses can also be compared to Ziade et al. [30], who also compare their results with several other authors. Their results are within small error to those obtained here, and the error margin somewhat matches the generalities from the results by other authors listed by Ziade et al..

3.1.2 Optical properties

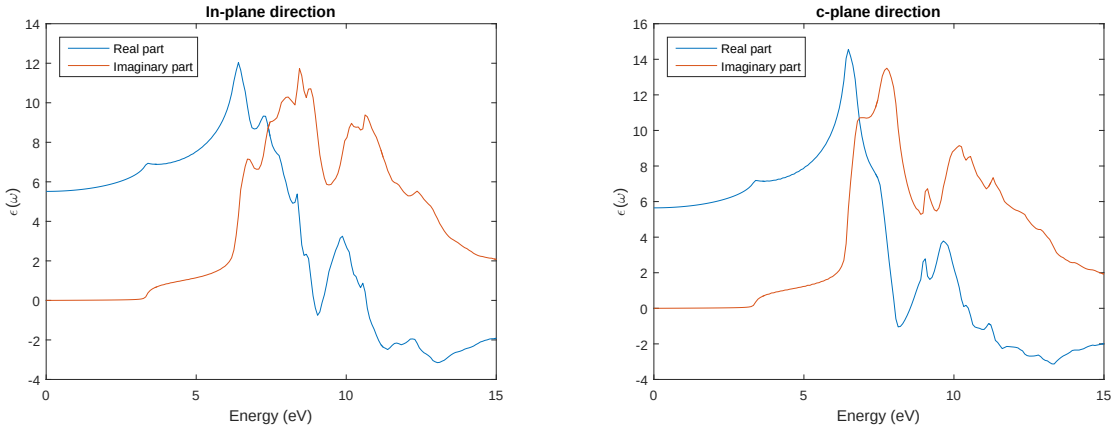
The optical properties of wurtzite GaN are calculated in a similar manner to that for zinc blende GaN, which is discussed in appendix C. However, since wurtzite is anisotropic, the impulse matrix moments must be evaluated separately to obtain the optical response for different directions of the crystal. The in-plane direction, however, contains symmetry such that it may again be evaluated only in the irreducible zone as

$$|P^z_{vc}|^2 = 24(|P^x_{vc,IBZ}|^2 + |P^y_{vc,IBZ}|^2)/2 \tag{3.8}$$

whereas for the c-axis it must be evaluated only along the z-direction of the irreducible zone, i.e.

$$|P^z_{vc}|^2 = 24|P^z_{vc,IBZ}|^2 \tag{3.9}$$

after which the remaining calculations follow those for zinc blende GaN, where the optical response is likewise calculated from eq. A.31. Since the unit cell of wurtzite contains twice as many atoms as that of zinc blende, it also contains twice as many valence bands. In order to reduce the computational time, all eight valence bands were included along with the 12 lowest most conduction bands, in the same manner as Wang et al. [28], who also argue that the contribution from higher conduction bands is negligible.



(a) (b)

FIGURE 3.6: *Calculated optical dielectric function for wurtzite GaN with a broadening factor of $\hbar\Gamma$ =0.05 eV.* **a**: *shows the in-plane polarization and* **b**: *shows the c-plane polarization. The IBZ was discretized in 109800 k-points.*

The calculated dielectric functions for either direction are shown in figure 3.6. Wang et al. also calculated the dielectric function for the in-plane direction and their result agrees well with

the spectrum shown here. The features are of same magnitude and all of Wang et al.'s features are reproduced. Some extra minor features are even obtained here, which Wang et al. did not see, and this could arguably be caused by differences in broadening factors (as for the zinc blende case). The spectrum for the c-plane is quite similar to that of the in-plane as should be expected. The value for the dielectric constant at low energy found here is Re(ϵ) =5.52 for the in-plane direction and Re(ϵ) =5.65 for the c-plane direction. These values agree well with those obtained by Wang et al. [28], who obtained 5.40 and 5.62 for the in-plane and c-plane directions, respectively. Dugdale, on the other hand obtained Re(ϵ) =5.03 for either direction. When compared to experimental work, Dugdale notes that the values for the in-plane direction differ slightly from 5.29 to 5.35 [26], and are thus in good agreement with the values obtained here. The overall spectrum for the in-plane direction also compares well to measured refractive index of wurtzite GaN [28], reproducing most of the features.

Excitons

The above evaluation of the dielectric properties does not correctly account for their exact optical properties. In order to fully account for the dielectric properties of a material, exciton effects must be considered as well. A simple theory was given in section 2.1.3, where a free electron-hole pair was considered bound to each other (a situation very similar to a hydrogen atom). The binding energy of such an exciton was then derived and is given in eq. 2.24, repeated here

$$R^* = \left(\frac{\mu}{m_0 \epsilon_r^2}\right) \times 13.6\,\text{eV} \tag{3.10}$$

and this binding energy (given that the thermal energy does not overcome it) results in a number of distinct energy levels below the conduction band edge (following a series as in eq. 2.23). The binding energy may be found from the calculated effective masses $1/\mu = 1/m_e + 1/m_h$ and the dielectric constant. However, this may not be adequate since the effects of excitons influence the dielectric function as well. This should thus have been accounted for in the model, in order to give the full description.

In wurtzite GaN the three topmost valence bands must be considered as well as the exciton k-vector. The excitons associated with the three topmost bands are usually denoted the A, B, and C, exciton, respectively (A being the topmost band of the three). The A exciton is the most common it has the lowest energy transition, that is, to the topmost valence band. It is also typically the A exciton that is seen in RT PL measurements as part of the quenched band edge luminescence. As argued in section 2.1.3, the dielectric constant of the exciton environment is often calculated as an interpolation between the static and optical dielectric constants, and Gil [27] noted that the relative dielectric constant for excitons should be approximated as an average value by

$$\frac{1}{\epsilon} \approx \frac{11}{16\epsilon_s(0)} + \frac{5}{16\epsilon_\infty(0)} = 7.753 \tag{3.11}$$

where $\epsilon_s(0) = 9.5$ [31] is the static dielectric constant for GaN and $\epsilon_\infty(0)$ is the optical dielectric constant modelled above. If this averaging is not carried out, the exciton binding energies are heavily overestimated. Considering again eq. 3.10, this leads to a binding energy for the A exciton with $\vec{k}$ ‖c-axis (the most common configuration in PL measurements of GaN) of R^* =29.1 meV. This value must thus be the same for the B exciton ($\vec{k}$ ‖c-axis) due to degeneracy, and for the C exciton ($\vec{k}$ ‖c-axis) R^* =14.8 meV. [6]

Finally, in a full description of excitons the electron and hole spins should be considered as well. Different spin combinations may lead to a difference in energy, and furthermore, the possibility of a few exciton energy states associated with transitions between the same energy bands, observable e.g. in a high resolution PL spectrum. Also the k-vector (and hence the polarization of the exciting light) of the exciton influences its energy state, and often it has to be considered, whether it is parallel or perpendicular to the c-axis. Nonetheless, keeping in mind the simplicity of the exciton description given here, the calculated binding energies for wurtzite GaN show that exciton effects must be considered even at room temperature. This is another important property of GaN, and even one that manifests itself as part of the band edge luminescence, and thus it is accountable for many of the optical properties and e.g. highly interesting applications in optoelectronic devices. [6, 10, 27]

3.2 Photoluminescence

In this section it is of interest to consider the most general luminescence seen in PL measurements on typical GaN samples grown by some of the most conventional methods (e.g. MOCVD, HVPE, and MBE). This is a topic much debated in literature, and the most accepted sources responsible for the luminescence have experienced several changes in the past 30 years due to the revelation of increased information on the topic. Several of the methods deployed as evidence in the evaluation of luminescent bands in GaN are beyond the scope of this report. However, it is still of essence to outline some of the latest results within the area, in order to gain the correct understanding of the luminescent properties of GaN, since PL measurements are one of the main methods deployed in the characterization of GaN films in this thesis. [6]

The main purpose of PL spectroscopy in this thesis is the characterization and determination of defects in GaN along with their optical properties. Unintentionally doped (UID) GaN grows as n-type due to native defects introduced during growth, no matter the growth method. Initially this led people to suspect nitrogen vacancies as the culprit for the n-type nature, mainly because they saw a much lower impurity concentration compared to the electron concentration. However, further studies have now shown that oxygen is the main defect in GaN, because it has a negative formation energy (gallium oxide has a much larger negative formation enthalpy than GaN). Oxygen is also present during any type of growth, and the negative formation energy would mean that any oxygen available during growth will be incorporated as a substitutional donor for nitrogen. This is a general problem in any GaN film, because donor-type defects might be compensated through the formation of acceptor-type defects. Therefore most luminescence bands are formed due to transitions by acceptor-like defects except for the NBE band. [1, 6, 32, 33]

The most characteristic properties of a given defect in GaN are the position of its energy level within the GaN band gap and its ability to "capture" photo generated holes required for optical transitions with electrons from the conduction band. How these properties are involved in recombination statistics for an e-A^0 transition is illustrated in figure 3.7, where the generation rate and subsequent recombination rates of free carriers are shown. Initially, the acceptor is negatively charged, because it lies below the Fermi level and therefore it accepts free electrons stemming from ionized donors, which are abundant in an n-type semiconductor. Thus when an electron-hole pair is generated by optical excitation, and the negatively charged acceptor can now become neutral by binding an excited hole from the valence band. Finally, equilibrium

conditions are reached as the acceptor once again accepts an electron from the conduction band by emission of light and hence, it will once again be negatively charged. [1, 6, 32, 33]

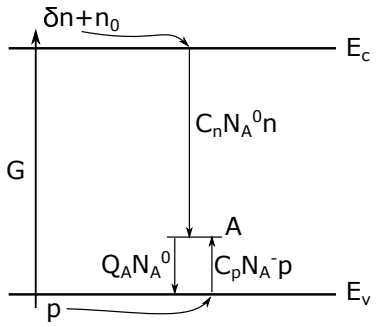


Figure 3.7: *Schematic over the e-A^0 transition, one of the most common optical defect related transition in n-GaN. Transitions rates between the energy bands and the acceptor energy state are shown as well. Inspired by [6].*

In figure 3.7, G is the generation rate of e-h pairs with concentrations $n = \delta + n_0$ (n_0 being the equilibrium concentration of electrons) and p for excited electrons and holes, respectively. $C_p N_A^- p$ is the rate by which the acceptor captures holes with C_p [$\mathrm{cm}^{-1}\,\mathrm{s}^{-1}$] being the hole capture coefficient, and N_A^- the concentration of ionized acceptors. Other processes compete with the acceptor related transition rate, e.g. the exciton dissociation rate and the nonradiative recombination rate. The exciton dissociation rate is normally considered in a full description of optical transitions, however, at low temperatures it may be considered negligible. Another important rate shown in the figure is the rate of thermal activation of holes $Q_A N_A^0$ with Q_A [s^{-1}] being the probability of thermal activation described in an Arrhenius manner ($\propto \exp(-E_A/kT)$) with E_A being the activation energy or *ionization energy*. N_A^0 is the concentration of acceptor bound holes, which are the ones that can "accept" electrons from the conduction band e.g. by optical transitions. [6, 33, 34]

If the excitation intensity is low enough for the concentration of ionized acceptors to be much larger than the concentration of acceptor bound holes, $N_A^0 << N_A^- \approx N_A$, and steady-state conditions are assumed, the rate equation for the concentration of holes in the valence band is

$$\frac{\partial p}{\partial t} = G - \sum_{i=1}^{N} C_{A,i} N_{A,i}^- p + \sum_{i=1}^{N} Q_{A,i} N_{A,i}^0 = 0 \tag{3.12}$$

where the sum runs over the total number of channels N. Similarly the rate equation for concentration of donor bound holes is

$$\frac{\partial N_{A,i}^0}{\partial t} = C_{p,i} N_{A,i} p - \frac{N_{A,i}^0}{\tau_{R,i}} - Q_{A,i} N_{A,i}^0 = 0 \tag{3.13}$$

with $\tau_{R,i}$ as the recombination lifetime with the i'th channel ($1/\tau_{R,i} = C_{n,i} n$ from the figure). The rate at which a defect captures a carrier happens much faster for minority carriers compared to majority carriers. Thus for n-type GaN the efficiency for a given recombination is proportional to the hole capture rate, meaning that the low temperature quantum efficiency for the i'th channel is mainly determined by its ability to capture holes as

$$\eta_i(0) = \frac{\text{hole capture rate of the i'th acceptor}}{\text{total rate of holes leaving the valence band}} = \frac{C_{A,i} N_{A,i} p}{\sum_{j=1}^{N} C_{A,j} N_{A,j} p} \tag{3.14}$$

If eqs. 3.12, 3.13, and 3.14 are combined, an expression for the concentration of donor bound holes become

$$N_{A,i}^0 = \frac{\eta_i(0)\left(G + \sum_{j=1}^{N} Q_{A,j} N_{A,j}^0\right)}{\tau_{R,i}^{-1} + Q_{A,i}} \tag{3.15}$$

and when this system of equations is solved, the temperature dependent PL intensity for the i'th acceptor may be expressed as [34]

$$I_{A,i}(T) = \frac{N_{A,i}}{\tau_{R,i}} = \frac{I_0}{1 + (1 - \eta_i)\tau_R Q_{A,i}} \tag{3.16}$$

where I_0 is the PL intensity prior to quenching (introduced by using that $I_0 = \eta_i G$). The probability for thermal activation of the i'th acceptor must be given by the hole capture coefficient, the degeneracy factor, g, the DoS for holes in the valence band, N_v, and the Arrhenius term, i.e.

$$Q_{A,i} = C_{p,i} g^{-1} N_v \exp\left(-\frac{E_{A,i}}{kT}\right) \tag{3.17}$$

because the probability must naturally depend on the ability to capture a hole and the amount of holes available for capture in the valence band, which is described by the density of states for holes as

$$N_v = 2\left(\frac{m_h kT}{2\pi\hbar^2}\right)^{3/2} \tag{3.18}$$

and the hole capture coefficient is defined as

$$C_{p,i} = \sigma_{p,i} v_p = \sigma_{p,i}\sqrt{\frac{8kT}{\pi m_h}} \tag{3.19}$$

where $\sigma_{p,i}$ is the hole capture cross section, and v_p is the mean thermal velocity for holes. The activation energy in eq. 3.17 is here the important parameter of interest along with the hole capture cross section, and both can be determined through PL spectroscopy for a specific type of defect by making a linear fit to eq. 3.16 in the region of thermal quenching. The hole capture coefficient additionally requires the knowledge of τ_R, which can only be determined through time-resolved PL spectroscopy. This technique was not used here, and therefore only the activation energy was determinable in this thesis. [6, 33, 34]

In the following sections, a discussion of commonly observed luminescence bands in GaN is presented. However, the fact that this overview is by no means exhaustive is indicative of the myriad of different luminescence characteristics in epitaxially growth GaN.

Near Band Edge Luminescence

The near band edge (NBE) luminescence of GaN covers its main luminescent properties that lie near the GaN band gap energy in a PL spectrum. Figure 3.8 shows the room temperature NBE luminescence around ~365 nm from a freestanding GaN sample. The peak is rather dull in RT measurements, and generally it is hard to extract any information from it, except that it is GaN. The peak is expected to be a mix of free exciton transitions that has not completely

quenched at RT, and also band to band recombination of free carriers. At lower temperatures several exciton lines may be observable even for the same type of exciton bound to different impurities (usually shallow donors). A high resolution spectrometer is often required to observe these extremely sharp exciton lines that typically have a FWHM of less than 1 meV. Observation of several exciton lines usually indicates high crystalline quality. With increasing temperature some of the impurities will ionize. This makes the appearance of the free excitons in GaN more probable (although its energy level is almost equivalent to that of the (D^0, X) state), and it will usually result in quenching of the luminescent exciton lines seen at low temperatures. The NBE luminescence is usually a good first-hand fingerprint of the sample, revealing information about its band gap energy along with valuable information about the crystalline quality. Crystalline quality can often be evaluated (with much precaution!) from the FWHM of the band edge luminescence, however, several oscillations (either representing more exciton lines or e.g. phonon replicas) can result in a rather wide NBE peak. Reshchikov [35] notes that RT PL measurements for high quality freestanding GaN often display a wide NBE luminescence band, and therefore evaluation of the NBE bandwidth may be misleading, when only RT PL measurements are considered. [6, 33]

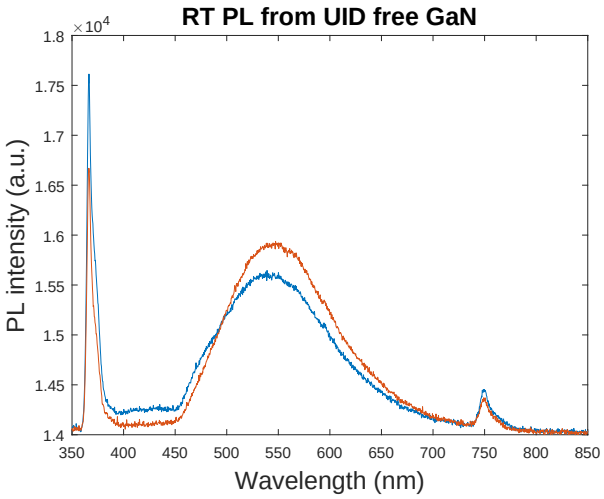


Figure 3.8: *Room temperature PL spectra from the freestanding HVPE (0001) GaN crystal. The two spectra are taken with the 150 l/mm grating. For these spectra an optical filter was used to limit the PL intensity of the NBE luminescence.*

Yellow Luminescence

The most reviewed defect luminescence from GaN is the so called yellow luminescence (YL), which has been a topic of much debate regarding the quality of GaN films. Figure 3.8 shows the YL from the freestanding HVPE sample used in this project. The spectra are taken with an optical filter, absorbing most of the PL intensity below 385 nm. As can be seen from the figure, varying the excitation intensity slightly changes the YL band, and actually it is blue-shifted. Additionally, its peak intensity at the NBE is reduced as the excitation intensity increases, while the YL intensity grows. The exact chemical origin of the YL is still unclear, and the confusion is likely connected to the fact that many experiments with doped GaN have been carried out to find the defect responsible, but an enhancement of the YL has been found for almost all different dopants [36]. While this points to the origin being a native compensating defect, many experimental results suggest that it is related to some kind of carbon defect, possibly in combination with other impurities or native defects [37, 38]. Recent calculations using hybrid

density functional theory by Lyons et al. [39] show that carbon substituting nitrogen (C_N) does not act as a shallow donor (as was previously believed), and that it instead has an ionization energy of 0.90 eV. Based on their calculations, they further suggest that the C_N could be responsible for the YL in GaN films. [6, 32, 38]

However, in a more recent study by Demchenko et al. [32], the authors found theoretical and experimental agreement that the YL from GaN arises due to the electrically neutral C_N-O_N complex. They found that the C_N-O_N complex not only matches observations from PL measurements, it is also the carbon complex with the lowest formation energy in n-type GaN. By the term complex it is meant that the C_N and O_N atoms are combined in a manner, where they form a deep defect state in the band gap of GaN, which may then act as a recombination centre. The energy state of the complex is then decided by the depth and position (relative to the band edge) of the potential well formed by it. Demchenko et al. further studied other possibilities (e.g. the $V_{Ga}O_N$ complex and the C_N) using the theoretical and experimental methods reported in the letter, justifying their results. Demchenko et al. and Reshchikov et al. [38] state that a common misunderstanding is that the YL band is associated with the gallium vacancy-oxygen ($V_{Ga}O_N$) complex, however, they disprove of that belief in their most recent papers (although some of their later articles did actually support that belief as well, however, mainly due to the results of other groups, see e.g. the review [6] where a thorough discussion of papers supporting the $V_{Ga}O_N$ complex as a source for YL is given). Reshchikov et al. did an extensive experimental study of the carbon-related defect bands in GaN samples grown by MOCVD and HVPE. They found that both the C_N and the C_N-O_N complex could contribute to the YL band, and that they have only slightly different maxima, thus causing their luminescence bands to overlap. The C_N-O_N complex is found in all GaN samples, and it is the major contributor to the YL band with a peak intensity at 2.2 eV. The C_N is only found in GaN samples of high quality (e.g. grown by HVPE) and has its peak intensity at 2.1 eV. [6, 32, 38]

Nonetheless, it is quite clear that the YL band may arise from a number of different sources, and extra care should be taken, when samples grown under much different conditions are compared. This means that YL from a sample produced by MBE may not arise due to the same reasons as the YL from a similar sample made by e.g. MOCVD or HVPE [38]. One of the main reasons for the interest in the YL from GaN is due to its apparent relation to nonradiative recombination sites. Therefore it is often directly compared to the amount of band edge luminescence, when considering the quality of GaN films. This comparison requires some caution, however, because the YL band is intensity dependent. [6]

Green Luminescence

The green luminescence (GL) band is usually related to high quality GaN samples, because it is mainly observed in samples grown by HVPE. These samples often contain a low density of carbon and oxygen defects. According to Reshchikov et al. [38], this serves as the ideal conditions for the C_N impurity. They also note that the C_N state can at high excitation intensities be saturated with holes. This allows for the emergence of another energy state, where the C_N impurity may capture two holes, and thus the GL band emerges with a peak intensity of 2.4 eV at the expense of the YL band. This can e.g. be verified by doing intensity dependent PL measurements, where an increasing intensity should show that the GL band increases relative to the YL band related to the C_N impurity. This behaviour was also seen for the freestanding HVPE sample used in this report and it is shown in figure 3.8. Here the excitation intensity was varied, and the

spectrum clearly shows that a higher excitation intensity (the blue graph) leads to a decrease in the YL maxima and instead a blue-shifted shoulder emerges. This observation, however, may be a challenge in some samples if the YL band e.g. from the C_N-O_N complex overlaps with the GL band, and other methods such as time-resolved PL spectroscopy may be needed to verify the existence of the GL band. Another GL band (GL2) is sometimes seen on MBE samples, peaking at roughly 2.35 eV, which is likely due to native complexes involving Ga. However, this band saturates at very low excitation intensities, and might thus be difficult to observe. [6, 38]

Ultraviolet Luminescence

An ultraviolet luminescence (UVL) band is commonly seen in GaN samples grown by MOCVD, HVPE, and MBE with a peak intensity of around 3.26 eV. The UVL band will usually be more intense than e.g. the YL, however, it quenches at temperatures below room temperature. The UVL band has been found to increase in intensity with increasing Mg_{Ga} concentration by intentionally doping GaN with Mg (a standard for generation of p-type GaN), and the UVL band is thus suspected to be related to the Mg_{Ga} acceptor. In the case of Si doping (given that it substitutes N), this would cause an equivalent UVL band to that of the Mg dopant. Since the UVL band is seen in all the growth schemes, it is likely to be present as a native defect during growth. At low temperatures, the UVL band is often seen to transform into a DAP. [6]

Blue Luminescence

A few blue luminescence (BL) bands have been observed in different GaN samples. A commonly observed BL band has its maximum at 2.9 eV and it is frequently argued to be related to the Zn_{Ga} acceptor. Because the Zn_{Ga} acceptor has a large hole capture cross section, it is commonly observed even for low concentrations of Zn. The BL is never observed in samples grown by MBE, however, both HVPE and MOCVD samples display the BL band [6]. Thus it must be attributed to the use of impure gasses. The BL band could also be observed in the HVPE sample used to illustrate the previously described bands, however, due to its low intensity compared to the YL band it seems rather flat in figure 3.8, but can be identified through an increase of PL intensity in the blue region as the excitation intensity was increased. This behaviour is expected because the saturation of the YL band leaves more photons to excite other states. For lower temperatures (below 150 K) another BL band can be seen in some samples with its maximum at 3.0 eV, denoted the BL2 band. The origin of this band is still at debate, however, Reshchikov et al. [38] found that it is unrelated to the YL band. [6, 38]

Red Luminescence

A red luminescence (RL) band (1.5 eV to 2 eV) can be observed in some GaN samples as well, and its peak intensity differs. This band is often disguised by the YL band, and often requires low excitation intensities in order to be resolved in a PL spectrum. Although differing in peak intensity, a similar RL band is often seen in MBE and HVPE samples (the similarity being its saturation at low excitation intensities [6]). Due to its saturative behaviour it is expected to be due to a deep acceptor in n-type GaN. Reshchikov and Morkoc argue that it is unrelated to other luminescent bands, because the intensity of the remaining bands increases similarly after saturation of the RL band. Another RL band (RL2) can be found in samples grown by MBE under large Ga-rich conditions (e.g. when Ga droplets are seen at the surface of the sample after growth). Samples grown under such conditions usually display weak NBE behaviour, and their

PL spectrum is highly dominated by defects, and a usual fingerprint is the emergence of the RL2 band at low temperatures, possibly due to localized defects caused by the extreme growth conditions. [6]

3.3 Raman scattering

This section concerns itself with Raman scattering in GaN and how Raman spectroscopy can be used as a characterization technique of the vibrational properties of a GaN samples. The unit cell of wurtzite GaN contains four atoms with a total of 8 allowed vibrational modes, out of which the 6 optical modes (atoms of the lattice move as a response of an electromagnetic field, i.e. opposite charges move in opposite directions) are depicted in figure 3.9(a). Figure 3.9(b) shows the phonon dispersion for wurtzite GaN. The important modes related to Raman scattering events are those centred at the Γ-point. The designations LO and TO refer to longitudinal and transverse modes with respect to the vibration direction. [19, 40]

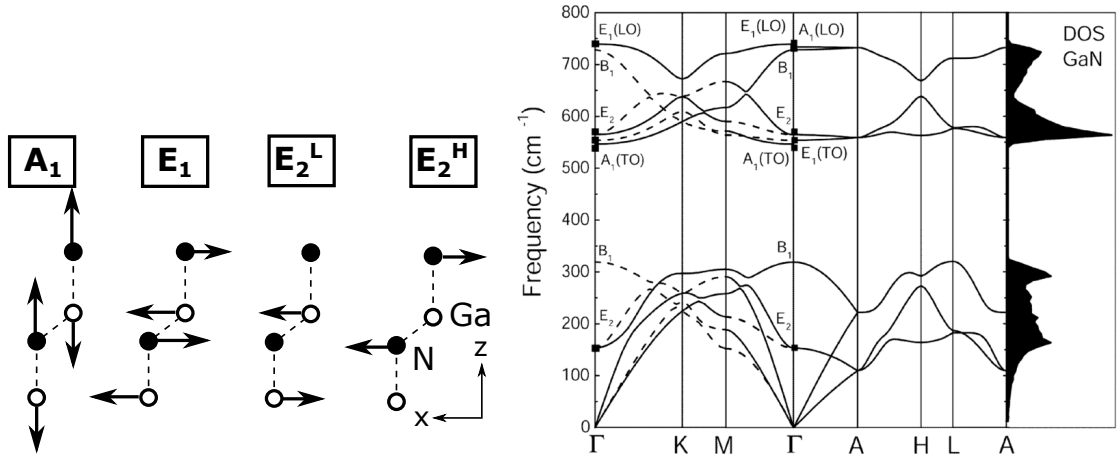


FIGURE 3.9: ***a**: Illustrates the Raman active optical phonon modes in wurtzite. L and H refer to lower and higher branches in figure **b**, respectively. **b**: Phonon dispersion for wurtzite GaN. Calculated by Davydov et al. [41].*

The scattered intensity was given in eq. 2.38, repeated here

$$I_s \propto |\mathbf{e}_i \cdot \overline{R} \cdot \mathbf{e}_s|^2 \tag{3.20}$$

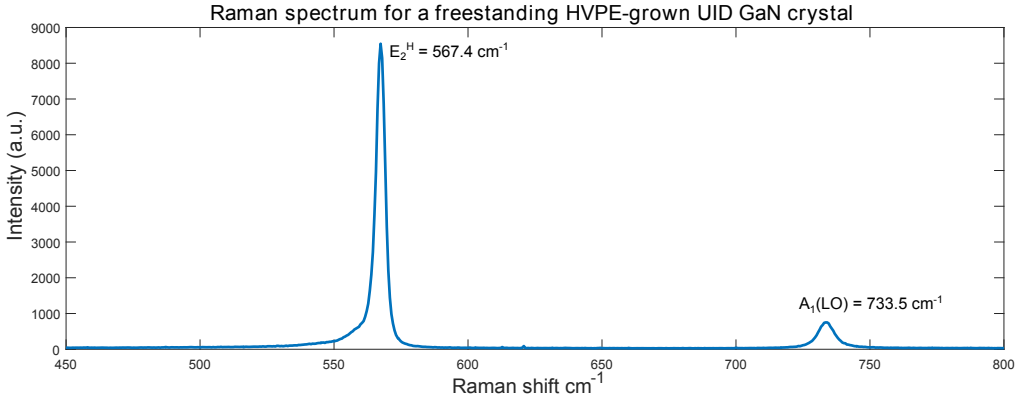
where $\overline{R}$ is the Raman tensor and $\mathbf{e}_i$ and $\mathbf{e}_s$ are the polarization of the incident and scattered light, respectively. Each phonon mode has its own specific Raman tensor ultimately defining the designated polarization configuration of a given phonon mode. Considering the Raman active optical phonon modes of wurtzite given in figure 3.9, the Raman tensor of the E_1 mode only contains off-diagonal elements, and conversely the A_1 mode only has diagonal elements, whereas the E_2 modes have both off-diagonal and diagonal elements. This means that the Raman tensor only permits the observation of a phonon mode in some specific scattering geometries. These are summarized in table 3.2, where the Porto notation $a(bc)d$ is used to define geometry and polarization of the scattered and incident light as follows: a and d are the wavevector direction for incident and scattered light, respectively. b and c are then the polarization direction for the incoming and scattered light, respectively. [19, 40]

Raman geometry	$y(xx)\bar{y}$	$y(zz)\bar{y}$	$y(zx)\bar{y}$	$z(xy)\bar{z}$	$z(xx)\bar{z}$
Raman mode	$A_1(TO)$, E_2^L, E_2^H	$A_1(TO)$	$E_1(TO)$	E_2^L, E_2^H	$A_1(LO)$, E_2^L, E_2^H

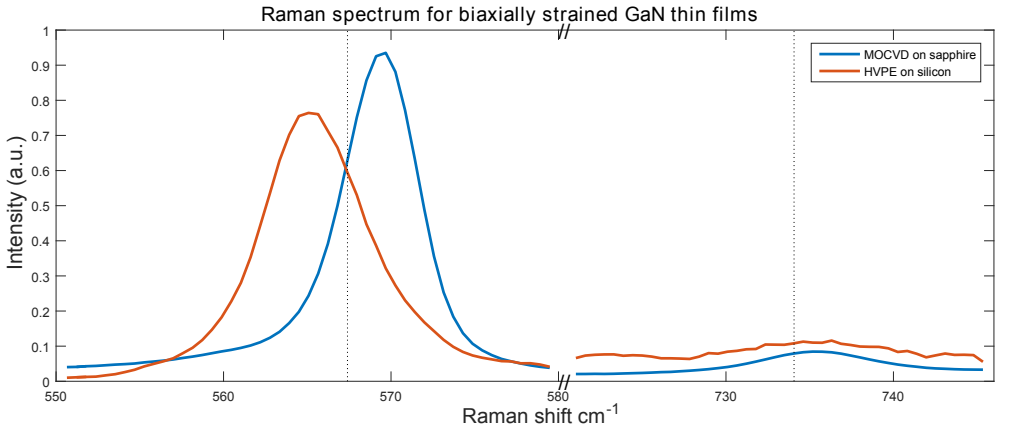
Table 3.2: *Showing the Raman modes for different Raman geometries. The $\bar{a}$ notation is equivalent to $-a$.*

Determination of stress

In this report, only the $z(xx)\bar{z}+z(xy)\bar{z}$ geometries are being used due to the orientation of the as-grown films, and therefore the following only considers this geometry. Figure 3.10 shows the Raman spectrum taken by illuminating a freestanding GaN sample by the 488 nm line of an Ar-ion laser. The Raman shifts observed here are throughout this thesis assumed to be ideal for the $A_1(LO)$ and E_2^H modes, and are used as reference points for comparison with strained samples.



(a)



(b)

Figure 3.10: *Raman spectra of* **a***: a freestanding unstrained un-intentionally doped HVPE-grown GaN crystal and* **b***: GaN thin films grown by MOCVD on sapphire and HVPE on silicon, respectively. The shift in the peak positions of the E_2^H mode due to strain is apparent.*

Raman spectroscopy can be used to determine the state of stress in a sample due to shifts of these peaks. Stress is mainly induced by the substrate on which the GaN film is grown, primarily due to lattice mismatch and difference in thermal expansion of substrate and film.

Here the stress is primarily biaxial given that the substrate is (0001) oriented (or (111) oriented for a cubic geometry), and since the E_2^H mode is very sensitive to biaxial stress (see figure 3.9), a shift of its frequency with respect to the freestanding GaN sample in figure 3.10(a) is readily observed (conveniently this mode typically also gives the strongest Raman signal). [19, 40]

In the case of GaN on sapphire, the thermal expansion coefficient is larger for sapphire and will thus induce compressive strain in the GaN layer upon cooling (the GaN film is usually grown at elevated temperatures). This strain can then be studied with Raman spectroscopy, where the E_2^H mode should be shifted to a higher cm^{-1} than the unstrained value. Conversely, if the substrate is Si(111), which has a smaller thermal expansion coefficient than GaN, the GaN layer is expected to be tensile strained when cooled, as can be seen in figure 3.10(b). Uniaxial stress along the c-axis can be determined from the A_1(LO) mode (see figure 3.9), however, it is not nearly as sensitive to stress as the E_2^H mode. Note that the thermal expansion differs for in-plane and c-axis directions in anisotropic crystals such as wurtzite, and thus the direction of the shift can differ between the two modes with respect to the unstrained Raman line. [19, 40]

LO phonon-plasmon coupling

Phonon modes can couple to plasmon modes, and this coupling can be investigated with Raman spectroscopy. In GaN it is the A_1 phonon modes that couple to the plasmon modes, and this coupling is enhanced with an increase in the amount of free carriers. This can be observed using Raman spectroscopy, where the A_1(LO) mode shifts to higher cm^{-1} and broadens. Additionally, the other branch of the A_1 mode is affected, and it could be observed as a broad peak for weak coupling, and then it enhances as the free carrier concentration increases. However, this mode is forbidden in the geometry applied here. [19, 40]

3.4 Dislocations

This section is devoted to giving a brief description of threading dislocations (TD) in wurtzite GaN, and how these affect the optical and electrical properties of GaN samples. When a GaN layer is grown on a substrate through some epitaxial method, the lattice mismatch with the substrate causes strain in the GaN layer. The strain is typically relaxed within the first few monolayers, however, the relaxation is achieved through formation of defects and dislocations.[5, 42, 43]

Two common types of dislocations are illustrated in figure 3.11. **The edge dislocation** is often seen at the interface between a substrate and an epitaxial layer such as GaN, where lattice mismatch between the respective unit cells results in strain. This stretches (or compresses) some of the GaN unit cells. In order for the higher lying layers to grow without strain, the strain must be relaxed in the lower lying layers, and this can happen through the formation of edge dislocations. The relaxation is achieved through the formation of dangling bonds next to the dislocation. The dislocation usually propagates along the interface in the directions parallel to it (i.e. out of the paper) until it reaches a natural stop (e.g. when it reaches the grain boundary, or a surface defect of the substrate). The growth of GaN orients itself in the least strained manner, where it can still bond to the substrate, and therefore the edge dislocation is commonly oriented along a crystal plane of the substrate. The edge dislocations located at the interface generally tend to lower the internal energy of the crystal. [5, 42, 43]

The screw dislocation (figure 3.11(b)), however, is quite different in nature. It is formed when one crystal plane is matched to two adjacent crystal planes as illustrated in the figure by the red marking. This dislocation is usually formed due to shear stress, e.g. during 3D growth when islands merge and the crystal planes are displaced relative to each other. The dislocation direction will be such that it continues vertically in the example in the figure. Again, the missing lines between unit cells of the figure represent dangling bonds. Finally, there are dislocations of mixed edge and screw character, and these may thus come in many forms and are therefore also more commonly observed in a crystal such as GaN. [5, 42, 43]

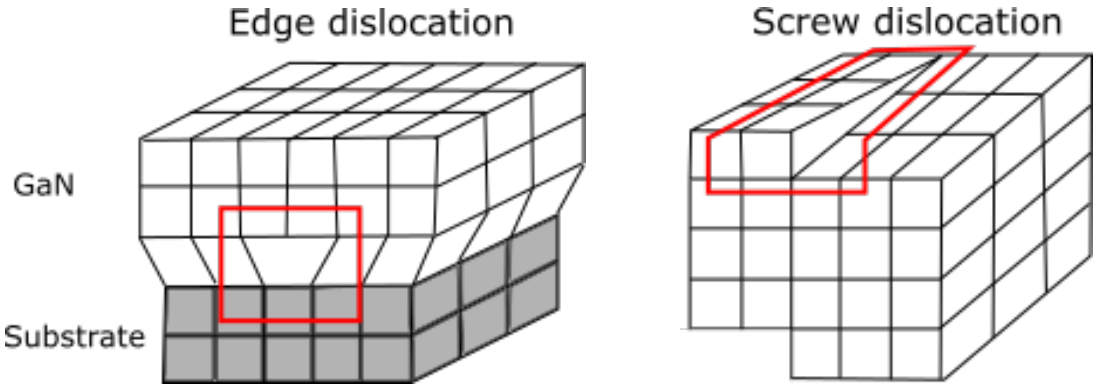


Figure 3.11: *Two common misfit dislocations are shown. The cubes represent unit cells and the red squares mark the dislocation. Missing lines between adjacent cubes resemble missing bonding between unit cells.*

It is important to note, however, that the effects of structural defects do not seem to be as severe for GaN-based devices as for most other semiconductors. The dislocation density present in GaN thin films used for optoelectronic devices is much higher (e.g. $10^8 \mathrm{cm}^{-3}$) than what would cause failure in any non-nitride semiconductor, around $10^5 \mathrm{cm}^{-3}$. As dislocations act as non-radiative recombination centers, these not only reduce the quantum efficiency of a device, but also generate heat, giving rise to additional structural and point defects. It is interesting, then, that nitride LEDs and lasers can function with dislocation densities this high, and much research has been dedicated to finding out why. One suggestion is that the average distance between dislocations in the layer is simply larger than the diffusion lengths for electrons and holes, and thus they rarely interact with one another. In any case, dislocations in GaN have been shown to adversely affect the electrical and optical properties of the thin films, and thus a reduction of the density of these is advantageous. [44, 45]

Threading dislocations (TDs) emerge from the ending points of the misfit dislocations at the interface, from an impurity or defect in the crystal, or during island coalescence. Threading dislocations can thread all the way to the surface of the film, and they are therefore considered a major obstacle for the electrical and optical properties of GaN, since their dangling bonds serve as electron scattering centres, and may even bind to impurities or collect excess Ga atoms, which can lead to leakage currents in electronic devices. Conversed to the misfit edge dislocation, TDs increase the internal energy of the film. Therefore it is of essence to reduce the density of TDs in GaN samples. In the simplest manner, this is accomplished by growing thicker GaN films. It has been found experimentally (also explained by theoretical models) that the TD density in GaN is inversely proportional to the GaN film thickness [43]. A simple reason for this is that

when a TD intercept another TD, they can combine into a single TD or even be annihilated. Thus by growing thicker films the density reduces towards the surface of the sample. A real image of this is given in figure 3.12, where the combination of two TDs into one is shown along with the decreased TD density towards the Ge surface. [5, 42, 43]

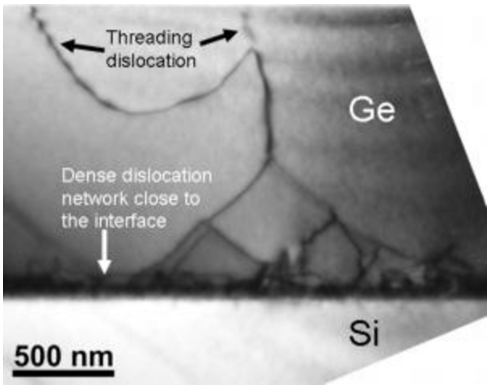


FIGURE 3.12: *TEM image of a Ge/Si sample showing some TDs. Taken from [46].*

Transmission electron microscopy (TEM) is often used to observe the TDs. However, the TDs also affect the crystalline properties, which can be seen in both PL and Raman spectroscopy. In PL they will reduce the quantum efficiency of the GaN film [6], and due to the relationship between TDs and film thickness, it can be expected that the quantum efficiency increases with film thickness as well. TDs also induce local strain fields in the GaN film, and this can be seen in both Raman and PL spectroscopy. TDs in GaN often induce tensile stresses. In PL measurements, stress induces a shift of the NBE PL. In Raman spectroscopy the state of stress can be measured. In the case of GaN/sapphire the strain is expected to be compressive, however, for very thin films (e.g. 30-100 nm) it can sometimes be seen that the strain is tensile. This is thus due to large densities of TDs. The FWHM of a rocking curve measured with XRD will be wider for larger TD densities. All these effects were observed using PL, Raman, and XRD by H. Shih et al. [47]. [42, 47]

3.5 Origin of the 2DEG at the AlGaN/GaN interface

One of the main interests in high quality GaN films is caused by its potential as part of the AlGaN/GaN heterostructure, in which the formation of a 2DEG places GaN as a promising material for HEMT devices. This quite specific property of AlGaN/GaN heterostructures is also one of the main motivations for working with GaN in this thesis, and therefore it is desired to give a brief description of the underlying physics related to the formation of the 2DEG.

