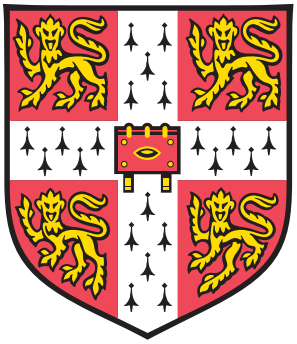
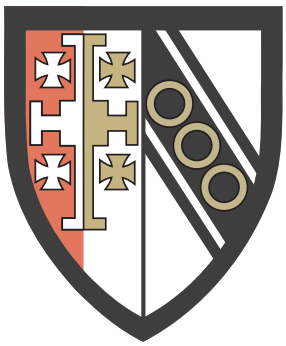
University of Cambridge



Selwyn College



Multi-scale modelling of III-nitrides: from dislocations to the electronic structure

David Holec

Selwyn College, Cambridge

A thesis submitted for the degree of
Doctor of Philosophy at the University of Cambridge

July 2008

Multi-scale modelling of III-nitrides: from dislocations to the electronic structure

David Holec

Gallium nitride and its alloys are direct band gap semiconductors with a wide variety of applications. Of particular importance are light emitting diodes and laser diodes. Due to the lack of suitable lattice-matched substrates, epitaxial layers contain a high density of defects such as dislocations. To reduce their number and to design a device with desired specifications, multilayered systems with varying composition (and thus material properties) are grown. Theoretical modelling is a useful tool for gaining understanding of various phenomena and materials properties.

The scope of the present work is wide. It ranges from a continuum theory of dislocations treated within the linear elasticity theory, connects the continuum and atomistic level modelling for the case of the critical thickness of thin epitaxial layers, and covers some issues of simulating the electronic structure of III-nitride alloys by means of the first principle methods.

The first part of this work discusses several topics involving dislocation theory. The objectives were: (i) to apply general elasticity approaches known from the literature to the specific case of wurtzite materials, (ii) to extend and summarise theoretical studies of the critical thickness in heteroepitaxy. Subsequently, (iii) to develop an improved geometrical model for threading dislocation density reduction during the growth of thick GaN films.

The second part of this thesis employs first principles techniques (iv) to investigate the electronic structure of binary compounds (GaN, AlN, InN) and correlate these with experimentally available N K-edge electron energy loss near edge structure (ELNES) data, (v) to apply the special quasi-random structures method to ternary III-nitride wurtzite alloys aiming to develop a methodology for modelling wurtzite alloys and to get quantitative agreement with experimental N K-edge ELNES structures, and (vi) to theoretically study strain effects on ELNES spectra.

Preface

This thesis is submitted to the University of Cambridge for the degree of Doctor of Philosophy. It is an account of the research that I have undertaken under the supervision of Professor C.J. Humphreys in the Department of Materials Science and Metallurgy at the University of Cambridge between August 2005 and July 2008. It is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. The dissertation is less than 60,000 words in length and has not been submitted to any other institution for a degree.

David Holec

Acknowledgements

My greatest thanks belong to Professor Colin Humphreys for countless numbers of fruitful discussions, suggestions, comments and for his continual support of my work. During the course of my PhD I had the opportunity to discuss my ideas and results with very many people which subsequently helped me to direct my research. For that I would like to acknowledge Dr. Jonathan Barnard, Prof. Mick Brown, Dr. Peter Cherns, Dr. Pedro Costa, Dr. Ranjan Datta, Dr. Martin Friák, Prof. Cécile Hébert, Dr. Kevin Jorrisen, Dr. Menno Kappers, Dr. Liverios Lymperakis, Clifford McAleese, Dr. Shelly Moram, Dr. Rachel Oliver, Michal Petrov, Dr. D.V. Sridhara Rao, Yucheng Zhang, and Prof. Jian-Min Zuo.

The calculations on the staircase structures as presented in section 3.4.3 were partly inspired by notes of Prof. Peter Hirsch. The calculations in chapter 9 resulted from a collaboration with the group of Prof. Jörg Neugebauer at the Max-Planck Institut für Eisenforschung in Düsseldorf, Germany. I acknowledge the kind permission to use their computational facilities as well as the support of many people there, in particular Sixten Boeck and Wolfgang Vogt.