The formation mechanisms of the 2DEG have been intensely studied, and it has been established that the large accumulation of carriers at the AlGaN/GaN interface is achieved due to the polarization fields. However, the 2DEG is also formed for AlGaN/GaN structures in the absence of intentional doping, and due to that fact, it is still debated where the charge carriers are then supplied from. Before addressing this topic, a brief explanation of the polarization fields in GaN and AlGaN is required. [2, 48, 49, 50]

Wurtzite GaN is a polar semiconductor due to the different bonding polarity between Ga and N (Ga is more electron rich compared to N) and the stacking nature of a wurtzite crystal. The difference in electron negativity results in a net dipole moment of each unit cell of the

wurtzite lattice, pointing from the Ga-atoms to the N-atoms (see e.g. the unit cell of wurtzite in figure 3.1(b)). This would in the ideal case for wurtzite mean that $u = (a/c)^2$ (this topic was described in more detail in section 3.1, see e.g. figure 3.2 for clarity), however, u is often called a *non-ideality factor*, because it has some other value than $(a/c)^2$. This results in a spontaneous polarization field P_{SP} along the $[000\bar{1}]$ direction due to the displacement of the Ga atoms with respect to the N atoms in the unit cell. GaN is the III-nitride semiconductor where the value of u lies closest to the ideal value, and therefore the polarization field is smallest in GaN. When GaN is heavily doped with Al, the spontaneous polarization field will thus increase towards that of AlN. [2, 48, 49, 50]

Another important polarization field is the piezoelectric field P_{PE}, which is caused by strain in the crystal. This field depends on the amount of strain and the type of stress. Biaxial tensile strain e.g. in AlGaN would result in a piezoelectric polarization field that points in the $[000\bar{1}]$ direction, because bond length between cations and anions is increased due to an increase of the lattice constant a, and thus an increase of u. Conversely, biaxial compressive strain would generate a piezoelectric field in the opposite direction. The whole polarization scheme is illustrated in figure 3.13, where the AlGaN layer is tensile strained because of lattice mismatch with the relaxed GaN buffer. The polarization fields of the AlGaN layer can then be controlled by adjusting the content of Al in the layer. [2, 48, 49, 50]

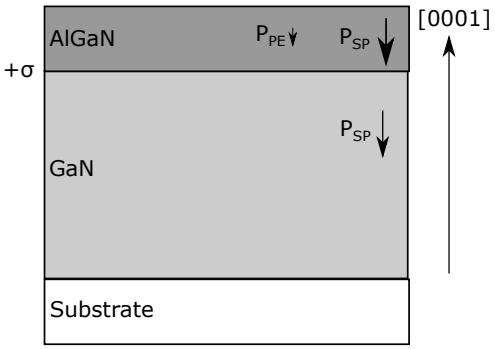


Figure 3.13: *Polarization fields in a GaN/AlGaN heterostructure. The GaN layer is assumed to be relaxed, while the AlGaN layer is tensile strained.*

The gradient of the polarization fields equals a polarization induced charge density ρ_P. Thus at the abrupt interface between GaN (bottom) and AlGaN (top) the polarization can change, inducing a sheet charge density (σ) given by

$$\sigma = P(\text{top}) - P(\text{bottom}) = (P_{SP}(\text{top}) + P_{PE}(\text{top})) - (P_{SP}(\text{bottom}) + P_{PE}(\text{bottom}))$$

and it is this sheet charge density that, if it is positive, can attract electrons and confine them to form the 2DEG. Conversely, if it is negative, holes are confined into a two-dimensional hole gas. [2, 48, 49, 50]

So now one question remains, where do the electrons come from? The simpler case is when the AlGaN layer is n-doped. The situation is shown in figure 3.14(a). The polarization field across the AlGaN layer results in a potential drop across the layer, which affects the band diagram. Then the electrons in the conduction band moves toward the part of the semiconductor, where the Fermi level is lowest. In figure 3.14(b) the AlGaN layer is now put in contact with the GaN layer. As the Fermi levels are aligned across the interface, this results in band bending, and the electrons can thus accumulate in the GaN layer as shown in the figure. However, even in intrinsic AlGaN/GaN heterostructures a 2DEG can be formed. Here the electron supplier is

more debated, but the most accepted theory assumes that donor states exists at the surface of the AlGaN layer [48, 49]. This could indeed explain the formation of the 2DEG, and the situation becomes similar to that of n-doped AlGaN, except the electrons are donated from surface states. This situation is illustrated in figures 3.14(c) to 3.14(d) in the same manner as it was for n-type AlGaN. [2, 48, 49, 50]

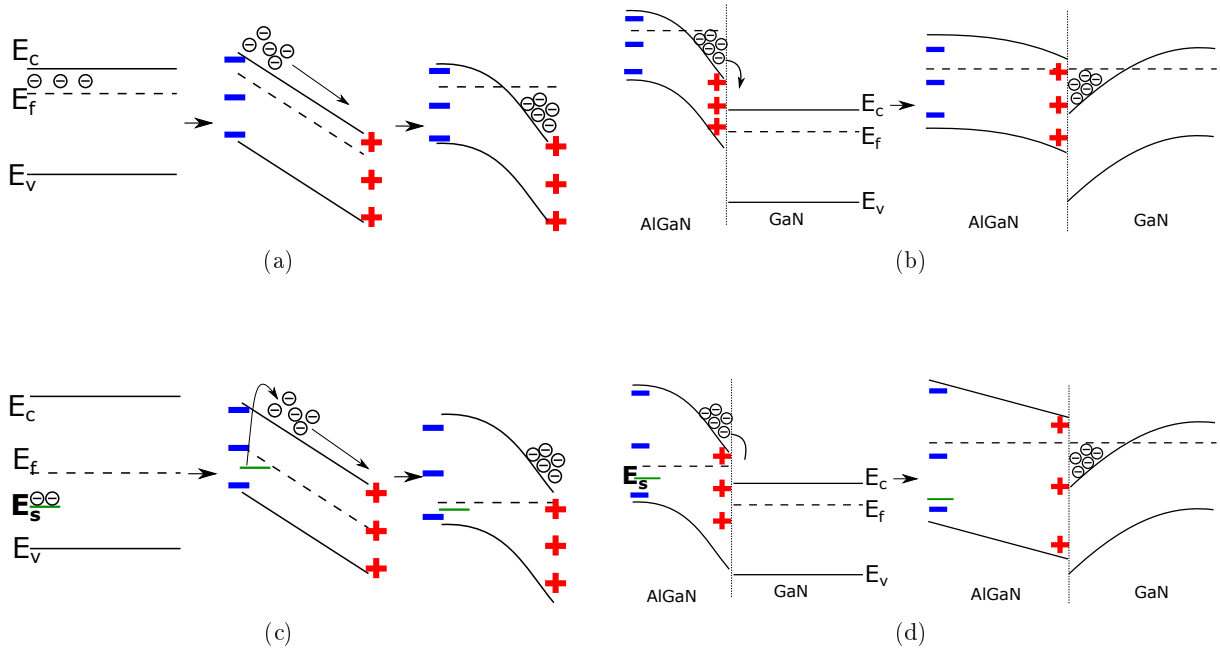


FIGURE 3.14: *Band diagram scheme over the formation of a two-dimensional electron gas at the AlGaN/GaN interface.* **a** *and* **b** *show the process for n-doped AlGaN layer.* **c** *and* **d** *show the process for intrinsic AlGaN layer, which involves surface states. Inspired by [48].*

GaN growth by Molecular Beam Epitaxy

4

In this section, the thin film deposition method known as molecular beam epitaxy (MBE) will be described and compared to other nitride synthesis methods. The operational principles of the method, as well as its advantages and disadvantages, are overviewed. In particular, the technology of GaN growth using MBE is described, and the adjustable parameters and their influence on the as-grown films are discussed, in order to elucidate the types of defects that are commonly found in the films, and to discuss the suitability of different substrates and growth techniques to grow high quality GaN.

The technology of GaN crystal growth dates back more than fifty years, and scientific and commercial interest in the material has driven the exploration of the capabilities of growth techniques to grow high quality GaN thin films. Among these techniques are metal-organic chemical vapour deposition (MOCVD), hydride vapor phase epitaxy (HVPE), and molecular beam epitaxy (MBE). Research into these methods have greatly improved the attainable film quality, but must still deal with two overall issues: difficulty in the incorporation of stochiometric amounts of nitrogen during growth, and the lack of suitable substrates for GaN epitaxy. The first problem results in the need for high flow rates of the nitrogen-carrying species, due to a high vapour pressure. In HVPE, this is circumvented by growing GaN at high substrate temperatures and decomposing a nitrogen-carrying molecule such as ammonia. In MBE, the growth can be carried out at lower substrate temperatures by increasing the reactivity of nitrogen by using a source capable of generating either an atomic nitrogen plasma, or molecular excited species (however, e.g. ammonia can also be used in MBE). The substrate issue has been the target of investigation for many years, with commercial GaN devices most often employing silicon carbide, sapphire, native GaN, and more recently, silicon. [1, 44]

4.1 Thin film deposition methods

In the following section, the setup and operational principles of MBE are described, while other common GaN epitaxy methods are briefly discussed.

4.1.1 Molecular Beam Epitaxy

The concept of MBE was initially conceived as a method for the growth of GaAs/$Al_xGa_{1-x}As$ heterostructures, but has later been developed for the synthesis of a wide range of materials. In its essence, MBE is simply a controlled evaporation technique which takes place in ultra high vacuum (UHV). Due to UHV conditions, low temperatures, and slow film growth rates, MBE

allows precise control of growth parameters, such that doping concentrations, film properties and crystal morphology can be reproducibly manipulated. This means that multilayer structures, e.g. with different doping concentrations and/or alternating bandgaps can be grown with MBE, and as such, the development of MBE paved the way for sharp semiconductor material "tailoring". Thus, it is a very versatile deposition technique. [51, 52, 53]

In the following, a general description of some important vacuum equipment is given. The vacuum system used for MBE in this thesis is illustrated in D.1.

Vacuum system

The term "molecular beam epitaxy" is actually somewhat misleading, as the species evaporated by the sources may be either molecular or atomic. In any case, the aim is to direct the source material onto the substrate without any inter-molecular collisions on the way (hence the term "beam"), as well as having a controlled growth environment on the substrate with low contamination from the chamber ambient. To do this, and to have precise control over flux ratios and other growth parameters, an UHV environment must be created. The kinetic theory of gasses provides useful relationships between e.g. the gas pressure, the mean free path of molecules, and the impingement rate of molecules on a surface. An approximate figure of the latter is given by [54]

$$F = \frac{p}{\sqrt{2\pi m k_B T}} \tag{4.1}$$

where F is the flux, p is the pressure, m is the mass of the molecule, k_B is the Boltzmann constant, and T is the temperature. Thus, if the pressure is not low enough, a large amount of residual gas molecules are impinging on the surface at any time. This is generally not an issue for stable solids surfaces, but for a newly deposited surface, unsaturated chemical bonds may lead to greater adsorption of residual molecules, which can interfere with the growth process, or lead to incorporation of the molecule. Fortunately, in most cases, the sticking coefficient of residual gasses on substrates is relatively low, but MBE still requires UHV pressures below roughly 10^{-9} mbar to prevent contamination. At these pressures, the mean free path of residual gas molecules is much larger than the chamber dimensions, which means that the walls in the vacuum chamber largely determine the gas phase properties, as the majority of molecules inside the chamber are attached to the chamber walls as adsorbates. [51, 53, 54]

Water from air adsorbs particularly well from air on the stainless steel walls of the chamber due to its high dipole moment. Thus, in order to achieve pressures lower than 10^{-8} mbar, it is standard procedure the perform a bakeout of the chamber, whereby the water rapidly evaporates from the inner walls instead of being released continually, thus allowing the pressure to drop into the UHV regime after the chamber returns to room temperature. It is additionally necessary to degas any equipment which works by the heating of filaments (ion gauges, evaporation sources, substrate heaters) to at least their operating temperature, to prevent outgassing during deposition. [51]

In order to obtain and maintain an UHV environment, a variety of different pumps can be used. Generally, there exist two types of high vacuum pumps: capture pumps, such as ion getter pumps and titanium sublimation pumps, and throughput pumps, such as the turbomolecular pump. Capture pumps work by letting residual gas in the vacuum chamber bind chemically to a some element released by the pump (often titanium, as pure titanium is highly reactive). Turbomolecular pumps are mechanical pumps, which work by shuffling gas molecules from the

UHV side of the chamber into the backing side. A series of inclined rapidly rotating blades and oppositely inclined static blades ensure that the probability that gas molecules transfer from the UHV side to the backing side is much larger than the opposite occurrence. While this process lowers the pressure on the UHV side, it requires that, on average, the gas molecules only interact with the blades in the pump without encountering another gas molecule, which means that the pressure on the backing side must be around 0.1 mbar or lower for the pump to work. To establish this backing pressure, roughing pumps are used, of which rotary vane pumps and sorption are common types. An advantage of turbomolecular pumps is that they are well suited for establishing UHV conditions, while capture pumps are typically most suited for maintaining them, as these merely trap gasses inside the chamber. However, mechanical pumps also generate vibrations, which can be detrimental in some systems. Thus, the type of pumps employed should be chosen to accommodate the applications of the chamber. [51, 52, 53]

Naturally, another key component in the MBE system are the material sources. For metal evaporation, three types of sources are generally used: Knudsen effusion cells, evaporation cells and electron beam sources. Knudsen cells consist of a small-volume crucible with a lid into which a small hole is bored, which is heatable by a wounded filament. These have the advantage of very uniform evaporation rates, as the aperture of the crucible is small enough that vapour leaving it causes negligible changes in the internal pressure of the cell. However, while Knudsen cells allow very high reproducibility, they suffer from small evaporation rates, leading to low deposition rates. Evaporation cells, in which a crucible without a lid is heated, offer much higher evaporation rates but might be more difficult to control. In electron beam sources, a filament generates a high electron flux which is focused at the surface of the evaporant, thus raising its surface temperature, causing evaporation. It can, however, also be difficult to achieve uniform flux using these. Thus, choosing the right evaporation method might be a balance between evaporation rate, reproducibility, and uniformity of the molecular beams, which are of pivotal importance for the formation of film quality, thickness uniformity, film composition. [52, 55]

As mentioned earlier, molecular nitrogen is a very stable compound due to the N_2 triple bond having a cohesive energy of 9.8 eV, and as this means that it cannot be thermally dissociated, it does not contribute to the growth of nitride semiconductors. This means that the nitrogen must be activated through some other process, either by delivering it through a carrier gas (most often NH_3 which is then cracked at the sample surface), or through the cracking of molecular nitrogen by e.g. a radiofrequency (RF) plasma source (this is known as plasma-assisted MBE, or PAMBE). The latter method produces an activated nitrogen flux consisting of metastable excited molecular nitrogen ($N_2{}^*$), atomic nitrogen (N), as well as ionic species. The ionic species are generally thought to be detrimental to growth, and are easily removed using an electric field. The nitrogen atoms, having five valence electrons, has no activation barrier for the bonding with group III atoms. The excited nitrogen molecules can be adsorbed on a solid surface through dissociative chemisorption. The ratio of these two species can be controlled through the flow rate and the RF power, and the influence of composition of the N plasma on the growth of III-V semiconductors is important to consider for high quality thin film growth. [1, 55, 56]

Finally, the UHV conditions in MBE allows for *in-situ* growth characterisation techniques such as reflective high-energy electron diffraction (RHEED) and Auger electron spectroscopy (AES). Especially RHEED is celebrated as ideal for the MBE configuration, due to its capabilities for real-time study of the deposition of individual monolayers and the crystallographic properties.

The technique works by letting high energy electrons scatter elastically from the crystal surface at grazing angle of incidence, and observing the pattern of scattered electrons on a detector opposite the electron gun. Through this, evidence of the growth mode of the thin film can be obtained, as well as precise information about the deposition rate and crystal structure of the growing film. [51, 55]

4.1.2 Vapour Phase Epitaxy

Another common method for III-V semiconductor synthesis is vapour phase epitaxy (VPE). In contrast to MBE, GaN is here grown by chemical reactions between gas precursors rather than physical deposition from a beam. The nitrogen is usually supplied through ammonia, while the metal is supplied either by the chemical reaction of an organometallic compound such as trimethylgallium (TMG, $Ga(CH_3)_3$), or through a reaction with a metal chloride created by reacting hydrogen chloride with metallic Ga at an elevated temperature. In the former case, the method is known as metal-organic vapour phase epitaxy (MOVPE) or metal-organic chemical vapour deposition (MOCVD), while in the latter it is known as hydride vapour phase epitaxy (HVPE). Due to the high growth rates achievable through MOVPE as well as the possibility for mass production, this has been the growth method of choice for commercial III-nitride devices. [1, 57]

Because ammonia is a relatively stable compound, the deposition temperatures are considerably higher than in MBE ($>$900°C), in order for the precursors to achieve the appropriate reactivities for decomposition at the surface and generation of easily removable reaction side products. However, this high temperature, along with the high vapour pressure of nitrogen over GaN, leads to inevitable nitrogen loss from the growing films, necessitating a very low ratio of the III/V precursors ($<$1:2000). The higher substrate temperatures along with the low vacuum conditions, as well as the differences in the system setups themselves, heavily influence both the thermodynamics and kinetics of the thin film growth process, and makes the results of MBE and VPE growth markedly different. Both approaches have been shown to be able to produce high quality nitride films, but the characteristics of the grown films are highly growth method dependent. [1, 57]

4.2 Surface processes

In MBE, the thin films are grown at temperatures (600-900°C) which are well below half of the GaN melting temperature ($T_M \approx 2500$°C), and the composition of the grown epilayers are dependant on the flux rates of the source elements. This means that MBE growth takes place under nonequilibrium conditions governed primarily by kinetics rather than thermodynamics, and it is therefore appropriate to give a description of the surface processes involved in the MBE technology of GaN growth. The processes are adsorption and desorption from the substrate surface, surface diffusion, incorporation into the substrate or epilayer lattice, and decomposition of the incorporated species. These competing processes are shown in figure 4.1 and as their interplay is temperature dependent, they limit the temperature window available for GaN growth, which is again highly dependent on vacuum conditions and the delivery mechanism of activated nitrogen. For the following discussions of surface processes, it will be assumed that the substrate is atomically smooth GaN, which makes it simpler to discuss growth processes in isolation. [1, 7, 55]

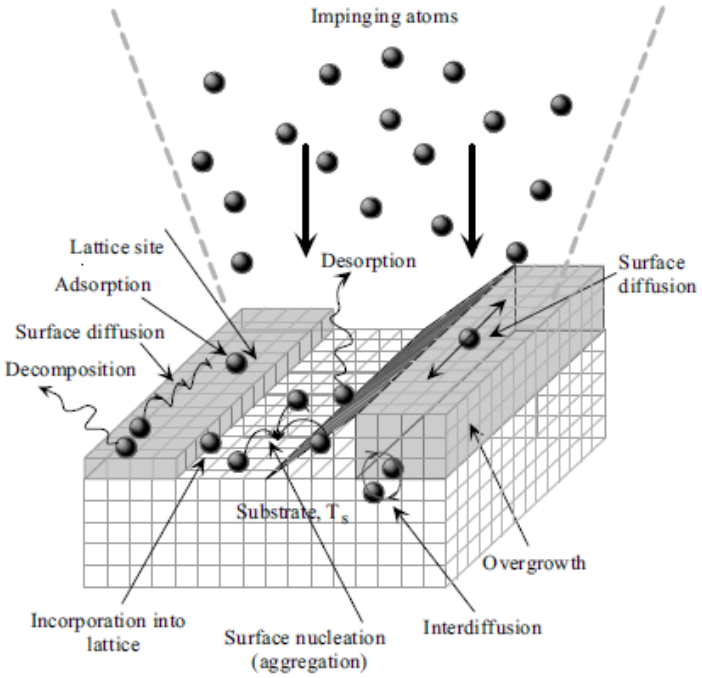


FIGURE 4.1: *Illustration of the surface processes relevant to MBE growth. These are adsorption, desorption, diffusion, incorporation, and decomposition. Taken from [1].*

Adsorption

Adsorption is the process in which an atom or molecule impinging on a surface can adhere to it by overcoming an activation barrier. The activation energies, as well as the strength of the adsorption, is dependent on the type of molecule-surface interactions, and are generally considered in two regimes: *physisorption*, in which no chemical bonds are formed between the species (interaction through van der Waals forces, which are on the scale of a 100 meV) and *chemisorption*, where the adsorbate and the surface form strong chemical bonds and thus drastically alter their electronic configurations, forming hybrid orbitals and allowing charge transfer between the species. These interactions are on the scale of electron volts. Typically in growth processes, the quantity of interest is the *adsorption rate*, which is determined by the flux of material to the surface (eq. 4.1) and the probability that it sticks to the surface, i.e. $r_a = SF$, where S is the sticking coefficient. The sticking coefficient depends on several factors: first, the impinging molecule has some impact energy, which must be larger than the activation barrier for adsorption to occur - this follows a standard Arrhenius dependence. Secondly, chemisorption can be highly orientation dependent with respect to the dangling bonds at the surface, and the surface contains adsorption sites which are more or less favourable, which must also be accounted for. Thirdly, in order for adsorption to occur, the adsorption site must not be occupied by another adsorbate. This leads to the following term for the adsorption rate

$$r_a = \sigma F(\theta) \frac{p}{\sqrt{2\pi m k_B T}} \exp\left(-E_{act}/k_B T\right) \tag{4.2}$$

where σ is the condensation coefficient related to the steric and impact requirements, $F(\theta)$ is the occupancy factor dependent on available adsorption sites, and the exponential term describes the adsorption activation barrier. [1, 51]

The adsorption of the Ga and N depends on parameters such as substrate temperature and III/V flux ratio. For common MBE substrate temperatures on Ga-polar GaN, adsorption has

been investigated by Heying et al. [58] for a fixed N-flux of 2.5 nm/min under varying the Ga-flux. The results are summarised in figure 4.2, where three different growth regimes are categorized. In the N-rich growth regime, the growth rate is limited by the Ga arrival rate, as all excess N desorbs from the surface to yield stoichiometric GaN. In the Ga-rich regime, however, even though the N-flux limits the GaN growth rate, excess Ga can accumulate on the surface in the form of a Ga adlayer, but this is heavily affected by Ga flux and substrate temperature. As such, two Ga-rich regimes are identified: an intermediate regime where a Ga adlayer up to 2.5 monolayer thick is found on the surface, and a Ga-droplet regime, with at least a bilayer of Ga coverage plus additional Ga droplets. The adlayer does not solidify at MBE growth temperatures, but rather behaves like a liquid film. The boundary between the two Ga-rich regimes follows an Arrhenius dependence with an activation energy of 2.8 eV, corresponding to the activation energy for Ga desorption from liquid Ga. Unless otherwise noted, the following discussion will assume that growth takes place under Ga-rich conditions. [1, 7, 25]

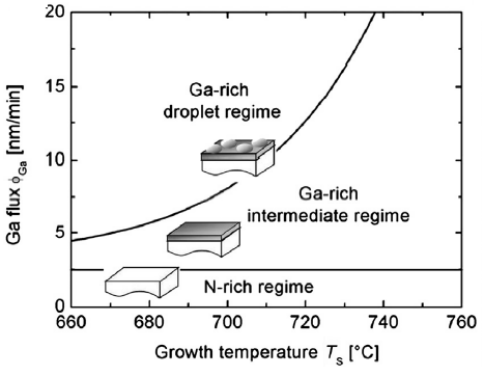


Figure 4.2: *Diagram of the growth scheme of GaN as a function of temperature and Ga flux for an N-flux of 2.5 nm/min. Taken from [7].*

Desorption

Naturally, once a molecule has been adsorbed on a substrate, it can desorb again if it is provided sufficient thermal energy to overcome the activation barrier for desorption. The desorption probability is dependent on the bonding strength to the surface, which again depends interatomic interactions and surface structure. Using the kinetic framework, the desorption rate can be defined very similar to the adsorption rate in eq. 4.2, except with all the factors being related to desorption, and the adsorption activation energy being replaced by the desorption activation energy. [1]

The quantity which is often of particular interest, and which is rather easily measurable, is the adsorbate residence time, defined as the average lifetime of an adsorbate on the surface before desorption. This also follows an Arrhenius form, i.e.

$$\tau = \tau_0 \exp\left(\frac{E_{des}}{k_B T}\right) \tag{4.3}$$

from which it is apparent that the lifetime of an adsorbate is sensitive to the temperature of the substrate. The desorption energy of Ga from liquid Ga was mentioned earlier, although reports vary on the exact values. For N, it is likely that it desorps as a dimer, which implies that the desorption rate is limited by the diffusion of two N atoms. If two N atoms meet, they will immediately desorb due to the strong exothermic N_2 bond formation. [1, 55]

Diffusion

The process of surface diffusion is integral to the growth of thin films. Due to the finite temperature of the substrate, adatoms can be mobile on the surface, where there motion can be described by random walk. For any adatom to diffuse to an adjacent lattice position, it must overcome the diffusion activation barrier. However, because this activation energy is highly dependent on the very local energy landscape, some lattice sites can require much higher energy to diffuse from. These sites are called nucleation sites, as it can be very unlikely that an adatom at these sites diffuses further at the given growth temperature. These sites are defects in the surface lattice, the most common of which in MBE is the step edge. As such, step edges are often the starting point for further nucleation and growth of adsorbate layers and islands. [1, 54]

The average diffusion length is an important parameter for epitaxial growth. It can be defined as the root-mean-square displacement of an adatom from its initial impact site before desorption, and is written as

$$x = (D\tau)^{1/2} = (D_0\tau_0)^{1/2} \exp\left(\frac{E_{des} - E_{dif}}{2k_BT}\right) \tag{4.4}$$

where D is the diffusion coefficient. At high substrate temperatures, the diffusion length may be larger than the width of the substrate terraces, which means adatoms can diffuse to the step edge and be incorporated into the lattice here - this is called *step-flow* growth. At lower temperatures, the diffusion length might be lowered enough that reacting atoms meet and nucleate on a terrace rather than at a step edge, which can result in rough surface morphologies. However, as the substrate temperature increases, so does the desorption rate, which might actually lower the diffusion length. Estimating the diffusion length is thus not a straightforward task. [1, 59]

The diffusion of Ga and N on GaN was investigated by Zywietz et al. [60]. They found that Ga atoms have an approximately four times lower diffusion barrier than N atoms on Ga-terminated surfaces. This is because the Ga atoms interact with the substrate through metallic Ga-Ga bonds, which are not very strong, as evidenced from the low melting point of Ga, which is why a liquid-like film is formed. Ga adatoms are thus very mobile at typical growth temperatures, while N atoms are not. It was also found that the presence of N atoms on the surface can dramatically reduce the Ga adatom mobility. This is important for epitaxy, as a long diffusion length can lead to step-flow growth, smooth surface morphology and low stacking fault density, which can evidently be obtained under Ga-rich conditions, where only a small amount of excess N is present. Additionally, N has a high incorporation ratio in this regime, as it is much more likely for a fast-moving Ga atom to capture N atoms than it is for another slower moving N atom, which would lead to desorption. In the N-rich regime, however, the much lower mobility of Ga leads to nucleation on terraces, a higher density of stacking fault, and rougher surface morphologies, all of which has been supported by experimental evidence. [1, 60]

Incorporation

GaN growth is a metastable process, as there is competition between the forward reaction of incorporation and the reverse reaction of decomposition. In the incorporation process, the evaporated species react with one another and the substrate surface to form films, a process which is determined not only by the thermodynamics of surface and adsorption energies, but also by kinetics in MBE. The growth mode of the films is often regarded in three distinct regimes shown in figure 4.3: Frank van der Merwe (layer-by-layer) growth, Volmer-Weber (island) growth, and

Stransky-Krastanov (layer-plus-island) growth. In the layer-by-layer growth mode, new layers only grow after the previous one is completed, making the growth two-dimensional. In the other limiting case, island growth mode, three-dimensional islands are formed directly on the substrate. In the intermediate case, a few monolayers are formed on the substrate surface, after which subsequent growth continues in the form of islands. This is often the case for heteroepitaxy of materials with mismatched lattice constants. The manner in which the deposited layers grow is dependent mainly on the competition between surface diffusion and desorption. As these parameters are affected by surface structure, temperature and material flux, this means that the growth mode of the epilayer can largely be controlled by the substrate temperature and the III/V ratio. [1, 51]

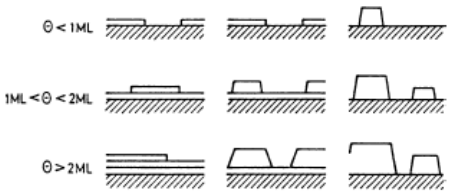


Figure 4.3: *Illustration of the three surface growth regimes. First column: layer-by-layer growth. Second column: layer-plus-island growth. Third column: island growth. θ denotes the surface coverage in monolayers. Taken from [51].*

At typical GaN growth temperatures in MBE, Ga does not fully incorporate into the growing layers even with an excess of N. However, at lower temperatures, the Ga adatom lifetime is long enough that near-unity incorporation ratios can be achieved, as Ga atoms will almost always encounter N atoms before desorbing. As low substrate temperatures severely limit Ga diffusion, this will, however, result in an increase of defects and stacking faults. [1]

Decomposition

Finally, the decomposition process denotes the breaking of the bonds of the film material into either its original constituents, or some new product. Despite the strong Ga-N bond, the formation enthalpy for GaN is rather small, indicating that it easily decomposes. The decomposition energy is supplied from the heated substrate and at atmospheric pressures, the decomposition temperature for GaN is around 800°C, while it is lower in vacuum. The GaN decomposition temperature in vacuum has been reported in a wide range, likely due to differences in pressure, gas flow, and the GaN films themselves. Grandjean et al. [61] found the decomposition of GaN to be negligible below 750°C, and increase rapidly above 800°C. This is in good agreement with Fernandez-Garrido et al. [62], who also note that the decomposition rate of GaN can be heavily suppressed by exposing the decomposing surface to active N, which allows the recapture of Ga atoms. From RHEED measurements, GaN has been found to decompose in a layer-by-layer manner. [1, 62]

4.3 GaN growth with PAMBE

4.3.1 Growth optimisation

Within the framework of GaN homoepitaxy, the main issues arising from foreign substrates are not a concern, and thus what is important is optimisation of the growth parameters. Once the substrate has been treated, the two *in-situ* controllable parameters are substrate temperature and flow rates. Extensive research has gone into establishing the optimal growth conditions for GaN thin films with smooth surfaces and desirable electrical and optical characteristics, leading to the understanding of the growth regimes in figure 4.2. The conventional wisdom has been that the best thin films are created on the border between the Ga-droplet and intermediate regimes where step-flow growth is achieved, leading to low defect densities, smooth surfaces, and the highest electron mobilities. However, this growth scheme requires very precise control over the substrate temperature and III/V ratio to stay within this narrow optimal window, which often warrants modulated growth techniques with periodical shuttering of either flux to desorb excess Ga, and attain the desired surface coverage of Ga. Another disadvantage of Ga-rich growth is that Ga from the metallic adlayer tends to decorate vertically aligned defects, such that these can act as leakage paths in a GaN based device structure. [7]

In order to tackle these issues, Koblmüller et al. [63] has conducted PAMBE growth experiments on Ga-face GaN at temperatures above the onset of thermal decomposition (>750°C). The use of high N flux was shown to limit the decomposition of growing GaN layers enough to expand the growth window considerably, leading to the growth diagram in figure 4.2 for Ga-rich growth being supplemented by one for N-rich growth shown in figure 4.4 for an N flux of 5 nm/min. Naturally, for high temperatures and very low Ga flux, decomposition is faster than incorporation, so no growth can take place. It was, however, found that the higher substrate temperature facilitated sufficient Ga adatom diffusion to allow for 2D-growth, even without the presence of the Ga adlayer. An apparent advantage of this N-rich/high-temperature growth regime is that growth is less sensitive to III/V ratio than the Ga-rich/low-temperature regime, while still offering GaN layers with reduced density of dislocations and smooth surfaces. [7, 63]

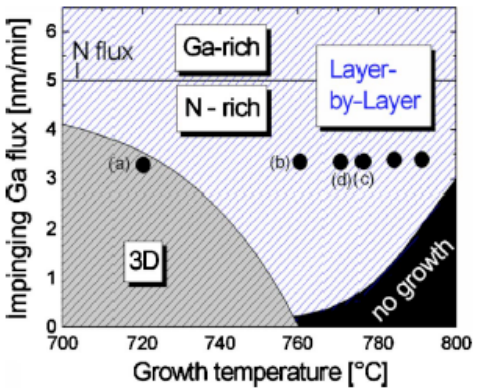


FIGURE 4.4: *Diagram showing the PAMBE high-temperature growth scheme for GaN as a function of temperature and Ga flux for an N-flux of 5 nm/min. Taken from [63].*

4.3.2 Growth on foreign substrates

The growth patterns and surface processes outlined above have assumed homoepitaxy, which works well as an idealised system, but is infeasible for commercial growth of GaN devices due to the high cost and lack of large size GaN substrates. For this reason, new avenues to produce high quality GaN thin films on foreign substrates are being continuously explored. However, the films grown on these substrates are not nearly of the quality of the homoepitaxially grown films due lattice- and thermal expansion coefficient mismatch and non-ideal nucleation, as evidenced by wide x-ray rocking curves, rough surfaces, and high electron concentrations (n-type doping). The large lattice mismatch leads to a 3D growth mode, as well as a high density of threading dislocations, which, as mentioned earlier, are detrimental to the electrical and optical properties of the films. Threading dislocation density reduction is thus a very important challenge in GaN heteroepitaxy. [1, 7]

In order to improve the quality of the GaN films, a two-step growth scheme employing *nucleation* layer(s) has been conceived. These layers are also frequently denoted *buffer* layers in the literature, but as subsequent layers are also often denoted buffer layers in device structures, this terminology is needlessly ambiguous. Since the advent of low-temperature grown nucleation layers in the mid 1980's driven by the work of Amano et al. [64], these have been considered an indispensable tool for growing high quality GaN and AlGaN films. Films grown on appropriate nucleation layers have markedly improved crystallinicity and optical and electrical properties, as evidenced by e.g. x-ray diffraction, photoluminescence, and Hall measurements. The nucleation layer can have several purposes, as depending on whether it is AlN or GaN, and what the substrate is, it can also contribute to determining epilayer polarity and can help reduce diffusion of impurities from the substrates into the epilayer.

The main role of the nucleation layer is to optimize the transition between the substrate and the following epilayers. Because GaN is often very highly lattice mismatched with respect to the substrate ($\sim 16\%$ for sapphire and $\sim 17\%$ for Si), the interfacial free energy between the substrate and the film is large. Initially, the films are grown pseudomorphically under strain, but after only a few monolayers, the strain energy exceeds the elastic limit in the layer and the film will relax through formation of dislocations or twisting and/or a transition to the Stransky-Krastanov growth scheme [65]. The result is highly defected films with rough surfaces after the islands coalesce through lateral growth. The introduction of a low-temperature nucleation layer allows for a high density of nucleation sites due to limited diffusion during growth, resulting in many smaller islands. Once these islands coalesce during further growth, a smoother surface is provided, since the rms roughness of the layer is roughly proportional to the separation of islands [7]. Further growth of the epilayer is then initiated on this layer, which hopefully overgrow the large amount of defects found in the nucleation layer. In this vein, the goal is actually just to make the system behave more like homoepitaxy, such that pseudo-2D growth can be accomplished for the epilayer. [1, 7, 66]

As the low temperature allows the nucleation layer to assume the least-defected structure necessary to relieve the interfacial stress by the twisting of islands with respect to the substrate and prismatic growth, the structural quality of the layer may consequently be very poor. It may even contain some zincblende domains and in any case a large amount of misfit dislocations and disorder, but this is acceptable, as it is merely supposed to provide a template for further growth. In any case, it has been shown that during subsequent annealing, the nucleation layer can be converted into a single relaxed crystal of wurtzite [67]. Early experiments with nucleation

layers in MBE, however, showed not as dramatic an improvement for the epilayers as for e.g. MOVPE, which has been theorised to be due to the lower available temperatures in MBE (due to lower N partial pressure and low growth rates) not allowing for complete recrystallization of the nucleation layer [68]. However, the annealing of the nucleation layer still has a profound effect of the surface morphology of the epilayer. [1, 7, 44]

It is important to note that while the growth parameters and final structure of the nucleation layer are interesting in and of themselves, the important part is primarily what effect the layer has on the properties of the subsequent epilayers. Whether an AlN, GaN, or another type of nucleation layer is preferable depends on several parameters. It might seem that GaN nucleation layers are the obvious candidate, as they avoid introducing another material into the system which can lead to creation of new of defects. However, the growth schemes for the AlN and GaN layers vary in their temperature and flow dependencies, and are largely dependent on the type of substrate, and as such, the influence of different types of nucleation layers on the entire device structure is a complicated matter. In any case, if suitable growth conditions are used, many different types of nucleation layer schemes have been shown to greatly improve the performance of subsequently deposited GaN epilayers.

Even employing the two-step growth method mentioned here, a large issue with GaN heteroepitaxy, particularly with MBE, remains the large density of threading dislocations. These can often propagate through the nucleation layer and epilayer through the coalescence boundaries between coalesced islands, which are numerous due to the high nucleation density of the nucleation layer. One way to reduce the density of threading dislocations is to simply grow the epilayer thicker, so a higher number of dislocations inclined with respect to the c-axis meet the wafer edge. However, on substrates such as SiC, thick layers have a tendency to crack due to thermal expansion mismatch during cooling, meaning that the thickness cannot be increased indefinitely (although it can be greatly increased with the use of interlayers). [1, 7]

Other growth methods than PAMBE, such as ammonia-MBE or MOVPE, are usually able to produce GaN layers with threading dislocation densities an order of magnitude lower. This is primarily because of the higher growth temperatures available to them, reducing the nucleation density of the epilayer and thus resulting in fewer coalescence boundaries. According to the growth regimes defined in figure 4.2, for conventional Ga-rich growth in PAMBE, raising the growth temperature beyond $\sim 750°C$ is not feasible. Another approach for threading dislocation density reduction, as mentioned in section 3.4, is the possibility for mutual annihilation of threading dislocations, if the film can be designed such that dislocations collide. One way to do this was demonstrated by Manfra et al. [69], who successfully reduced threading dislocation density and improved HEMT device performance by using two different III/V ratios during growth of the GaN epilayer. In the first layer, a lower III/V ratio is used, which results in a faceted surface, and thus serves to incline the propagation direction of the threading dislocations away from the c-axis, making it possible for them to meet and annihilate. As this faceted surface is not ideal for further epitaxy, another epilayer is grown as usual near the Ga-droplet border on top, resulting in a total epilayer both with reduced threading dislocation density and a smooth surface. [7, 69]

Reduction of threading dislocations can also be achieved through the use of *interlayers*. These are grown at conditions which may vary greatly with respect to the epilayer, and may consist of several types of material. Interlayers can also help prevent cracking during the cool-down process in the growth chamber, as they can allow the epilayer to be grown oppositely strained

with respect to the strain generated during cool-down, leaving the epilayer unstrained at room temperature. Both these effects were demonstrated by e.g. Frayssinet et al. [70], who used an SiN interlayer to reduce the threading dislocation density in the epilayer, and an AlN interlayer to grow GaN compressively strained. As with nucleation layers, the appropriate interlayer growth parameters are highly dependent on material, substrate, and growth method. [1, 70]

A simpler way to control the residual stress from thermal expansion coefficient (TEC) mismatch was demonstrated by Aidam et al. [71]. They showed that up to 4.2 µm can be grown on Si(111) without cracks by simply using by using an AlN nucleation layer. Since GaN is grown compressively strained with respect to AlN, it is possible to perfectly compensate for the lower TEC of Si by growing the layer to a thickness where compressive strain from lattice mismatch and tensile strain from TEC mismatch perfectly balance. This approach reduces the amount of layers needed to grow crack-free, thick GaN layers on highly TEC mismatched substrates.

Polarity control

One of the advantages of GaN growth by MBE compared to other growth methods, and one which largely paved the way for interest in the MBE growth of GaN, is that MBE allows the control of the polarity of the GaN epilayer by varying the growth parameters on substrates where can be difficult by other growth methods. There are several reasons why it is of interest to control the polarity, the most significant for electronic device application being the inversion of the polarisation-induced electric fields in N-polar heterostructures with respect to their Ga-polar counterparts. The influence of this is, for example, that the possible electrical contacts are different for structures of opposite polarity, and that in a HEMT the confinement of electrons in the 2DEG in the off-mode is stronger in N-polar structures, which can be favourable. The polarity also seems to strongly affect the surface structure of device layers, which makes polarity a necessary parameter to control. [1, 72]

The polarity control method is highly substrate dependent. For substrates with a structure very similar to GaN and AlN, e.g. 4 H- or 6 H-SiC which, like the nitrides, belong to space group number 186, the polarity of the epitaxial layer is entirely determined by the polarity of the substrate if grown in the c-direction. As such, growing epitaxial layers of either polarity is, in theory, easily achieved by having either Si-face or C-face SiC. The growth schemes employed to achieve high quality epitaxial films on SiC are also very similar for either polarity. On substrates which do not share the packing of the III-nitrides, such as sapphire, the polarity is largely determined by the chemical termination of the sapphire surface, the III/V ratio, and the initiation conditions such as nitridation and choice of nucleation layer parameters. Improper conditions can lead to a large concentration of polarity inversion domains and stacking faults due to differences in growth rates for different domains, again highlighting the importance of polarity control. [1, 7, 72]

Investigation of the effects of different procedures on GaN polarity has yielded mixed results. It has been found by Mikroulis et al. [73] that direct growth of GaN on sapphire usually leads to N-polarity, while the polarity when grown directly on nitrided sapphire is dependent on the temperature of nitridation. This was argued to be related to the completeness of the AlN nitridation layer on the sapphire, with a complete layer leading to Ga-polarity. The effects of the nucleation layer on the polarity of the epilayer was investigated by Huang et al. [74], where it was found the high temperature AlN and GaN nucleation layers lead to Ga-face and N-face

epilayers, respectively, and that the polarity of the nucleation layer also determines the polarity of the epilayer. For low temperature nucleation layers, the polarity was found to be growth rate dependent. While these results are somewhat ambiguous in the literature, it seems that the presence of a complete AlN nucleation layer leads to Ga-face epilayers, as already mentioned. This can be explained through the relative bond strengths between Al and Ga with N and O. Typically, sapphire substrates for epitaxy are oxygen terminated, which leads to the assumption that the first deposited layer will be composed of group III metal, as shown in figure 4.5. The polarity inversion stems from the difference in the strength of the III-N and III-O bonds. Because the Ga-N bond is stronger than the Ga-O bond, the first Ga deposited on the surface will be singly bonded to the O and triply bonded to N to satisfy the tetrahedral configuration. Thus, the first Ga deposited already belongs to the nitride layer, leading to N-polarity (left side of the figure). However, the Al-N bond is weaker than the Al-O bond, so the first layer of Al deposited will be triply bonded to O and singly bonded to N, and is thus incorporated into the sapphire matrix rather than the nitride layer. Thus, the subsequent N layer is singly bonded to Al and triply bonded to Ga, and thus belongs to the nitride layer which causes Ga-polarity (right side on the figure). [72, 75]

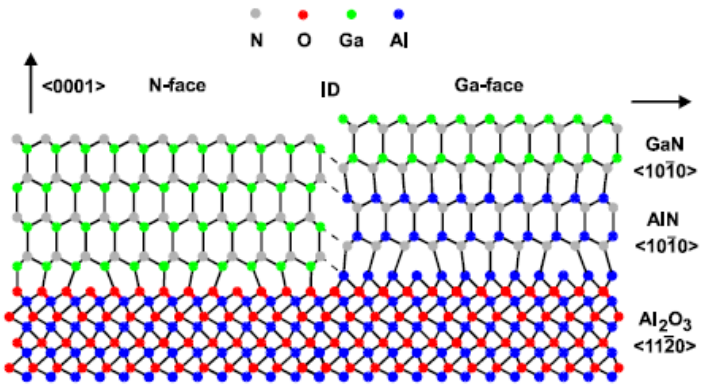


FIGURE 4.5: *Schematic showning how the introduction of an initial AlN layer can reverse the polarity of the epilayer. Taken from [75].*

This model explains the effects of high temperature nucleation layers, but not ambiguousness of the low temperature ones. Yoshikawa and Xu [76] argues that the polarity conversion from N-face to Ga-face by introduction of an AlN layer is achieved by having a 2 monolayer thick Al platform which kinetically favours Ga-face GaN, as N-polarity nitrogen positions require the involvement of more Al atoms. This can perhaps explain why higher growth rates of low temperature nucleation layers favours Ga-face polarity, while slower rates seem to favour N-face. In any case, the polarity for low temperature nucleation layers is likely an interplay between other growth parameters and nitridation conditions. [1, 72]

Determining the polarity of the epilayer can be done in several ways. The surfaces of opposite polarity exhibit characteristic surface reconstructions, which can be identified by RHEED to examine polarity. A simpler method is wet etching of the surface by either sodium hydroxide or potassium hydroxide. This method is based on the differences in chemical resistance towards the etchant for the two polarities, as N-face III-nitrides is easily etched in either etchant, while Ga-face III-nitrides are only easily etched at structural defects. [75, 77]

4.4 Substrates for GaN epitaxy

There are many considerations which must be made when considering the type of substrate employed for GaN heteroepitaxy. As mentioned, the most common candidates are (with usual quality of the grown epilayers in decreasing order) SiC, sapphire, and Si, with the technology for growth on the latter not being well developed yet. Among the many parameters necessary to consider for the substrate are: lattice matching, thermal expansion coefficient mismatch, differences in chemical composition leading to e.g. poor wetting of films on the substrate, and polarity/non-polarity of the substrate influencing the polarity/inversion domains in epitaxial films. It is also preferred that the substrate is non-reactive at the growth temperatures. Additionally, the thermal and electrical conductivities of the substrate, as well as its optical transparency, are also of pivotal importance for device structures. [78]

While Si and SiC markedly different in their crystalline structure, many of the same precautions must be taken when growing a thin film stack on them. In both cases, it has been found that GaN nucleation layers lead to poor quality films, both structurally and electrically, while AlN nucleation layers provide a good template for further growth. This has several reasons: first, and most importantly, the AlN layer provides a barrier for the diffusion of Si into the GaN layer (also relevant for SiC, as it contains a high density of surface impurities of Si). Secondly, AlN with its low lattice mismatch towards GaN provides a superior template to SiC, and especially Si, and can even be used for stain balancing to prevent cracking in Si from TEC mismatch, as discussed earlier. Finally, while GaN has poor wetting properties on both SiC and Si, resulting in 3D growth modes and highly defected films, AlN wets these substrates rather well, resulting in much smoother layers. Additionally, a large issue for Si substrates is the very high solubility of Si in Ga, which causes a surface reaction known as the meltback etching effect, where large voids in the Si substrate under the GaN film are produced. [7, 79]

However, the suitable parameters for the growth of the AlN nucleation layer might be different for Si and SiC substrates. On SiC, while growth of the nucleation layer under Al-rich conditions leads to the best surface morphology as expected, these conditions have shown to causes severe buffer leakage in HEMT devices. The origin of this is the solubility of Si impurities in the Al adlayer during growth, which causes the migration of Si into the GaN layers, making it highly n-type. By growing, or at least nucleating, the AlN layer under N-rich conditions, there is no Al adlayer, and Si migration can be considerably suppressed. The same reasoning can be applied to Si, where the Al adlayer have also resulted in meltback etching, while N-rich growth avoids this. However, the nitridation of the Si surface, converting it into amorphous Si_3N_4 with deleterious effects on the final epilayer quality, has led to some authors regarding Al adlayer wetting on the Si surface as necessary [80]. [1, 7, 79]

4.4.1 Sapphire

Although there are several properties of sapphire (α-Al_2O_3) that are detrimental to it serving as a substrate for GaN epitaxy, it is still one of the most explored substrates for this purpose due to its stability at high temperatures, availability of high quality crystals at a reasonable price, and a mature growth technology. The III-nitrides can be grown on several planes of sapphire with different results, but the most common substrates for epitaxy are c-plane, which will be discussed here. The lattice mismatch between GaN and sapphire is very large, as the sapphire lateral lattice parameter is 33% larger than that of GaN. However, because there exists a smaller cell of Al atoms that N can be bonded to, which is rotated by 30° with respect to the sapphire

unit cell, the actual lattice mismatch can be found as

$$\frac{\sqrt{3}a_{GaN} - a_{sapphire}}{a_{sapphire}} = 0.16 \tag{4.5}$$

which also means that the lattice strain induced in the GaN layer becomes compressive instead of tensile. 16% is, however, still a very large mismatch that results in a Frank-van-der-Merwe critical thickness of only a few monolayers, and associated island growth mode. As such, epitaxy on sapphire is heavily dependent on a suitable nucleation layer, which for sapphire is usually either GaN or AlN, as discussed earlier. Prior to nucleation layer deposition, a *nitridation* procedure is usually carried out. The goal of this is to aid the nucleation layer in easing the transition between the substrate and the epilayer by converting the Al_2O_3 through $Al_2O_{1-x}N_x$ to AlN, upon which the nucleation layer can grow. The parameters of nitridation are thus very important for the structure of both the nucleation layer and the epilayers. [1]

While a consensus has been established that nitridation is an important step in achieving high quality epilayers, the exact mechanisms involving the transformation of Al_2O_3 to AlN in an MBE environment are not well understood. Similarly, the literature is conflicted as to which parameters are optimal, and they seem to depend heavily on the subsequent growth parameters. Kaneko et al. [81] found that incomplete nitridation was advantageous, as it allowed a lower nucleation density for the subsequent nucleation layer, leading to lower density of dislocations and better quality of the epilayer, while Namkoong et al. [77] found that complete nitridation allowed for the best epilayers. Mikroulis et al. [73] found that high temperature is needed for significant nitridation to occur, while Widmann et al. [82] found that a 200°C nitridation temperature yielded a well-ordered nitride layer that led to high quality epilayers. While good quality epilayers have certainly been grown where high nitridation temperatures have been employed, much evidence supports the superiority of low temperature nitridation. The difference lies in the adsorption mechanism of nitrogen, which may react either with Al to form AlN or O to form NO. At low temperatures, NO adsorbs on an on-top site on Al, while at higher temperatures it adsorps on a bridge site. The latter has a higher interaction energy with the surface, and does not easily desorp, leaving the Al atoms blocked for AlN formation, while the former readily desorps, allowing additional N to react and form AlN. This, in turn, can lead to better uniformity in the nucleation layer, the absence of inversion domain boundaries induced by oxygen in the AlN layer, and a smoother surface. However, as was argued with nucleation layers, the importance lies in the nitridation layer effect on epilayer quality, and thus any nitridation parameters that optimise the properties of subsequent layers can be considered suitable. [77, 82]

Another issue in the epitaxy of GaN on sapphire, which was touched upon earlier, is the thermal expansion coefficient mismatch between the species. While the coefficients vary with temperature, the higher TEC in sapphire can lead to a high dislocation density and residual compressive strain once the structure is cooled to room temperature. While this issue is not as dramatic for the GaN on sapphire system as it is on e.g. Si, the poor thermal conductivity of sapphire is a significant disadvantage for its use in electronic devices, especially high-power application. Despite this, sapphire remains a commonly used substrate due to reasons listed earlier. [83]

5 Results

This chapter presents the results obtained from various characterization methods used in this thesis. The aim is to describe the most general properties of GaN, and also to demonstrate how various characterization methods can be applied to GaN, and what to expect from the outcome. The chapter begins with an AFM surface analysis, in which special attention is given with respect to MBE grown samples in this thesis. Electrical characterization of mobility and carrier concentration is then considered for a few selected samples by Hall measurements. This is followed by a Raman investigation of stress and identification of various materials by their measured Raman shifts. XRD was also used for consideration of crystalline quality and material identification. Finally, PL studies were used to characterize and identify the nature of defects and excitons in GaN. All characterization methods and sample parameters are listed in appendix D.1.

5.1 Surface morphology

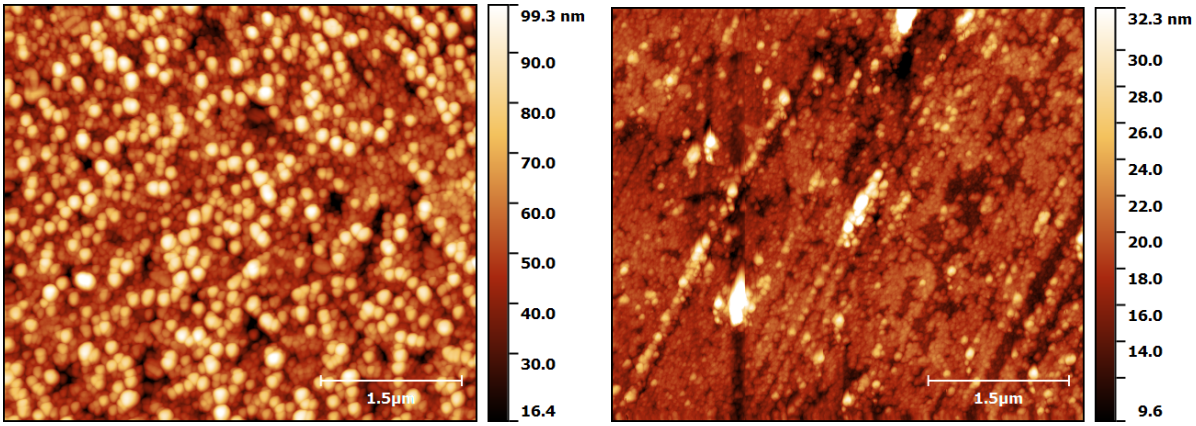
This section presents the results of the investigation of the surface morphology of the samples by AFM. First, the data for the MBE samples is presented, along with a description of the parameters under which these samples are grown. Next, the MOCVD and HVPE samples are analysed.

MBE samples

As discussed earlier, different III/V ratios define different growth regimes in PAMBE, with distinct surface morphologies as a consequence. It is expected that N-rich growth leads to low Ga adatom mobility, resulting in a high nucleation rate, 3D growth and consequently rough surfaces. On the contrary, the high adatom mobility in the Ga-rich regime should lead to a 2D growth mode and smooth surfaces.

Sample MBE 30-11 was grown directly on untreated sapphire at a substrate temperature of 750 °C, a nitrogen flow of 2 sccm, and a gallium source temperature of 1100 °C. The nitridation was carried out during ramp up of the sample temperature. The flow conditions were attempted to be kept constant for all samples presented here. Considering the above, and noting that the surface appears to be very similar to the one under N-rich conditions by L. He [55], it would seem that sample MBE 30-11 shown in figure 5.1(a) was made under N-rich conditions. However, it was found that slightly increasing the III/V ratio above that which was used for this sample resulted in Ga-droplets on the surface, which means that MBE 30-11 was made in the Ga-rich

intermediate regime near the Ga-droplet border (see figure 4.2). The fact that the surface has characteristics similar to N-rich growth can thus likely be attributed to the lack of a nucleation layer on this sample, as the referenced samples by L. He were grown on MOCVD GaN templates.



(a) (b)

Figure 5.1: *AFM images of* **a***: sample MBE 30-11 and* **b***: MBE 22-03.*

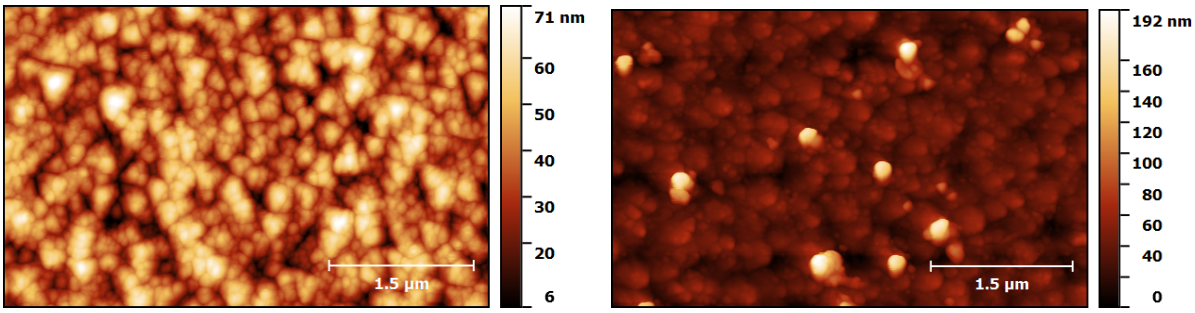
As mentioned in appendix D, the gallium crucible cracked at one point. Once a new crucible was installed and filled, it was found that gallium from the cracked crucible had likely been deposited on the heating filaments of the source, which reduced the attainable temperature of the tip filament below 1100 °C as well as the temperature uniformity.

The remaining MBE samples were grown after this, which had drastic influences on their properties. For sample 22-03, the preparation procedure of Namkoong et al. [77] was roughly followed, with a nitridation procedure for 1 hour at 200 °C (for this and all remaining samples on annealed sapphire), nucleation layer deposition under the same flow as that during growth of the epilayer, annealing of the nucleation layer for 10 min at 750 °C, and finally epilayer deposition at approximately 615 °C. While the lower substrate temperature during deposition could result in a higher growth rate due to less decomposition, it is evident from table 5.1 that the rate is dramatically reduced compared to sample 30-11, due to the issues with the gallium source. However, as evidenced from the RMS roughness of the AFM image in figure 5.1(b), the surface is much smoother than on 30-11. This vast improvement could be due to two things: lower growth temperature resulting in a growth regime closer to the Ga-droplet border (despite the much reduced Ga-flow as evidenced by the reduced growth rate), or due to the advantageous effect of the atomically flat, nitrided substrate and nucleation layer template. The latter seems most likely, as sample 30-11 was already found to be within the appropriate III/V ratio regime for optimized growth.

MBE samples	**30-11**	**22-03**	**30-03**	**01-04**	**04-04**
Growth rate ($\mathrm{nm\,min^{-1}}$)	3.3	1.29	0.97	0.93	0.58
RMS roughness (nm)	15.03	3.47	9.46	11.59	18.50

Table 5.1: *Table of the RMS surface roughness for the investigated MBE samples. As seen from the growth rate, the condition of the Ga source worsened with use.*

Sample 30-03 was grown in the same manner as 22-03, except the substrate temperature was raised to 750 °C during epilayer growth. This dramatically increased the roughness (see table 5.1), likely due to degradation of the Ga source giving a lower III/V ratio, as well as the higher substrate temperature yielding a growth regime further away from the Ga droplet border. Sample 01-04 was grown similar to 30-11, except with a much lower growth rate. Compared to 30-11, the AFM image shown in figure 5.2(a) clearly shows increased island width, but not increased height (as also seen from the roughness), indicating enhanced surface mobility on this sample. This is curious, as it is grown under a substantially lower III/V ratio. A plausible explanation is that the atomically flat sapphire substrate for this sample enhanced the mobility, which along with the lower growth rate (possibly increasing the average diffusion time) outweigh the decreased Ga-adatom mobility due to N-rich conditions.



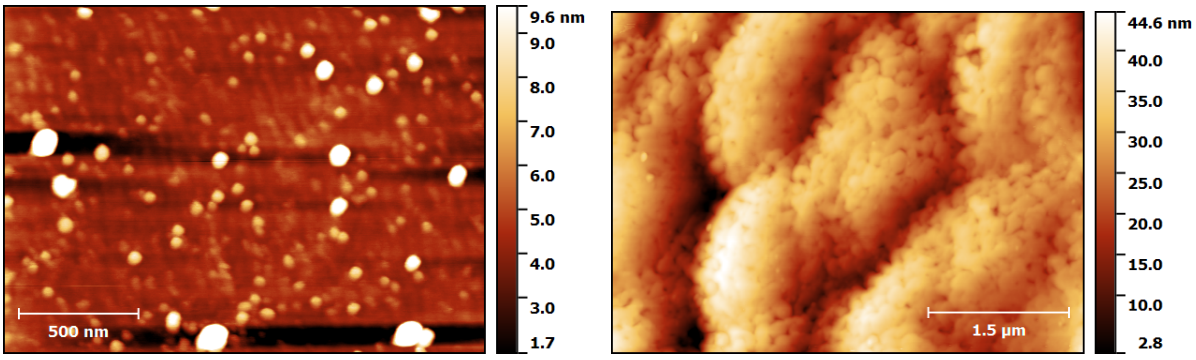
(a) (b)

FIGURE 5.2: *AFM images of* **a***: sample MBE 01-04 and* **b***: MBE 04-04.*

The final MBE sample is shown in figure 5.2(b), which was made with the same conditions as 22-03. It is evident that the roughness is much higher than that for 22-03. As the growth rate is very low for an N-flow of 2 sccm, the increase in roughness is likely due to extremely N-rich conditions, which is further supported by the appearance of facets in the AFM image.

HVPE and MOCVD samples

The surface morphology of the HVPE Si sample is shown in figure 5.3(a). While many islands are apparent on the image, the overall roughness of the sample is low compared to the MBE samples, as shown in table 5.2. Despite the poor quality of this image, surface steps are apparent, and the height of these correspond 0.2-0.4 nm, which fits well with step height of a (0001) GaN monolayer [58], although the image quality makes it difficult to measure.



(a) (b)

Figure 5.3: *AFM images of* **a***: sample HVPE Si and* **b***: MR 1*

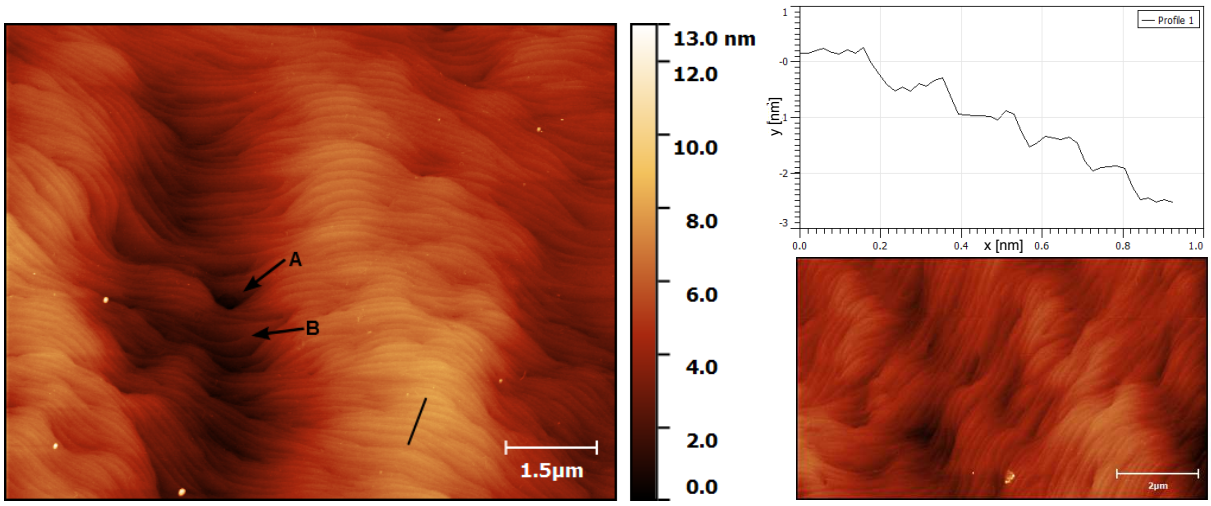
Sample MR 1 is shown in figure 5.3(b), where the surface morphology is markedly different, with a rather large RMS roughness. Large terraces, with very large terrace to terrace height, cover the surface, with the terraces having a high degree of mosaicity. From this, it seems that this sample suffers from unoptimized growth conditions which has lead to island growth, but the uncommon surface morphology makes it difficult to conclude much without further investigation. The surface morphology could also be the result of aqua regia etching, which was performed to remove gold coating.

Sample	**HVPE Si**	**MR 1**	**PC 5274**	**PC 5276**	**PC 5260**
RMS roughness (nm)	1.27	6.47	1.53	0.88	1.97

Table 5.2: *Table of the RMS surface roughness for the investigated HVPE and MOCVD samples.*

Figure 5.4(a) shows an image of sample PC 5274, where a much more common MOCVD-grown GaN surface is observed, compared to e.g. Koleske et al. [84]. The surface on sample PC 5276 in figure 5.4(b) has a similar structure, but a lower RMS roughness. Both surfaces are much smoother than sample MR 1, with clear indication of layer-by-layer growth. They are observed from the profile in figure 5.4(b) to be covered by bi- and trilayer steps, while monosteps are commonly observed on these types of samples. The steps appear to be "pinned", curving out from termination points at either end. This is argued by Heying et al. [58] to be the result of pure screw dislocations pinning the steps at their end points, one of which is possibly seen at point A with a radius of a few hundred nm. At point B, the termination of a step on a terrace is observed, which is also seen by both Heying et al. and Krtschil et al. [85], and is argued to be due to a mixed dislocation terminating at the surface. It is not known why the commonly observed depressions due to the dislocations are not observed, as the step shape otherwise indicate their presence, but this could be due to the large cantilever tip curvature

radius. If this is the case, there may be also numerous on-terrace depressions ∼30 nm wide due to edge dislocation termination (as was also observed by e.g. Heying et al.), that can not be seen in these images.



(a) (b)

FIGURE 5.4: *Surface morphology of* **a***: sample PC 5274 and* **b***: sample PC 5276. The profile in figure* **b** *corresponds to the black line on figure* **a**.

In figure 5.5, an image of the surface of a HEMT structure sample (PC 5260) is shown. While the structure itself is similar, this sample is somewhat rougher than the previous ones, and the step curvature radius is much smaller, indicating a higher density of the dislocations terminating at the surface. Additionally, rather large holes are present in the layer of up to 5 nm depth which, considering that the GaN cap has an estimated thickness of 3 nm, might be a serious issue.

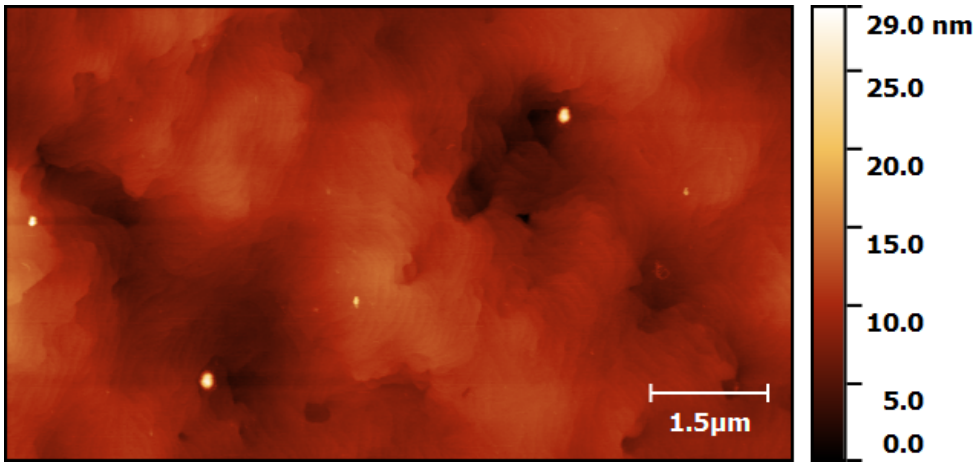


FIGURE 5.5: *AFM image of sample PC 5260.*

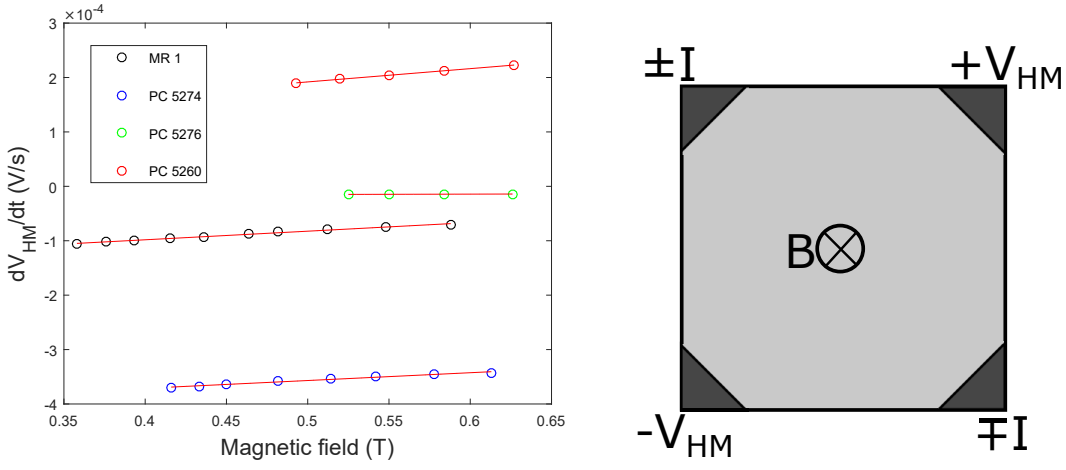
5.2 Electrical characterization

The sheet resistance, R_s, sheet carrier density, and Hall mobility were all measured by using a combined van der Pauw and Hall method. Measurements were made on five samples, and the results are listed in table 5.3. The three PC samples all gave very reliable data, and the fits for their dV_{HM}/dt-B graphs are shown along with the measurement geometry in figure 5.6. The fits are very linear, and the same observation was seen for the I-V_{HM} graphs, and all samples were found to be n-type.

The MR 1 sample showed some non-linear behaviour in its I-V_{HM} graph. Thus the reliability of the results obtained for the MR 1 sample are more questionable. The non-linear I-V behaviour was not removed by improving the contacts, so it is possible that it is related to the large sample roughness, as this could make the application of ohmic contacts more difficult. Nevertheless, most errors would indeed be expected to be cancelled by proper data treatment, and the fit obtained for the dV_{HM}/dt-B graph is adequately linear. For the MBE 30-11 sample it was not possible to apply a current through the sample, and therefore no data was collected for it. It is likely that the sample roughness resulted in poor bonding of the indium contacts. However, it is also possible that it is a matter of sample quality, yielding a very large sample resistivity.

Sample	R_s ($\Omega/\square$)	n_s (cm^{-2})	N_d (cm^{-3})	μ_n ($\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$)
MBE 30-11	-	-	-	-
MR 1	4.18×10^3	1.91×10^{13}	1.00×10^{17}	78.13
PC 5274	2.26×10^3	2.22×10^{13}	1.85×10^{17}	124.89
PC 5276	78.11	3.75×10^{14}	3.12×10^{18}	213.25
PC 5260	541.42	1.02×10^{13}	-	1.13×10^3

Table 5.3: *Van der Pauw and Hall data obtained from electrical characterization of a few GaN samples (sample 5260 has a HEMT on top of it). It was not possible to apply a current through the MBE 30-11 sample. The results are obtained from an average for the geometry in figure 5.6(b) and a geometry, where the applied current and measured voltage are switched.*



(a) (b)

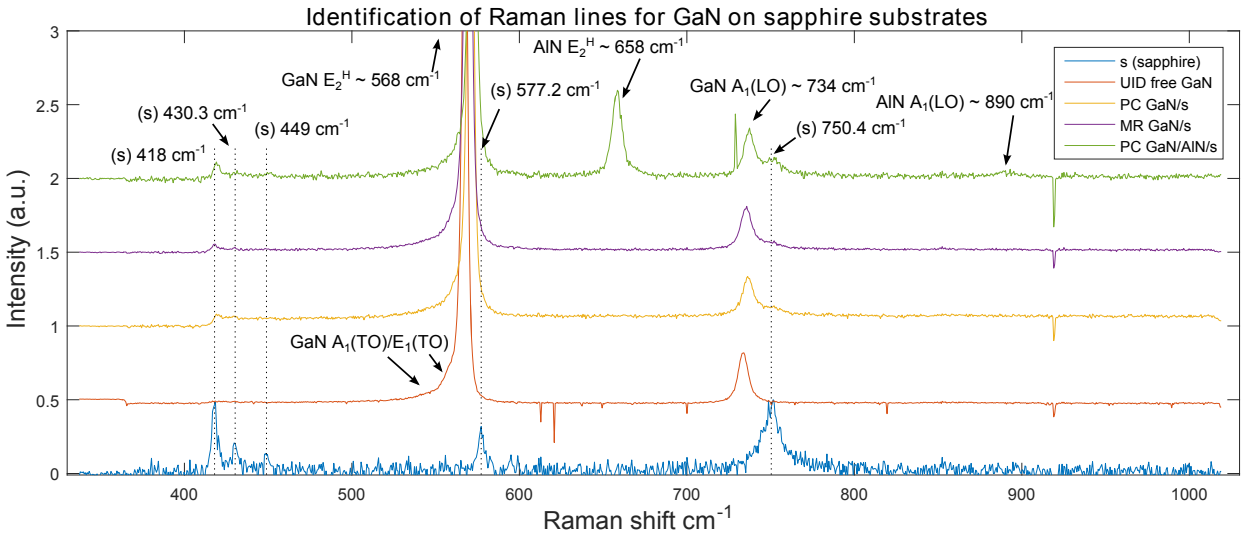
Figure 5.6: **a**: *shows the linear fits made for data points obtained from Hall effect measurements. The red lines are the linear fits for each sample.* **b**: *shows the measurement geometry for* **a**.

Comparing the results for the MR 1 and PC 5274 samples show that while the two samples have comparable carrier densities, the PC 5274 sample has a significantly lower resistivity than the MR 1 sample. This clearly indicates that the PC 5274 sample has fewer structural scattering sites and therefore its mobility is also larger. Furthermore, the Si doped PC 5276 has an even larger mobility than PC 5274. This could indicate that the Si-doped sample has a reduced number of scattering sites, but as the Si doping is the only known sample difference between the samples, this seems unlikely. Another explanation is proposed through comparison with Lee et al. [86]. They found a similar increase in mobility for samples of comparable Si doping concentrations with the ones found in PC 5276, and through XRD measurements they confirmed that this increase was not related to structural improvements. However, while it was confirmed by Chin et al. [87] that the introduction of additional impurities increase the impurity scattering, they showed that this only lowered the mobility for a fixed compensation ratio, N_A/N_D. If the number of deep compensating acceptor impurities is not increased proportionally by the introduction of Si dopant atoms (shallow donors), this would lead to a higher degree of conduction band conduction relative to defect band conduction, which could account for the improved the carrier mobility.

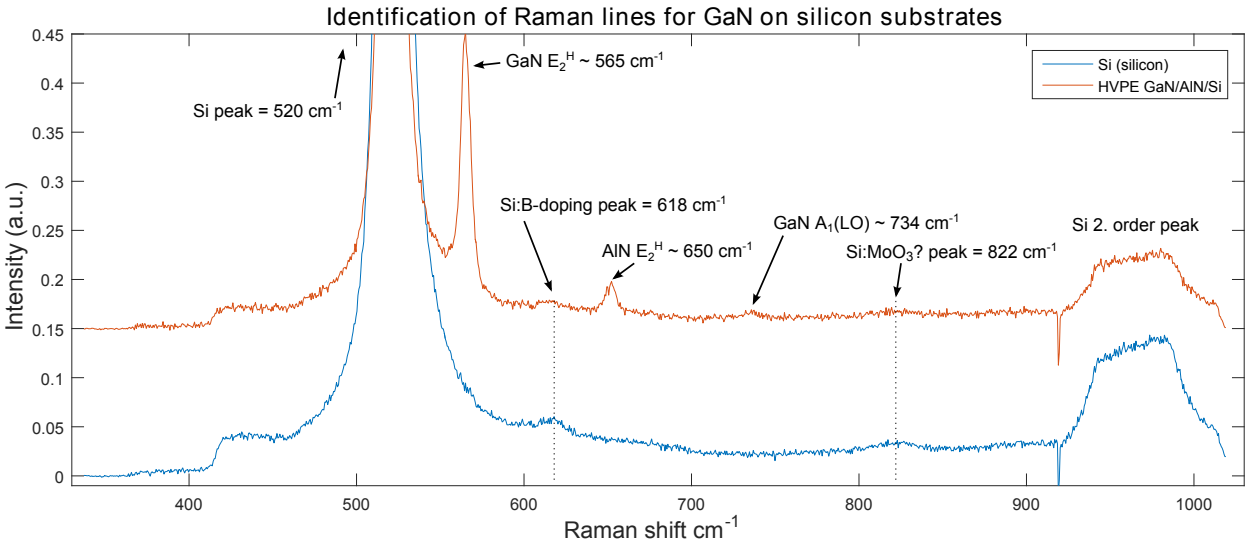
The Hall effect was also measured for the PC 5260 sample to observe the effect of a 2DEG. This sample clearly shows a much higher mobility as would be expected for a HEMT structure. The result is far from the predicted theoretical limit of $\approx$2000 $\mathrm{cm^2\,V^{-1}\,s^{-1}}$ [88], yet the value is comparative to the results by other authors using MOCVD for HEMT growth [1].

5.3 Raman spectroscopy

As was discussed in sections 2.2 and 3.3, Raman spectroscopy is a versatile technique well suited to studying many properties of epitaxially grown GaN. In order to get an overview of the spectral features observed in different measurements, spectra from several samples as well as an identification of their Raman peaks are shown in figure 5.7.



(a)



(b)

Figure 5.7: *Spectra identifying the Raman peaks observed from different samples. The largest intensity peaks are off the scale in order to clarify the rest of the spectral features. The sharp negative peaks are due to pixel errors. The spectra are normalized, and the sapphire spectrum in (a) was then divided by 8 for illustrative purposes. The spectra are shifted vertically for clarity.*

The positions of the peaks are generally in good agreement with the values given in the literature. Depending on the source, the peak positions can vary typically up to 1-2 cm^{-1},

which can be attributed both to differences between the investigated samples, and the accuracy of the spectroscopic setup employed. The accuracy in the setup used for our investigation was found to be within $1\,\mathrm{cm}^{-1}$, making possible the detection of quite small line shifts due to e.g. residual strain.

In figure 5.7(a), it is evident that the sapphire substrate is responsible for five Raman peaks within this energy range. Of these, the two peaks with the highest energy shifts at $577.2\,\mathrm{cm}^{-1}$ and $750.4\,\mathrm{cm}^{-1}$ can be troublesome for the identification of position and width of the GaN E_2^H and A_1(LO) modes, respectively. This is a larger problem for the weaker and broader the GaN peaks. Besides the GaN peaks, the corresponding peaks of higher energy are also observed for AlN on the PC GaN/AlN/s sample at $658\,\mathrm{cm}^{-1}$ and $890\,\mathrm{cm}^{-1}$ for the E_2^H and A_1(LO) modes, respectively, although the A_1 mode is very faint. It is interesting to note that the GaN E_2^H peaks are broadened towards the low phonon-frequency side, which is likely due to the presence of the GaN A_1(TO) and E_1(TO) modes located at $533\,\mathrm{cm}^{-1}$ and $561\,\mathrm{cm}^{-1}$, respectively. While these modes are assumed to be forbidden in the geometry used here, it is argued by Harima et al. [40] that it is possible that due to the large solid angle of the microscope objective lens, not only strictly backscattered light is collected. This means that it is possible to observe mixed E_1 and A_1 modes, yielding quasi-TO or quasi-LO modes that propagate at a slight angle relative to the optical axis.

In figure 5.7(b), the silicon substrate has two very distinct Raman peaks - one strong, sharp peak at $520\,\mathrm{cm}^{-1}$ and a broad second order (two-phonon) peak at $975\,\mathrm{cm}^{-1}$. Two other faint peaks are apparent, the first at $618\,\mathrm{cm}^{-1}$ which is believed to be caused by boron doping of the silicon wafer. This is supported by Fukata et al. [89], who continuously observed this boron-related peak in silicon nanowires. The origin of the faint peak at $822\,\mathrm{cm}^{-1}$ is less obvious, but might be due to MoO_3, as Mo is a common contaminant in processed Si wafers, likely from the production equipment itself [90]. Also apparent are the GaN E_2^H and A_1(LO) and AlN E_2^H peaks, which show that the films are clearly tensile strained.

5.3.1 Residual strain and epilayer quality

MBE and HVPE samples

Raman spectra for the MBE samples and the HVPE sample are presented in figure 5.8. In general, the signal was low for these samples, which can partly be attributed to crystalline quality, but also simply because the films are thinner than the depth resolution. In particular, the signal was weak for the bottom four samples in the figure, and the GaN E_2^H peak is very broad, even compared to other films of similar thickness. This indicates a very poor crystalline quality of these samples. Samples MBE 01-04 and MBE 30-03 also show an additional peak at $\sim 555\,\mathrm{cm}^{-1}$, which could correspond to the TO phonon mode of the zinc-blende phase. This will be discussed further in the coming sections.

For sample MBE 30-11, it is apparent from the E_2^H shift that the film is compressively strained, as expected on sapphire, while samples MBE 03-11, MBE 25-11, and HVPE Si are tensile strained, as expected on silicon. However, the bottom four samples in the figure seem to be tensile strained as well, even though they are grown on sapphire. This effect was discussed in section 3.4, where it was argued that this can be due to a high concentration of threading dislocations, which again highlights the poor quality of these samples. Another interesting point is that none of the top four samples have any discernible A_1(LO) peak.

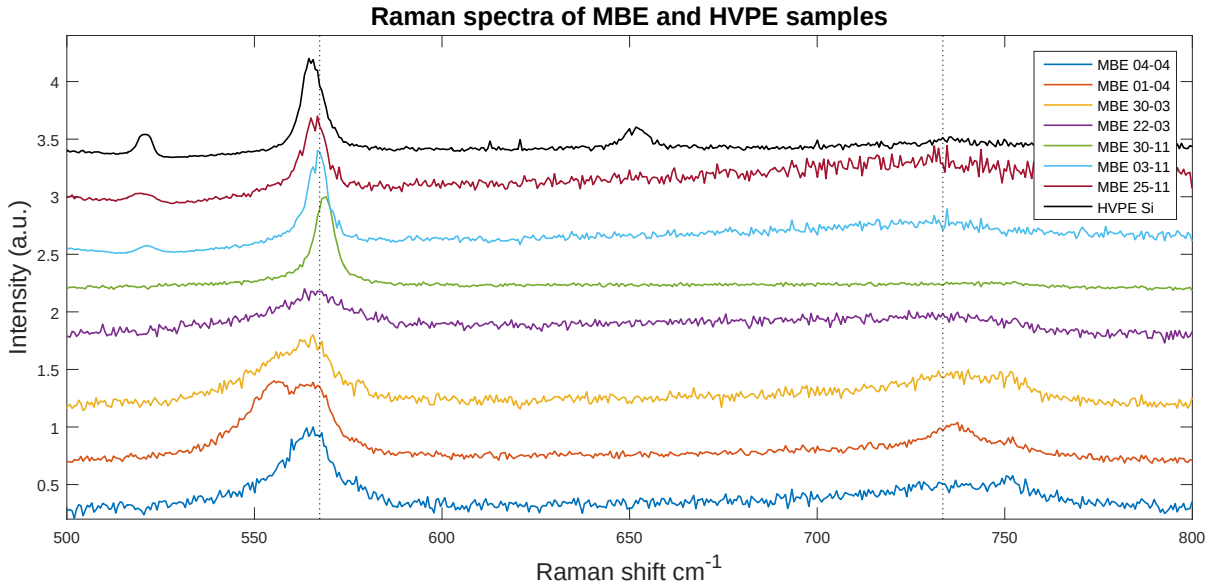


Figure 5.8: *Spectra of MBE and HVPE samples. The dotted lines show the positions of the unstrained E_2^H and $A_1(LO)$ peaks, respectively. The silicon peak was removed for clarity on the samples on silicon substrates. The spectra are normalized and shifted vertically for clarity.*

In order to quantify the residual strain found from E_2^H peak shift, an empirical equation from Zhao et al. [91] is used to relate the shift to the stress

$$\sigma = \frac{\Delta\omega}{k} \left(\text{GPa} \cdot \text{cm}^{-1}\right) \tag{5.1}$$

where σ is the biaxial stress in the epilayer, $\Delta\omega$ is the peak shift, and k is an experimental parameter (4.3 for Si substrates, and 2.56 for sapphire, due to Zhao et al.). The stress is always calculated relative to the UID free GaN sample. This, along with the peak maxima and FWHMs, are found in table 5.4 for the samples in figure 5.8, which have a peak that is well-defined enough to make a Gaussian fit (excludes the bottom four samples). For all samples, due to the enlargement of the peak towards lower cm^{-1}, the fits may be slightly shifted, but this effect was found to be smaller than the resolution of the spectra. As expected, the FWHM of all samples were larger than that for the UID free GaN sample.

	MBE 30-11	MBE 03-11	MBE 25-11	HVPE Si	UID free GaN
Peak pos. (cm^{-1})	568.6	566.8	566.0	565.4	567.4
FWHM (cm^{-1})	9.5	9.0	10.6	9.5	6.4
σ ($\text{GPa}\cdot\text{cm}^{-1}$)	0.4877	−0.1298	−0.3385	−0.4648	

Table 5.4: *Table of the GaN E_2^H peak positions, Gaussian FWHM, and calculated stresses.*

MOCVD samples

Continuing with the MOCVD samples, it is apparent in figure 5.9 that the Raman peaks from these samples generally had a sharper and stronger peak signal. This is partly due to the samples being thicker, and seen from table 5.5 they also have smaller FWHM than the MBE samples, likely due to better crystallinicity. The table also shows the residual stress in the films relative to the UID free GaN sample, and this is larger than that found on any of the MBE samples. This is possibly due to a stronger correspondence between the substrate and the epilayers, resulting

in the need for accommodation of the TEC mismatch upon cooling. The strain is likely also increased simply because the films in MOCVD are grown at higher temperatures than in MBE.

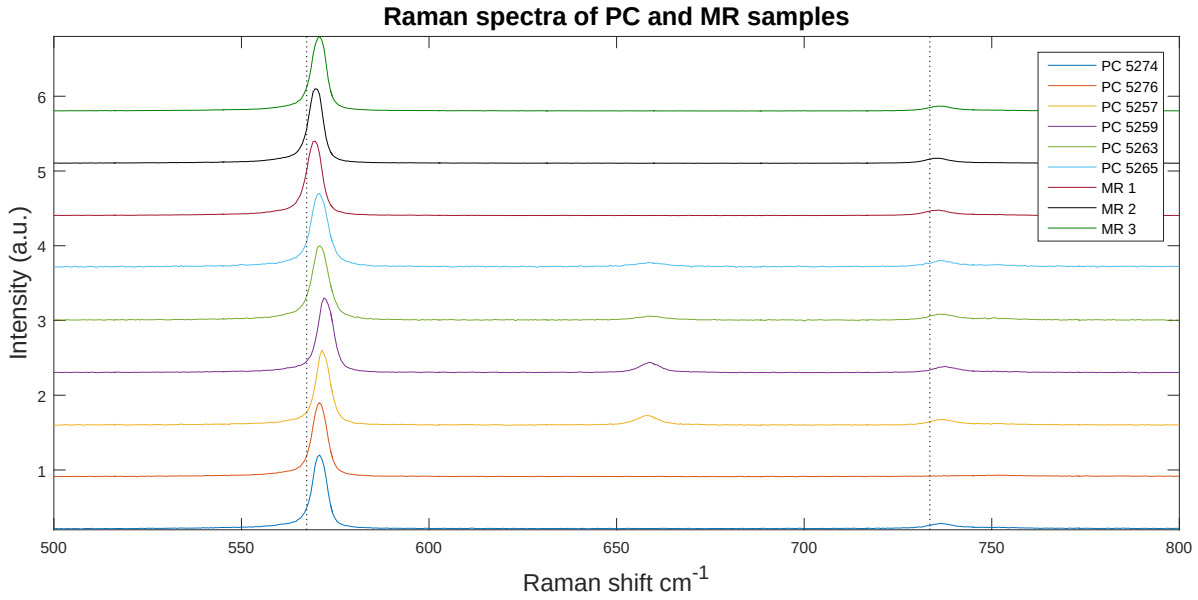


FIGURE 5.9: *Spectra of the MOCVD samples named PC and MR. The dotted lines show the positions of the unstrained E_2^H and $A_1(LO)$ peaks, respectively. The spectra are normalized and shifted vertically for clarity.*

When taking the averages of the samples grown on the PC HEMT structure versus the PC samples without HEMTs, it seems that those with HEMTs are more compressively strained and have larger FWHM, although individual sample data is somewhat ambiguous in this regard. The additional compressive strain is likely due to the GaN epilayer being compressed from both sides, but the increase in FWHM more likely has to do with the difference between the nucleation layers, perhaps indicating that in these samples, the GaN nucleation layer results in superior quality films. Comparing the MR samples and the PC non-HEMT samples, the MR samples are less strained, something which may both be related to the registry with the substrate and to the epilayer thickness.