I would like to acknowledge the financial support from the EU project PARSEM (contract number MRTN-CT-2004-005583).

I am very thankful to those who read my thesis and my reports at various stages of my research and helped me that way to improve the texts by pointing out unclear spots and factual mistakes as well as typos and grammar mistakes. These were Dr. Jonathan Barnard, Dr. Richard Beanland, Dr. Ron Broom, Dr. Pedro Costa, Dr. Martin Friák, Mark Galtrey, Prof. Colin Humphreys, Carol Johnston, Dr. Menno Kappers, Dr. Nicole van der Laak, Cliford McAleese, Dr. Shelly Moram, Dr. Rachel Oliver, Dr. D.V. Sridhara Rao, Joy Sumner and Dr. Dandan Zhu.

Many thanks belong to all the people from our group as well as from outside the department which have made my time in Cambridge so enjoyable, stimulating and rewarding – simply a great life experience!

Last but not least, I would like to thank my beautiful wife Jitka for her willingness to follow me here to our new home in Cambridge, her endless support in all my doing and for making it so easy for me to become and to be a proud husband and dad. Děkuji!

David Holec

Cambridge, July 2008

„Pro každý problém existuje úhel pohledu, z něhož se daný problém zdá být zanedbatelný.“
“For every problem there is a perspective from which the problem seems negligible.”

Aleš Návrat

List of publications

The following is a list of my publications directly related to the topics of this dissertation. These papers were submitted for publication during the course of this research work.

Critical thickness calculations for InGaN/GaN
D. Holec, P.M.F.J. Costa, M.J. Kappers and C.J. Humphreys
J. Cryst. Growth, 303(1):314–317, 2007 (doi:10.1016/j.jcrysgro.2006.12.054).

Calculations of equilibrium critical thickness for non-polar wurtzite InGaN/GaN systems
D. Holec and C.J. Humphreys
Mater. Sci. Forum, 567–568:209–212, 2007.

Detailed theoretical study of ELNES spectra of $Al_xGa_{1-x}N$ using Wien2k and Telnes programs
D. Holec, P.M.F.J. Costa, P.D. Cherns and C.J. Humphreys
Comput. Mater. Sci., 2008, in press (doi:10.1016/j.commatsci.2008.01.029).

Electron energy loss near edge structure (ELNES) spectra of AlN and AlGaN: a theoretical study using the Wien2k and Telnes programs
D. Holec, P.M.F.J. Costa, P.D. Cherns and C.J. Humphreys
Micron, 2008, in press (doi:10.1016/j.micron.2007.10.013).

First principle calculations of N K-edge electron energy loss near edge structure of ternary III-nitride alloys
D. Holec and C.J. Humphreys
Solid State Sci., submitted.

Equilibrium critical thickness for misfit dislocations in III-nitrides
D. Holec, Y. Zhang, D.V.S. Rao, M.J. Kappers, C. McAleese and C.J. Humphreys
J. Appl. Phys., submitted.

Contents

List of symbols and abbreviations

3DAP	3D atom probe
$\boldsymbol{a}_1$, $\boldsymbol{a}_2$, a	hexagonal lattice vectors along the $[2\bar{1}\bar{1}0]$ and $[\bar{1}2\bar{1}0]$ directions and their length
A, A_{cont}	prelograthmic term in the dislocation energy expression (from the atomistic and continuum calculations, respectively)
AES	Auger electron spectroscopy
AFM	atomic force microscopy
APW	augmented plane waves
β	collection semi-angle
$\boldsymbol{b}$, b	Burgers vector and its length
b_s, b_e	screw and edge components of the Burgers vector
$\boldsymbol{c}$, c	hexagonal lattice vectors along the $[0001]$ direction and its length
CBED	convergent beam electron diffraction
c_{ij}	elastic constants
CT	critical thickness
DDSCS	double differential scattering cross-section
DFT	density functional theory
DOS	density of states
d.u.	distance units
$\boldsymbol{\varepsilon}$, ε_{ij}	strain tensor and its (i,j)th component
$\boldsymbol{\varepsilon}^m$	strain tensor due to lattice mismatch (misfit strain)
$\boldsymbol{\varepsilon}^d$	dislocation strain tensor
ε_{ijk}	permutation operator
$\mathcal{E}$	energy
E_{cut}	plane wave cut-off energy
$\mathcal{E}_d$	elastic part of the dislocation self-energy
$\mathcal{E}_{\text{core}}$	dislocation core energy
E_g	band gap
$\mathcal{E}_k$	energy of a dislocation kink