	PC 5257	**PC 5259**	**PC 5263**	**PC 5265**	**PC 5274**	**PC 5276**
Peak pos. (cm^{-1})	571.7	572.4	571.0	570.7	570.8	570.8
FWHM (cm^{-1})	7.2	7.5	8.1	8.8	7.3	7.5
σ ($GPa \cdot cm^{-1}$)	1.6788	1.9348	1.3914	1.3066	1.3119	1.3231

	MR 1	**MR 2**	**MR 3**
Peak pos. (cm^{-1})	569.4	569.8	570.6
FWHM (cm^{-1})	7.6	7.0	7.2
σ ($GPa \cdot cm^{-1}$)	0.7636	0.9174	1.2420

TABLE 5.5: *Tables of the GaN E_2^H peak positions, Gaussian FWHM, and calculated stresses.*

5.3.2 Effect of carrier concentration

As was discussed in section 3.3, phonons in GaN are able to couple to plasmons, which can have a large influence in the prominence of the phonon modes in a Raman spectrum. In particular, higher carrier concentration is related to a shift and broadening of the A_1(LO) peak. This was evidenced in figure 5.9, where sample PC 5276 has no discernible A_1(LO) signal, and this mode is clarified in figure 5.10.

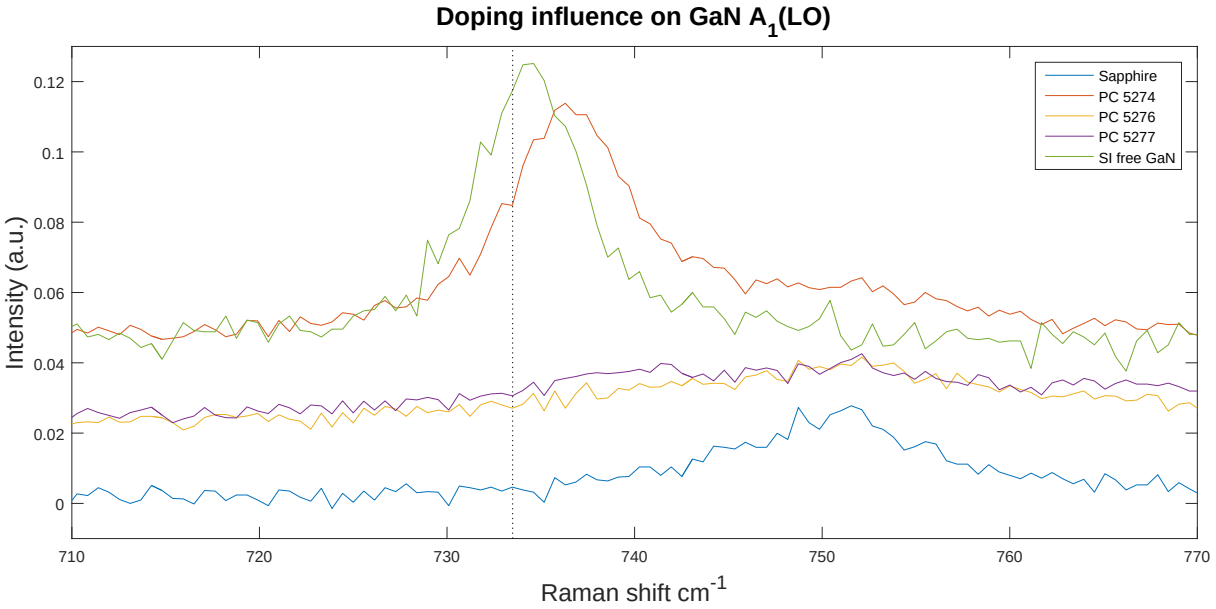


Figure 5.10: *Raman spectra showing the A_1(LO) of selected samples.*

Compared to sample PC 5274, which is undoped but is otherwise similar, the twin samples PC 5276/5277 have either no, or a very flat A_1(LO) peak. It is also evidenced that the SI free GaN sample, which has a low carrier concentration, has none of this effect, and obviously no sapphire signal. Comparison with measurements done by Kozawa et al. [92] on Si-doped GaN indicates that for the A_1(LO) signal to disappear (or at least be very faint), a carrier concentration above $10^{18}\,\mathrm{cm}^{-3}$ is required. This is in agreement with the carrier concentration obtained with the Hall measurements.

The A_1(LO) was also absent on most MBE samples. While no Hall data was obtained for these samples, it is still possible that the magnitude of the carrier concentration could explain the absence of the peak. Given that this is the case, it seems likely that the resistance in these samples is also large, because a low crystalline quality is expected.

5.4 X-ray diffractometry

In order to investigate the crystalline phases present in the GaN epilayers, XRD analysis was carried out on selected samples. In figure 5.11(a), the representative XRD spectrum from sample MR 1 is shown, where very sharp peaks belonging to the (0002) and (0004) planes of wurtzite GaN are immediately identified at $\sim$ 34.7[2Θ] and $\sim$ 73.0[2Θ], respectively. Also apparent are the sapphire (0006) and (000(12)) peaks at $\sim$ 41.8[2Θ] and $\sim$ 90.8[2Θ], respectively. As evidenced by figure 5.11(b), these peaks were by far the most prominent for all samples, with sample HVPE Si also showing an AlN nitide (0002) peak at $\sim$ 36.2[2Θ]. Several peaks which are comparatively very small are also seen at e.g. $\sim$ 37.4[2Θ] and $\sim$ 48.0[2Θ] for all samples, but on these graphs they are only visible for the UID free GaN and MR 1 samples. These peaks correspond well with the reflections from the $(10\bar{1}1)$ and $(10\bar{1}2)$ planes of wurtzite GaN. Their presence could indicate a slight misalignment in the measurement setup.

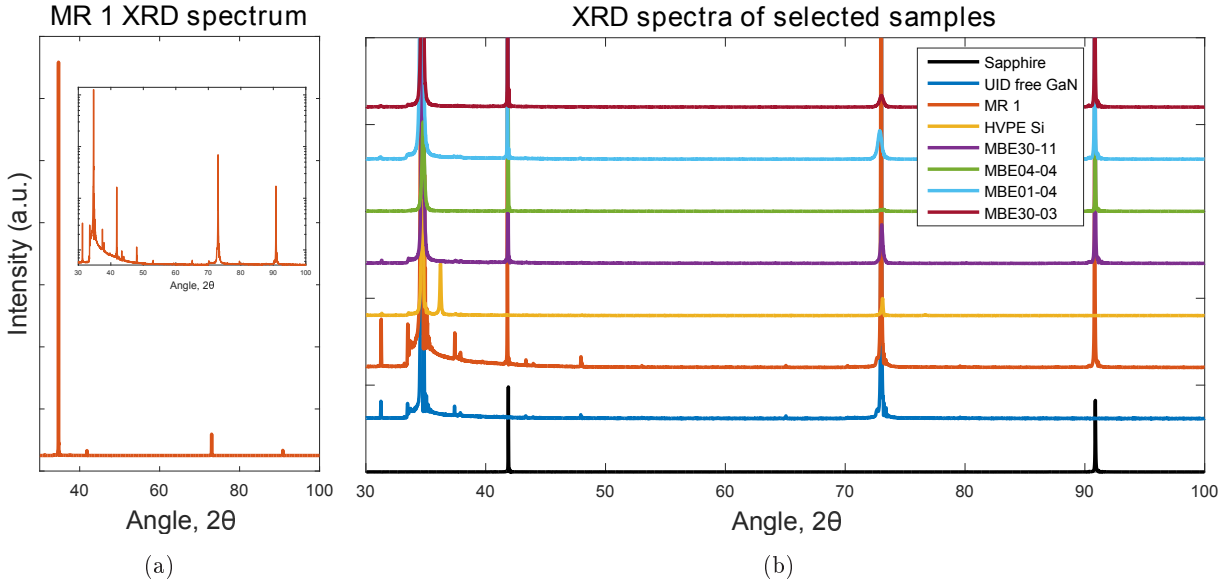
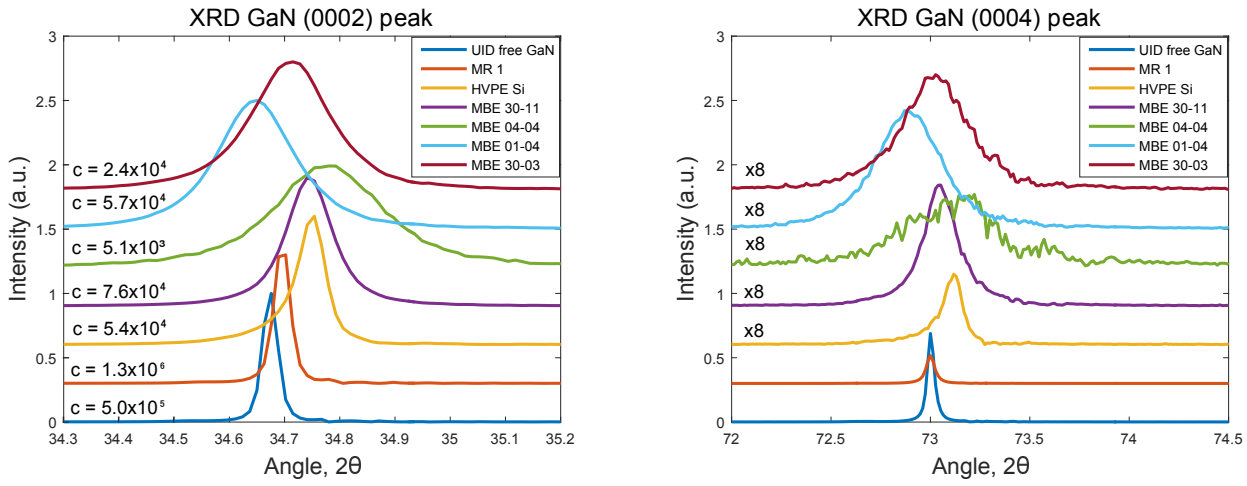


FIGURE 5.11: *XRD spectra of* **a***: raw, stripped data from sample MR 1 with an inset of the same spectrum on a semilogarithmic scale and* **b***: selected samples with the largest peaks off the scale. The spectra are vertically shifted for clarity.*

Generally, sharp and narrow peaks are indicative of high crystalline quality [93], which was also evidenced in these XRD measurements. In figure 5.12, it is apparent that the UID free GaN and MR 1 samples are of superior crystallinicity compared to the rest of the samples, and that the peaks of MBE samples 04-04, 01-04, and 30-03 are particularly wide. While the width of the rocking curve in this geometry indicates the tilt of domains in this geometry (and thus also the density of dislocations with a screw component), the width seen here is simply indicative of differences the spacing between the (0001) planes. The wide peaks seen on the MBE samples are thus likely due to a strain gradient in the c-direction. A high density of dislocations could explain the peak width, as these can cause strong local strain fields. While the peaks are also apparently shifted slightly with respect to one another, it is difficult to conclude much from this, as the sample thickness also has an influence on the constructive interference angle, and this was not corrected for.



(a) (b)

Figure 5.12: *Close-up view of the wurtzite GaN peaks.* **a** *shows the (0002) peak with all spectra normalised, where the inset c denotes the counts at the peak maximum.* **b** *shows the (0004) peak, with all spectra normalised with respect to the (0002) maxima. The top five spectra were subsequently enhanced for clarity.*

It is curious that sample MR 1 has a higher number of counts at the peak maximum than the UID free GaN crystal. However, in figure 5.12(b) where the spectra are normalised with respect to the (0002) peak, it is evident that the (0004) reflection from the UID GaN sample is relatively much stronger than this reflection from MR 1, which could indicate a higher degree of long range order in the UID free GaN. Both GaN peaks show that sample MBE 04-04 seems to be of particularly poor quality, and the shoulder at $\sim 73.6[2\Theta]$ has not been identified. No apparent peaks were found on the MBE samples that were not also apparent on the UID free GaN, indicating that only the wurtzite phase is present. Therefore the suggested zinc blende origin of the Raman peak at 555 cm^{-1} seen in figure 5.8 for the MBE 01-04 sample is likely incorrect. It is thus more likely that the peak origin is due to an enhancement of the $E_1(TO)$ and $A_1(TO)$ peaks of the wurtzite phase.

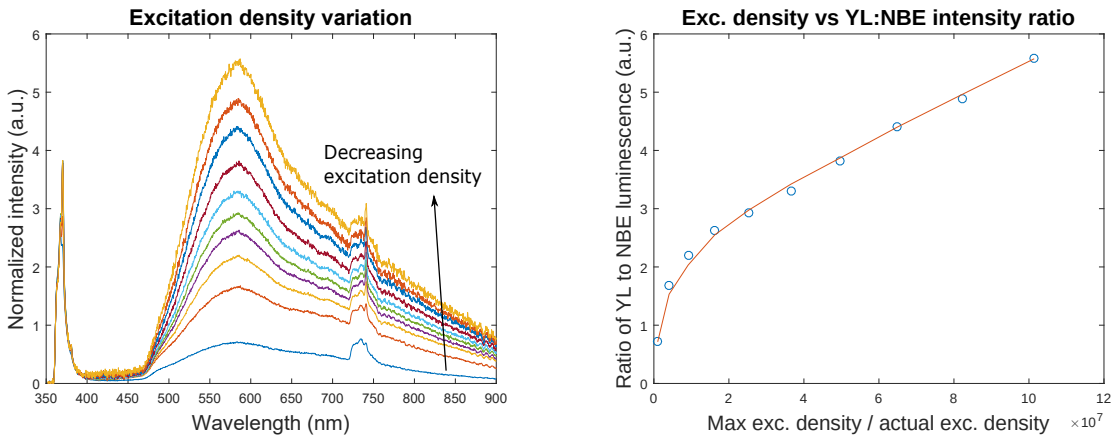
5.5 Photoluminescence spectroscopy

5.5.1 Ratio of yellow to NBE luminescence

In the most general case of RT PL characterization of GaN, the PL spectrum will consist of two distinct luminescence bands, namely the NBE and YL bands. These bands are very different in nature, and therefore it is only natural to consider how these bands are linked to each other.

YL intensity dependency on excitation density

Figure 5.13(a) shows several PL spectra obtained by gradually decreasing the excitation power density (varied by increasing the spot size by adjusting the sample position relative to the objective) using PL geometry 1 from appendix D.3.6. Note also that the 2. order diffraction from the grating is shown in the spectrum, i.e. the NBE luminescence is replicated at twice its wavelength. This PL geometry gave rise to some artefacts, along with a decreased NBE luminescence, but if these effects are considered to have a constant influence on each measurement, the spectra in the figure still illustrate an important result, namely how the YL intensity varies relative to the NBE intensity.



(a) (b)

FIGURE 5.13: **a**: *PL spectra accumulated with decreasing excitation density on the MBE 30-11 sample. The spectra are normalized with respect to the band edge maxima (somewhat masked by a nearby Raman line).* **b**: *data points for the YL band maxima relative to the NBE maxima are plotted with respect to the the maximum excitation density relative to the actual excitation density. The line is shown guide the eye.*

This is shown in figure 5.13(b), where the YL band maximum normalized with respect to the band edge (as shown in **a** of the figure, and it is normalized in order to account for losses) is plotted with respect to the normalized excitation density, meaning that it is the maximum excitation density divided by the actual excitation density. It is expressed in this manner, since the excitation density was varied by increasing the distance between the objective and the sample, thus increasing the laser spot size on the sample. By assuming a Gaussian laser profile, the beam radius at the focal point of the objective is

$$r_{00} = \frac{\lambda}{\pi \mathrm{NA}} = 248.3\,\mathrm{nm} \tag{5.2}$$

with a wavelength of $\lambda = 351$ nm and numerical aperture of NA=0.45. The increase of the laser spot radius as the sample is moved away from the focal point is found from geometrical

considerations (in an ideal case, where the angle is not changed with distance) to be

$$r_L = L \tan\theta + r_{00} \tag{5.3}$$

where L is the distance between the sample and the focal point of the objective, and the angle is related to the numerical aperture and refractive index, n, by $\mathrm{NA} = n \sin\theta$. The power density variation plotted along the horizontal axis of figure 5.13(b) is then calculated as

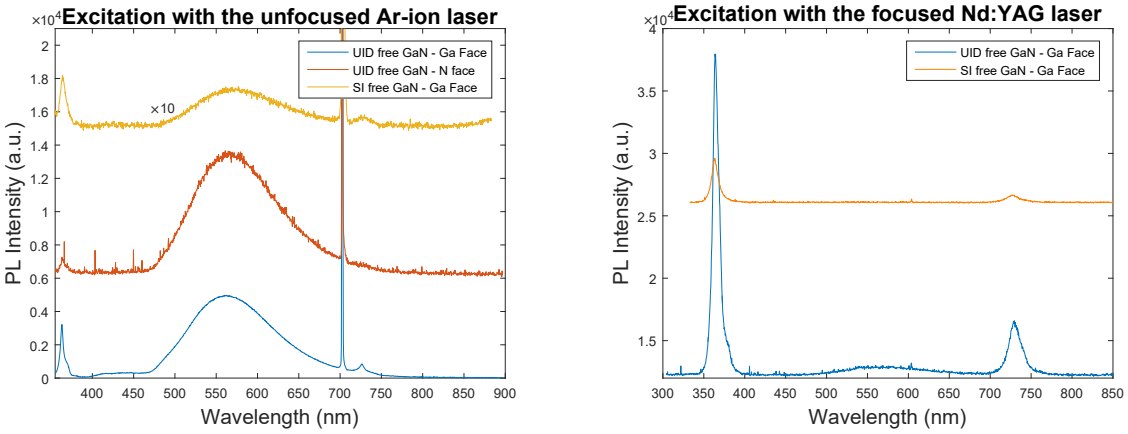
$$\frac{\text{Max. exc. density}}{\text{Actual exc. density}} = \left[\frac{P}{\pi r_{00}^2} \Big/ \frac{P}{\pi r_L^2} \right] = \frac{r_L^2}{r_{00}^2} \tag{5.4}$$

where P is the laser power. For the two plots in figure 5.13, the distance, L, between sample and objective was varied in steps of 50 µm. The variation in figure 5.13(b) is somewhat consistent with what other authors have found for the YL band. Reshchikov and Morkoc [6] report that the YL band has a square-root dependence on the excitation intensity for high excitation intensities, and then it becomes linear as the excitation intensity is decreased. This behaviour can be explained by saturation of the defects responsible for the YL band at high excitation intensities. This is an important consideration, and it clearly illustrates the importance of the excitation intensity, which should always be considered, when comparing the YL:NBE ratio. Furthermore, the region where this dependency changes from a square-root dependence to a linear dependence may vary depending on sample properties.

Unfocused vs focused excitation

Figure 5.13 gives an illustration of the relationship between the YL band and the excitation density. However, the NBE luminescence obtained from these spectra is disturbed by artificial effects due to the filter and beam splitter used in that geometry. Furthermore, since the energy of the laser line is close to the band gap energy, and also due to the laser geometry using an objective for focusing, some parasitic Raman lines could be seen in the spectra as well. Therefore other optical geometries were deployed. Geometry 2 utilizes a diffuse setup, where the Ar-ion laser is unfocused on the sample resulting in a spot size of several millimetres. Figure 5.14(a) shows the spectra for both sides of the two freestanding GaN crystals. The spectra are accumulated similarly, except that electron multiplying was used for amplification of the UID N face and SI Ga face, which dramatically enhances the signal. Despite their appearance in the spectrum, the PL efficiency is thus much higher for the Ga faced side (no luminescence from the N-face of the SI GaN sample was measured), and also the ratio of YL:NBE is larger for the N face side of the UID sample. This can be explained by the growth of the samples, since they were grown with Ga-polarity. Hence more structural defects are expected on the backside, lowering the PL efficiency. Similarly, the YL:NBE ratio indicates a larger concentration of point defects.

The Ga face side of both samples was also excited under steady-state conditions by the focused Nd:YAG laser in geometry 3. The spectra for these two measurements are shown in figure 5.14(b), where the NBE luminescence dominates the spectrum. Only little YL is seen for the UID sample, and the SI sample shows no YL. The two figures thus further illustrate the importance of the excitation density. In the literature, the quality of GaN samples is often justified from the YL:NBE ratio in PL measurements. As the results given here indicate, this justification must be treated carefully, especially when the excitation density is not supplied. [6, 38]

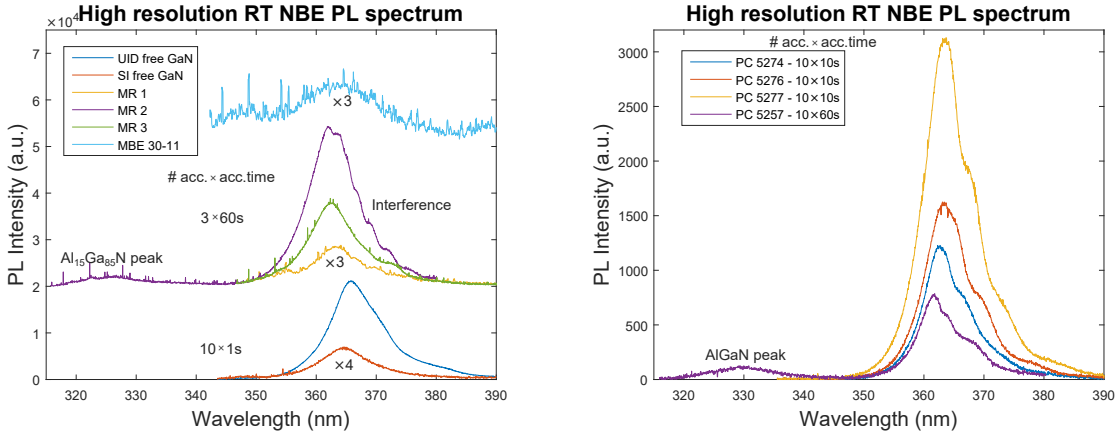


(a) (b)

FIGURE 5.14: *PL spectra of the two freestanding GaN samples.* **a***: excited with the unfocused Ar-ion laser in geometry 2. The UID N Face and SI Ga face spectra are taken with electron multiplying of 150 in order to enhance the measured signal. The sharp line around 700 nm is the 2. order diffraction of the laser line due to the grating. It should be noted that the laser line sometimes caused saturation effects in the spectrometer, which in this instant was included in the NBE band (note the very sharp tip in the NBE band).* **b***: excited with the focused Nd:YAG laser in geometry 3.*

5.5.2 NBE luminescence excited by focused Nd:YAG laser

It is now of interest to consider the NBE luminescence, its structure, intensity, and where it peaks for different samples. It was found that the NBE luminescence is easier to resolve with high excitation densities, and since good agreement of the NBE shape and peak energy was found between CW excitation with the Ar-ion laser and pulsed excitation with the Nd:YAG laser (seen e.g. in figure 5.14), this section is concerned with focused excitation using the Nd:YAG laser.



(a) (b)

FIGURE 5.15: *PL spectra obtained by excitation of GaN samples by the focused Nd:YAG laser. The accumulation time and number of accumulations are listed in both figures.* **a***: shows some of the difference between the HVPE, MOCVD, and MBE grown samples. The spectra are shifted vertically for clarity.* **b***: shows the difference between a few of the polish MOCVD samples.*

HVPE grown freestanding GaN samples

The NBE luminescence at RT for the free GaN samples are shown in figure 5.15(a). The NBE luminescence is much more prominent for the UID crystal compared to the SI one, suggesting a much higher quantum efficiency. The SI sample has a peak energy 10 meV blue-shifted from the UID one. Both also display a prominent tail on the low-energy side of the band, likely due to an LO phonon replica of the NBE.

MR MOCVD grown samples

The three MR samples are shown in figure 5.15(a). The two samples containing an AlGaN layer give more prominent NBE bands compared to the one without an AlGaN layer. A faint, broad AlGaN peak could be identified only for the $Al_{0.15}Ga_{0.85}N$ sample, probably because $Al_{0.30}Ga_{0.70}N$ would require a higher excitation energy. The peak energy of the NBE is blue-shifted for samples containing AlGaN on top, possibly due to more strained top layer from the AlGaN/GaN interface.

MBE grown samples

The NBE band of sample 30-11 is shown in figure 5.15(a). The signal is weak, likely due to sample quality. Yet this is the only sample with any NBE signal out of the MBE grown samples.

PC MOCVD grown samples

The PL spectra for four of these samples are shown in figure 5.15(b). The GaN samples with a GaN NL generally exhibit much more NBE luminescence compared to those with a HEMT structure on top. Only one spectrum is shown for a sample with a HEMT structure, since these had quite similar NBE bands. Sample 5276 and 5277 are both Si doped, and they have same band maxima, slightly shifted from the undoped sample 5274 by 10 meV, which in turn is shifted from the HEMT samples also by 10 meV. The shift for the PC 5274 and 5276 samples may be explained as follows. Raman measurements showed that their E_2^H modes have equal Raman shifts, so the difference in NBE maxima cannot be stress related. The carrier density for the undoped, and Si doped sample was found to be 1.85×10^{17} and 3.12×10^{18}, respectively. Therefore, only band gap shrinkage is a likely explanation for this NBE maxima variation between the undoped and Si doped samples. Indeed a shift of $\sim$10 meV was also found by In-Hwan Lee et al. [94], who studied the band gap shrinkage for similar carrier concentrations for UID and Si doped GaN.

Summary

The freestanding GaN samples generate more NBE luminescence compared to other samples. The samples containing the AlGaN layers tend to have their NBE peak at a higher energy compared to those without AlGaN. The MBE grown sample has a broad, faint, NBE peak, suggesting poor crystalline quality and a high concentration of defects. MR MOCVD samples with AlGaN layers generate more NBE luminescence compared to bare GaN, which is studied further with respect to passivation later. The tail emerging around 370-380 nm is only visible in HVPE grown freestanding GaN and PC MOCVD grown samples, and shifts along with the NBE maxima, which strongly implies that it is a phonon replica.

An empirical formula describing the linear variation of the band gap energy with stress is used to see if stress can explain some of the difference in NBE maxima for the investigated

samples. The UID free GaN is expected to be bulk and thus unstrained. Therefore its maxima is used as the zero point for this equation. The band gap variation with stress is [91]

$$E_g = 3.389 + 0.0211\sigma \tag{5.5}$$

and it is compared to the band maxima values in figure 5.16. Each data point represent the band maxima of a sample compared to the stress obtained from Raman measurements. It is thus found that stress variation between the samples cannot explain the entire difference in NBE maxima, but it is definitely part of the reason, especially when considering the sample grown on Si, which is the only one shifted to the left of the UID position in the figure.

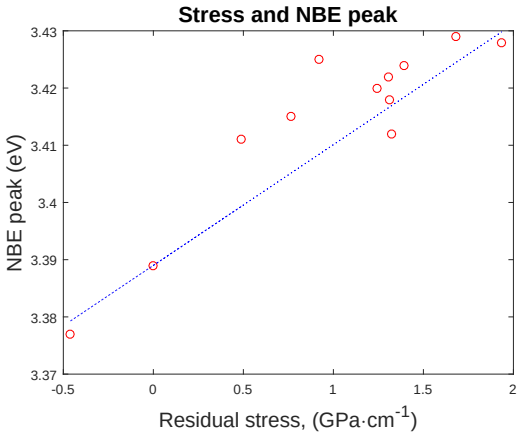
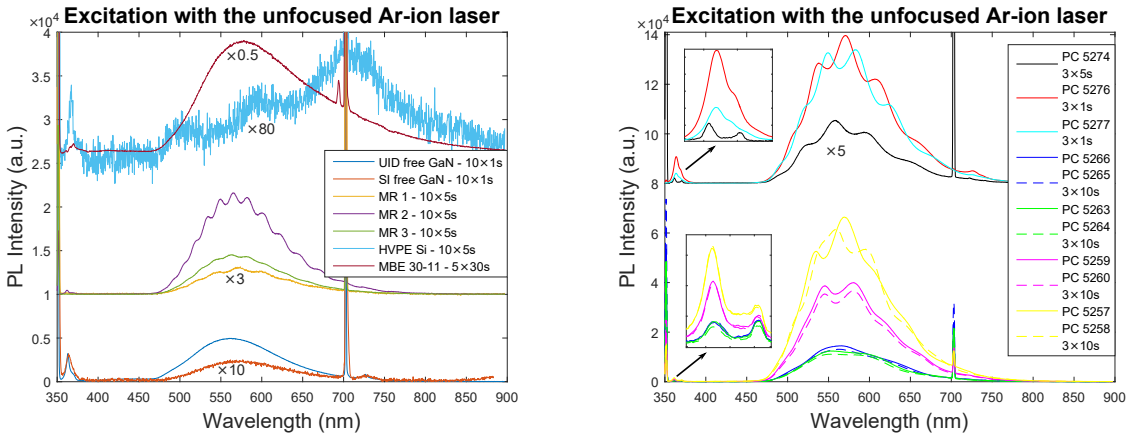


FIGURE 5.16: *Scatter plot of the residual stress against the NBE peak position. The dotted line is a fit to equation 5.5.*

5.5.3 Excitation with the unfocused Ar-ion laser

In this section, mainly the defect luminescence of the GaN samples is considered. The spectra for these are shown in figure 5.17, excluding samples that gave no luminescence.



(a) (b)

FIGURE 5.17: *PL spectra obtained by excitation of GaN samples by the unfocused Ar-ion laser. The accumulation time and number of accumulations are listed in both figures.* **a**: *shows some of the difference between the HVPE, MOCVD, and MBE grown samples. The spectra are shifted vertically for clarity.* **b**: *shows the difference between all the polish MOCVD grown samples. Stippled lines refer to twin samples of the same color. The spectra are shifted vertically for clarity, and the two insets are included for better comparison of the NBE luminescence.*

HVPE free GaN

The spectra for the two freestanding GaN samples are shown in figure 5.17(a). Both crystals have clear YL bands. The UID crystal gives much more luminescence, however, its YL:NBE ratio is larger than for the SI crystal. The SI sample is Fe doped, and this could have reduced the amount of other compensating defects, possibly explaining the reduced YL:NBE ratio.

MR MOCVD grown samples

The spectra for these are also shown in figure 5.17(a). All samples give clear YL bands with some interference. Only the MR 2 sample shows any NBE luminescence, and thus makes it difficult to compare any YL:NBE ratios. The two samples containing AlGaN gives more overall luminescence compared to the one without.

HVPE Si and MBE samples

These samples gave almost no luminescence with the unfocused Ar-ion laser, and only the HVPE Si and MBE 30-11 spectra are shown in figure 5.17(a). The HVPE Si sample gave a noticeable signal. It displays what would seem like several defect related luminescence bands, including GL, YL, and RL bands. Each band are of comparable size to that of the NBE band, indicating a small (defect band):NBE ratio. The MBE 30-11 sample has a large YL:NBE ratio and shows a clear YL band like other GaN/s samples. The YL band tail seems to stretch further towards larger wavelengths, being a potential hint of a disguised defect band, possibly the RL band. Note also a sharp peak around 694 nm, which was identified as the well known Cr^{3+} peak (ruby) in the sapphire substrate [95]. This peak was only seen in PL from very thin samples, where some of the exciting light still reaches through the GaN layer.

PC MOCVD grown samples

The spectra for all of these samples are shown in figure 5.17(b), including twin samples (allegedly made with same parameters). The three GaN/s samples (5274, 5276, 5277) all display broad YL bands. For PC 5276 and 5277 the NBE luminescence is clearly seen, and show a smaller YL:NBE ratio compared to other MOCVD grown samples. Interference in the YL band might also be observed. Here the period is larger than for the MR MOCVD samples (note that the PC GaN layers are thinner than the MR ones), and also the peaks in the YL band are sharper compared to previous results.

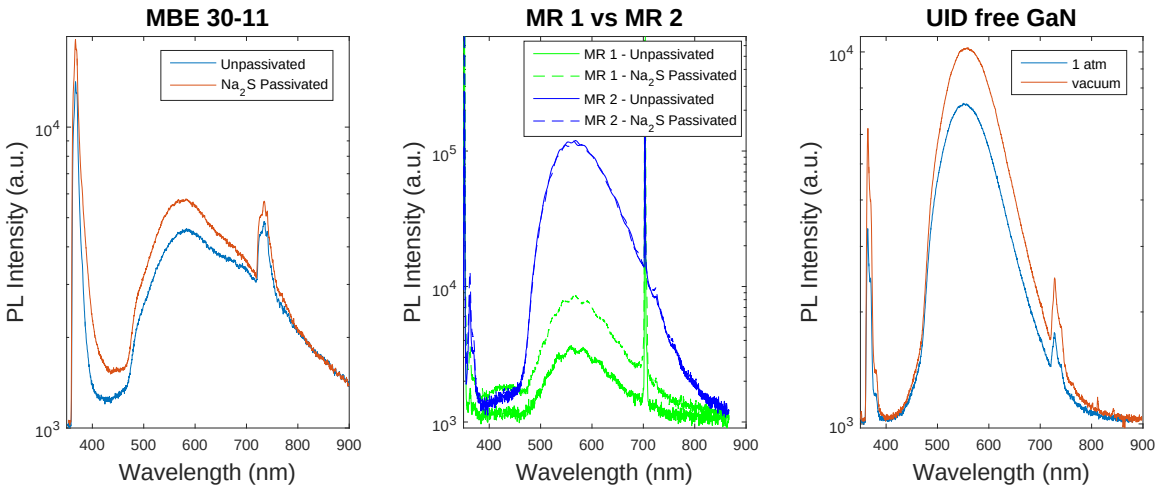
The PC HEMT samples are also shown. These samples display less luminescence compared to those without a HEMT structure. The NBE bands are quite faint with a large YL:NBE ratio. Some of the YL bands here also display what might look like interference, most prominently seen for the PC 5259, 5260, 5257 and 5258 samples. These samples also seem to have more prominent defect bands, however, the YL:NBE ratio does not seem to vary much for the respective samples. In these spectra the twin samples are also shown, and out of these only the 5257 and 5258 samples appear to have different shapes of the YL band. It is also seen that there may be slight variations in the accumulation of sample luminescence, and therefore small intensity variations seen between twin samples cannot necessarily be attributed to small variations in sample properties.

Summary

Freestanding GaN samples display the smallest YL:NBE ratio, followed by the PC 5276 sample. The PC 5276 and 5277 samples could seem to generate more YL compared to the freestanding ones. All samples display very similar YL bands, confirming that this band is the most dominating defect band despite varying sample growth and composition properties. For the HVPE Si and MBE grown samples, these require different excitation mechanics to be properly resolved. It was found that for the MR samples, those with AlGaN on top had an overall higher PL intensity, while the opposite was the case for the PC samples.

5.5.4 Passivation of surface states

The interaction of surface states in the optical properties of GaN is investigated here. It is relevant to consider surface states as a possible candidate for non-radiative recombinations. This effect is shown in figure 5.18(a), where the MBE 30-11 sample is shown before and after passivation. The PL intensity was increased by a factor 1.32 except for the RL band, proving the effect of the passivation procedure, and also indicating that the RL band saturated before passivation. The sample was then rinsed in water in order to remove the passivating layer, and the removal was confirmed by observation of an identical PL spectrum to that prior to passivation.



(a) (b) (c)

FIGURE 5.18: *RT PL spectra comparing the influence of surface states on the optical properties.* ***a****: shows the passivation effect on the MBE 30-11 sample. The spectra are taken using geometry 1.* ***b****: shows the passivation effect on the MR 1 and MR 2 samples. For MR 2 the passivation had no effect. The spectra here are taken using geometry 2.* ***c****: shows the effect of varying ambient conditions on the UID freestanding GaN sample. Spectra are taken using geometry 4.*

The PL intensity for the MR 1 and MR 2 samples was found to differ a lot in RT PL measurements. MR 1 is a bare GaN sample, whereas the MR 2 sample has an AlGaN layer on top. Therefore it was investigated, whether surface states on the GaN layer could be one cause of this large PL intensity difference. The results are shown in figure 5.18(b), obtained by doing three measurements for each spectrum with subsequent removal and repositioning of the sample for statistical purposes (a slight error is expected depending on the sample position in

the optical geometry). The PL intensity for the MR 1 sample increased by a factor 2.83 after passivation, while the intensity for the MR 2 sample did not change. This is expected, since the band gap energy of $Al_{0.15}Ga_{0.85}N$ is larger than the excitation energy, inhibiting excitation and subsequent trapping of carriers at the AlGaN surface. It is indeed possible that the AlGaN layer has a passivating effect on the GaN epilayer, which could explain to some extent the large difference in PL intensity for the GaN layers in the two samples.

A final consideration of this section is that the PL intensity increased by a factor 1.5 when the bulk UID GaN sample was placed in vacuum conversed to air at 1 atm, the result is shown in figure 5.18(c). A similar increase of PL magnitude was observed by Reshchikov and Morkoc [6]. They further saw a similar effect by placing a sample in e.g. hydrogen or nitrogen ambient compared to air and oxygen atmospheres. They suggested two possibilities: either there is an increase in band bending near the surface due to oxygen adsorbates, or the oxygen adsorbates act as nonradiative recombination sites for excited carriers.

5.5.5 Cathodeluminescence

Cathodeluminescence (CL) was used mainly as a second choice method to confirm the absence of PL peaks on many of the MBE grown samples, but no luminescence was obtained from CL either.

CL was also used to investigate depth profiling in the MR 1-3 samples. The spectra are shown in figure 5.19, where the electron acceleration voltage was varied for each sample from 5 keV to 20 keV corresponding to a depth from 0 to ∼160 nm and 0 to ∼1600 nm, respectively. Remarkably, these samples all displayed characteristic BL bands, which were not seen in any RT PL measurements. As the probing depth was increased, a red-shift of the NBE maxima was seen with a shift of MR 1: 19.8 meV, MR 2: 6.6 meV, and MR 3: 13.3 meV. The shift could indicate that the deeper sample layers are more strained, possibly due to local strain fields from an increased density of structural defects. The red-shift of the NBE maxima with increased depth could explain some of the outliers seen in figure 5.16, since PL only probes the top ∼200 nm, whereas the depth resolution of the Raman setup was found to be around 1 μm. Indeed the three MR samples lie on the high energy side of the fit, and stress variation with depth is thus a plausible explanation for these outliers. MR 2 and 3 varies less with the depth, probably due to larger strain field from the AlGaN/GaN interface. Note also the $Al_{0.3}Ga_{0.7}N$ peak for the MR 3 sample around 300 nm, which was not seen in PL excitation with the Nd:YAG laser.

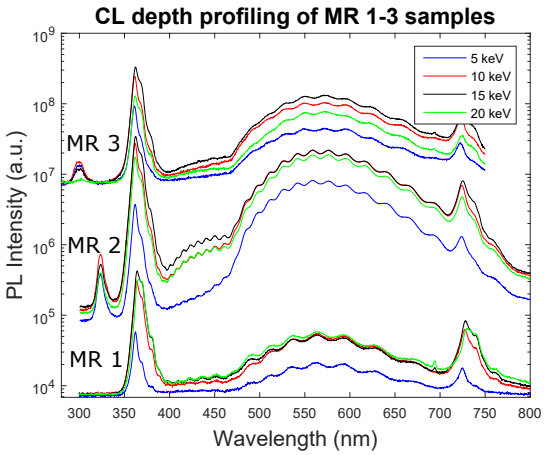


Figure 5.19: *RT CL spectra of the MR 1-3 samples at different electron acceleration voltages.*

It was found that increasing the probing depth resulted in an overall increase of the intensity of the luminescence spectrum. The ratio of YL:NBE luminescence also increased slightly with an increase in depth, suggesting an increased density of impurities as well.

5.5.6 Temperature variation

PL spectroscopy with varying temperature was used to account for the temperature dependency of most defect related bands, and also to resolve bands that had quenched at RT. In order to link substantial properties of a luminescence band to a specific defect, it is generally necessary to investigate the band in excitation dependent, time-resolved, and temperature dependent spectroscopy. In this thesis only the temperature was varied, and therefore the identification of the origin of luminescence characteristics will in most cases be assisted by comparison with the in-depth studies of other authors. The spectra for all samples investigated in this section are shown at T =21K in figure 5.20. The most noticeable changes with respect to RT are the appearance of more defect bands, and an overall large increase in luminescence intensity. The bands and their luminescence intensity are investigated with respect to temperature below.