EB	energy balance (model)
EELS	electron energy loss spectroscopy
EFTEM	energy filtered transmission electron microscopy
ELNES	electron energy loss near edge structure
ELOG	epitaxial lateral overgrowth
F	Airy stress function
$\boldsymbol{F}$, F_i	force acting on a dislocation
ϕ	Airy stress function; angle describing dislocation line orientation
$\phi(\boldsymbol{r})$	Kohn-Sham wave functions
Φ	many-body wave function
$\boldsymbol{F}_l$	dislocation line tension
F_P	Peierls force
GGA	generalised gradient approximation
h	layer thickness
$\hat{H}$	Hamiltonian (operator)
h_c	critical thickness
HRTEM	high resolution transmission electron microscopy
HVPE	hydride vapour phase epitaxy
IR	infrared
$\boldsymbol{k}$, $\boldsymbol{k}_0$, k	electron wave vector
$K_{\max}$	maximum k-vector used for the expansion into basis set
$\boldsymbol{l}$	dislocation line direction
$l_{\max}$	maximum momentum number for the expansion of wave functions
LAPW	linear augmented plane waves
LD	laser diode
LDA	local density approximation
LED	light emitting diode
μ	shear modulus
MB	(model of) Matthews and Blakeslee
MBE	molecular beam epitaxy
MD	misfit dislocation
MOCVD	metalorganic chemical vapour deposition
MOSS	multibeam optical stress sensor
MTS	muffin tin sphere
ν	Poisson's ratio
$\Pi_{(k,m)}(\sigma)$	structural correlation coefficient
PDOS	projected density of states
PL	photoluminescence

PLM	photoluminescence microscopy
$\boldsymbol{q}$, q	momentum transfer
ρ	density of (threading) dislocations
$\rho(\boldsymbol{r})$	electron density
$\rho(\sigma)$	probability of finding the configuration σ in a given ensamble
r_0, r_c, r_{core}	dislocation core radius (inner cut-off radius)
R_{MT}	maximum muffin tin sphere radius
σ	configuration of a supercell
$\boldsymbol{\sigma}^d$	dislocation stress tensor
s_{ij}	elastic compliances
SCC	structural correlation coefficient
SCM	scanning probe microscopy
SCPA	site-coherent potential approximation
SEM	scanning electron microscopy
SQS	special quasi-random structure
θ	angle between the Burgers vector and the dislocation line
$\hat{T}$	operator of the kinetic energy
T_k	line tension in a kink
T_i	line tension in ith segment of a dislocation
TD	threading dislocation
TDD	threading dislocation density
TEM	transmission electron microscopy
t.u.	time units
u	internal parameter defining the wurtzite structure
$\boldsymbol{u}$, u_i	(deformation) displacements
UPS	ultraviolet photoelectron spectroscopy
UV	ultraviolet
v	dislocation velocity
$\hat{V}$	operator corresponding to (various) potentials
VCA	virtual crystal approximation
$\mathcal{W}$	work done by the stress field
x	composition of an alloy
XANES	X-ray absorption near edge structure
XPS	X-ray photoelectron spectroscopy
XRD	X-ray diffraction

Chapter 1

Introduction: gallium nitride and related alloys (III-nitrides)

GaN and its alloys with aluminium or indium are in their stable form, wurtzite direct band gap semiconductors that have become the most important since silicon. Optical devices based on GaN emit bright light in a wide range of wavelengths[1] while electronic devices can operate at high frequencies and at high temperatures. Light emitting devices are currently the main applications of III-nitrides. Light emitting diodes (LEDs) and laser diodes (LDs) based on GaN and its alloys can, in principle, cover the range from deep ultraviolet (UV) through visible to the far infrared (IR) region (see Fig. 1.1). Large scale displays, modern traffic lights or public lighting nowadays use GaN-based LEDs. Significant effort is being invested in development of efficient and reliable LEDs which emit comfortable white light, as there is a huge market potential for replacing bulbs and fluorescent tubes currently used. Apart from the long lifetime, the other advantages are substantial energy savings as well as a reduction of CO_2 emissions through greater efficiency. Short-wavelength GaN LDs enabled development of the Blu-ray technology which is

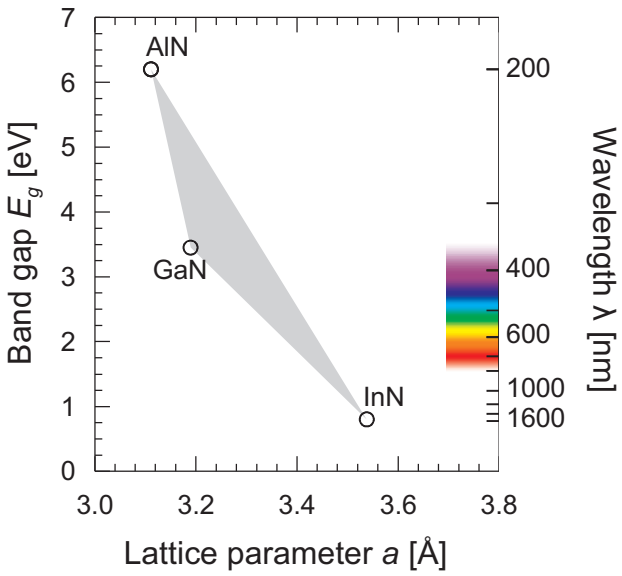


Figure 1.1: Lattice parameters against band gaps of wurtzite forms of the III-nitrides. The range of the visible spectrum is shown on the wavelength axis.

[1] Gallium nitride itself emits short-wavelength light in the near ultra-violet range.

now becoming a standard to replace DVDs. If standard red laser diodes (working at $\lambda = 650\,\text{nm}$) in DVD recorders are replaced with GaN-based laser diodes emitting blue light ($\lambda = 405\,\text{nm}$), more data (up to $\approx 25\,\text{GB}$) could be written on every single disk due to the shorter wavelength of the writing beam. Other exciting applications include solar cell arrays for satellites and power amplifiers at microwave frequencies (as GaN can operate at much higher temperatures and voltages than GaAs).

The above outline of everyday applications of GaN-based devices demonstrates how exciting a material gallium nitride is.

1.1 Short history of GaN

The great interest in GaN has continued now for about 15 years and was started by the commercial production of blue GaN LEDs at Nichia Laboratories in Japan, the achievement of Shuji Nakamura and his co-workers. However, the history of GaN goes much further into the past.

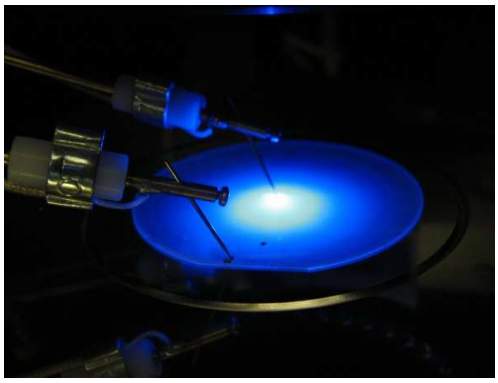


Figure 1.2: An example of the blue light extracted from an InGaN-based LED: an electroluminescence test performed on a blue emitting LED wafer grown on a sapphire substrate. (Courtesy of C. Salcianu, University of Cambridge.)

Small needles of GaN were synthesised by Juza and Hahn (1938), followed by a synthesis of small crystals by Grimmeiss and Koelmans (1959). Maruska and Tietjen (1969) used a chemical vapour deposition method to obtain the first large-area GaN layer. As a substrate they used sapphire and they demonstrated GaN to be a direct-transition-type semiconductor with a band gap of about 3.39 eV. This immediately led to an increased research interest in blue LEDs based on GaN which was reflected in the increased number of publications on the topic: about 3–4 per year before compared with ∼30 per year after their paper (Akasaki and Amano, 2006). Although several good papers followed in subsequent years determining various properties of GaN, further development struggled (when compared with other III-V semiconductors, for example GaAs or InP) mainly because of difficulties in obtaining high quality epitaxial layers.