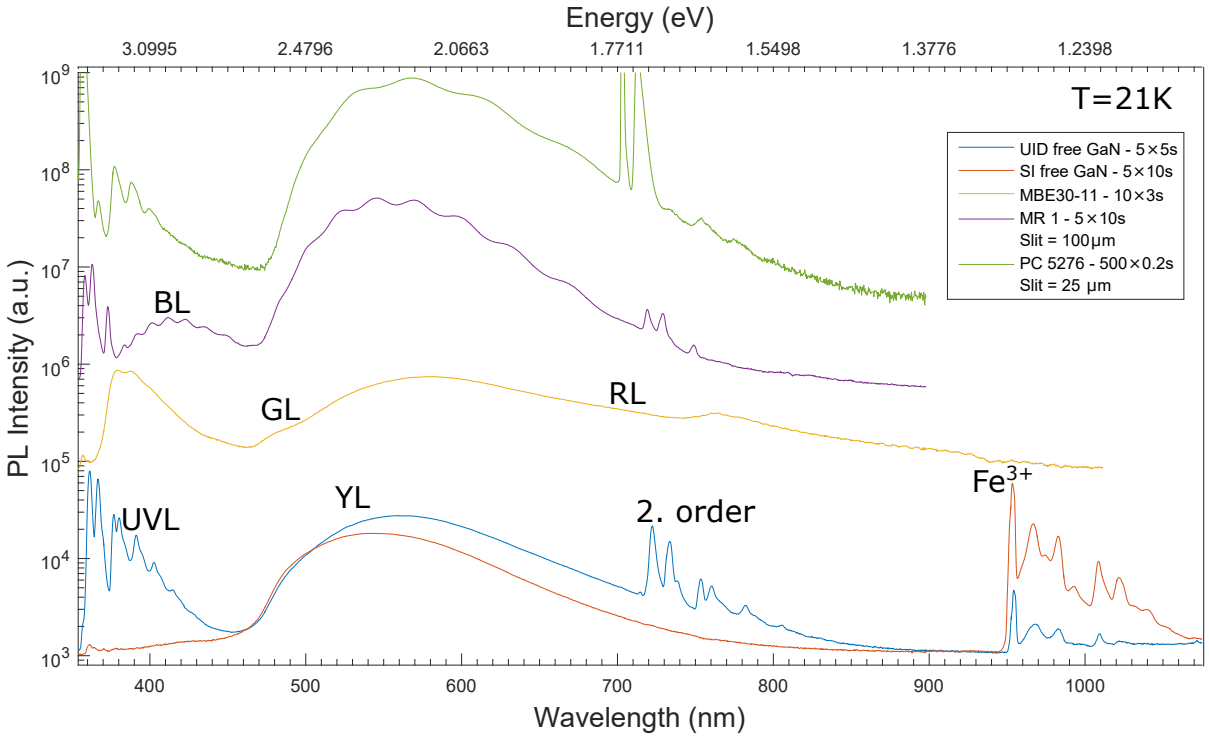


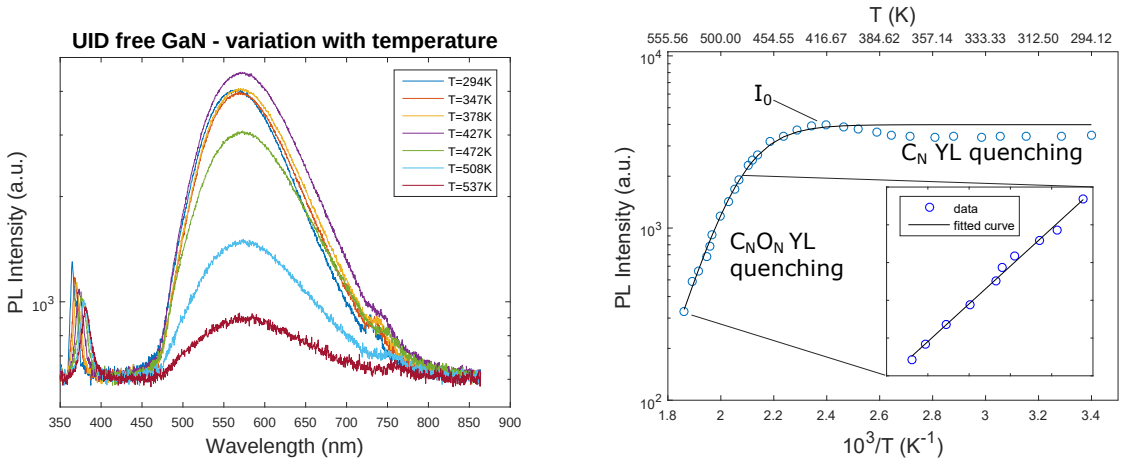
FIGURE 5.20: *Low temperature PL spectra for all samples that were cooled. Several defect bands are observed and labelled. The PL spectrum for the PC5276 sample was collected without the laser line filter, because it was possible to obtain more PL signal with respect to the laser line intensity for this sample, requiring a smaller slit. The spectra are shifted vertically for clarity.*

YL band

When considering the YL band, the intensity dependent behaviour of the supposed C_N defect must be considered for bulk GaN samples. Geometry 3 was therefore used for excitation of the UID freestanding GaN sample, while the temperature was gradually increased through a resistively heated sample holder. In this manner the thermal quenching of the YL band could be investigated. This is shown in figure 5.21. As the temperature was increased from 294K

to 380K the YL band remained unchanged. Then its band peak intensity increased, which could be explained by the quenching of the C_N defect, because this would reduce the amount of available defect traps and thus increase the amount of holes available for the C_NO_N complex to trap (these are the two defects claimed to be responsible for the YL band). This is somewhat consistent with the findings of Reshchikov et al. [38], who used temperature varied and time-resolved PL and estimated that the YL/GL band due to the C_N defect quenches and completely vanishes in this temperature range. This is also consistent with the findings here, where the YL band continues its constant maxima after the supposed C_N quenching. The position of the band maximum varies within $+/-2.5$ nm across the entire temperature range, while the NBE band is redshifted by 147 meV from 294K to 537K. This energy independence with temperature is a typical behaviour for defects with a strong electron-phonon coupling [96]. This can be explained through the CC model in section 2.1.4. If it assumed that the CC model ground and excited state curves do not change with temperature, then the defect is localized enough to be independent of the thermal change of the lattice environment with varying temperature. The variation of the energy position of the band maximum is thus only due to local variations of electronic ground and excited state phonon modes with temperature.

From the quenching of the YL band shown in figure 5.21(b), a fit (the fitting method is described in appendix D.3.6) yielded an activation energy of E_A =861 meV, using the value of I_0 to be the intensity right after the C_N YL band had completely quenched (note that in this context it is assumed that the C_N defect only captures one hole and the band is thus yellow, not green). This value of E_A is in good agreement with values reported elsewhere [34, 38]. In this particular figure the complete fit is also shown to demonstrate, where the value I_0 lies and how the actual fit looks like. The complete fit is henceforth omitted, since it would require time-resolved PL to obtain the involved constants (they were used as fitting parameters in this example) in the denominator of the fitting expression.



(a) (b)

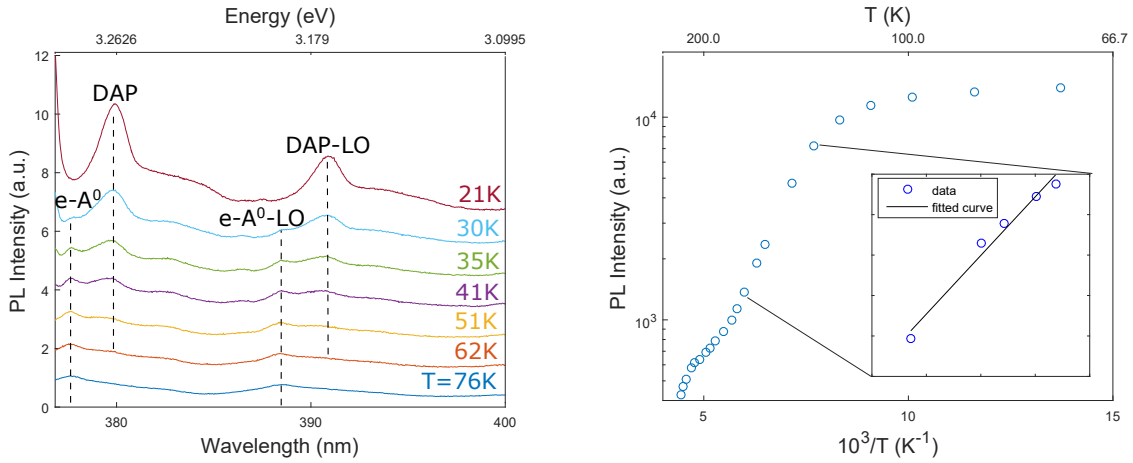
Figure 5.21: *Temperature-induced intensity variation of the YL band of the UID freestanding GaN sample.* **a:** *shows the PL spectra obtained at different temperatures.* **b:** *shows the variation of the YL band maximum with temperature. The region of quenching is shown, and from it a fit yields an activation energy of* $E_A = 861\,meV$.

The behaviour of the YL band during cooling was consistent for all samples, namely that the band intensity gradually increased with decreasing temperature. This increase could be

due quenching of non-radiative channels, which would lead to an increase of all other bands by making more holes available for the remaining bands, or due to an increase in the hole capture coefficient.

UVL band

The UVL band was first investigated for the UID freestanding GaN sample, since it displayed the sharpest features in this region compared to other samples (see figure 5.20). At elevated temperatures, the band is supposedly generated by an Mg acceptor [6]. As the temperature lowers, shallow donors become deionized, allowing for DAP transitions. This behaviour was also observed here and is shown in figure 5.22(a). Also shown are the first LO replicas displaced by 91.7 meV with respect to their main peaks. Generally, several LO replicas were observable at low temperature, indicating good long range crystalline order. The e-A^0 peak energy lies 192.8 meV from the main exciton line ((D^0,X_A), not shown in the figure). The energy separation between the e-A^0 and DAP band maxima is 17.3 meV. Reshchikov and Morkoc [6] obtained a larger separation, but presumably their excitation density was lower compared to the one used here. Larger excitation density is expected to blue-shift a DAP band maximum, because the probability of exciting acceptors and donors with shorter spatial separation increases, which in turn increases the attractive Coulomb potential between the two impurities, as argued in section 2.1.3. Another parameter of importance is the concentration of donors involved in the DAP transition. If this parameter varies between samples, it also affects the effective donor-acceptor distance affecting the effective Coulomb potential. Having knowledge of the donor concentration would allow for calculation of the effective Coulomb potential, and hence make it possible to estimate the ionization energy of the donor involved in the DAP transition.



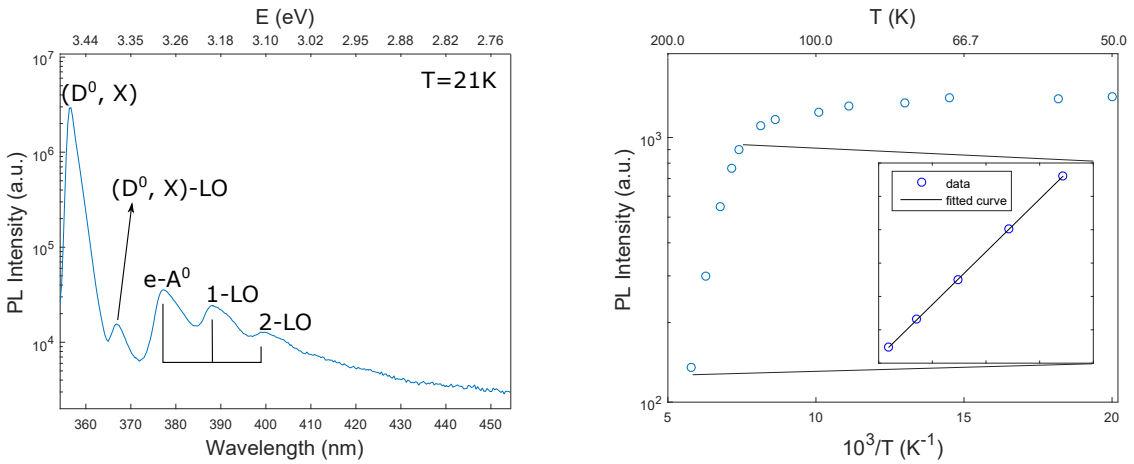
(a) (b)

FIGURE 5.22: **a**: *High resolution PL spectra at varying temperatures obtained by excitation of the free UID GaN sample by the focused Ar-ion laser using geometry 4. Shown here is the UVL band, and how it transforms from an e-A^0 transition to a DAP transition as the temperature decreases.* **b**: *shows the temperature dependence of the e-A^0 related UVL band. The region of quenching is shown, and from it a fit yields an activation energy of $E_A = 119$ meV.*

Also studied here is the influence of temperature on the e-A^0 transition of the UVL band. This is shown in figure 5.22(b), where the fit results in an activation energy of $E_A = 119$ meV. The quenching temperature region matches the observations made by other authors [34, 38], however, they estimated activation energies in the range 170-180 meV. This is unlikely due to

difference in acceptor species, as no common acceptor in n-type GaN has this low activation energy. The discrepancy could possibly be reduced if more measurements were made in the given temperature range. The result did not change by normalizing each spectra with respect to the YL band maxima, ruling out the possibility of geometrical changes with temperature. A different possibility is that the temperature measurement during cooling was inadequate and had a non-linear relationship to the actual temperature, e.g. if the thermocouple was not properly connected to the sample stage as was sometimes the case.

The PC 5276 sample also displayed a clear UVL band. A closer view of its UVL band at 21K is shown in figure 5.23(a). This spectrum showed no sign of a transformation from a defect band to a DAP band. The e-A^0 peak energy of 3.285 eV for this sample is very close to the e-A^0 peak energy of 3.282 for the bulk UID. The energy difference with respect to the (D^0,X) peak is 190.6 meV, which is slightly smaller compared to the bulk UID GaN sample. If strain and doping induced band gap shrinkage are the cause of the variation in e-A^0 peak position between the two samples, it would still be expected that the distance between the (D^0,X) and e-A^0 peaks would be equal for the two samples. Since it is not equal, there may be a small difference in activation energy for the same defect in the two samples. Also noteworthy, is the large intensity difference between the (D^0,X)-LO relative to its ZPL compared to the e-A^0-LOs relative to their ZPL, indicating a much stronger phonon coupling for the e-A^0 defect.



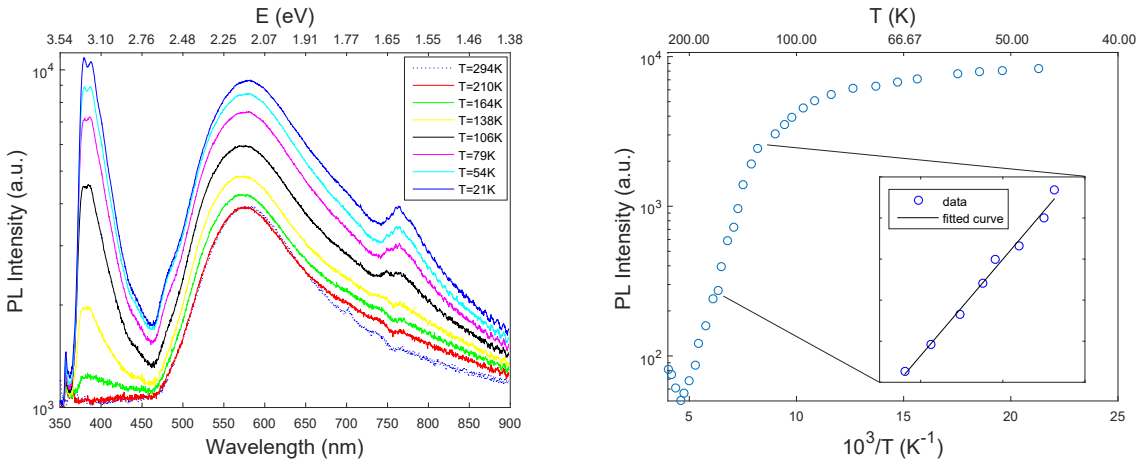
(a) (b)

Figure 5.23: ***a**: Low temperature PL spectra obtained by excitation of the MOCVD grown PC 5276 GaN:Si sample by the focused Ar-ion laser using geometry 4 without the laser line filter. **b**: shows the temperature dependence of the e-A^0 related UVL band. The region of quenching is shown, and from it a fit yields an activation energy of $E_A = 136$ meV.*

The temperature dependence for the e-A^0 peak of this UVL band was also investigated. Its intensity variation with temperature is shown in figure 5.23(b), and a linear fit in the quenching region yields an activation energy of $E_A = 136$ meV. Here a much better fit was obtained compared to that of the bulk UID GaN. The value of E_A obtained from the quenching of the band is somewhat lower than the energy separation between the (D^0,X) and e-A^0 ZPLs of 187 meV. Foremost, this difference can be explained from the temperature independence of the e-A^0 peak (see e.g. figure 5.22(a)), conversed to the temperature dependence of the band gap energy which for the PC 5276 sample was found to decrease by 30 meV from 21 K to 200 K, where the UVL band becomes totally quenched. If the temperature dependencies of the valence band DoS and hole capture coefficient in eq. 3.17 are neglected, the fit used in the quenching

region of figure 5.23(b) is a pure Arrhenius and yields an energy of 164 meV, matching the energy separation of the (D^0,X) and e-A^0 ZPL at 200 K. This illustrates why the temperature dependence of the coefficients in the activation energy should always be considered, since the pure Arrhenius would lead to an overestimation of the activation energy.

The reason why the e-A^0 peak did not transform into a DAP transition at low temperatures is most likely related to the availability of donors in the PC 5276 sample, which was found to be large from Hall measurements. Two shallow donors are usually considered in GaN, namely Si_{Ga} and O_N with ionization energies of 30.8 meV and 32.4 meV, respectively. Reshchikov and Morkoc [6] found an ionization energy for the shallow donor involved in the UVL DAP matching that of the O_N ionization energy. However, since this sample is Si doped, it is natural to consider Si as a candidate as well. The effect of having a much larger concentration of donors would increase the availability of donors for DAP transitions and also enhance the Coulomb potential due to shorter D-A distances. This would be expected to cause a blueshift of the DAP ZPL. Furthermore, Si_{Ga} has a lower ionization energy than O_N, meaning that if it is the Si_{Ga} donor that should yield the DAP transition, its peak energy would lie closer to the acceptor energy compared to O_N, and even more so if the concentration of Si dopants is much larger than O dopants. So one possibility is that the DAP transition energy is actually larger than the e-A transition energy. Another explanation for the lack of a DAP transition is simply that the Si concentration is high enough and the donor atom distance small enough that the wave functions overlap, creating a defect band accompanied by a decrease in ionization energy. Thus, it is possible that the temperature of 21K is simply not small enough to freeze out a significant amount of Si atoms. A third possibility, which has O as the shallow donor, is that an increase in Si concentration also increases the formation energy for O, which results in a low O concentration for DAP formation.



(a) (b)

FIGURE 5.24: ***a**: Low temperature PL spectra obtained by excitation of the MBE 30-11 GaN sample by the focused Ar-ion laser using geometry 4. **b**: shows the temperature dependence of the UVL band. The region of quenching is shown, and from it a fit yields an activation energy of $E_A = 101\,meV$.*

The MBE grown sample investigated by cooled PL displayed a prominent broad band in the UVL region of the spectrum. How this band varies with temperature is shown in figure 5.24. The band appears below $T \approx$164K similarly to the other samples. The activation energy is found to be 101 meV, which is significantly lower compared to the UID and PC 5276 samples. The band is seen to transform into two sharper peaks at low temperature with maxima at 3.27

eV and 3.20 eV for $T = 21$K. The peak at 3.27 eV could indeed correspond to the e-A peak position seen for the UID. The other peak, however, matches neither the DAP or any LO replica. Assuming that it is unrelated to the e-A UVL peak, the variation of this peak with temperature could likely disrupt the the fit made in figure 5.24(b), and thus explain the discrepancy in the value of E_A obtained here. The origin of the 3.20 eV peak remains unknown.

No UVL band was observed on neither the freestanding SI nor the MR 1 GaN samples. The SI GaN sample has been made semi insulating by Fe doping, which could kill other recombination channels either by capturing their holes, or hindering the defect formation in the first place. It is likely that this could explain the absence of the UVL band on this sample. The MR 1 sample was the only sample displaying a prominent BL band. Since no doping is expected for this sample, a plausible reason could be that the defect responsible for the BL band in this sample minimizes transitions related to the defect of the UVL band. This could e.g. be due to a much larger hole capture cross section for the BL band defect compared to the UVL band defect.

BL band

The cooled MOCVD grown MR 1 sample displayed a very prominent BL band peaking around 2.9 eV. The effect of varying temperature on this band is shown in figure 5.25. The red and blue data points are refer to two different PL accumulation conditions, differing because it was necessary to change the accumulation conditions as the temperature was decreasing. The quenching of the BL band always takes place above 200 K in accordance with findings of others [6]. The blue data points were accumulated below 200 K, and therefore they represent the onset of the band quenching rather than the linear quenching region. Thus the fit was only be made for three points and could thus suffer from statistical inaccuracy. Nevertheless, the activation energy was found to be E_A =299 meV, which is slightly lower than the values reported (340 to 380 meV [6]) for undoped GaN samples containing the BL band. The best fitting acceptor for this activation energy would be Zn_{Ga} or Ca_{Ga}, where Zn is a natural candidate since it exists as a contaminant in gallium source materials.

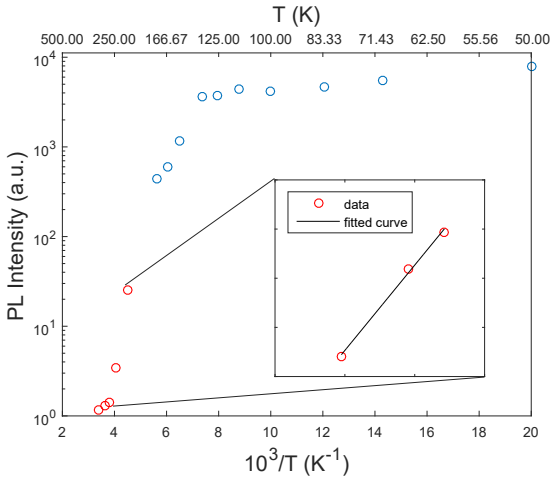


Figure 5.25: *Intensity variation with temperature for the BL band on the MOCVD grown MR 1 sample. Blue and red dots are data collected under different accumulation conditions, but corrected to match the same intensity scale. The activation energy was found to be E_A = 299 meV.*

GL and RL bands

The GL and RL bands can be seen on several of the samples, and also e.g. the variation with temperature seen in figure 5.24(a) for the RL band is clearly different from the YL band. Nevertheless, these bands are largely disguised by the YL band, and any meaningful investigation of the GL and RL bands would require an artificial removal of the YL band, so that it could be removed from the spectra. Other experimental methods could also be applied, e.g. time-resolved PL spectroscopy and excitation dependent spectroscopy, since it is expected that PL lifetime differ from band to band and also the RL band often saturates at very low excitation densities. These bands were not studied further in this thesis.

Fe band

A number of sharp lines followed by a series of broader peaks can be seen in the NIR part of the spectrum in figure 5.20, and these lines are attributed to Fe defects. These peaks were only observed for the two freestanding GaN samples, where the SI one is made semi insulating through Fe doping. Therefore, the SI free GaN sample was investigated further with respect to these peaks. Figure 5.26 shows the temperature variation for these peaks, where the transformation from RT to 21K is shown in **a**, and a few high resolution spectra are shown in **b**. The disturbance caused by the filter has been removed from these figures (see e.g. the transmittance spectrum in figure D.8). Note also that the sharp line around 1.31 eV in figure 5.26(a) actually consists four peaks, only resolvable with a high resolution grating. The four peaks are shown for a few different temperatures, where it can be seen that the peaks quickly broadened with increasing temperature, which at some point makes two of the peaks merge. The sharpest line is located at 1.299 eV and the three other lines were all also observed by Dumcenco el. al. [97], who assigned the sharpest line to the ZPL to an internal transition of an Fe_{Ga} deep acceptor in its 3+ charge state. The following three lines lying on the high energy side of the ZPL were attributed to transitions from states due to splitting of the state responsible for the ZPL transition.

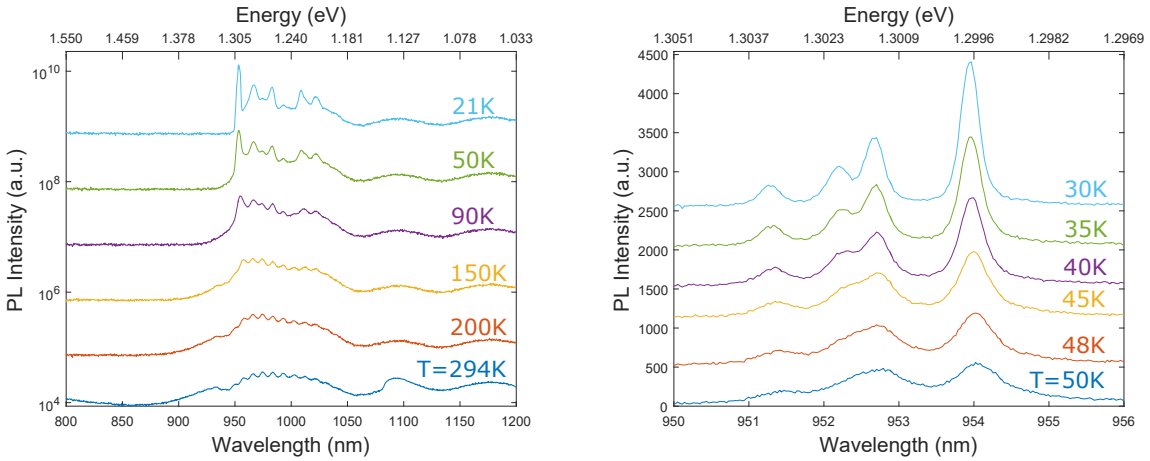


FIGURE 5.26: *Temperature varied PL spectra of the SI freestanding GaN sample obtained with* **a***: the 150 l/mm grating,* **b***: the 1200 l/mm grating. The disturbance caused by the filter transmittance spectrum has been removed.*

Excitons

In order to study the NBE PL at low temperatures, it was found necessary for all samples to insert the filter described in D.8, except for sample PC 5276, which was luminescent enough that the NBE could be distinguished from the laser line. However, this sample had a very broad exciton related peak at low temperature, and therefore only one peak could be resolved for it (the peak is shown in figure 5.23(a)). This could either be related to the Si doping or the use of a larger spectrometer slit.

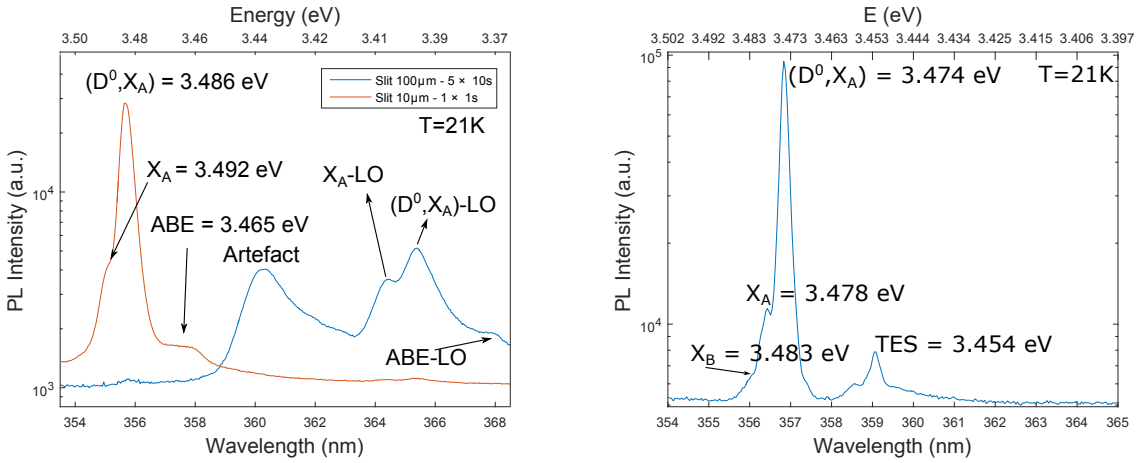


Figure 5.27: *Low-temperature PL spectra of* **a***: MR 1 and* **b***: sample UID free GaN.*

The largest exciton peaks in figures 5.27(a) of the MR 1 sample and 5.27(b) of the UID free GaN sample have been assigned to the shallow donor bound A exciton, with the shallow donor believed to be either Si_{Ga} or O_N, or a combination of the two (the ionization energies of these are $\sim$ 30.8 meV and $\sim$ 32.4 meV, respectively). According to Hayne's rule (see section 2.1.3), and using $b = 0.14$ and $a = 0$ as the fitting parameters (due to Reshchikov and Morkoc [6]), the difference in E_{DBX} for Si_{Ga} and O_N is ~ 0.22 meV, and thus it is obvious that it cannot be resolved here. It should be noted, however, that the fitting parameters used in Hayne's rule vary much in the literature. In general, comparison with PL spectra with extremely high resolution ($>$0.1meV) of other authors show that the excitonic peaks resolved here likely consist of several individual peaks, which are not resolved. This is both due to the temperature employed here, and due to lack of spectral resolution.

In figure 5.27(a), the red spectrum was collected without the filter, while the blue was collected with the filter. This shows that the first peak observed with the filter inserted is merely the tail of the much stronger peaks lying below the filter cut-off. Three exciton peaks are observed, assigned to transitions due to a free A exciton, a shallow donor bound exciton, and an acceptor bound exciton, respectively, along with their first phonon replicas. The position of the ABE peak is difficult to make out exactly, but applying Hayne's rule with the fitting parameters $a = 0$ and $b = 0.06$ (due to Reshchikov and Morkoc [6]), and using the activation energy for the blue band found above, the energy difference between X_A and ABE should be $\sim$ 18meV, which is somewhat different from the observed peak difference. It is noted that the phonon coupling of the ABE seems to be much stronger than that of the DBE, as evidenced from the relative sizes of their 1-LO phonon peaks, which is in agreement with the finding of Reshchikov and Morkoc [6].

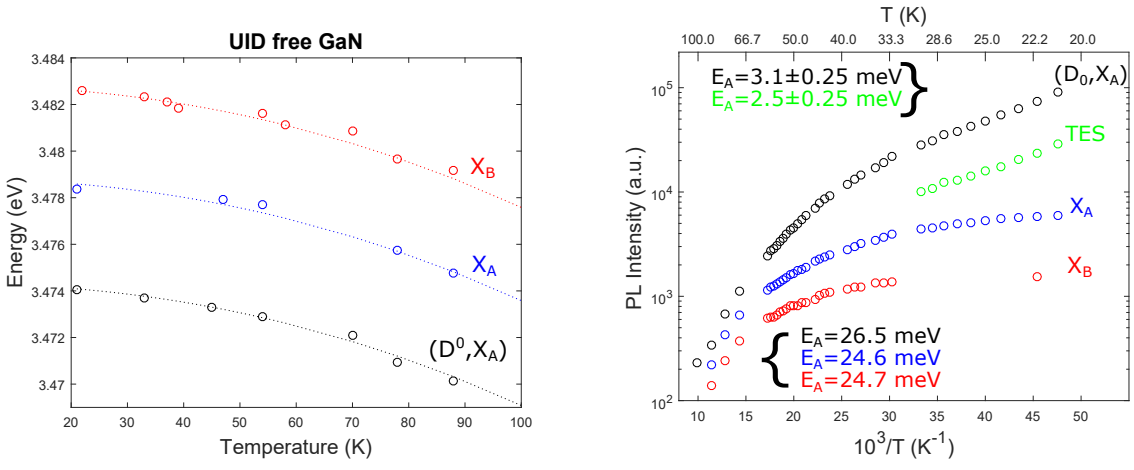
In figure 5.27(b), some notable differences are observed compared to 5.27(a). First, a smaller width of the exciton lines indicate much higher structural quality. Secondly, no ABE peak is expected to be seen on the UID free GaN. This is because no BL was seen on this sample, since it has been argued to be related to the same acceptor as the ABE transition. For both samples the position of the (D^0,X_A) and X_A peaks are shifted because of reduced strain. The relative positions of the peaks for the two samples differ by 2 meV, likely due to inaccuracy in determination of peak positions on the MR 1 sample, due to wider peaks. According to Hayne's rule above, the energy difference between (D^0,X_A) and X_A should be $\sim$ 4.4meV, which is within experimental error of the observed peak difference, if the (D^0,X_A) peak is due to a combination of Si and O. Also identified on the UID free GaN sample is the free B exciton, the position of which is in excellent agreement with the findings of Reynolds et al. [98].

While no perfect argument could be found for the appearance of the lines around $\sim$ 3.454 eV in figure 5.27(b), a likely possibility is that they are the so-called two-electron satellite (TES) transitions, which can occur if an exciton bound to a neutral donor leaves the donor in an excited state, following the description by Freitas et al. [99]. The energy difference between the (D^0,X_A) peak and the TES peak is thus the difference between the ground and excited states of the donor. While the resolution used here does not allow discerning individual TES lines, the relatively broad peak between $\sim$ 16 and $\sim$ 31 meV (roughly the binding energy of the donor electron) could be due to this.

In order to further investigate the properties of the excitons involved in the NBE PL spectrum, the dependence of the photon energy of the peaks on the UID free GaN sample with temperature is shown in figure 5.28(a). Here, the same kind of dependence is found for the three exciton peaks, with the dotted lines being fitted to the Varshni equation

$$E_0(T) = E_0 - \frac{\alpha T^2}{\beta - T} \tag{5.6}$$

where E_0 is the expected energy at $T = 0$K. These relatively good fits indicate that the peaks are indeed due to excitons, as they have the same temperature dependence as the band gap, and are in excellent agreement with the results of Reynolds et al. [98].



(a) (b)

FIGURE 5.28: *Temperature dependencies for the UID freestanding GaN sample.* **a***: shows the temperature dependence of the photon energy for the three excitons.* **b***: shows the temperature dependence of the intensity for all observed exciton peaks. The TES has been vertically shifted by a factor of 10 for easier comparison.*

For the PC 5276 sample NBE luminescence was obtained from RT to 21 K. If the Varshni equation is fitted to the NBE energy shift with temperature a poor correspondence is seen. However, as previously suggested the NBE maximum on this sample is further shifted due to band gap shrinkage, which is also temperature dependent. Sarkar and Ghosh [100] saw a similar trend and therefore they combined the temperature dependence of the band gap shrinkage with the Varshni equation and found a much better fit.

Figure 5.28(b) shows the temperature dependence of the PL intensity for the all the observed exciton peaks on sample UID free GaN. First, a clear difference is seen between the free excitons compared to the donor bound and TES excitons in the temperature region below $T = 30$K. Here the free excitons have little variation with intensity, but the others seem to have an exponential increase with lowering temperature. As the shallow donor is expected to experience the onset of thermal quenching as low as $T = 5$K, it is expected from this data that the donor is already quenching at the temperatures between 20K and 30K. The activation energies for this donor and the TES were calculated in the usual manner, which means that it is assumed that the exciton capture coefficient of the donor has the same temperature dependence as the carrier capture coefficients from earlier. The fact that these energies are very close indicate the correct assignment of the TES peak, as these would certainly be expected to quench along with the shallow donor responsible for them. The activation energy is somewhat lower than the ~ 4meV difference between the peak positions of the free A exciton and the (D_0,X_A). If it is assumed that free excitons lose their kinetic energy upon being bound to the donor atom, thermal energy is lost to the lattice and not included in the (D^0,X_A) transition (but it is included for X_A), which causes the exciton peaks to be further apart by $k_BT/2 \approx 0.9$ meV. Adding this to the activation energy for the shallow donor, the peak positions fit well.

The other temperature range with a pronounced quenching was between 70K and 100K. For the fits yielding the activation energies found here, the temperature dependence of the carrier capture coefficients was omitted, since the exciton is a free particle and is thus not "captured". Similar temperature dependence on the intensity for all the excitons in this region indicates that it is the same type of quenching, namely exciton dissociation. The reason why E_A is slightly higher for the donor bound exciton is most likely that the shallow donors are not yet completely ionized, and their quenching adds a small part to the slope in this region. Note also that the binding energies obtained here are in reasonable agreement with the ones calculated from the band structure in section 3.1, whereas slight discrepancy is found with respect to the value of 26.3 meV obtained experimentally by Volm et al. [101].

It has been found that the binding energies of X_A and X_B are equal within experimental error. Figure 5.29 illustrates simply why this is the case, and why the valence splitting due to spin-orbit coupling results in the only difference in peak position between the excitons.

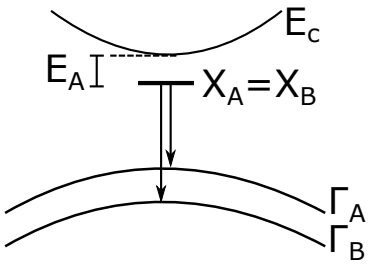


Figure 5.29: *Scheme of a plausible explanation for the activation energies for X_A and X_B.*

By considering in figure 5.28(b) that the excitons are all in the quenching region below 100K, it would seem that at room temperature some exciton transitions are still probable, but that judging from the NBE intensity generally seen in RT measurements, it is likely combined with a large degree of e-h recombinations.

5.6 Discussion

One of the main purposes with this thesis was the investigation of how different characterization methods could be applied to studying the properties of GaN thin films. In this vein it is natural to discuss the advantages and disadvantages of these methods.

AFM was the method of choice for surface characterization, as the surface roughness and uniformity are very important for GaN based devices. Furthermore, information regarding growth patterns are often reflected in the surface morphology. However, the optical and structural properties are not necessarily mirrored in the surface as was e.g. seen from the difference between samples MBE 30-11 and MBE 22-03. The former was very rough compared to the latter, but when considering structural qualities from Raman spectroscopy, it was evidenced that the MBE 30-11 sample was vastly superior. A similar example can be given for the MR 1 and PC 5274 samples, where the much rougher sample had similar FWHM of the Raman peak compared to the smoother one. Therefore, caution should be taken when considering the surface morphology as a measure of sample quality, and it generally requires knowledge regarding the growth parameters, if meaningful information about sample quality is to be extracted from surface morphology. Threading dislocations extending to the surface were observed frequently by other authors. These were not readily found with AFM during this thesis, and it would thus be relevant to ascertain that the obtained AFM images do indeed show every detail of the surface. This could be done e.g. through other scanning probe techniques or TEM.

Hall measurements yield very concrete results, and the reliability of the Hall setup is therefore of utmost importance. There is no doubt that the Hall setup used in this thesis can be improved for more reliable electrical characterization. Firstly, the use of a well defined calibration sample would give an idea regarding the error margin. Secondly, a more statistical collection of data, e.g. by doing a series of measurements with subsequent removal and deposition of contacts could improve reproducibility, and employment of appropriate van der Pauw patterns could drastically limit the error margin. Thirdly, establishment of more accurate means to measure the magnetic field along with higher quality magnets would greatly improve the reliability of the results. Nevertheless, while the Hall parameters obtained here require caution when compared to the findings of others, the results are still conclusive with respect to each other. When considering devices such as HEMT structures, Hall measurements are the most definite way to characterize the quality of a 2DEG. Therefore, the above mentioned improvements are advantageous to ensure high reliability of the Hall setup.

No obvious correspondence was observed between the measured Hall mobility and the stress analysis from Raman spectroscopy. The strain measured due to a shift of the E_2^H mode was difficult to attribute to an improved sample quality, as it was argued that dislocations can shift the peak in the opposite direction, thereby minimizing the total biaxial stress in the sample. It is therefore important to consider the width of the Raman peak, but this quantity must also be treated with caution. While it is easily used to discern between high and low quality samples, individual comparison between two higher quality samples is difficult. It is likely

that a higher resolution could provide more accuracy regarding the obtained FWHM values, allowing the distinguishing of samples with comparable peak widths. In order to increase the depth resolution, two-phonon and resonant Raman measurements could also have been applied, and especially resonant Raman measurements would have enabled more local probing of strain and structural quality near the 2DEG interface. Nevertheless, the ease of use of conventional Raman spectroscopy is advantageous when it comes to the quick first-hand identification of the GaN films. Additionally, it was shown how the behaviour of the A_1(LO) mode can be used to analyse carrier concentration. For identification purposes, Raman spectroscopy is beautifully complimented by XRD. This technique readily confirms the crystalline phase of the GaN layer. By using a rocking curve geometry, the type of structural defects present can be further investigated.

RT PL results seemed to provide equivalent conclusions regarding sample quality, when compared to Hall measurements. RT PL generally only provided two characteristic measures; PL generation efficiency and YL:NBE ratio. A general trend was observed that higher mobility equalled more efficient PL generation, along with an improved YL:NBE ratio. While conclusions regarding the PL intensity from a sample must be made with caution, it can be used in conjunction with other methods to indicate the overall quality of the GaN film. The PL generating efficiency is not only important for optical devices, it is also related to crystal structure and sample quality. The YL:NBE ratio was found to differ greatly between samples, suggesting variations in the probability for radiative transitions from defects, a property that evidently would affect the mobility of a sample. Obviously, any comparison of YL:NBE ratios must be carefully accounted for, because the ratio was found to depend on the excitation density. Thus for PL spectroscopy, it is a necessity to ensure a high reproducibility of the measurements. This reproducibility was readily found to be higher, when the laser was not focused onto the samples. Then, even for slight variations in the PL accumulation, the ratio would not be affected, due to preservation of the excitation density. Another consideration is the choice of excitation source. Here an Ar-ion laser was used, which posed some difficulties regarding the separation of the laser line with respect to the NBE luminescence. This problem was especially severe for cooled PL, and it could have been circumvented by using e.g. another excitation source, or a suitable long-pass filter with a sharper cut-off.

It was argued how the total PL intensity is an important property to investigate. In this respect, it was found from surface passivation that especially for samples containing e.g. an AlGaN layer, relative sample comparison might be difficult. This is especially true, since the measurements provided here did not reveal anything about the total effect of surface states on the optical properties. Several passivation attempts are available in the literature, with the most promising one by far being that by Song et al. [102]. They used a CH_3CSNH_2 solution for surface passivation of GaN and found that their PL intensity increased by a factor 100, and moreover found that it reduced contact resistance four orders of magnitude and promote the formation ohmic Ti/Al contacts to GaN. Their study shows that surface states can have an even greater influence on the optical properties of GaN than what was seen here. Furthermore, the study given here also shows that the degree of the passivating effect could be sample dependent, but in order to properly verify this, further study is required.

In order to investigate the origin of luminescent scattering centers in the films, temperature varied PL was applied. By analysing several different samples, the activation energy for some commonly observed defect bands was accounted for. While this provides some insight into the

nature of the defects by comparison with the literature, it is obviously of importance to analyse the concentrations of different defects. This can be done if the lifetime of the defect in the excited state is found through time-resolved PL, which can be used to find the hole capture coefficient for the acceptor. By comparison with the dependence of the PL intensity of the excitation density, it is then possible the calculate the concentration of different types of radiative defects, and this method has the advantage compared to e.g. SIMS of being completely non-destructive, as well as relatively simple once an appropriate setup has been built. These experiments are an obvious candidate for further sample analysis.

The observation of individual, relatively narrow exciton lines from the UID free GaN sample is evidence of high crystalline quality. Furthermore, from the activation energies and the nature of the defect bands, assignments of the chemical origin for these defects were generally suggested for the samples. Knowing the chemical origin of a scattering site is necessary if it is attempted to reduce either its quantity or its capture rates, both of which could reduce its influence on the electrical and optical properties of the sample. Investigation of how this can be achieved requires precise control over the growth parameters, along with a statistically significant sample quantity. For the MBE setup, however, it is suggested that before impurity reduction schemes are considered, optimization of the growth scheme is first prioritized. An illustration of this argument is the BL band observed in the MR samples. While this band can readily be related to impure gasses, its incorporation into GaN may vary greatly depending on the growth conditions, and it may be thus be redundant to attempt to reduce this without first finding appropriate growth parameters.

It was found that the produced MBE GaN samples were generally of poor quality. With the exception of sample MBE 30-11, none of them had showed any luminescence, they all had wide Raman peak FWHMs, as well as broad XRD (0002) reflections. This is likely a result of highly N-rich growth conditions and an unstable Ga-source. However, the atomically flat, nitrided substrates and the employment of a nucleation layer seemed to have very advantageous effects on the surface roughness, as expected from theory. As evidenced from section 4, GaN growth is in general a very complicated process, and thus further investigation of the optimal parameters for GaN growth on sapphire with PAMBE will necessarily have to include a much more controlled variation of growth parameters, e.g. of III/V ratio, substrate temperature, and nucleation layer parameters.