Yoshida et al. (1983) introduced a two-step growth method using a buffer layer which resulted in a great improvement in crystalline quality of GaN. Amano et al. (1989)

obtained *p*-type conductivity in GaN using Mg as a dopant. Nakamura et al. (1993) announced the successful realisation of the first *p*-GaN/*n*-InGaN/*n*-GaN double-heterostructure blue LEDs and shortly after that Nichia Chemical Industries started their commercial production. Hand-in-hand with the industrial interest went money for research projects which was reflected in a sudden increase in the number of papers published in the field (∼2000 papers in year 2000 according to the review by Akasaki and Amano (2006)).

The above mentioned achievements are only a few milestones in the history of GaN and its alloys. Many more details can be found in reviews by Akasaki and Amano (2006) and Jain et al. (2000) or in books such as the one by Gil (1998).

1.2 GaN-based light emitting diodes

GaN has a wide range of optoelectronic applications, chief amongst which are light emitting diodes (LEDs) and laser diodes (LDs). Compared to a traditional light source, the LED has many prominent advantages including its low energy usage, long lifetime and small size, making it a potential competitor in the domestic and business lighting market. Because lighting accounts for about 20 % of total electricity consumption, national programs promoting LEDs for lighting are underway in the US, Japan, Korea and China, whose main motivation is large-scale energy savings, providing the benefits of reduced oil imports and lower greenhouse gas emissions.

There are two fundamental reasons to choose the III-nitride system for light sources. Foremost is that AlN, GaN and InN have respective direct band gap energies of 6.2 eV, 3.4 eV and 0.7 eV at room temperature, which cover the entire visible spectrum from the UV to the IR range (see Fig. 1.1). This is in a contrast to the (Ga,Al)As based LEDs (1.5–2.2 eV) or (Al,Ga,In)P based LEDs (1.3–2.5 eV) that cover "only" the IR to green region. The other main advantage of the III-nitride semiconductors is their strong chemical bond, which makes the nitrides very stable and resistant to degradation under high electric currents and at high temperatures.

The basic structure of an LED is based on a *p*–*n* junction, while higher efficiency designs make use of a quantum well (QW). In the *p*–*n* junction LED, electrons from the *n*-side of a *p*–*n* junction and holes from the *p*-side are injected across the junction by the application of a forward bias. These minority carriers recombine with majority carriers by spontaneous transitions across the band gap (see Fig. 1.3a). In QW LEDs, the electrons and holes injected across the *p*–*n* junction become confined within the well, filling

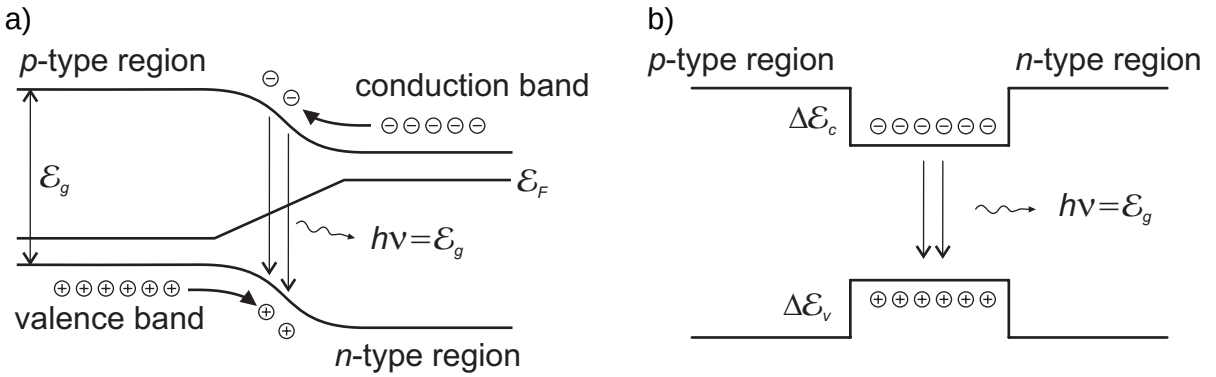


Figure 1.3: Schematic diagram of operating principles of two basic kinds of LEDs: (a) a p–n junction based LED and (b) a quantum well based LED. Wavelength of the light is related to the energy gap $\mathcal{E}_g$. $\mathcal{E}_F$ is the Fermi energy.