Conclusion

6

The aim of this thesis was the investigation of GaN thin films and the imperfections which arise due to heteroepitaxial growth, as these hinder device performance by acting as scattering sites and recombination centers for charge carriers. Defects responsible for deep-lying impurity levels introduced unintentionally during growth have been found to be a major obstacle in this regard, which was studied in this thesis through photoluminescence.

Many of the PL defect bands that are commonly observed in GaN samples were also seen and studied here, and it was argued that the very broad spectral emission of these defects is due to strong electron-phonon coupling. In this vein, it was found that different localized impurities and complexes thereof can account for most of the luminescence characteristics in GaN. It was argued that the notorious YL band could be due to the C_N substitutional impurity and its complexes, and this assignment fit well with the results obtained here. Other defect bands studied and discussed include UVL, BL, GL, RL, and Fe-induced NIR luminescence. The excitonic transitions in GaN were also studied, with a good overall correspondence between the experimentally obtained binding energies and those obtained from a combination of the exciton theory and the calculated band structure.

Utilizing characterization methods such as Raman spectroscopy, X-ray diffractometry, and atomic force microscopy, many properties of the studied GaN samples were accounted for, particularly residual strain, structural defects, and surface morphology. The films studied here were found to be strained primarily due to TEC mismatch, but also due to structural defects. The growth of wurtzite GaN was achieved by PAMBE, but it was found that these samples suffer both from low structural quality as well as poor surface morphology. Generally, the FWHM of the Raman lines, as well as the total PL efficiency and YL:NBE ratios of room temperature PL, gave a first-hand indication of sample quality. In addition, it was found that an interplay between the different characterization methods was necessary in order to avoid making the wrong conclusions about sample properties, as conventional wisdom from analysis of data obtained by one characterisation method might not agree with that from another.

Hall measurements were carried out to analyse the electron mobility and carrier concentrations. In general, agreement was found between the observed GaN film properties from other methods and the expected electrical properties. Especially the comparison with the PL results showed that electrical and optical properties are highly related. Evidence of the advantageous effects on electron mobility of the 2DEG in HEMT structures was also supplied.

By analysing the plentiful literature on the subject, it was found that several schemes can be employed to improve the quality of the grown GaN films. The appropriate growth parameters are highly dependent on the growth method, as well as the intended purpose of the film. However, it was generally found that for sapphire heteroepitaxy, taking advantage of proper nitridation and nucleation layers is pivotal to achieving high quality films. This was also indicated experimentally, where PAMBE was used to grow several samples, but in order to understand the intricacies of MBE growth of GaN and find optimal growth parameters for GaN buffers in HEMTs, more research is required. In the same vein, a more systematic analysis of samples grown by different methods could elucidate the capabilities of the MBE setup employed here.

It is concluded that, while the characterization methods employed here gave vital insight into the nature of the defects in GaN thin films, the properties of defects in GaN are a complicated matter. Further study utilizing additional characterization schemes could supplement those used here, and this could pave the way for a more in-depth understanding, particularly of the specific nature of the structural defects inherent in the films, as well as the concentrations of the deep luminescence centers.

Bibliography

[1] Hadis Morkoç. *Handbook of Nitride Semicondcutors and Devices: Materials Properties, Physics and Growth.* Wiley, 2008.

[2] Jr-Tai Chen. *MOCVD growth of GaN-based high electron mobility transistor structures.* Doctorate thesis, Linköping University, 2015.

[3] Tyler J. Flack, Bejoy N. Pushpakaran, and Stephen B. Bayne. GaN Technology for Power Electronic Applications: A Review. *Journal of Electronic Materials*, 16(4435):1–10, 2016.

[4] Shinichi Hoshi, Hideyuki Okita, Yoshiaki Morino, and Masanori Itoh. Gallium Nitride High Electron Mobility Transistor (GaN-HEMT) Technology for High Gain and Highly Efficient Power Amplifiers. *Oki Technical Review*, 74(3):90–93, 2007.

[5] J. S. Speck and S. J. Rosner. The role of threading dislocations in the physical properties of GaN and its alloys. *Physica B*, 273-274:24–32, 1999.

[6] Michael A. Reshchikov and Hadis Morkoç. Luminescence properties of defects in GaN. *Applied Physics Reviews*, 97(061301):95, 2005.

[7] Tatau Nishinga and Thomas F. Kuech. *Handbook of Crystal Growth - Thin films and Epitaxy.* Elsevier, 2015.

[8] Ivan Pelant and Jan Valenta. *Luminescence Spectroscopy of Semiconductors.* Oxford university press, 2012.

[9] Rohit P. Prasankumar and Antoinette J. Taylor, editors. *Optical Techniques for Solid State Materials Characterization.* CRC Press, 2012.

[10] Peter Yu and Manuel Cardona. *Fundementals of Semiconductors.* Springer, 2010.

[11] Swati Jain. Photoluminescence study of cadmium zinc telluride. Master's thesis, West Virginia University, 2001.

[12] Thomas Garm Pedersen. *Electric, Optical and Magnetic Properties of Nanostructures.* Aalborg University, 2015.

[13] J. R. Haynes. Experimental proof of the existence of a new electronic complex in silicon. *Physical Review Letters*, 4(7):361–363, 1960.

[14] J. R. Haynes. Experimental observation of the excitonic molecule. *Physical Review Letters*, 17(16):860–863, 1966.

[15] X B Zhang, T Taliercio, S Kolliakos, and P Lefebvre. Influence of electron–phonon interaction on the optical properties of III nitride semiconductors. *Journal of Physics: Condensed Matter*, 13:7053–7074, 2001.

[16] Rudolph Erasmus. *Optical properties and mechanical stress state in cubic boron nitride and diamond*. PhD thesis, University of the Witwatersrand, 2012.

[17] John Ferraro and Kazuo Nakamoto. *Introductory Raman spectroscopy*. Academic press, 1994.

[18] Robert Loudon. The Raman effect in crystals. *Advances in Physics*, 50:7:813–864, 2001.

[19] Willes Weber and Roberto Merlin. *Raman Scattering in Materials Science*. Springer, 2000.

[20] Donald A. Neamen. *Semiconductor physics and devices Basic principles 4th Edition*. McGraw Hill, 2012.

[21] Hong Ye. *Electrical Characterization of GaN:Si and AlGaN:Si*. PhD thesis, Chalmers University of Technology, 2011.

[22] Casey Joe Kirkpatrick. *Gate Stack Design for Threshold Voltage Control of Gallium Nitride Power Transistors*. PhD thesis, North Carolina State University, 2013.

[23] L. J. van der Pauw. A method of measuring the resistivity and Hall coefficient of lamellae of arbitrary shape. *Philips technical review*, 20:220–224, 1958.

[24] Keithley Instruments Inc. Performing van der Pauw Sheet Resistance Measurements Using the Keithley S530 Parametric Tester. *Application Note Series*, (Number 3180), 2012.

[25] B. Heying, I. Smorchkova, C. Poblenz, C. Elsass, P. Fini, S. Den Baars, U. Mishra, and J. S. Speck. Optimization of the surface morphologies and electron mobilities in GaN grown by plasma-assisted molecular beam epitaxy. *Applied Physics Letters*, 77:2885–2887, 2000.

[26] David Dugdale. *Electronic structure calculations on nitride semiconductors and their alloys*. PhD thesis, Durham University, 2000.

[27] Bernard Gil. *Physics of Wurtzite Nitride and Oxides*. Springer, 2014.

[28] R. Wang, P. P. Ruden, J. Kolnik, I. Oguzman, and K F. Brennan. Dielectric Properties of Wurtzite And Zincblende Structure Gallium Nitride. *Journal of Phys. Chem. Solids*, 58:913–918, 1997.

[29] B. Rezaei, A. Asgari, and M. Kalafi. Electronic band structure pseudopotential calculation of wurtzite III-nitride materials. *Physica B*, 371:107–111, 2005.

[30] Pierre Ziade, Christophe Palermo, Antonio Khoury, Roland Habchi, Myriam Rahal, and Luca Varani. Comparative Analysis of Nitrides Band Structures Calculated by the Empirical Pseudopotential Method. *Universal Journal of Materials Science*, 3:58–72, 2014.

[31] K. Reimann, M. Steube, D. Fröhlich, and S.J. Clarke. Exciton binding energies and band gaps in GaN bulk crystals. *Journal of Crystal Growth*, 189/190:652–655, 1998.

[32] D. O. Demchenko, I. C. Diallo, and M. A. Reshchikov. Yellow Luminescence of Gallium Nitride Generated by Carbon Defect Complexes. *Physical Review Letters*, 110:087404–1–5, 2013.

[33] Hadis Morkoç. *Handbook of Nitride Semiconductors and Devices: Electronic and Optical Processes in Nitrides*. Wiley, 2008.

[34] M. A. Reshchikov and R. Y. Korotkov. Analysis of the temperature and excitation intensity dependencies of photoluminescence in undoped GaN films. *Physical Review B*, 64(115205):1–11, 2001.

[35] Michael A. Reshchikov. Evaluation of GaN by photoluminescence measurement. *Phys. Status. Solidi C*, 8(7-8):2136–2138, 2011.

[36] J. I. Pankove and J. A. Hutchby. Photoluminescence of ion-implanted GaN. *Journal of Applied Physics*, 47, 1976.

[37] S. O. Kucheyev, M. Toth, M. R. Phillips, J. S. Williams, C. Jagadish, and G. Li. Chemical origin of the yellow luminescence in GaN. *Journal of applied physics*, 91(9):5867–5874, 2002.

[38] M. A. Reshchikov, D. O. Demchenko, A. Usikov, H. Helava, and Yu. Makarov. Carbon defects as sources of the green and yellow luminescence bands in undoped GaN. *Physical Review B*, 90(235203):15, 2014.

[39] J. L. Lyons, A. Janotti, and C. G. Van de Walle. Carbon impurities and the yellow luminescence in GaN. *Applied Physics letters*, 97:152108–1–3, 2010.

[40] Hiroshi Harima. Properties of GaN and related compounds studied by means of Raman scattering. *Journal of Physics: Condensed Matter*, 14:R967–R993, 2002.

[41] V. Davydov, Yu. Kitaev, I. Goncharuk, A. Smirnov, J. Graul, O. Semchinova, D. Uffmann, M. Smirnov, A. Mirgorodsky, and R. Evarestov. Phonon dispersion and Raman scattering in hexagonal GaN and AlN. *Physical Review B*, 58(19):12899–12907, 1998.

[42] M. A. Moram, C. S. Ghedia, D. V. S. Rao, J. S. Barnard, Y. Zhang, M. J. Kappers, and C. J. Humphreys. On the origin of threading dislocations in GaN films. *Journal of Applied Physics*, 106(073513):1–9, 2009.

[43] A. E. Romanov, W. Pompe, G. Beltz, and J. S. Speck. Modeling of Threading Dislocation Density Reduction in Heteroepitaxial Layers. *Phys. stat. sol.*, 198:599–613, 1996.

[44] S. Jain, M. Willander, and R. Van Overstraeten. *Compound Semiconductors: Strained Layers and Devices*. Springer Science, 2000.

[45] J. S. Speck and S. J. Rosner. The role of threading dislocations in the physical properties of GaN and its alloys. *Physical B*, 273-274:24–32, 1999.

[46] L. Souriau, T. Atanasove, V. Terzieva, A. Moussa, M. Cyamax, R. Loo, M. Meuris, and W. Vandervorst. Characterization of Threading Dislocations in Thin Germanium Layers by Defect Etching: Toward Chromium and HF-Free Solution. *Journal of The Electrochemical Society*, 2008.

[47] Huan-Yu Shih, Makoto Shiojiri, Ching-Hsiang Chen, Sheng-Fu Yu, Chung-Ting Ko, Jer-Ren Yang, Ray-Ming Lin, and Miin-Jang Chen. Ultralow threading dislocation density in GaN epilayer on nearstrain-free GaN compliant buffer layer and its applications in hetero-epitaxial LEDs. *Nature: Scientific Reports*, 5(13671):1–11, 2015.

[48] He Xiao Guang, Zhao De Gang, and Jiang De Sheng. Formation of two-dimensional electron gas at AlGaN/GaN heterostructure and the derivation of its sheet density expression. *Chin. Phys. B*, 24(6):067301–1–5, 2015.

[49] J. P. Ibbetson, P. T. Fini, K. D. Ness, S. P. DenBaars, J. S. Speck, and U. K. Mishra. Polarization effects, surface states, and the source of electrons in AlGaN/GaN heterostructure field effect transistors. *Appl. Phys. Lett.*, 77(250), 2000.

[50] O. Ambacher, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, W. J. Schaff, L. F. Eastman, R. Dimitrov, L. Wittmer, M. Stutzmann, W. Rieger, and J. Hilsenbeck. Two-dimensional electron gases induced by spontaneous and piezoelectric polarization charges in N- and Ga-face AlGaN/GaN heterostructures. *Journal of Applied Physics*, 86(6):3222–3233, 1999.

[51] Hans Lüth. *Solid Surfaces, Interfaces and Thin Films.* Springer-Verlag, 2010.

[52] Stephen A. Campbell. *Fabrication Engineering at the Micro- and Nanoscale.* Oxford University Press, 2008.

[53] Robin F.C. Farrow. *Molecular Beam Epitaxy: Application to key materials.* Noyes Publications, 1995.

[54] Harald Ibach. *Physics of Surfaces and Interfaces.* Springer, 2006.

[55] Lei He. *III-nitride Semiconductors Grown by Plasma Assisted Molecular Beam Epitaxy.* PhD thesis, Virginia Commonwealth University, 2004.

[56] E. Iliopoulos, A. Adikimenakis, E. Dimakis, K. Tsagaraki, G. Konstantinidis, and A. Georgakilas. Active nitrogen species dependence on radiofrequency plasma source operatingparameters and their role in GaN growth. *Journal of Crystal Growth*, 278:426–430, 2005.

[57] Mark Daniel Losego. *Interfacing Epitaxial Oxides to Gallium Nitride.* PhD thesis, North Carolina State University, 2008.

[58] B. Heying, E. J. Tarsa, C. R. Elsass, P. Fini, S. P. DenBaars, and J. S. Speck. Dislocation mediated surface morphology of GaN. *Journal of Applied Physics*, 85:6470–6476, 1999.

[59] John Venables. Atomic Processes in crystal growth. *Surface Science*, 299:798–817, 1994.

[60] Tosja Zywietz, Jörg Neugebauer, and Matthias Scheffler. Adatom diffusion at GaN (0001) and (000-1) surfaces. *Applied Physics Letters*, 73:487–489, 1997.

[61] N. Grandjean, J. Massies, F. Semond, S. Yu. Karpov, and R. A. Talalaev. GaN evaporation in molecular-beam epitaxy environment. *Applied Physics Letters*, 74:1854–1856, 1999.

[62] S. Fernández Garrido, G. Koblmüller, E. Calleja, and J. S. Speck. In situ GaN decomposition analysis by quadrupole mass spectrometry and reflection high-energy electron diffraction. *Journal of Applied Physics*, 104:033541–1–6, 2008.

[63] G. Koblmüller, F. Wu, T. Mates, and J. S. Speck. High electron mobility GaN grown under N-rich conditions by plasma-assisted molecular beam epitaxy. *Applied Physics Letters*, 91:221905–1–3, 2007.

[64] H. Amano, N. Sawaki, I. Akasaki, and Y. Toyoda. Metalorganic vapor phase epitaxial growth of a high quality GaN film using an AlN buffer layer. *Applied Physics Letters*, 48:353–355, 1986.

[65] G. Koblmüller, R. Averbeck, H. Riechert, Y.-J. Hyun, and P. Pongratz. Strain relaxation dependent island nucleation rates during the Stranski–Krastanow growth of GaN on AlN by molecular beam epitaxy. *Applied Physics Letters*, 93:243105–1–3, 2008.

[66] C. A. Parker. *Growth and Critical Layer Thickness Determination of Indium Gallium Nitride Films Grown on Gallium Nitride*. PhD thesis, North Carolina State University, 2001.

[67] Naoki Ohshima, Hiroo Yonezu, Seiji Yamahira, and Kangsa Pak. Initial growth of GaN on sapphire(0 0 0 1) using an amorphous buffer layer formed at room temperature by RF-MBE. *Journal of Crystal Growth*, 189/190:275–281, 1998.

[68] M.O. Manasreh. *III-V Nitride Semiconductors: Defects and Structural Properties*. Elsevier Science, 2000.

[69] M. J. Manfra, N. G. Weimann, J. W. P. Hsu, L. N. Pfeiffer, K. W. West, and S. N. G. Chu. Dislocation and morphology control during molecular-beam epitaxy of AlGaNÕGaN heterostructures directly on sapphire substrates. *Applied Physics Letters*, 81:1456–1458, 2002.

[70] Yvon Cordier Eric Frayssinet and, H. P. David Schenk, , and Alexis Bavard. Growth of thick GaN layers on 4-in. and 6-in. silicon (111) by metal-organic vapor phase epitaxy. *Physical Status Solidi*, 8:1479–1482, 2010.

[71] Rolf Aidam, Elke Diwo, Nicola Rollbühler, Lutz Kirste, and Fouad Benkhelifa. Strain control of AlGaN/GaN high electron mobility transistor structures on silicon (111) by plasma assisted molecular beam epitaxy. *Journal of Applied Physics*, 111:114516–1–6, 2012.

[72] Man Hoi Wong, Stacia Keller, Sansaptak Dasgupta Nidhi, Daniel J Denninghoff, Seshadri Kolluri, David F Brown, Jing Lu, Nicholas A Fichtenbaum, Elaheh Ahmadi, Uttam Singisetti, Alessandro Chini, Siddharth Rajan, Steven P DenBaars James S Speck1, and Umesh K Mishra. N-polar GaN epitaxy and high electron mobility transistors. *Semiconductor Science and Technology*, 28:074009–1–22, 2013.

[73] S. Mikroulis, A. Georgakilas, A. Kostopoulos, V. Cimalla, E. Dimakis, and Ph. Komninou. Control of the polarity of molecular-beam-epitaxy-grown GaN thin films by the surface nitridation of Al 2 O 3 (0001) substrates. *Applied Physics Letters*, 80:2886–2888, 2002.

[74] D. Huang, P. Visconti, K. M. Jones, M. A. Reshchikov, F. Yun, A. A. Baski, T. King, and H. Morkoc. Dependence of GaN polarity on the parameters of the buffer layer grown by molecular beam epitaxy. *Applied Physics Letters*, 78:4145–4147, 2001.

[75] M. Stutzmann, O. Ambacher, M. Eickhoff, U. Karrer, A. Lima Pimenta, R. Neuberger, J. Schalwig, R. Dimitrov, P. J. Schuck, and R. D. Grober. Playing with Polarity. *Physica Status Solidi B*, 208:505–512, 2001.

[76] A. Yoshikawa and K. Xu. Polarity control of GaN grown on sapphire substrate by RF-MBE. *Thin Solid Films*, 412:38–43, 2002.

[77] Gon Namkoong, W. Alan Doolittle, April S. Brown, Maria Losurdo, Pio Capezzuto, and Giovanni Bruno. Role of sapphire nitridation temperature on GaN growth by plasma assisted molecular beam epitaxy: Part I. Impact of the nitridation chemistry on material characteristics. *Journal of Applied Physics*, 91:2499–2507, 2002.

[78] S. A. Kukushkin, A. V. Osipov, V. N. Bessolov, B. K. Medvedev, V. K. Nevolin, and K. A. Tcarik. Substrates for epitaxy of gallium nitride: New materials and techniques. *Rev.Adv.Mater.Sci*, 17:1–32, 2008.

[79] M. Sobanska, K. Klosek, Z. R. Zytkiewicz, J. Borysiuk, B. S. Witkowski, E. Lusakowska, A. Reszka, and R. Jakiela. Plasma-assisted MBE growth of GaN on Si(111) substrates. *Cryst. Res. Technol.*, 47:307–312, 2012.

[80] P. Chen, R. Zhang, Z.M. Zhao, D.J. Xi, B. Shen, Z.Z. Chen, Y.G. Zhou, S.Y. Xie, W.F. Lu, and Y.D. Zheng. Growth of high quality GaN layers with AlN buffer on Si(1 1 1) substrates. *Journal of Crystal Growth*, 225:150–154, 2001.

[81] K. Kaneko, N. Iizuka, and N. Suzuki. Initial growth conditions for MBE-grown GaN using high-temperature AlN buffer layer. *Phys. stat. sol.*, 2:2232–2235, 2005.

[82] F. Widmann, G. Feuillet, B. Daudin, and J. L. Rouvière. Low temperature sapphire nitridation: A clue to optimize GaN layers grown by molecular beam epitaxy. *Journal of Applied Physics*, 85:1550–1555, 1999.

[83] Lei Zhang, Jiaoxian Yu, Xiaopeng Hao, Yongzhong Wu, Yuanbin Dai, Yongliang Shao, Haodong Zhang, and Yuan Tian. Influence of stress in GaN crystals grown by HVPE on MOCVD-GaN/6H-SiC substrate. *Scientific Reports*, 4:1–6, 2014.

[84] D.D. Koleske, S.R. Lee, M.H. Crawford, K.C. Cross, M.E. Coltrin, and J.M. Kempisty. Connection between GaN and InGaN growth mechanisms and surface morphology. *Journal of Crystal Growth*, 391:85–96, 2014.

[85] A. Krtschil, A. Dadgar, and A. Krost. Decoration effects as origin of dislocation-related charges in gallium nitride layers investigated by scanning surface potential microscopy. *Applied Physics Letters*, 82:2263–2265, 2003.

[86] In-Hwan Lee, In-Hoon Choi, Cheul-Ro Lee, Sung-Jin Son, Jae-Young Leem, and Sam Kyu Noh. Mobility enhancement and yellow luminescence in Si-doped GaN grown by metalorganic chemical vapor deposition technique. *Journal of Crystal Growth*, 182:314–320, 1997.

[87] V. W. L. Chin, T. L. Tansley, and T. Osotchan. Electron mobllities in gallium, indium, and aluminum nitrldes. *Journal of Applied Physics*, 75:7365–7372, 1994.

[88] Rüdiger Quay. *Gallium Nitride Electronics 123.* Springer, 2008.

[89] N. Fukata, J. Chen, T. Sekiguchi, N. Okada, K. Murakami, T. Tsurui, and S. Ito. Doping and hydrogen passivation of boron in silicon nanowires synthesized by laser ablation. *Applied Physics Letters*, 89:2103109–1–3, 2006.

[90] J. L. Benton, D. C. Jacobson, B. Jackson, J. A. Johnson, T. Boone, D. J. Eagleshama, F. A. Stevie, and J. Becerro. Behavior of Molybdenum in Silicon Evaluated for Integrated Circuit Processing. *Journal of the Electrochemical Society*, 146:1929–1933, 1999.

[91] D. G. Zhao, S. J. Xu, M. H. Xie, S. Y. Tong, and Hui Yang. Stress and its effect on optical properties of GaN epilayers grown on Si(111), 6HSiC(0001), and c-plane sapphire. *Applied Physics Letters*, 83:677–679, 2003.

[92] T. Kozawa, T. Kachi, H. Kano, Y. Taga, and M. Hashimoto. Raman scattering fro LO phonon-plasmon coupled modes in gallium nitride. *Journal of Applied Physics*, 75:1098–1101, 1994.

[93] Xiao-Yong Gao, Song-You Wang, Jing Li, Yu-Xiang Zheng, Rong-Jun Zhang, Peng Zhou, Yue-Mei Yang, and Liang-Yao Chen. Structural and Optical Investigation of GaN Grown by Metal-Organic Chemical Vapor Deposition. *Journal of the Korean Physical Society*, 44:765–768, 2004.

[94] In-Hwan Lee, J. J. Lee, P. Kung, F. J. Sanchez, and M. Razeghi. Band-gap narrowing and potential fluctuation in Si-doped GaN. *Applied Physics Letters*, 74(1):102–104, 1999.

[95] R. Chapoulie, C. Capdupuy, M. Schvoerer, and F. Bechtel. Cathodoluminescence and Crystal Growth of Sapphire. *Phys. Stat. Sol.*, 171:613–621, 1999.

[96] Shigeo Shionoya, Takao Koda, Koh Era, and Hiroko Fujiwara. Nature of Luminescence Transitions in ZnS Crvstals. *Journal of the Physical Society of Japan*, 19:1157–1167, 1964.

[97] D. O. Dumcenco, S. Levcenco, Y. S. Huang, C. L. Reynolds Jr., J. G. Reynolds, K. K. Tiong, T. Paskova, and K. R. Evans. Characterization of freestanding semi-insulating Fe-doped GaN by photoluminescence and electromodulation spectroscopy. *Journal of Applied Physics*, 109(123508):1–6, 2011.

[98] D. C. Reynolds, J. Hoelscher, C. W. Litton, and T. C. Collins. Temperature dependence of free excitons in GaN. *Journal of Applied Physics*, 92:5596–5598, 2002.

[99] J. A. Freitas, W. J. Moore, B. V. Shanabrook, G. C. B. Braga S. K. Lee, S. S. Park, and J. Y. Han. Donor-related recombination processes in hydride-vapor-phase epitaxial GaN. *Physical review B*, 66:233311–1–4, 2002.

[100] Niladri Sarkar and Subhasis Ghosh. Temperature dependent band gap shrinkage in GaN: Role of electron-phonon interaction. *Solid State Communications*, 149:1288–1291, 2009.

[101] D. Volm, K. Oettinger, T. Streibl, D. Kovalev, M. Ben Chorin, J. Diener, B. K. Meyer, J. Majewski, L. Eckey, A. Hoffmann, H. Amano, I. Akasaki, K. Hiramatsu, and T. Detchprohm. Exciton fine structure in undoped GaN epitaxial films. *Physical Review B*, 53:16543–16550, 1996.

[102] June O Song, Seong-Ju Park, and Tae-Yeon Seong. Effects of sulfur passivation on Ti/Al ohmic contacts to n-type GaN using CH3CSNH2 solution. *Applied Physics Letters*, 80(17):3129–3131, 2002.

[103] Dragica Vasileska. *Emperical Pseudopotential Method*. PhD thesis, Arizona State University.

[104] Conyers Herring. A New Method for Calculating Wave Functions in Crystals. *Physical Review*, 57:1169–1177, 1940.

[105] S. Logothetidis, J. Petalas, M. Cardona, and T. D. Moustakas. Optical properties and temperature dependence of the interband ttransition of cubic and hexagonal GaN. *Physical Review B*, 50(24):18017–18029, 1994.

[106] J. Cui, A. Sun, M. Reshichkov, F. Yun, A. Baski, and H. Morkoç. Preparation of Sapphire for High Quality III-Nitride Growth. *MRS Internet Journal of Nitride Semiconductor Research*, 5(01):1–6, 2000.

[107] F. Cuccureddu, S. Murphy, I.V. Shvets, M. Porcu, H.W. Zandbergen, N.S. Sidorov, and S.I. Bozhko. Surface morphology of c-plane sapphire (alpha-alumina) produced by high temperature anneal. *Surface Science*, 604:1294–1299, 2010.

[108] Seon Young Moon, Jun Ho Son, Kyung Jin Choi, Jong-Lam Lee, and Ho Won Jang. Indium as an efficient ohmic contact to N-face n-GaN of GaN-based vertical lightemitting diodes. *Applied Physics Letters*, 99(202106):1–3, 2011.

[109] Semrock. 355 nm EdgeBasic best-value long-pass edge filter. Technical report, Part number: BLP01-355R-25, 2015.

[110] Semrock. 365 nm edge BrightLine single-edge dichroic beamsplitter. Technical report, Part number: FF365-Di01-25x36, 2015.

[111] G.L. Martinez, M.R. Curiel, B.J. Skromme, and R.J. Molnar. Surface Recombination and Sulfide Passivation of GaN. *Journal of Electronic Materials*, 29(3):325–331, 2000.

Linear response theory

A

In this appendix the objective will be to derive an equation that can describe the response of a system due to a first order perturbation. This is done by following the methodology by T. G. Pedersen [12], and thus the whole section follows the methods listed in it. Starting by considering the time dependent Schrödinger equation on the form

$$i\hbar\frac{\partial\Psi(\vec{r},t)}{\partial t} = \left(\hat{H}_0 + \frac{1}{2}\hat{H}_1 \mathrm{e}^{-i\omega t} + \frac{1}{2}\hat{H}_1^\dagger \mathrm{e}^{+i\omega t}\right)\Psi(\vec{r},t) \tag{A.1}$$

where $\hat{H}_0$ is for the unperturbed case, and $\hat{H}_1$ is then due to a first order perturbation, where $\dagger$ is the Hermitian conjugate, ω is the frequency, and t is the time. $\hat{H}_1$ may be understood as a quantity that is proportional to the perturbation itself. As usual the wavefunction is written on the form

$$\Psi(\vec{r},t) = \sum_n a_n(t)\psi_n(\vec{r})\mathrm{e}^{-E_n t/\hbar} \tag{A.2}$$

where a_n is some unknown time-dependent coefficient. This is only possible if $\hat{H}_0$ is Hermitian, meaning that the wavefunction can exactly be written as a complete set of eigenfunctions. Inserting this in eq. A.1

$$\begin{aligned}&\sum_n \left(i\hbar\frac{\partial a_n}{\partial t}\psi_n \mathrm{e}^{-iE_n t/\hbar} + a_n\psi_n E_n \mathrm{e}^{-iE_n t/\hbar}\right)\\ =&\sum_n a_n\left(\hat{H}_0 + \frac{1}{2}\hat{H}_1\mathrm{e}^{-i\omega t} + \frac{1}{2}\hat{H}_1^\dagger \mathrm{e}^{+i\omega t}\right)\psi_n \mathrm{e}^{-E_n t/\hbar}\end{aligned} \tag{A.3}$$

The second term on the left of the equal sign $a_n\psi_n E_n \mathrm{e}^{-iE_n t/\hbar} = \hat{H}_0\Psi$ and thus vanishes on either side of the equal sign, reducing eq. A.3 to

$$\frac{1}{2}\sum_n a_n\left(\hat{H}_1\mathrm{e}^{-i\omega t} + \hat{H}_1^\dagger \mathrm{e}^{i\omega t}\right)\psi_n \mathrm{e}^{-E_n t/\hbar} = i\hbar\sum_n \frac{\partial a_n}{\partial t}\psi_n \mathrm{e}^{-iE_n t/\hbar} \tag{A.4}$$

Taking the inner product of this expression, from which the orthogonality of the eigenstates is used, $\langle\psi_m\,|\,\psi_n\rangle = \delta_{nm}$. Eq. A.4 then becomes

$$\begin{aligned}i\hbar\sum_n \frac{\partial a_n}{\partial t}\delta_{nm}\mathrm{e}^{-iE_n t/\hbar} &= \frac{1}{2}\sum_n a_n\left(\langle\psi_m|\,\hat{H}_1\,|\psi_n\rangle\,\mathrm{e}^{-i\omega t} + \langle\psi_m|\,\hat{H}_1^\dagger\,|\psi_n\rangle\,\mathrm{e}^{i\omega t}\right)\mathrm{e}^{-iE_n t/\hbar}\\ i\hbar\frac{\partial a_m}{\partial t}\mathrm{e}^{-iE_m t/\hbar} &= \frac{1}{2}\sum_n a_n\left(\langle\psi_m|\,\hat{H}_1\,|\psi_n\rangle\,\mathrm{e}^{-i\omega t} + \langle\psi_m|\,\hat{H}_1^\dagger\,|\psi_n\rangle\,\mathrm{e}^{i\omega t}\right)\mathrm{e}^{-iE_n t/\hbar}\\ i\hbar\frac{\partial a_m}{\partial t} &= \frac{1}{2}\sum_n a_n\left(\langle\psi_m|\,\hat{H}_1\,|\psi_n\rangle\,\mathrm{e}^{-i\omega t} + \langle\psi_m|\,\hat{H}_1^\dagger\,|\psi_n\rangle\,\mathrm{e}^{i\omega t}\right)\mathrm{e}^{iE_{mn}t/\hbar}\end{aligned} \tag{A.5}$$

where the delta function only allowed one index of the sum at the left hand side of the first line, and $E_{mn} = E_m - E_n$. It is now the objective to evaluate the coefficients a_n. This can be done by considering the Taylor expansion $a_n = a_n^{(0)} + a_n^{(1)} + ...$, where each term is then proportional to the perturbation, i.e. if it is a perturbing electric field $a_n^{(0)} \propto E^0$, $a_n^{(1)} \propto E^1$, $a_n^{(2)} \propto E^2$, etc. Taking a theorem from polynomial series

$$\sum_p b_p x^p = \sum_p c_p x^p \Longrightarrow b_p = c_p \tag{A.6}$$

which means that the total power of the perturbation on either side of the expression must be equal. Noting that eq. A.5 the right hand side already contains one power of perturbation. Using this along with the Taylor series of a_n in eq. A.5 thus yields

$$i\hbar \frac{\partial a_m^{(p)}}{\partial t} = \frac{1}{2} \sum_n a_n^{(p-1)} \left(\langle \psi_m | \hat{H}_1 | \psi_n \rangle \, \mathrm{e}^{-i\omega t} + \langle \psi_m | \hat{H}_1^\dagger | \psi_n \rangle \, \mathrm{e}^{i\omega t} \right) \mathrm{e}^{iE_{mn}t/\hbar} \tag{A.7}$$

however, since this section is only concerned with first order perturbations, any higher perturbations are ignored. Thus starting with the unperturbed case where $p = 0$, it is obvious from the differential part that $\partial a_m^{(0)}/\partial t = 0$. Thus only $p = 1$ is used, hence

$$i\hbar \frac{\partial a_m^{(1)}}{\partial t} = \frac{1}{2} \sum_n a_n^{(0)} \left(\langle \psi_m | \hat{H}_1 | \psi_n \rangle \, \mathrm{e}^{-i\omega t} + \langle \psi_m | \hat{H}_1^\dagger | \psi_n \rangle \, \mathrm{e}^{i\omega t} \right) \mathrm{e}^{iE_{mn}t/\hbar} \tag{A.8}$$

This expression can then be integrated with respect to time, say from $-\infty$ to a time t. This means that there is thus no perturbation anywhere during the lower limit of the integration, and thus the only interesting part is at a time t, where the perturbation is turned on. Also to account for losses a damping factor $\exp(\Gamma t)$ can be multiplied at the right hand side of the equation. The integration thus yields

$$\begin{aligned}
\int_{-\infty}^{t} \frac{\partial a_m^{(1)}}{\partial t} \mathrm{d}t &= \frac{1}{2i\hbar} \sum_n a_n^{(0)} \int_{-\infty}^{t} \left(\langle \psi_m | \hat{H}_1 | \psi_n \rangle \, \mathrm{e}^{-i\omega t} \mathrm{e}^{iE_{mn}t/\hbar} + \langle \psi_m | \hat{H}_1^\dagger | \psi_n \rangle \, \mathrm{e}^{i\omega t} \mathrm{e}^{iE_{mn}t/\hbar} \right) \mathrm{e}^{\Gamma t} \mathrm{d}t \\
a_m^{(1)} &= \frac{1}{2i\hbar} \sum_n a_n^{(0)} \left(\langle \psi_m | \hat{H}_1 | \psi_n \rangle \, \frac{\mathrm{e}^{-i\omega t} \mathrm{e}^{iE_{mn}t/\hbar}}{iE_{mn}/\hbar - i\omega + \Gamma} + \langle \psi_m | \hat{H}_1^\dagger | \psi_n \rangle \, \frac{\mathrm{e}^{i\omega t} \mathrm{e}^{iE_{mn}t/\hbar}}{iE_{mn}/\hbar + i\omega + \Gamma} \right) \frac{\hbar}{i} \mathrm{e}^{\Gamma t} \\
a_m^{(1)} &= -\frac{1}{2} \sum_n a_n^{(0)} \left(\frac{\langle \psi_m | \hat{H}_1 | \psi_n \rangle \, \mathrm{e}^{-i\omega t}}{E_{mn} - \hbar\omega - i\hbar\Gamma} + \frac{\langle \psi_m | \hat{H}_1^\dagger | \psi_n \rangle \, \mathrm{e}^{i\omega t}}{E_{mn} + \hbar\omega - i\hbar\Gamma} \right) \mathrm{e}^{iE_{mn}t/\hbar}
\end{aligned} \tag{A.9}$$

where the factor from the integration was rewritten by $iE_{mn}/\hbar - \omega/ + \Gamma = i(E_{mn} - \hbar\omega - i\Gamma)/\hbar$, and the damping factor was removed again in the final line by choosing its value such that $\exp(\Gamma t) = 1$. This corresponds to the beginning of a perturbation and further means that the time domain in which the perturbation theory is used cannot completely be ignored.

It is now possible to consider any time-independent hermitian operator $\hat{X}$, from which its expectation value $\langle \Psi | \hat{X} | \Psi \rangle$ corresponds to an observable response due to a perturbation of the system. Using again eq. A.2, the expectation value then becomes

$$\langle \Psi | \hat{X} | \Psi \rangle = \sum_{n,m} a_n^* a_m \langle \psi_n | \hat{X} | \psi_m \rangle \, \mathrm{e}^{iE_{nm}t/\hbar} \tag{A.10}$$

and since only the first order cases of a are of interest, $a_n^* a_m = (a_n^{(0)*} + a_n^{(1)*})(a_m^{(0)} + a_m^{(1)}) = a_n^{(0)*} a_m^{(0)} + a_n^{(1)*} a_m^{(0)} + a_n^{(0)*} a_m^{(1)}$, where the $a_n^{(1)*} a_m^{(1)}$ term was ignored, since it is a second order

term. Thus eq. A.10 may be expanded to

$$\langle\Psi|\,\hat{X}\,|\Psi\rangle \approx \sum_{n,m}(a_n^{(0)*}a_m^{(0)} + a_n^{(1)*}a_m^{(0)} + a_n^{(0)*}a_m^{(1)})\,\langle\psi_n|\,\hat{X}\,|\psi_m\rangle\,\mathrm{e}^{iE_{nm}t/\hbar} \tag{A.11}$$

where the a-terms still need evaluation. a_n says something about the probability of occupying a given state ψ_n. This may be further understood from how a_n relates to the wavefunctions, see eq. A.2, i.e.

$$1 = \langle\psi\,|\,\psi\rangle = \sum_n |a_n|^2 \tag{A.12}$$

and thus this corresponds to an unperturbed situation with $a_n = a_n^{(0)}$. $|a_n^{(0)}|^2$ is thus the probability for the unperturbed system to occupy a given state, i.e. this quantity can be related to the Fermi function $|a_n^{(0)}|^2 = f(E_n)$ in thermal equilibrium. Furthermore, it can be shown from density functional theory [12], that

$$a_n^{(0)*}a_m^{(0)} = \begin{cases} f(E_n) & \text{if } n = m \\ 0 & \text{otherwise} \end{cases} \tag{A.13}$$

which if used along with the preceding arguments and the expression for the first order perturbation (eq. A.9) may help reformulating eq. A.11 to

$$\begin{aligned}\langle\Psi|\,\hat{X}\,|\Psi\rangle \approx \sum_n f(E_n)\,\langle\psi_n|\,\hat{X}\,|\psi_n\rangle \\ -\frac{1}{2}\sum_{m,n} f(E_n)\,\langle\psi_n|\,\hat{X}\,|\psi_m\rangle\left(\frac{\langle\psi_m|\,\hat{H}_1\,|\psi_n\rangle\,\mathrm{e}^{-i\omega t}}{E_{mn}-\hbar\omega-i\hbar\Gamma} + \frac{\langle\psi_m|\,\hat{H}_1^\dagger\,|\psi_n\rangle\,\mathrm{e}^{i\omega t}}{E_{mn}+\hbar\omega-i\hbar\Gamma}\right) \\ -\frac{1}{2}\sum_{m,n} f(E_m)\,\langle\psi_n|\,\hat{X}\,|\psi_m\rangle\left(\frac{\langle\psi_m|\,\hat{H}_1^\dagger\,|\psi_n\rangle\,\mathrm{e}^{i\omega t}}{E_{mn}-\hbar\omega+i\hbar\Gamma} + \frac{\langle\psi_m|\,\hat{H}_1\,|\psi_n\rangle\,\mathrm{e}^{-i\omega t}}{E_{mn}+\hbar\omega+i\hbar\Gamma}\right)\end{aligned} \tag{A.14}$$

where the $\exp(iE_{mn}t/\hbar)$ vanished by multiplication with $\exp(iE_{nm}t/\hbar)$. This expression may be written in a shorter version as

$$\langle\Psi|\,\hat{X}\,|\Psi\rangle \approx \sum_n f(E_n)\,\langle\psi_n|\,\hat{X}\,|\psi_n\rangle + \frac{1}{2}X(\omega)\mathrm{e}^{i\omega t} + \frac{1}{2}X^*(\omega)\mathrm{e}^{i\omega t} \tag{A.15}$$

where

$$X(\omega) = -\sum_{m,n} f_{nm}\frac{\langle\psi_m|\,\hat{H}_1\,|\psi_n\rangle\,\langle\psi_n|\,\hat{X}\,|\psi_m\rangle}{E_{mn}-\hbar\omega-i\hbar\Gamma} \tag{A.16}$$

in which $f_{nm} = f(E_n) - f(E_m)$. Eq. A.16 is completely general for linear response theory. [12]

A.1 Optical response

It is now the purpose to consider a more specific perturbation, namely the interaction of an optical electric field $\vec{\mathcal{E}} = \hat{z}\mathcal{E}$ with an electric dipole with momentum $-e\vec{r}$. The perturbation is thus given by $\hat{H}_1 = e\vec{\mathcal{E}}\cdot\vec{r} = e\mathcal{E}z$ and the response is then the dipole moment density $\hat{X} = -ez/\Omega$, Ω being the volume. Here it was implicitly chosen that the induced dipole moment is parallel to the incident electric field by considering only the z-component of the dipole. It is thus the

polarization that has to be considered, where by considering eq. A.16 the polarization may then be written as

$$P(\omega) = -\frac{1}{\Omega}\sum_{m,n} f_{nm}\frac{\langle\phi_m| e\mathcal{E}z |\psi_n\rangle \langle\psi_n| - ez |\psi_m\rangle}{E_{mn} - \hbar\omega - i\hbar\Gamma} = \frac{e^2}{\Omega}\mathcal{E}\sum_{m,n} f_{nm}\frac{|\langle\psi_m| z |\psi_n\rangle|^2}{E_{mn} - \hbar\omega - i\hbar\Gamma} \tag{A.17}$$

from which the susceptibility $\chi(\omega)$ is rewritten by using $P(\omega) = \epsilon_0\chi(\omega)\mathcal{E}$, such that

$$\chi(\omega) = \frac{e^2}{\epsilon_0\Omega}\sum_{m,n} f_{nm}\frac{|\langle\psi_m| z |\psi_n\rangle|^2}{E_{mn} - \hbar\omega - i\hbar\Gamma} \tag{A.18}$$