up and causing a local increase in the concentration of holes, n_h, and electrons, n_e (see Fig. 1.3b). The radiative recombination rate is proportional to $n_e n_h$. Therefore, in a heterostructure LED the amount of radiative recombination is larger than in the case of the p–n junction based LED. This leads to an increased number of photons emitted for the same current across the junction. Research conducted in The Cambridge Centre for Gallium Nitride is focused on QW LEDs. A typical structure of a QW based LED includes an InGaN/GaN or AlGaN/GaN single or (more often) multiple quantum wells active layer sandwiched between n-type and p-type nitride materials (as shown in Fig. 1.4).

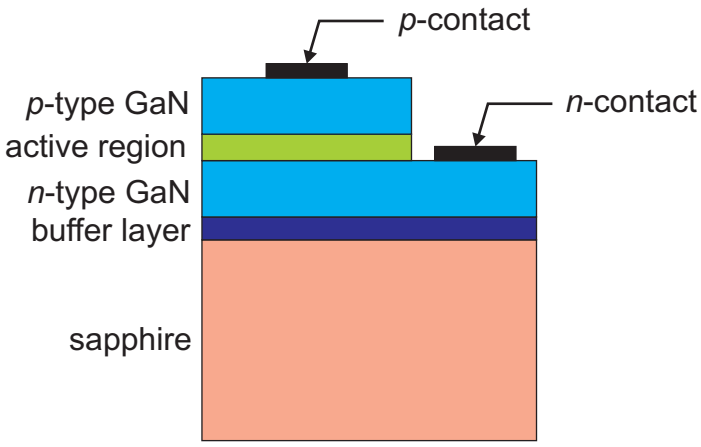


Figure 1.4: Schema of GaN based LED structure.

1.3 Physical properties of III-nitrides

1.3.1 Crystal structure

Like most other semiconductor materials, nitrides have tetrahedrally coordinated atomic arrangements that result in either cubic (zincblende) or hexagonal (wurtzite) lattice structures. For AlN, GaN and InN the zincblende structure is metastable while the wurtzite variant is stable and easier to grow. Therefore, most research has been focused on the wurtzite form which, as a consequence, has given better results up to date for optoelectronic applications. The atomic arrangement of the nitrides can be viewed as

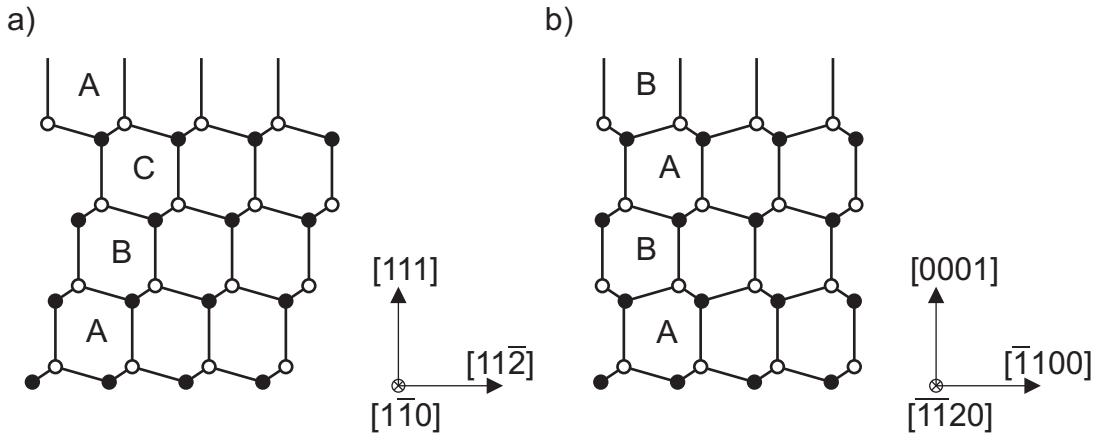


Figure 1.5: Atomic arrangement in tetrahedrally coordinated nitrides: a) cubic zincblende and b) hexagonal wurtzite lattice (after Ponce, 1998).

consisting of two hexagonal layers. One layer is occupied by nitrogen while the other contains the group III elements. The zincblende structure occurs when the hexagonal double-layers are stacked in a periodic $\ldots ABCABC \ldots$ sequence while the wurtzite structure follows an $\ldots ABABAB \ldots$ arrangement (see Fig. 1.5).

The three-dimensional arrangement of wurtzite nitrides is shown in Fig. 1.6 where white colour represents nitrogen atoms while orange represents the group III atom sites (Al, Ga or In).

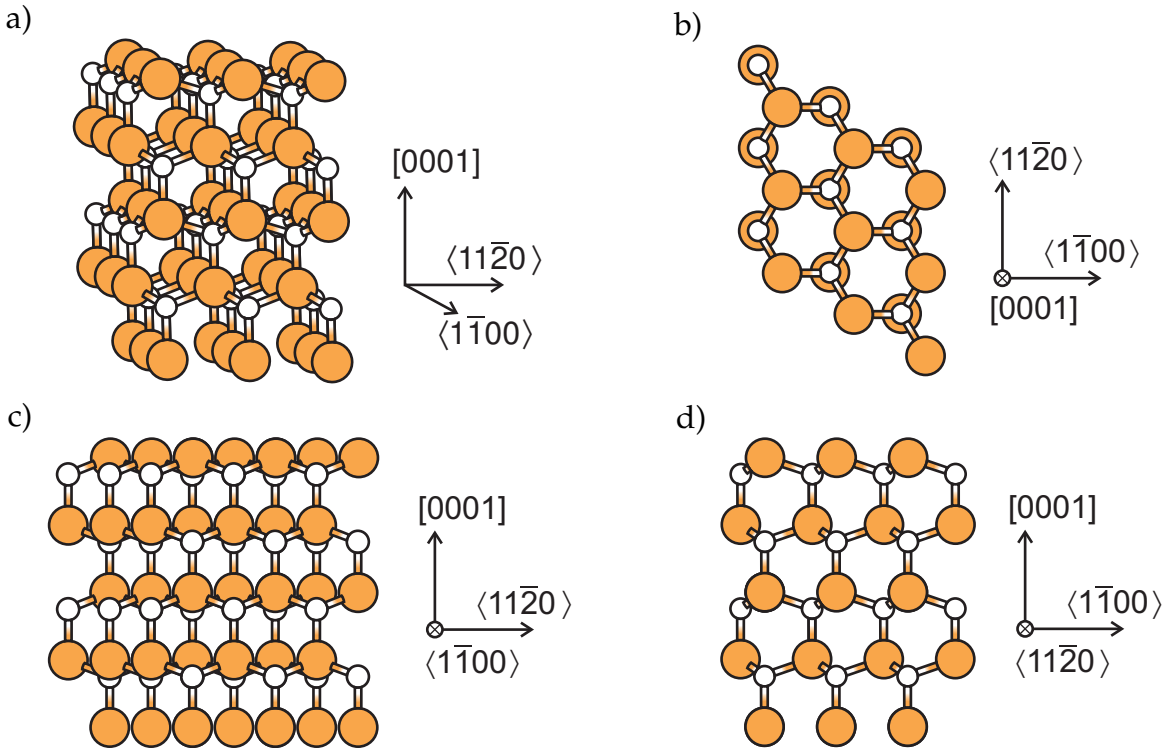


Figure 1.6: A 3D picture of the wurtzite structure together with some important projections.