If it is the optical response of a semiconductor that is considered, then the energy bands are divided into valence and conduction bands, where the lowest unoccupied states are separated from the highest occupied states by an energy gap E_g. If the temperature is assumed such that $E_g >> kT$, the Fermi level then lies in the middle of the gap. Rewriting the double sum from eq. A.18 as

$$\sum_{m,n} = \sum_{m\in v}\sum_{n\in v} + \sum_{m\in c}\sum_{n\in v} + \sum_{m\in v}\sum_{n\in c} + \sum_{m\in c}\sum_{n\in c} \tag{A.19}$$

For the first and last of these sums the occupancy factor $f_{nm} = f(E_n) - f(E_m) \approx 0$. Eq. A.18 is then rewritten as

$$\chi(\omega) = \frac{e^2}{\epsilon_0\Omega}\left(\sum_{m\in c}\sum_{n\in v} f_{nm}\frac{|\langle\psi_m| z |\psi_n\rangle|^2}{E_{mn} - \hbar\omega - i\hbar\Gamma} + \sum_{m\in v}\sum_{n\in c} f_{nm}\frac{|\langle\psi_m| z |\psi_n\rangle|^2}{E_{mn} - \hbar\omega - i\hbar\Gamma}\right) \tag{A.20}$$

now the indexes of the second term are exchanged and utilizing that $f_{mn} = -f_{nm}$, $E_{nm} = -E_{mn}$ and the indexes in $\langle\psi_m| z |\psi_n\rangle$ may be exchanged without any other intervention since the summation is over a complete set of eigenfunctions. Eq. A.20 is then rewritten as

$$\chi(\omega) = \frac{e^2}{\epsilon_0\Omega}\left(\sum_{m\in c}\sum_{n\in v} f_{nm}\frac{|\langle\psi_m| z |\psi_n\rangle|^2}{E_{mn} - \hbar\omega - i\hbar\Gamma} - \sum_{n\in v}\sum_{m\in c} f_{nm}\frac{|\langle\psi_n| z |\psi_m\rangle|^2}{-E_{mn} - \hbar\omega - i\hbar\Gamma}\right) \tag{A.21}$$

Due to the sums $|\langle\psi_m| z |\psi_n\rangle|^2 = |\langle\psi_n| z |\psi_m\rangle|^2$, and utilizing that $f_{nm} \approx 1$ at low temperature, since then the occupancy of either band is $f(E_n) = 1$ and $f(E_m) = 0$. Eq. A.21 is then reduced to

$$\chi(\omega) = \frac{e^2}{\epsilon_0\Omega}\sum_{m\in c}\sum_{n\in v} |\langle\psi_m| z |\psi_n\rangle|^2\left(\frac{1}{E_{mn} - \hbar\omega - i\hbar\Gamma} - \frac{1}{-E_{mn} - \hbar\omega - i\hbar\Gamma}\right) \tag{A.22}$$

from which the fractions in the parenthesis are further simplified first by multiplying by (-1) in the second fraction in the parenthesis, and finding a common denominator as

$$\left(\frac{E_{mn} + \hbar\omega + i\hbar\Gamma}{(E_{mn} + \hbar\omega + i\hbar\Gamma)(E_{mn} - \hbar\omega - i\hbar\Gamma)} + \frac{E_{mn} - \hbar\omega - i\hbar\Gamma}{(E_{mn} - \hbar\omega - i\hbar\Gamma)(E_{mn} + \hbar\omega + i\hbar\Gamma)}\right)$$
$$= \frac{2E_{mn}}{(E_{mn} + \hbar\omega + i\hbar\Gamma)(E_{mn} - \hbar\omega - i\hbar\Gamma)} \tag{A.23}$$

where $(E_{mn} + \hbar\omega + i\hbar\Gamma)(E_{mn} - \hbar\omega - i\hbar\Gamma) = E_{mn}^2 - \hbar^2\omega^2 - 2i\hbar^2\omega\Gamma + \hbar^2\Gamma^2 = E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2$. Eq. A.22 then becomes

$$\chi(\omega) = 2\frac{2e^2}{\epsilon_0\Omega}\sum_{m\in c}\sum_{n\in v}\frac{E_{mn}|\langle\psi_m| z |\psi_n\rangle|^2}{E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2} \tag{A.24}$$

where the extra factor 2 is to account for spin (allowing 2 electrons to occupy each state). The matrix element $\langle\psi_m|\, z\, |\psi_n\rangle$ may be rewritten through the commutator relations between the operators of quantum mechanics. Consider the commutation of the Hamiltonian $\hat{H}_0$ with any Hermitian operator $\hat{o}$ as

$$\begin{aligned}\langle\psi_m|\,[\hat{H}_0,\hat{o}]\,|\psi_n\rangle &= \langle\psi_m|\,\hat{H}_0\hat{o}-\hat{o}\hat{H}_0\,|\psi_n\rangle \\ &= \langle\psi_m|\,\hat{H}_0\hat{o}\,|\psi_n\rangle - \langle\psi_m|\,\hat{o}\hat{H}_0\,|\psi_n\rangle \\ &= \langle\psi_n|\,\hat{o}\hat{H}_0\,|\psi_m\rangle^* - \langle\psi_m|\,\hat{o}\hat{H}_0\,|\psi_n\rangle \end{aligned} \tag{A.25}$$

where the * refers to the complex conjugate of that expression. It was here utilized that both of the operators are Hermitian. The operation of $\hat{H}_0$ on the eigenfunctions will then simply yield the eigenvalue, thus eq. A.25 becomes

$$\langle\psi_m|\,[\hat{H}_0,\hat{o}]\,|\psi_n\rangle = E_m\,\langle\psi_n|\,\hat{o}\,|\psi_m\rangle^* - E_n\,\langle\psi_m|\,\hat{o}\,|\psi_n\rangle \tag{A.26}$$

$$= E_{mn}\,\langle\psi_m|\,\hat{o}\,|\psi_n\rangle \tag{A.27}$$

Applying this to the matrix element $\langle\psi_m|\, z\, |\psi_n\rangle$, thus yields

$$\langle\psi_m|\,[\hat{H}_0,z]\,|\psi_n\rangle = E_{mn}\,\langle\psi_m|\, z\,|\psi_n\rangle \tag{A.28}$$

but it may also be rewritten by considering the commutation of the operators, i.e.

$$\begin{aligned}[\hat{H}_0,z]\psi &= -\frac{\hbar^2}{2m}\left(\left[\frac{\partial^2}{\partial z^2},z\right]\psi\right) \\ &= -\frac{\hbar^2}{2m}\left(\psi\frac{\partial^2 z}{\partial z^2}+2\frac{\partial z}{\partial z}\frac{\partial\psi}{\partial z}+z\frac{\partial^2\psi}{\partial z^2}-z\frac{\partial^2\psi}{\partial z^2}\right) \\ &= -\frac{\hbar^2}{m}\frac{\partial\psi}{\partial z} \\ &= \frac{\hbar}{im}\hat{p}_z\psi \end{aligned} \tag{A.29}$$

which in combination with eq. A.28 yield

$$\langle\psi_m|\, z\,|\psi_n\rangle = \frac{\hbar}{imE_{mn}}\,\langle\psi_m|\,\hat{p}_z\,|\psi_n\rangle \tag{A.30}$$

which is then inserted in eq. A.24 to give

$$\chi(\omega) = \frac{4e^2\hbar^2}{\epsilon_0 m^2\Omega}\sum_{m\in c}\sum_{n\in v}\frac{|\,\langle\psi_m|\,\hat{p}_z\,|\psi_n\rangle\,|^2}{E_{mn}(E_{mn}^2-\hbar^2(\omega+i\Gamma)^2)} \tag{A.31}$$

This expression can thus be used to obtain accurate numerical values for the susceptibility, given that the energy states and momentum matrix elements are known. [12]

Empirical Pseudopotential Method

B

In order to determine the electronic structure of a material, the parameters which need to be found are the energies and wavefunctions of single electron states as a function of position in the Brillouin zone. Generally, there are two approaches to this problem: *ab-initio* calculatios, which tend to be computationally expensive, but do not rely on experimental data, and *empirical* methods, where empirical input is needed to adjust the relevant parameters so that the result of the calculations match existing experimental or theoretical data. The Empirical Pseudopotential Method (EPM) is one of the latter methods, and like all band structure methods, it aims to solve the one-electron Schrödinger equation

$$\left[\frac{-\hbar^2}{2m_0}\nabla^2 + V(\mathbf{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{B.1}$$

where $V(\mathbf{r})$ is the potential felt by the electron due to all the other electrons and ionic cores. This lattice-periodic many-body potential has a strong atom-like nature near the cores, and is weak between the cores. The potential can be treated in different ways, where e.g. the tight-binding method models the electron as tightly bound to their respective cores, and at the other extreme, the nearly-free electron model assumes that the potential is weak everywhere. Somewhere in between this lies the EPM, which treats electrons as either core or valence. The core electrons are tightly bound, while the valence electrons, which are those primarily responsible for the properties of the material, are nearly free due to the screening of the nuclear charge by the core electrons. This means that only valence electrons are considered. [26, 103]

The valence electron wave functions must be constructed to be orthogonal to the core states, which are also mutually orthogonal. This is done by using the Orthogonalised Plane Wave method by Herring [104]. Here, the basis states are constructed from a set of plane waves, which are made to be orthogonalised to the wave functions of the core electrons, and thus, the true valence state wavefunction is expanded as a smooth plane wave function ϕ plus a sum over the occupied core states ϕ_n

$$|\psi\rangle = |\phi\rangle + \sum_{n}^{\text{core}} a_n |\phi_n\rangle \tag{B.2}$$

In the region between cores, $|\phi\rangle$ adequately describes the weak potential that is present, but near the cores, the orthogonalisation part of Φ forces the valence electron to occupy the next, non-occupied core state wavefunction, which effectively repels the valence electrons from the core. By requirement that ψ is orthogonal to the core states (which are mutually orthogonal as

well), it is obvious that

$$\langle \phi_n \,|\, \psi \rangle = 0 = \langle \phi_n \,|\, \phi \rangle + a_n \tag{B.3}$$

which fixes the expansion coefficients $a_n = -\langle \phi_m \,|\, \phi \rangle$, such that

$$|\psi\rangle = |\phi\rangle - \sum_{n}^{core} \langle \phi_n \,|\, \phi \rangle \, \langle \phi_n| \tag{B.4}$$

Operating on either side of eq. B.4 with the Hamiltonian $\left[\frac{-\hbar^2}{2m_0}\nabla^2 + V\right]$, where V is the attractive potential of the core, then yields

$$\begin{aligned} E\,|\phi\rangle - \sum_{n}^{\text{core}} E\,\langle \phi_n \,|\, \phi \rangle \, \langle \phi_n| &= \hat{H}\,|\phi\rangle - \sum_{n}^{\text{core}} \langle \phi_n \,|\, \phi \rangle \, E_n \, \langle \phi_n| \\ \implies \hat{H}\,|\phi\rangle + \sum_{n}^{core} (E - E_n) \, \langle \phi_n \,|\, \phi \rangle \, |\phi_n\rangle &= E\,|\phi\rangle \end{aligned} \tag{B.5}$$

Rewriting this allows the identification of the repulsion due to the orthogonalisation terms

$$\left(\hat{H} + V_R\right)|\phi\rangle = E\,|\phi\rangle\,, \qquad V_R = \sum_{n}^{\text{core}} (E - E_n)\,|\phi_n\rangle\,\langle\phi_n| \tag{B.6}$$

Splitting up the Hamiltonian in its kinetic and potential terms then gives

$$\left[\frac{-\hbar^2}{2m_0}\nabla^2 + V + V_R\right]\phi = \left[\frac{-\hbar^2}{2m_0}\nabla^2 + V_{ps}\right]\phi = E\phi \tag{B.7}$$

This is the equation for the pseudowavefunction ϕ, where the long range, real Coulomb potential has been added to the short ranged repulsive potential due to the orthogonalisation, yielding the *pseudopotenial* V_{ps}. Thus, while eq. B.7 can be thought of as the wave equation for the pseudowavefuntions ϕ, the energy eigenvalues are not pseudoenergies, but rather the true eigenvalues for the true wavefunction, ψ. [26, 103]

Naturally, the real one-electron potential is the Coulomb potential, which is strong near the core region, while the pseudopotential is weak near the core due to the cancellation term. Far from the core, the pseudopotential is long-ranged and attractive, as it should be identical to the unscreened Coulomb potential. This means that, while the true wavefunctions exhibit rapid spatial variation in the core region, the pseudowavefunctions are smooth and slow-varying in all regions. This is the real advantage of the pseudopotential method. Since a large number of plane waves are needed to represent a rapidly varying function, a computationally prohibitive number of basis functions are needed to express the true wavefunction. Due to the smoothness of the pseudowavefunction, it can be reasonably described in a basis including a much smaller number of plane waves. However, although the orthogonal plane wave method shows how to construct the pseudopotential from knowledge of the core states, the actual pseudopotentials used in calculations are usually made from some simple model potential, the exact expression for which is varied until the results of the calculations fit with experimental data. [26, 103]

The problem at hand is now to solve the Schrödinger equation B.7. First, as the pseudowavefunctions (henceforth denoted ψ) are required to have periodicity of the lattice, they can be expanded as a Fourier sum of plane waves

$$\psi_{\mathbf{k}} = U_{\mathbf{k}}(\mathbf{r})\mathrm{e}^{i\mathbf{G}\cdot\mathbf{r}}, \qquad U_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\Omega}\sum_{\mathbf{G}} U_{\mathbf{G}}\,\mathrm{e}^{i\mathbf{G}\cdot\mathbf{r}} \tag{B.8}$$

where $u_{\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{\mathbf{k}}(\mathbf{r})$ as $\mathbf{G}\cdot\mathbf{R} = p2\pi$ if $\mathbf{R}$ is a lattice vector (p integer), and Ω is the volume. Thus,

$$\psi_{\mathbf{k}} = \sum_{\mathbf{G}} U_{\mathbf{G}} \left(\frac{1}{\Omega} \mathrm{e}^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}} \right) = \sum_{\mathbf{G}} U_{\mathbf{G}} \phi_{\mathbf{G}}(\mathbf{k},\mathbf{r}) \tag{B.9}$$

with $\phi_{\mathbf{G}}(\mathbf{k},\mathbf{r})$ being the basis functions. It is readily seen that these basis functions are orthonormal

$$\int \phi^*_{\mathbf{G}}(\mathbf{k},\mathbf{r}) \phi_{\mathbf{G}'}(\mathbf{k},\mathbf{r}) \mathrm{d}^3 r = \frac{1}{\Omega} \int \mathrm{e}^{i(\mathbf{G}'-\mathbf{G})\cdot\mathbf{r}} \mathrm{d}^3 r = \delta_{\mathbf{G},\mathbf{G}'} \tag{B.10}$$

The potential is likewise periodic, i.e. $V(\mathbf{r}) = \sum_{\mathbf{R}} V(\mathbf{r}+\mathbf{R})$, and can thus also be Fourier expanded

$$V(\mathbf{r}) = \sum_{\mathbf{G}} V(\mathbf{G}) \mathrm{e}^{i\mathbf{G}\cdot\mathbf{r}} \tag{B.11}$$

with the expansion coefficients given by

$$V(\mathbf{G}) = \frac{1}{\Omega_{UC}} \int V(\mathbf{r}) \mathrm{e}^{-i\mathbf{G}\cdot\mathbf{r}} \mathrm{d}^3 r \tag{B.12}$$

where Ω_{UC} is the volume of the unit cell. [103]

The Hamiltonian for the pseudowavefunctions now becomes $\hat{H} = \left[\frac{-\hbar^2}{2m_0}\nabla^2 + \sum_{\mathbf{R}} V(\mathbf{r}+\mathbf{R}) \right]$. The kinetic energy can be written as

$$\begin{aligned} \left\langle \phi_{\mathbf{G}} \left| \frac{-\hbar^2}{2m_0}\nabla^2 \right| \phi_{\mathbf{G}'} \right\rangle &= -\frac{\hbar^2}{2m_0} \int \frac{1}{\Omega} \mathrm{e}^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \left[i(\mathbf{k}+\mathbf{G}') \right]^2 \mathrm{e}^{i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}} \mathrm{d}^3 r \\ &= \frac{\hbar^2}{2m_0} (\mathbf{k}+\mathbf{G})^2 \delta_{\mathbf{G},\mathbf{G}'} \end{aligned} \tag{B.13}$$

while the potential energy can be written

$$\begin{aligned} \left\langle \phi_{\mathbf{G}} \left| \sum_{\mathbf{R}} V(\mathbf{r}+\mathbf{R}) \right| \phi_{\mathbf{G}'} \right\rangle &= \sum_{\mathbf{R}} \frac{1}{\Omega} \int \mathrm{e}^{i(\mathbf{G}'-\mathbf{G})\cdot\mathbf{r}} V(\mathbf{r}+\mathbf{R}) \mathrm{d}^3 r \\ &= \frac{1}{\Omega_{UC}} \int \mathrm{e}^{i(\mathbf{G}'-\mathbf{G})\cdot\mathbf{r}} V(\mathbf{r}) \mathrm{d}^3 r = \frac{1}{\Omega_{UC}} \sum_{\mathbf{G}'} V(\mathbf{G}-\mathbf{G}') \end{aligned} \tag{B.14}$$

where it was utilized that all unit cells have the same contribution, i.e. $N/\Omega = 1/\Omega_{UC}$, allowing integration only over the volume of the unit cell, and $V(\mathbf{G}-\mathbf{G}')$ are the Fourier coefficients. It is noted that the potential energy contribution is independent of k. [26]

The Schrödinger equation can now be stated in matrix form using the pseudowavefunctions and pseudopotentials, yielding a standard matrix eigenvalue equation

$$\begin{bmatrix} H_{\mathbf{G}_1,\mathbf{G}_1} & H_{\mathbf{G}_1,\mathbf{G}_2} & \cdots & H_{\mathbf{G}_1,\mathbf{G}_N} \\ H_{\mathbf{G}_2,\mathbf{G}_1} & H_{\mathbf{G}_2,\mathbf{G}_2} & \cdots & H_{\mathbf{G}_2,\mathbf{G}_N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{\mathbf{G}_N,\mathbf{G}_1} & H_{\mathbf{G}_N,\mathbf{G}_2} & \cdots & H_{\mathbf{G}_N,\mathbf{G}_N} \end{bmatrix} \begin{bmatrix} U_{\mathbf{G}_1} \\ U_{\mathbf{G}_2} \\ \vdots \\ U_{\mathbf{G}_N} \end{bmatrix} = E \begin{bmatrix} U_{\mathbf{G}_1} \\ U_{\mathbf{G}_2} \\ \vdots \\ U_{\mathbf{G}_N} \end{bmatrix} \tag{B.15}$$

where $H_{\mathbf{G},\mathbf{G}'} = \left\langle \phi_{\mathbf{G}} \left| \hat{H} \right| \phi_{\mathbf{G}'} \right\rangle$. The size of this problem depends on the number of plane waves used in the expansion of the wave functions and potential. If N plane waves are used, then for

any given value of $\mathbf{k}$, N eigenvalues are found corresponding to the energies of the bands at that $\mathbf{k}$. It was found that the kinetic part of the Hamiltonian only enters the diagonal elements, while the potential part has a value in all elements. However, as the pseudopotential contribution for $\mathbf{G} = \mathbf{G}'$ is constant for all entries, this simply corresponds to a rigid shift of all bands, and is thus omitted, such that the diagonal only contains the kinetic part, and the off-diagonal contains the potential part.

The task is now to find expressions for the pseudopotential for the possible lattice types, which for the case of GaN are zincblende and wurtzite.

B.1 Zincblende lattice

In the zincblende lattice, there are two atoms per unit cell, one anion and one cation. If a two-atom basis is chosen with the origin as the midpoint between the two ions, the pseudopotential can be written

$$V(\mathbf{r}) = V_a(\mathbf{r}+\tau) + V_c(\mathbf{r}-\tau) \tag{B.16}$$

where τ is the atomic basis vector given by $\tau_1 = \tau = -\tau_2 = \frac{a}{8}[1\ 1\ 1]$. Substituting this into eq. B.12

$$V(\mathbf{G}) = \frac{1}{\Omega_{UC}} \int \left[V_a(\mathbf{r}+\tau) + V_c(\mathbf{r}-\tau)\right] \mathrm{e}^{-i\mathbf{G}\cdot\mathbf{r}} \mathrm{d}^3 r \tag{B.17}$$

$V(\mathbf{G})$ is frequently called the *form factor* in the pseudopotential method. Making the substitution $\mathbf{r}+\tau = \mathbf{r}'$, and using the displacement property of Fourier transforms $\mathcal{F}[V(\mathbf{r}\pm\tau)] = \exp(\pm i\mathbf{G}\cdot\tau)V(\mathbf{r})$,

$$V(\mathbf{G}) = \frac{1}{\Omega_{UC}} \int \left[V_a(\mathbf{r}')\mathrm{e}^{i\mathbf{G}\cdot\tau} + V_c(\mathbf{r}')\mathrm{e}^{-i\mathbf{G}\cdot\tau}\right] \mathrm{e}^{-i\mathbf{G}\cdot\mathbf{r}'} \mathrm{d}^3 r' \tag{B.18}$$

This can be rewritten using the form factors for the anion and cation, respectively, here approximating the atomic volumes to be equal

$$V_a(\mathbf{G}) = \frac{1}{2\Omega_{UC}} \int V_a(\mathbf{r}')\mathrm{e}^{-i\mathbf{G}\cdot\mathbf{r}'} \mathrm{d}^3 r' \quad , \quad V_c(\mathbf{G}) = \frac{1}{2\Omega_{UC}} \int V_c(\mathbf{r}')\mathrm{e}^{-i\mathbf{G}\cdot\mathbf{r}'} \mathrm{d}^3 r' \tag{B.19}$$

so the crystal form factor becomes

$$V(\mathbf{G}) = \frac{1}{2} \int \left[V_a(\mathbf{G})\mathrm{e}^{i\mathbf{G}\cdot\tau} + V_c(\mathbf{G})\mathrm{e}^{-i\mathbf{G}\cdot\tau}\right] \tag{B.20}$$

By now introducing symmetric and antisymmetric form factors, $V_S(\mathbf{G}) = \frac{1}{2}[V_a(\mathbf{G}) + V_c(\mathbf{G})]$ and $V_A(\mathbf{G}) = \frac{1}{2}[V_a(\mathbf{G}) - V_c(\mathbf{G})]$, and using Euler's formula, the expression for the crystal form factor becomes

$$\begin{aligned} V(\mathbf{G}) &= V_S(\mathbf{G})\cos(\mathbf{G}\cdot\tau) + iV_A(\mathbf{G})\sin(\mathbf{G}\cdot\tau) \\ &= V_S(\mathbf{G})S_S(\mathbf{G}) + iV_A(\mathbf{G})S_A(\mathbf{G}) \end{aligned} \tag{B.21}$$

where $S_S(\mathbf{G})$ and $S_A(\mathbf{G})$ are the symmetric and antisymmetric *structure factors* for zincblende. From this, it is seen that if the two atoms in the lattice are the same, as. e.g. for Si, there is no antisymmetrical form factor. On the other hand, if the potential difference between the two atoms is large, the antisymmetric part may be of large importance, which is the case for e.g. nitrides. [26]

A common, and appropriate, approximation is now made that the atomic potentials are spherically symmetric. Thus, the final expression becomes

$$V(\mathbf{G}) = V_S(G)S_S(\mathbf{G}) + iV_A(G)S_A(\mathbf{G}) \tag{B.22}$$

The fact that the form factors only depend on the magnitude of $\mathbf{G}$, which makes their evaluation much simpler. The distance to each equivalent reciprocal lattice site is an integer defined by $|\mathbf{G}|^2 = 0,\ 3,\ 4,\ 8,\ 11,\ \ldots$ in units of $\left(\frac{2\pi}{a}\right)^2$. The pseudopotential contribution to the Hamiltonian matrix, as given in eq. B.14, was seen to be depended on the difference between reciprocal lattice vector, which forms the same set of integers as those from the origin to any lattice site. Thus, the pseudopotential form factors are only required for these specific values of $|\mathbf{G}|^2$, corresponding to nearest-neighbour sites. Additionally, because $V(\mathbf{G})$ typically decreases as $1/G^2$ for large values of G, form factors with $G^2 > 11\left(\frac{2\pi}{a}\right)^2$ are omitted for zincblende, as this limits the number of plane waves needed in the expansion, and has little effect on the final result. [26, 30, 103]

The zincblende $\mathbf{G}$ vectors and their associated form factors for GaN used in this work are shown in table B.1. As noted earlier, the $V(0)$ has no influence on the shape of the band diagram, and is omitted. For $V_S(\sqrt{4})$ and $V_A(\sqrt{8})$, the symmetric and antisymmetric structure factors, respectively, are zero. Thus, the six values in the table, along with the lattice constant, are sufficient to differentiate between GaN and other zincblende structures.

$\mathbf{G}$	$(1,1,1)$	$(2,0,0)$	$(2,2,0)$	$(3,1,1)$
$\lvert\mathbf{G}\rvert^2$	3	4	8	11
$V_S(\sqrt{G})$	−0.32	0	0.03	0.06
$V_A(\sqrt{G})$	0.25	0.21	0	0.02

TABLE B.1: *Form factors for zincblende GaN. Adopted from [28].*

B.2 Wurtzite lattice

In the wurtzite structure, there are four atoms per unit cell. Again taking the origin of the unit cell to be the midpoint in the parallelogram formed by these atoms, the positions of the atoms according to basis vectors are

$$\begin{aligned}\mathbf{r}_1 &= \left[\frac{1}{6},\ \frac{1}{6},\ \frac{1}{2}\left(\frac{1}{2}+u\right)\right]a \\ \mathbf{r}_2 &= \left[\frac{1}{6},\ \frac{1}{6},\ \frac{1}{2}\left(\frac{1}{2}-u\right)\right]a\end{aligned} \tag{B.23}$$

such that anionic atoms are positioned at $\mathbf{r}_1$ and $-\mathbf{r}_2$, and the cationic atoms at $-\mathbf{r}_1$ and $\mathbf{r}_2$. The crystal form factor for wurtzite is then

$$V(\mathbf{G}) = \frac{1}{2}\int\left[V_a(\mathbf{G})\left(\mathrm{e}^{i\mathbf{G}\cdot\mathbf{r}_1} + \mathrm{e}^{-i\mathbf{G}\cdot\mathbf{r}_2}\right) + V_c(\mathbf{G})\left(\mathrm{e}^{-i\mathbf{G}\cdot\mathbf{r}_1} + \mathrm{e}^{i\mathbf{G}\cdot\mathbf{r}_2}\right)\right] \tag{B.24}$$

which is readily rewritten using eq. B.23 as

$$\begin{aligned}V(\mathbf{G}) &= \frac{1}{2}\left[V_a 2\cos\left(\frac{G_x}{6}+\frac{G_y}{6}+\frac{G_z}{4}\right)\mathrm{e}^{iG_z u/2} + V_c 2\cos\left(\frac{G_x}{6}+\frac{G_y}{6}+\frac{G_z}{4}\right)\mathrm{e}^{-iG_z u/2}\right] \\ &= \cos\left(\frac{G_x}{6}+\frac{G_y}{6}+\frac{G_z}{4}\right)\left(V_a\mathrm{e}^{iG_z u/2} + V_c\mathrm{e}^{-iG_z u/2}\right)\end{aligned} \tag{B.25}$$

where the a part of in the $\mathbf{G}$'s was eliminated in the dot product with $\mathbf{r}$'s. Now rewriting the term containing the exponentials as

$$\begin{aligned} V_a e^{iG_z u/2} + V_c e^{-iG_z u/2} &= V_a(\cos(G_z u/2) + i\sin(G_z u/2)) + V_c(\cos(G_z u/2) - i\sin(G_z u/2)) \\ &= \cos(G_z u/2)(V_a + V_c) + i\sin(G_z u/2)(V_a - V_c) \end{aligned}$$

which is inserted in eq. B.25 with the introduction of $V_S = V_a + V_c$, $V_A = V_a - V_c$, so that the structure factors become

$$\begin{aligned} S_S &= \cos\left(\frac{G_x}{6} + \frac{G_y}{6} + \frac{G_z}{4}\right)\cos(G_z u/2) \\ S_A &= \cos\left(\frac{G_x}{6} + \frac{G_y}{6} + \frac{G_z}{4}\right)\sin(G_z u/2) \end{aligned} \tag{B.26}$$

It is now convenient to adopt the methodology from zinc blende using the convention $a(\mathrm{ZB})=\sqrt{2}a(\mathrm{WZ})$, so that the set of reciprocal vectors become

$$\mathbf{G} = \frac{\sqrt{2}\pi}{a(\mathrm{WZ})}(G_x, G_y, G_z) \tag{B.27}$$

and thus the respective form factors can be calculated from the values in table B.2. Note that the values for $|\mathbf{G}|^2$ are for the ideal ratio between a and c, and that in the actual calculations, these values will be slightly different. [26, 27]

$\mathbf{G}$	$(1,0,0)$	$(0,0,2)$	$(1,0,1)$	$(1,0,2)$	$(2,1,0)$	$(1,0,3)$
$\lvert\mathbf{G}\rvert^2$	$2\frac{2}{3}$	3	$3\frac{5}{12}$	$5\frac{2}{3}$	8	$9\frac{5}{12}$
$V_S(\sqrt{G})$	−0.375	−0.305	−0.24	−0.07	0.045	0.07
$V_A(\sqrt{G})$	0	0.215	0.225	0.12	0	0.105

$\mathbf{G}$	$(2,0,0)$	$(2,1,2)$	$(2,0,1)$	$(0,0,4)$	$(2,0,2)$	$(1,0,4)$
$\lvert\mathbf{G}\rvert^2$	$10\frac{2}{3}$	11	$11\frac{5}{12}$	12	$13\frac{2}{3}$	$14\frac{2}{3}$
$V_S(\sqrt{G})$	0.06	0.09	0.045	0	0.035	0
$V_A(\sqrt{G})$	0	0.02	0.03	0.005	0.005	0.01

Table B.2: *Form factors for wurtzite GaN. Adopted from [28].*

Zincblende Modelling

C

The band diagram for zinc blende is calculated by discretizing a path along the high symmetry points of the irreducible Brillouin zone (500 k-points proved to be sufficient with this method). The high symmetry points are listed in units of $2\pi/a$. The path then moves from Γ =(0, 0, 0) to X=(1, 0, 0) to W=(1, 1/2, 0), then jumps to the U-point (since U and W are degenerate) and goes to L=(1/2, 1/2, 1/2) back to Γ and then to K=(3/4, 3/4, 0). The structure factors for the potential used for zinc blende are derived in appendix B.1. The structure factors are then readily calculated by finding any $|\vec{G}_{klm} - \vec{G}'_{klm}|^2$ equal to those listed in table B.1, for which the tabulated form factors are then used. The energies are then obtained by solving the eigenvalue problem in eq. 3.4 (this is accomplished using Matlab), and the calculated band diagram along with the corresponding density of states (DoS) are shown in figure C.1.

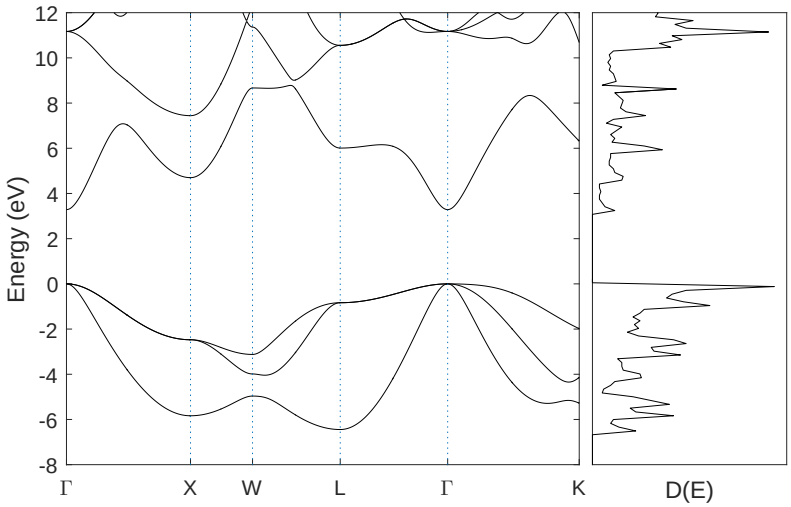


FIGURE C.1: *Calculated band structure for zinc blende GaN. The band diagram shows the three top most valence bands and a few of the lowest conduction bands. The corresponding density of states have been included as well.. The lattice constant used is a =4.5 Å.*

The calculated band structure in figure C.1 is in good agreement with the one obtained by Wang et al. [28], which is expected since the form factors used in the EPM listed in appendix B.1 are the same as those used by Wang et al.. The band gap energy obtained here is E_g =3.285 eV, which is slightly different from the one obtained by Wang et al. (Wang: E_g =3.26 eV). The available experimental and theoretical data for the band gap of zinc blende GaN typically lies in

the range E_g =3.2 eV to 3.35 eV [1, 28, 30]. The band gaps at the most critical points (locations with most degenerate bands) of the band diagram for zinc blende GaN are summarized in table C.1 along with the effective mass of the conduction band. The effective mass m_e^* is calculated by using the effective mass approximation

$$E_c(\vec{k}) = E_c(\vec{0}) + \frac{\hbar^2|\vec{k}|^2}{2m_e^*} \tag{C.1}$$

from which m_e^* is averaged over 20 k-points in the vicinity of the Γ-point. The value obtained here is in good agreement with the literature (see e.g. [28, 30]). For the hole masses two types of holes are typically considered. A light hole has a small effective mass and is thus referred to with respect to a band with a large gradient, since the effective mass is inversely proportional to the gradient of the band, and vice versa for heavy holes. The light and heavy effective hole masses are thus strongly dependent on the direction of the crystal, since the gradient of the bands differs. The effective hole masses are thus calculated in this respect, and the directions may be identified from the Brillouin zone scheme in figure 3.4(a) and compared with the curvature in the band diagram (e.g. the [111] direction corresponds to the path from Γ to L). To do the calculations the effective mass approximation is again utilized as

$$E_v(\vec{k}) = E_v(\vec{0}) - \frac{\hbar^2|\vec{k}|^2}{2m_h^*} \tag{C.2}$$

from which the results are also listed in table C.1. The light hole effective masses m_{lh}^* are in good agreement with those obtained by Ziade et al. [30]. Though the heavy hole effective masses m_{hh}^* tend to be lower than those calculated by Ziade et al., they also included many different values from the literature, which also tend to be lower than those obtained by Ziade et al., and in general in better agreement with those calculated here.

Critical points	Γ	L	X
$E_c - E_v$ [eV]	3.285	6.843	7.173

Effective mass	m_e^*	[100]	[111]	[110]
units of m_0	0.139	$m_{lh}^* = 0.162$ $m_{hh}^* = 0.871$	$m_{lh}^* = 0.132$ $m_{hh}^* = 2.235$	$m_{lh}^* = 0.141$ $m_{hh}^* = 0.866$

Table C.1: *Showing the energy difference between the lowest conduction band and the topmost valence band for a few critical points in the zinc blende band diagram. The calculated effective masses are given in units of the free electron mass* m_0.

The DoS are also calculated with respect to the band diagram. The occupancy is obviously largest for energy states, where the bands are flattest and more degenerate. The atomic orbitals for Ga and N are

Ga: $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} \mathbf{4s^2 4p}$ N: $1s^2 \mathbf{2s^2 2p^3}$

and the valence electrons are highlighted. Since there are one Ga and one N in the unit cell for zinc blende GaN, this means there are eight valence electrons. It can be discussed, whether the d-electrons from Ga participate in any bonding, however, their contribution is small compared to the others (see e.g. Dugdale [26], who studied the contribution from both valence and d-electrons and compared their contribution to the DoS). The Fermi level can then readily be visualized

from the DoS by filling up the states from the lowest energy levels until the area beneath the curves equals the amount of valence electrons in the unit cell. These are thus the valence bands and the Fermi level then lies in the middle of the gap before the empty conduction band states are reached.

It is now of interest to evaluate the optical response of zinc blende GaN. This is accomplished through linear response theory, from which an expression for the susceptibility $\chi(\omega)$ as a function of the frequency of light can be obtained. An introduction to linear response theory followed by the derivation of $\chi(\omega)$ are given in appendix A. The final result is given in eq. A.31, repeated here

$$\chi(\omega) = \frac{4e^2\hbar^2}{\epsilon_0 m^2 \Omega} \sum_{m\in c} \sum_{n\in v} \int \frac{|\langle\psi_m|\,\hat{p}_z\,|\psi_n\rangle|^2}{E_{mn}(E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2)} \mathrm{d}^3k \tag{C.3}$$

where the volume is calculated as $\Omega = \vec{a}_1 \cdot \vec{a}_2 \times \vec{a}_3$. Since zinc blende is an isotropic crystal the integral can be evaluated through a spatial average over the irreducible Brillouin zone, where the mean value of the impulse matrix moment can be evaluated separately for each direction in the irreducible Brillouin zone and divided by 3, i.e.

$$|P_{vc}^z|^2 = 48(|P_{vc}^x|^2 + |P_{vc}^y|^2 + |P_{vc}^z|^2)/3 \tag{C.4}$$

since the irreducible Brillouin zone can be repeated 48 times due to degeneracy of the energy states. The wavefunctions, however, are not equal for each zone, and must thus be evaluated through the mean value of the impulse matrix moment. The integral in eq. C.3 is then rewritten as

$$\left\langle \frac{(|P_{vc}^x|^2 + |P_{vc}^y|^2 + |P_{vc}^z|^2)/3}{E_{mn}(E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2)} \right\rangle_{IBZ} = \frac{1}{3N_{\vec{k}}} \sum_{\vec{k}} \frac{|P_{vc}^x|^2 + |P_{vc}^y|^2 + |P_{vc}^z|^2}{E_{mn}(E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2)} \tag{C.5}$$

where $N_{\vec{k}}$ is the number of k-points in the sum and the $\langle ... \rangle$ denotes the spatial average. Using this in eq. C.3 the final expression for the zinc blende GaN becomes

$$\chi(\omega) = \frac{4e^2\hbar^2}{3N_{\vec{k}}\epsilon_0 m^2 \Omega} \sum_n \sum_m \sum_{\vec{k}} \frac{|P_{nm}^x|^2 + |P_{nm}^y|^2 + |P_{nm}^z|^2}{E_{mn}(E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2)} \tag{C.6}$$

The impulse matrix elements may evaluated by using the Bloch functions and the definition of the impulse operator

$$\hat{\vec{p}} = \frac{\hbar}{i}\nabla \quad ; \quad \psi_{n\vec{k}} = \sum_{\vec{G}} \vec{c}_n(\vec{G})\phi_{\vec{G}}(\vec{r}) \quad ; \quad \phi_{\vec{G}}(\vec{r}) = \frac{1}{\sqrt{\Omega}} \mathrm{e}^{i(\vec{G}+\vec{k})\cdot\vec{r}} \tag{C.7}$$

where $\vec{c}_n(\vec{G})$ are the eigencoefficients. Thus the impulse matrix moment with respect to $\phi_{\vec{G}}$ becomes

$$\begin{aligned} \langle\phi_{\vec{G}}|\,\frac{\hbar}{i}\frac{\mathrm{d}}{\mathrm{d}x}\,|\phi_{\vec{G}'}\rangle &= \frac{1}{\Omega}\int \mathrm{e}^{-i(\vec{G}+\vec{k})\cdot\vec{r}}(G_x' + k_x)\mathrm{e}^{i(\vec{G}'+\vec{k})\cdot\vec{r}}\mathrm{d}^3r \\ &= \hbar(G_x + k_x)\delta_{\vec{G},\vec{G}'} \end{aligned} \tag{C.8}$$

and the k_x term vanishes in the final expression, since it does not have an effect on the eigencoefficients that have the property $\vec{c}_n^{\dagger} \cdot \vec{c}_m = \delta_{nm}$. However, n and m are never equal because they serve as indexes for the conduction band and valence band, respectively. The impulse matrix moments are then readily calculated as

$$P_{nm}^x = \hbar\vec{c}_n^{\dagger} \cdot P_{G,G}^x \cdot \vec{c}_m \tag{C.9}$$

where $P^x_{G,G}$ is a square matrix with its diagonal containing all x-values in G_{klm} and zeroes elsewhere. Similar expressions then follow readily for the y and z part. The energies are obtained along with the eigencoefficients by solving the eigenvalue problem in eq. 3.4 for the entire volume of the irreducible Brillouin zone. Finally, the real and imaginary parts of the dielectric function are plotted with respect to the photon energy in figure C.2, where it was used that $\epsilon = 1 + \chi$. The imaginary part of χ is in fact proportional to the absorption spectrum, and it is therefore zero until the photon energy is larger than the band gap. When the curvature of $\mathrm{Im}(\chi)$ changes rapidly this corresponds to locations where the energy bands flatten.

The spectrum is in good agreement with the one obtained by Wang et al., showing the same overall magnitude and features. Other authors (see e.g. Dugdale [26]) that have used different pseudopotentials or included e.g. spin-orbit coupling also achieve similar results, where the most prominent features are reproduced. The magnitude, however, may be somewhat different, e.g. resulting in a set of different values for $\mathrm{Re}(\epsilon(0))$, where e.g. Dugdale found $\mathrm{Re}(\epsilon(0))$ =6.1, which is larger than the value obtained by Wang et al. of $\mathrm{Re}(\epsilon(0))$ =5.16. In the calculations made here the result is $\mathrm{Re}(\epsilon(0))$ =5.35. Furthermore, Wang et al. observe only a single peak around 10-11 eV, whereas the result here shows two peaks in this region. Such details could likely be ascribed to the broadening factor. Wang et al. used a Gaussian function to describe the broadening, whereas in this report it was a fixed value. Logothetidis et al. [105] measured the dielectric properties of GaN samples produced by plasma-assisted molecular beam epitaxy with silicon and sapphire substrates. Several of their observed peaks for the imaginary part of the dielectric constant can be matched with those seen in the calculations given here. However, they also found peaks at 7.1 eV and 7.6 eV, which cannot be distinguished in the spectrum shown here. In the calculations made here, a very prominent peak can be observed around 7.45 eV, which is not observed in the experiments by Logothetidis et al., which could suggest that the two peaks they observed are shifted with respect to the one seen here. This is in accordance with Wang et al. [28] and Dugdale [26], and Wang et al. suggest that the 7.45 eV peak could have been red-shifted or inhibited due to e.g. strain in the samples grown by Logothetidis et al. This statement may be supported by Morkoc [1], who points out that the strain in GaN can be described by a deformation potential that tends to cause a shift and a splitting of the top most valence bands to higher energies.

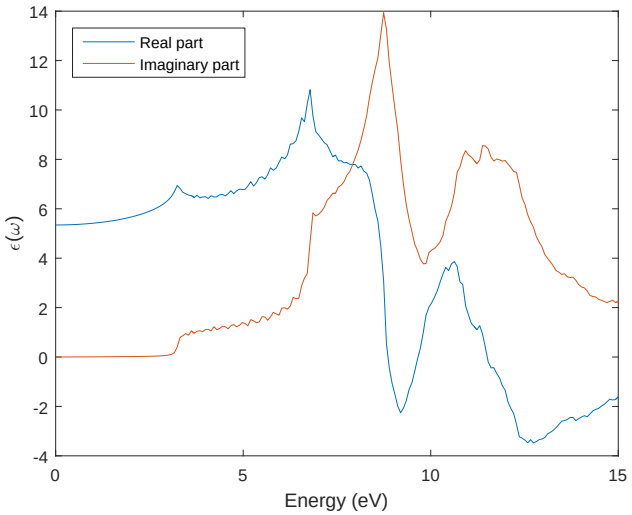


Figure C.2: *Calculated optical dielectric function for zinc blende GaN. The broadening was chosen as $\hbar\Gamma$ =0.05 eV. The IBZ was discretized in 36936 k-points.*

Experimental procedures

D

D.1 Molecular beam epitaxy: setup and growth

This section describes the MBE setup used for sample fabrication during this thesis along with some of the growth parameters and related investigations. A schematic over the MBE system is shown in figure D.1 (not drawn to scale, and the components are not all placed at their correct positions). The setup is comprised of a load-lock and a main chamber. The two vacuum chambers are pumped by turbo pumps, which are connected to the same backing pump. The load-lock is seen to the left of the figure. Here, the sample is mounted on a sample holder, which is then placed on the translator rod so it can be placed in the sample stage. Fresh samples are always degassed before any preparation. It should be mentioned that the nitrogen and gallium sources are aligned vertically with each other and positioned with the same relative angle to the sample stage. The Al source also has the same relative angle, but it is not aligned with the other sources. The He-Ne laser is placed underneath the chamber, such that the beam is nearly normally incident on the sample. All data (pressure, temperatures, reflectance) are periodically logged with respect to time.

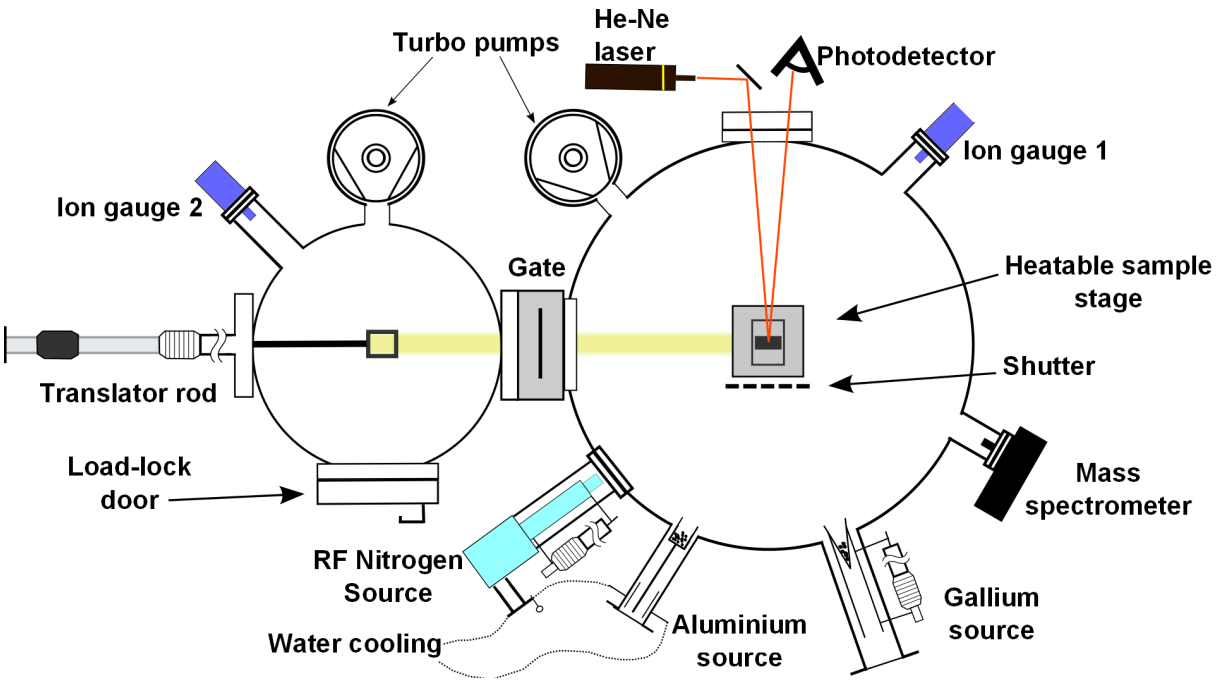


FIGURE D.1: *Schematic over the MBE system showing all components.*

Achieving UHV

The load-lock chamber is used for sample substitution without having to bake the main chamber. Whenever the main chamber was opened, the nitrogen source was run for 30 min with atomic plasma in order to remove water from the source. Then a bake-out was performed to remove adsorbed water and reduce the chamber pressure. All filaments were heated to 400 °C during bake-out, and the bake-out typically lasted for 16-24 hours. During bake-out the mass spectrometer is used to observe the partial pressure of water with respect to air and the total chamber pressure. After the bake-out all filaments (unless they are not to be used) are degassed slightly above their maximum operating temperature. This procedure typically gave a stable minimum pressure of 1.3×10^{-10} mbar at room temperature.

D.1.1 Sources

RF Nitrogen source

The RF Nitrogen source works by having a flow of molecular nitrogen (99.999% purity), which is then turned into a plasma of activated nitrogen that flows towards the sample. An oscillating electric field excites the vibrational state of the nitrogen molecules. When the plasma is ignited, two types of plasmas are obtainable, where one is just vibrationally excited N_2 and the other is atomic nitrogen. Here the procedure was to ignite the plasma with the RF source at 400 W and a nitrogen flow of 5 sccm. The plasma achieved here is ionized N_2, as is evident from the recorded spectrum (see figure D.2). The flow is then slowly reduced to 2 sccm, where it becomes primarily atomic plasma (see again figure D.2). [56]

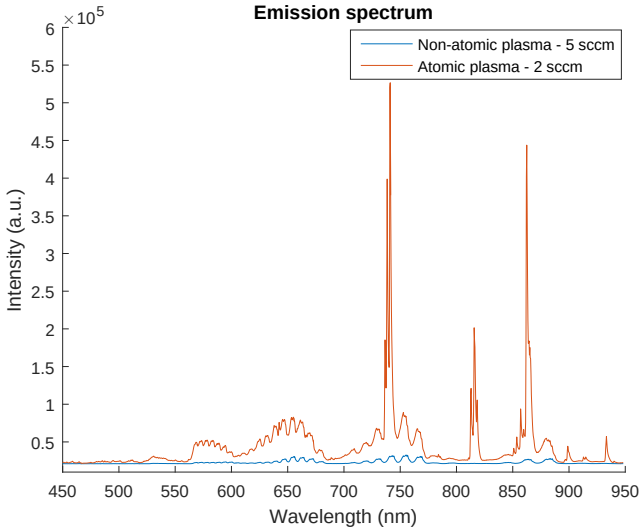


Figure D.2: *Emission spectrum from two types of nitrogen plasma ignited by the RF nitrogen source. The spectra were obtained by connecting an optical fiber to the source.*

Gallium effusion source

The Ga source is comprised of a conical PBN crucible containing the Ga (99.999% purity), which is heated by a shielded filament. The heating is controlled with two separate power suppliers, one heats the main part of the source and the other heats the tip of the source. Both are temperature controlled using built-in thermocouples with a constant heating (and cooling) rate of $\approx$10 °C min^{-1}. This, along with the conical shape of the crucible, should ensure a more

uniform Ga flux, and especially the hot lip should prevent Ga condensation into droplets, which can e.g. lower the flux and increase the density of oval defects in the film. The conical shape also help to preserve the Ga flux as the Ga melt is depleted from the Ga effusion source. Effusion means that the particle mean free path is larger than the crucible diameter, and therefore they should exit the source without making any collisions with the crucible walls.

An important consideration when handling Ga is the phase transition from liquid to solid Ga, in which the density of Ga decreases by ≈3.1%. This material expansion is dangerous to the crucible, because the liquid Ga first freezes at the top layer and thereby forms a tight cap suppressing expansion in the upwards direction of the crucible. This only leaves expansion towards the sides of the crucible as a possibility and can thus lead to crucible cracking, an unfortunate event that was also experienced during the work with this thesis. Fortunately, the phase transition occurs at 29 °C and it can thus be prevented by heating the Ga. Therefore, after we experienced crucible cracking, the Ga source was always kept at 50 °C.

Aluminium source

The Al source is a simple homebuilt design, where a cylindrical alumina crucible containing the Al is embraced and held by a winded wolfram wire filament. There is no thermocouple attached to it, but the power needed for Al deposition was found using a quartz resonator and also using the in-situ thickness measurement from the HeNe laser reflection. The size of the Al source limits its deposition rate, which was found to be ≈25 nm/h with a DC filament power of ≈370 W. This growth rate could be sufficient for deposition of AlN nucleation layers under very N-rich conditions, however, due to the breakdown of the Ga source, the growth of AlN nucleation layers was not further studied in this thesis.

D.1.2 Heatable sample stage

The sample stage is heated by a wolfram filament, and the temperature of the stage is measured with a thermocouple. The heating rate is controlled by a PID controller, which adjusts the time in which power is fed to the filament with respect to its temperature. When a sample is mounted in the stage, the sample temperature will differ from that of the filament. In order to estimate the sample temperature, an in-situ thickness measurement of a sapphire substrate was carried out by measuring the reflectance of the HeNe laser from a piece of sapphire polished on both sides. Thus by collecting the thickness at different temperatures and comparing these with the thermal expansion of sapphire, a rough estimate was obtained for the actual sample temperature with a maximum temperature of 750 °C, which in turn corresponded to a temperature of 1150 °C on the filament.

D.1.3 In-situ thickness measurements

A HeNe laser nearly normally incident on the sample is used to perform in-situ thickness measurements of the thin films during growth. The laser is positioned beneath the chamber, where a mirror is used to direct the beam through a window and onto the sample, from which the reflectance is measured with a photodetector. The measured intensity due to the reflection is logged with respect to time. As an example a plot is shown in figure D.3, which is recorded during a growth run of GaN on sapphire (only polished on one side). Assuming normal incidence the thickness and growth rate can then be found, i.e. as the reflection from the GaN surface

and the GaN/Sapphire interface interfere constructively, when

$$i\frac{\lambda}{n} = 2d \qquad \longrightarrow \qquad d = i\frac{\lambda}{2n} \tag{D.1}$$

where i is an integer, λ is the wavelength, n is the refractive index of the film (for GaN $n(\lambda = 633\,\text{nm}) = 2.38$), and d is the film thickness. Thus for the first intensity maxima, the GaN film thickness should be 133 nm, and taking this with respect to time gives a growth rate of ≈50 nm/h, which was later confirmed (roughly) with EDX (the method of which is described later). Due to the mentioned issues with the gallium source, however, the growth rate was not constant, and this figure is merely an example from one deposition.

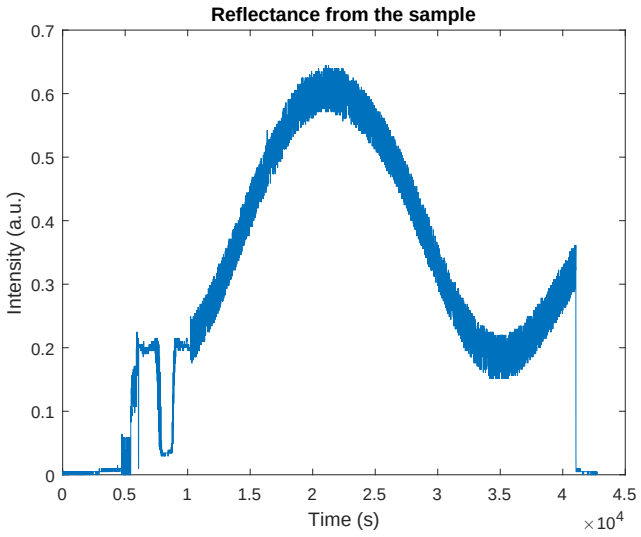


Figure D.3: *Reflectance measurement during the growth of a GaN sample.*

D.1.4 Growth of GaN

Having described the most vital parts of the MBE system, it is now natural to describe the growth of GaN. In MBE there are only two variable parameters; particle fluxes and sample/growth temperature. However, the particle fluxes can be manipulated in a number of ways. Both sources (N and Ga) can have their distance to the sample adjusted. For the N source, the flow of N_2 can be adjusted from 0 to 10 sccm, and the power used to generate the plasma was held constant at 400 W.

The Ga source is operated by having control over the temperature in Ga melt and at the tip of the source. A so called *hot lip* source is often used for Ga evaporation, where the temperature at the tip can seriously affect the Ga flux, because it helps reduce Ga condensation as it moves out of the crucible. Ga condensation can happen between adjacent Ga atoms/molecules or the walls of the crucible. When condensed at the walls, the Ga droplets decent to the Ga melt and this can result in spitting of Ga droplets from the source. These droplets along with the condensed forms of Ga (that did not stick to the crucible wall) are suspected to be a common source of the oval defects observed in Ga-compound films [53]. The degree of condensation can often be minimized by having the temperature of the source lip higher than the Ga melt.

The temperature at which Ga starts to evaporate from the source was determined to be ≈945 °C. This was determined during the growth of one GaN sample, where the N flux was directed onto the sample during the heating of the Ga source, and at the time where the chamber

pressure started to drop (corresponding to some of the nitrogen being used for growth) could then be compared to the Ga source temperature. The monitored pressure along with the monitored Ga source temperature is shown in figure D.4, and a drop in the pressure of $\approx 1 \times 10^{-5}$ mbar (around 2.25×10^4 s on the x-axis) occurs in the time period where the Ga source temperature goes from 945 °C to 1100 °C, after which the pressure stabilizes due to the source temperature reaching a constant value (the pressure before opening the nitrogen flow was in the range of 10^{-9} to 10^{-8} mbar).

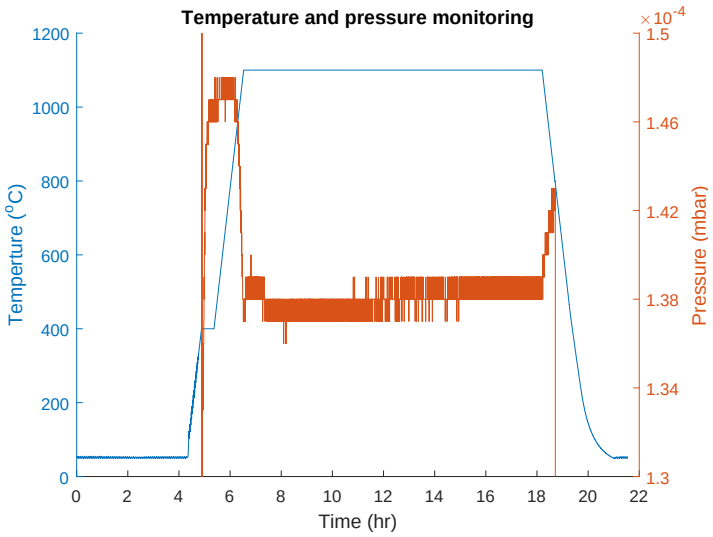


FIGURE D.4: *Ga effusion source temperature and chamber pressure are shown along the same time axis for the growth of one sample.*

All samples are grown without sample rotation. Since the N -and Ga fluxes are incident at the same angle on the sample, this naturally results in a difference in III/V ratio and thereby a thickness gradient across each sample. The thickness difference was found to up to around 20 nm.

D.2 Preparation of substrates

In this section the preparation of substrates for epitaxial growth is described. All substrate were sonicated (ultrasonic cleaning in acetone, DI water, and ethanol, each for 2-5 minutes) prior to any experimental use.

Sapphire substrates

The surface roughness of sapphire substrates was investigated with AFM, starting with a cleaned as-received sapphire substrate. A representative AFM image of the surface is shown in figure D.5(a). A root-mean-square (RMS) roughness of 0.669 nm was found, and no atomic steps are observed. For the epitaxial growth of GaN on sapphire, the surface of the substrate serves as the foundation for the overlying GaN layer. Cui et al. [106] reported a method for achieving atomically smooth (stepped) sapphire surfaces with low RMS roughness, and they saw that this severely improved the crystalline quality of the GaN layer. Thus it is of essence to achieve stepped sapphire surfaces for the work in this thesis as well. Cui et al. reported that a simple annealing in air at 1380 °C for 1 hour formed atomically flat surfaces.

In this thesis it was only possible to anneal at 1200 °C. Thus all annealing procedures were performed in air at 1200 °C with the substrate enclosed by an alumina cup. Annealing for 1 hour did not yield a stepped surface, and the RMS roughness was not improved. Therefore a sapphire substrate was annealed over night (about 14 hours) at 1200 °C, and this gave the AFM image in figure D.5(b), where atomic steps are clearly observed. The RMS for the entire image is 0.239 nm, and for a plane levelled image the RMS on a terrace is 0.11 nm. The surface could still seem to suffer from polishing damage and judging from the phase image (not shown here), most of the artefacts are also Al_2O_3. Cui et al. [106] also reported similar findings, and therefore they did an etching procedure prior to the annealing, which gave better results. Prior to etching, the substrates were cleaned with a degreasing procedure, where they were dipped first in TriChloroEthylene (TCE) for 5 minutes, and then rinsed in acetone, methanol, and DI water for 3 min each (repeated 3 times). The etching procedure was then to dip the substrates in a 300 °C 3:1 solution of H_2SO_4:H_3PO_4 for 20 min, and then followed by dipping the substrate in DI water for 5 minutes. This equivalent degreasing and etching procedure was followed in this thesis as well, and the result can be seen in figure D.5(c), that shows an AFM image of the surface with RMS roughness for the entire image of 0.167 nm, and the RMS on the terraces is 0.09 nm.

A similar sample was prepared without the degreasing procedure (thus it was only cleaned by the standard procedure mentioned in the beginning). The AFM image for this sample is shown in figure D.5(d) with an RMS for the entire image of 0.216 nm and for terraces it is 0.15 nm. The results are summarized in table D.1. Judging from these values it is quite important to degrease the substrates prior to the etching procedure, if the etching is to have a pronounced effect. Nevertheless, all three preparation methods seem to severely improve the substrate quality. The step height was measured to be ≈0.23 nm, which somewhat matches the theoretical value of 0.22 nm between two adjacent oxygen layers [107].

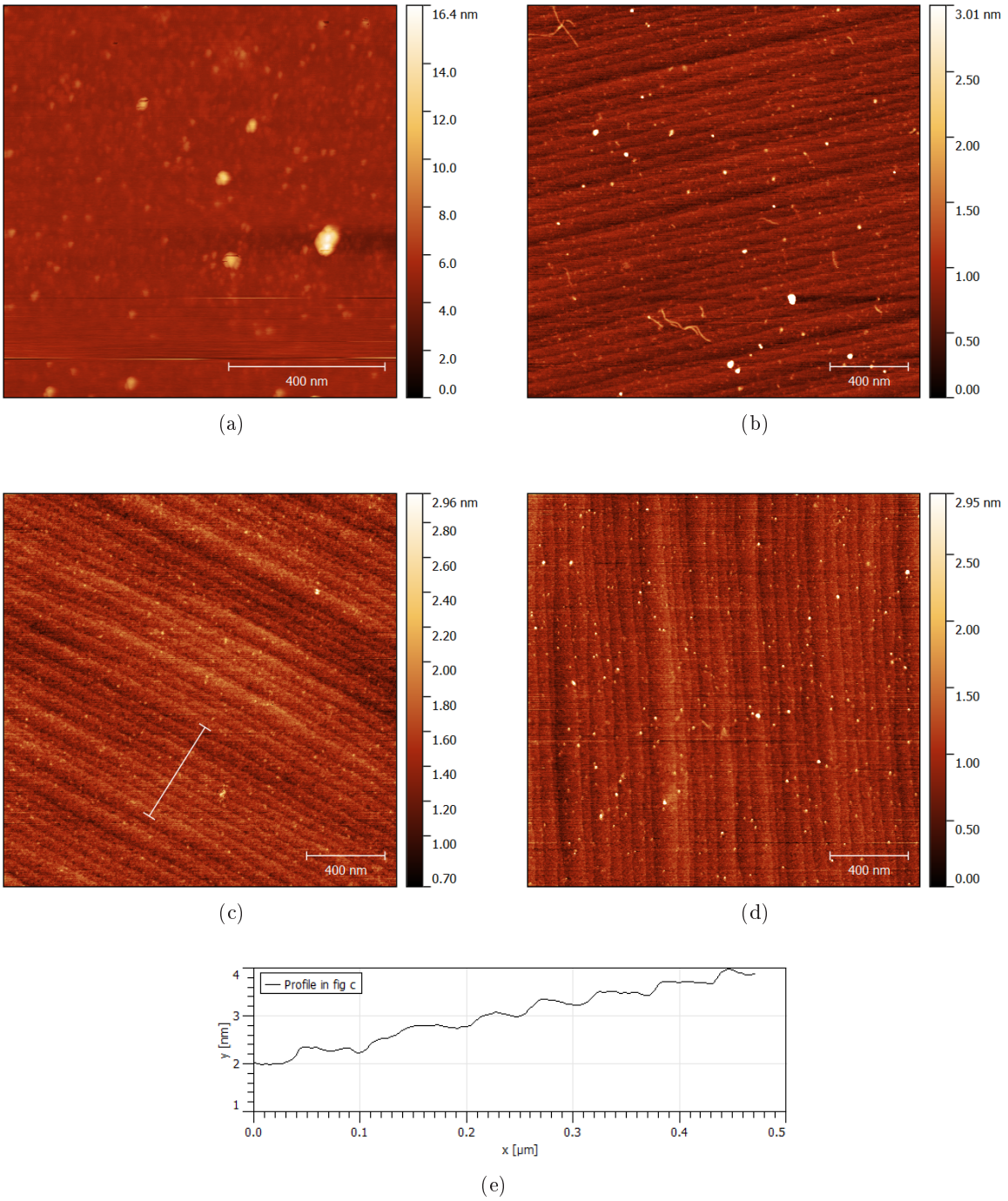


FIGURE D.5: *AFM surface images of sapphire substrates, that has undergone different treatments by which the specifics are described in the text above.* ***a****: sonication.* ***b****: sonication and annealing.* ***c*** *sonication, degreasing, etching, and annealing.* ***d****: sonication, etching, and annealing.* ***e****: shows the profile over a few steps in figure* ***c*** *for a plane levelled image.*

AFM image	**a**	**b**	**c**	**d**
Total RMS roughness (nm)	0.669	0.239	0.167	0.216
Terrace RMS roughness (nm)	-	0.11	0.09	0.15

TABLE D.1: *Summary of RMS values from the AFM images in figure D.5.*

D.3 Characterization methods

This section describes the technical details of the experimental equipment and how it was used for various characterizations schemes.

D.3.1 Atomic force microscopy

Atomic force microscopy (AFM) is a vital tool for characterizing the surface morphology of a sample. Several samples (both substrates and GaN samples) were characterized using an NT MDT AFM in semi-contact mode under ambient conditions. Surface charge on the GaN and sapphire samples was an obstacle during measurements. In order to overcome this, conductive tape was connected to the GaN layer and grounded, as well as using various conductive cantilevers. The most effective cantilevers were TiN coated, which dramatically reduced the polarisation effects, but the tip radius of these was 35 nm, which put a limit on the obtainable resolution.

D.3.2 Scanning electron microscopy and energy-dispersive X-ray spectroscopy

Scanning electron microscopy (SEM) was also used to characterize the surface structure of GaN samples. Here it was also a problem with polarization fields and charge building up on the surface during electron scanning, which was also circumvented by grounding the GaN film with conductive tape. The thickness of the as-grown GaN samples was identified using energy-dispersive X-ray (EDX) spectroscopy and also focused ion beam (FIB) cutting. The thickness was found with EDX by comparing the measurements to a model created with the CASINO software, which utilizes monte carlo simulations to track electron trajectories in solids. In this software the electron energy (in keV) is chosen to match the value of the microscope and then the sample composition is chosen with i.e. Al_2O_3 as substrate and GaN film on top. The GaN film thickness can then be adjusted and the ratio of Ga to Al counts is then compared to the ratio measured with EDX. Also a FIB was used to cut through the GaN layer and into the substrate. The thickness found in this manner always proved to match that found with EDX within 25 nm (keep in mind that the thickness was never uniform across MBE grown samples).

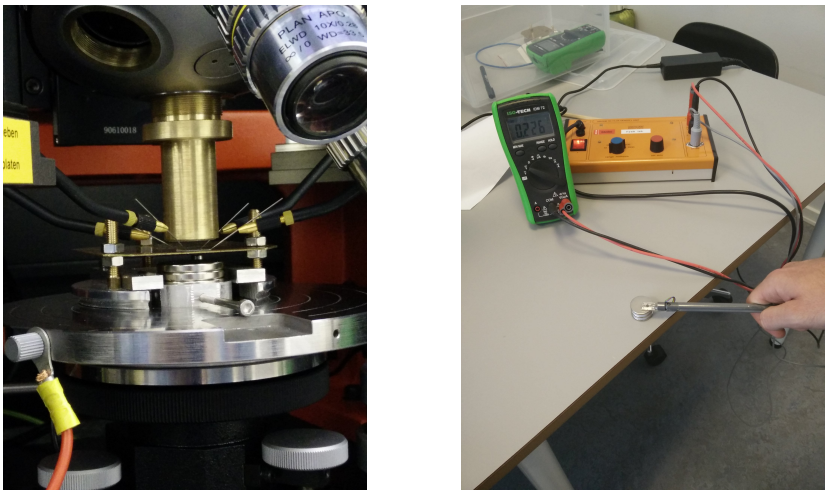
D.3.3 Sheet resistance and Hall effect

A four point probe (FPP) system was used to carry out electrical characterization of the GaN samples. A van der Pauw geometry would be ideal for achieving more accuracy in the measurements. However, the best GaN etch rate achieved with the available reactive ion etcher (RIE) was 1.5 nm/min using CHF_3 as precursor with a flow rate of 25 sccm, a chamber pressure of 10 mTorr, and a plasma power of 75 W. This etch rate was not viable for the formation of van der Pauw patterns, and thus the more simple geometry with small contacts at the corners of a squared sample was used. Alloyed indium contacts were formed using a soldering iron. Indium was used because it is rather easy to process, and more importantly it forms ohmic contacts with n-type GaN [108]. The ohmic nature of the contacts was confirmed with the FPP.

The sheet resistance was measured with the FPP in accordance with the van der Pauw method described in section 2.3.2, and the sheet resistance R_s was found using the van der Pauw equation. The sheet carrier density and Hall mobility were measured using the method described in section 2.3.3. It was found that in order to acquire accurate measurements, the

four wolfram probes from the FPP should not be moved between measurements with varying magnetic field. This was first circumvented by positioning a permanent magnet on the objective stage, which could then be moved close to the surface of the sample.

However, it is important that the magnetic field is uniform throughout the sample and moreover that it is perpendicular to the electric field through the sample. Therefore another permanent magnet consisting of three 2.5 cm diameter disc magnets was positioned slightly below the sample with. A small aluminium "table" was constructed, such that the sample could be fixed along with the probes on top of the table. The sample is then placed between two permanent magnets with north and south pole positioned towards each other, such that the magnetic field lines points from the topmost magnet to bottom magnets. This configuration is shown in figure D.6(a), and it is believed to ensure a stronger magnetic field strength along with perpendicular field lines through the sample. The magnetic field is then varied using the microscope stage, where the topmost magnet could be moved vertically with respect to the sample to known positions. After the Hall voltage had been measured the magnetic field was then measured at the 5-10 different vertical positions of the top magnet. The measurements were made using a magnetic flux density unit utilizing a hand held stick, and an example measurement is shown in figure D.6(b).



(a) (b)

FIGURE D.6: **a**: *shows the FFP system during Hall measurements on a GaN sample. Permanent magnets with opposite poles towards the sample are positioned at either side of it, with the top magnet being placed at the tip of the brass cylinder.* **b**: *shows the magnetic flux density unit during a test measurement on the disc magnets.*

The Hall setup was tested on a p-type B-doped silicon wafer with standard ohmic contacts formed by depositing and annealing aluminium contacts. The wafer should have a resistivity between 5 and 20 Ω·cm as stated by the manufacturer. The exact resistivity was found from van der Pauw measurements, and then compared to the sheet density measured with the Hall method. The doping concentration was measured to be $1.95{\cdot}10^{15}\mathrm{cm}^{-3}$, which corresponds to a resistivity of 7.00Ω·cm for the B-doped Si wafers. For all measurements it was typical that two different results were obtained depending on the direction, where the Hall voltage was measured. Remembering that no van der Pauw clovers were used for any of the sample, the differing results with respect to measurement direction is likely a result of the Hall field not being measured completely perpendicular to the direction of the current. Therefore the average value between

the two is always used, since the average is expected to cancel out this error. Furthermore, several measurements on the same GaN sample (MR 1) with removal and adhesion of contacts three times yielded very similar results within an accuracy of 5%. Similar results were also obtained for different 1 cm^2 pieces from the same sample.

D.3.4 Raman spectroscopy

Raman spectroscopy was used in this thesis for the analysis of biaxial strain in GaN films, which can give an indication of proper growth parameters as well as crystalline quality. The exciting light was the 488 nm line from an Ar-ion laser, and the spectral resolution of below 1 cm^{-1} was achieved through the use of two diffraction gratings. Typical Raman intensity for high quality GaN samples was around 150 counts per second, which is much larger than the background dark spectrum. Focus was achieved manually using an optical microscope, and the depth resolution is expected to be around one micron by observing the variation of the AlN and GaN peak ratio on samples containing both materials. The geometry used was $z(xx)\bar{z}$ with no polarization filter.

D.3.5 X-ray diffractometry

X-ray diffractometry (XRD) is one of the most common and versatile methods for investigating the crystalline properties of solids. The basic setup consists of an x-ray source, a sample stage, and an x-ray detector. By emitting soft x-rays, which have wavelengths comparable to the lattice constant, atomic planes in the crystal serve as diffraction gratings as the core electrons elastically scatter the beam, and thus by varying the angle of incidence, an interference pattern can be observed which is sensitive to crystal structure. Constructive interference occurs when Bragg's law is fulfilled

$$n\lambda = 2d\sin\theta \tag{D.2}$$

where n is the diffraction order, λ is the x-ray wavelength, d is the distance between atomic planes, and θ is the angle of incidence. From this simple condition, information can be extracted about crystal structure, orientation and strain. There are two primary scan modes in XRD when used for thin film investigation. In the first mode, the incidence angle is scanned while the source and the detector have the same angle towards the sample, which maintains the orientation of $\Delta k = k_s - k_i$ perpendicular to the sample surface. Through this, angles which result in constructive interference give rise to a sharp peak, which can be used to identify the structure. In the second mode, the source is fixed at some angle where a peak has been identified, while the detector angle is varied around the same angle, which slightly changes the orientation of the wave vector. The spectra from this mode is called a *rocking curve*, and shows how well-defined the constructive interference diffraction angle is. By measuring the diffraction from different planes in the crystal, information about the mosaicity of the crystal can be obtained from the width of the rocking curve, as a wide rocking curve for the (0001) plane indicates a large amount of tilt between domains, while wide rocking curve in the $(01-10)$ direction indicates a large amount of twisting. As such, the rocking curve can be seen as an indicator of crystalline quality.

The XRD equipment used in this project was a PANalytical Empyrean. The x-ray source was some copper target which emits x-rays with two different wavelengths. Since no filter was available for one of them, the diffraction signal from one of them was removed ("stripped") using the native software. Rocking curve equipment was not available, and thus the XRD technique was primarily used for identification of the presence of wurtzite GaN. All samples were measured

only in the geometry with Δk perpendicular to the sample surface. The sample was rotated around the c-axis in order to average contributions from in-plane directions.

D.3.6 Photoluminescence spectroscopy

Four different optical geometries labelled geometry 1-4 were used for PL measurements during this thesis. The geometries and their optical hardware are shown in figure D.7. For the Ar-ion laser, the penetration depth in GaN is $\sim 160 - 200$nm.

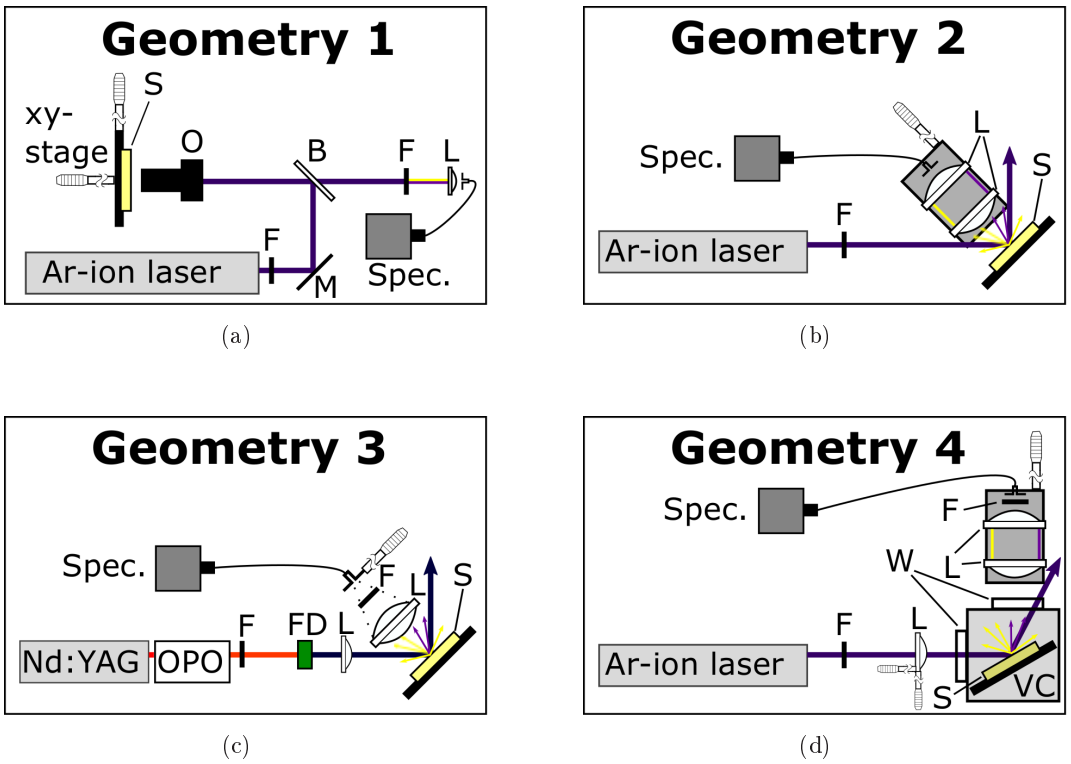


FIGURE D.7: *Optical geometries used for PL measurements during this thesis. The abbreviates are as follows; B: Beamsplitter; F: Filter; FD: Frequency doubler; L: Lens; M: Mirror; O: Objective; S: Sample; VC: vacuum chamber; W: window. For the Ar-ion laser, the 351 nm line is used. For the Nd:YAG laser, the 600 nm line is extracted and frequency doubled to 300 nm using an anisotropic crystal. In all geometries the luminescence from the sample is focused onto an optical fibre and thus coupled into the spectrometer.*

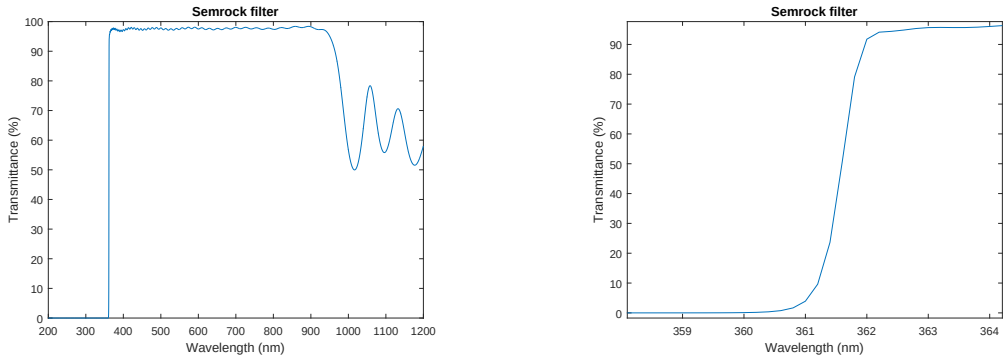
Geometry 1 is somewhat similar to the geometry of the Raman setup and therefore gives rise to parasitic Raman lines near the NBE, when measuring on GaN. Furthermore, the beamsplitter and laser line filter both had some obscure effects due to their absorption spectrum having nonzero values in the NBE wavelength region, see [109, 110] for more details on these components. Despite these unfavourable effects, geometry 1 could still be deployed for studying e.g. the yellow luminescence of GaN samples.

Geometry 2 is different from geometry 1, in that the laser is not focused on the sample, giving a spot size of several millimetres in diameter. This significantly reduced the magnitude of the Raman lines, which were rarely seen in this geometry. Also the geometry is designed for diffuse collection of the PL in order to avoid using the laser line filter and beamsplitter from geometry 1. Without the laser line filter, some laser line will thus be collected in this geometry due to diffuse reflectance from the GaN surface, e.g. due to surface roughness.

Geometry 3 utilizes the Nd:YAG laser, which was used to move the excitation energy further away from the band gap energy of GaN, avoiding significant Raman scattering, which, along with a 325 nm filter, made this geometry ideal for studying the NBE luminescence, since this was found to be easier to resolve under focusing.

Geometry 4 is used for cooled PL measurements. The sample is placed in a vacuum chamber with a helium compressor attached. The temperature is measured from two thermocouples located near the sample. Some lack of accuracy regarding the temperature measurement must be mentioned, probably due to poor thermocouple attachment. Minimum temperature reached was 21 K. The first lens was necessary here, since it is used to direct the laser beam onto the sample. A filter was needed to remove the laser line, due to a small solid angle with the accumulating optics. Note here that for low temperatures the filter cuts a significant amount of the NBE luminescence, see figure D.8(b).

The transmittance spectrum for **the laser line filter** used for the Ar-ion laser is shown in figure D.8.



(a) (b)

Figure D.8: *Transmittance spectrum for the filter used for cutting the Ar-ion laser line in some PL experiments.*

Determination of the activation energy

The activation energy, E_A, can be determined by first finding the region of thermal quenching, i.e. the region where a logarithmic intensity vs inverse temperature graph falls off linearly with increasing temperature. Here the slope can be directly related to the activation energy by using eq. 3.16 to isolate the activation energy as

$$\begin{aligned} I(T) &= \frac{I_0}{1 + (1-\eta_i)\tau_R C_p g^{-1} N_v \exp(-E_A/kT)} \\ \frac{I_0}{I(T)} - 1 &= (1-\eta_i)\tau_R C_p g^{-1} N_v \exp(-E_A/kT) \\ \ln\left(\frac{I_0}{I(T)} - 1\right) - \ln((1-\eta_i)\tau_R g^{-1} C_p N_v) &= -E_A/kT \end{aligned} \tag{D.3}$$

where any term unrelated to T can be removed in respect to the slope, which becomes

$$-E_A/kT = \ln\left(\frac{I_0}{I(T)} - 1\right) - \frac{3}{2}\ln(T) - \frac{1}{2}\ln(T) \tag{D.4}$$

where the slope is then found to be E_A/k.

Surface passivation of GaN

Martinez et al. [111] investigated the passivation of surface states on GaN by using various sulfide related compounds. From their work a recipe was adopted primarily due to availability of the chemicals. The passivation procedure was as follows. First the sample was cleaned by sonication. Then the sample was immersed for 1-2 min in a buffered oxide etch (BOE) in order to remove any oxide layer, followed by a rinse in water. Finally, the sample was immersed in the passivating solution of 1 M Na_2S for 5 min followed by a blow dry.

Cathodeluminescence

Cathodeluminescence (CL) was used as a supplementary characterization method to PL. CL was performed in the SEM, where an optical fibre was positioned in a small stage with a small focusing lens, which should focus light into the fibre. The SEM was then used for positioning and electron irradiation on the sample. Here both spot size and energy of the incident electron beam could be controlled. The spectrometer used to analyse the CL spectrum was the same as for PL measurements.

D.3.7 Sample overview

Sample name	Growth method	Composition	GaN doping	GaN thickness	AlN thickness	Nucleation layer	Substrate
UID free GaN	HVPE	GaN	-	∼350 µm	-	-	-
SI free GaN	HVPE	GaN	Fe	∼475 µm	-	-	-
HVPE Si	HVPE	GaN/AlN	-	-	-	AlN	silicon
MBE 03-11	MBE	GaN	-	∼430 nm	-	-	silicon
MBE 25-11	MBE	GaN	-	∼300 nm	-	-	silicon
MBE 30-11	MBE	GaN	-	∼670 nm	-	-	sapphire
MBE 22-03	MBE	GaN	-	∼360 nm	-	GaN	sapphire
MBE 30-03	MBE	GaN	-	∼305 nm	-	GaN	sapphire
MBE 01-04	MBE	GaN	-	∼655 nm	-	-	sapphire
MBE 04-04	MBE	GaN	-	∼295 nm	-	GaN	sapphire
MR 1	MOCVD	GaN	-	∼1900 nm	-	∼20 nm GaN	sapphire
MR 2	MOCVD	AlGaN/GaN	-	∼1900 nm	-	∼20 nm GaN	sapphire
MR 3	MOCVD	AlGaN/GaN	-	∼1900 nm	-	∼20 nm GaN	sapphire
PC 5257	MOCVD	HEMT/GaN/AlN	-	1200 nm	700 nm	20 nm AlN	sapphire
PC 5258	MOCVD	HEMT/GaN/AlN	-	1200 nm	700 nm	20 nm AlN	sapphire
PC 5259	MOCVD	HEMT/GaN/AlN	-	1200 nm	650 nm	20 nm AlN	sapphire
PC 5260	MOCVD	HEMT/GaN/AlN	-	1200 nm	650 nm	20 nm AlN	sapphire
PC 5263	MOCVD	HEMT/GaN/AlN	-	1200 nm	250 nm	20 nm AlN	sapphire
PC 5264	MOCVD	HEMT/GaN/AlN	-	1200 nm	250 nm	20 nm AlN	sapphire
PC 5265	MOCVD	HEMT/GaN/AlN	-	1200 nm	270 nm	20 nm AlN	sapphire
PC 5266	MOCVD	HEMT/GaN/AlN	-	1200 nm	270 nm	20 nm AlN	sapphire
PC 5274	MOCVD	GaN	-	1200 nm	-	20 nm GaN	sapphire
PC 5276	MOCVD	GaN	Si	1200 nm	-	20 nm GaN	sapphire
PC 5277	MOCVD	GaN	Si	1200 nm	-	20 nm GaN	sapphire

Table D.2: *Sample parameters. The thickness of the nucleation layers on the MBE samples is 15-30 nm, but not known exactly due to varying Ga flux between depositions. The HEMT structures are the same on all PC samples, and consist of 3 nm GaN/ 7 nm AlGaN/ 8 nm Si-doped AlGaN/ 6 nm AlGaN/ 1 nm AlN on top of the GaN epilayers. The MBE growth parameters are further accounted for in section 5.1.*