The wurtzite unit cell is generated by two lattice vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ with the same length a and separated by an angle of 120°, and the third one $\boldsymbol{c}$ of a different length c and perpendicular to both $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$. The Wyckhoff 2b atomic positions imply that the group III atoms are located at $(\frac{1}{3}, \frac{2}{3}, 0)$ and $(\frac{2}{3}, \frac{1}{3}, \frac{1}{2})$ within the unit cell and the nitrogen atoms have positions of $(\frac{1}{3}, \frac{2}{3}, u)$ and $(\frac{2}{3}, \frac{1}{3}, \frac{1}{2} + u)$; u depends on the specific material. The lattice constants and the parameter u used in this work are summarised in Table 1.1. The corresponding space group is $P6_3mc$ (nr. 186).

	a [nm]	c [nm]	u
AlN	0.311	0.498	0.382
GaN	0.319	0.519	0.377
InN	0.354	0.570	0.380

Table 1.1: Lattice parameters used in this work. The values are taken from papers by Dridi et al. (2003) and Wright and Nelson (1995).

1.3.2 Hetero-epitaxial relations

Four-index Miller notation is convenient for use with hexagonal structures. The important directions and planes are shown in Fig. 1.7. The vectors $\boldsymbol{a}$ generating the unit cell are aligned along the $\langle 11\bar{2}0 \rangle$ directions which are thus called the a-directions; similarly, the $[0001]$ direction is called the c-direction. Another direction of a special importance is the $\langle 1\bar{1}00 \rangle$ direction, otherwise called the m-direction (see Fig. 1.7). Planes perpendicular to those directions are labelled accordingly: the plane perpendicular to the c-direction (the *basal plane*) is called the c-plane, planes perpendicular to the a- and m-directions are called a- and m-planes, respectively. $\{1\bar{1}02\}$ planes are sometimes called the r-planes.

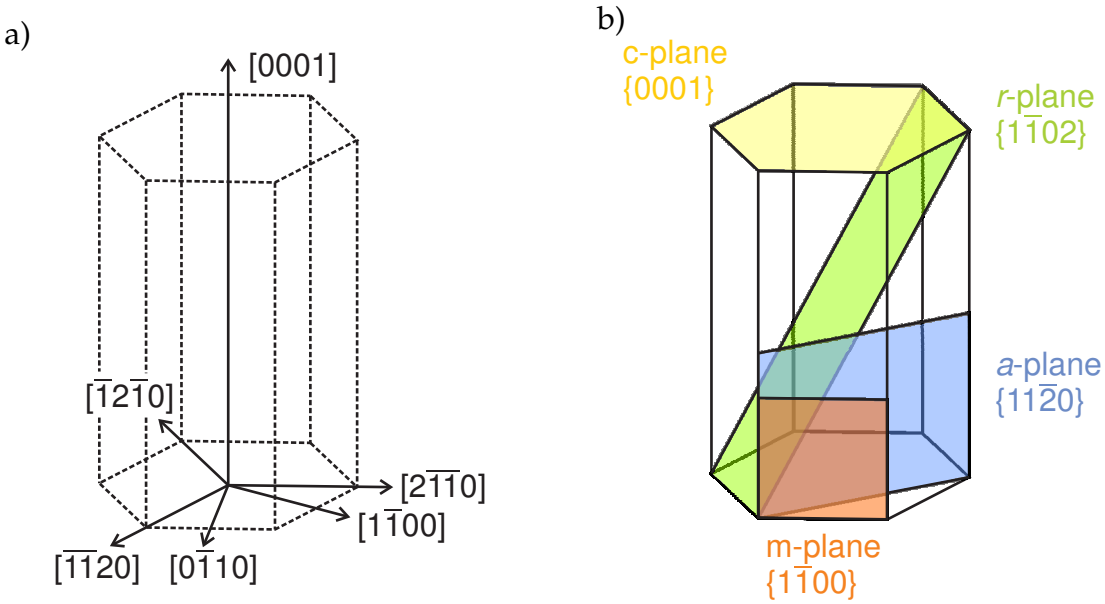


Figure 1.7: Directions and planes in hexagonal material.

The same notation is used also for sapphire, which is the most commonly used substrate for the epitaxial growth of GaN and related alloys. The orientation of the sapphire wafer determines the crystallographic orientation of the overgrown material. This significantly influences the device properties. The standard orientation is growth on *c*-plane sapphire with the overgrown *c*-plane GaN. The epitaxial relation is shown in Fig. 1.8. Note the 30° rotational offset of the unit cells (Gil, 1998).

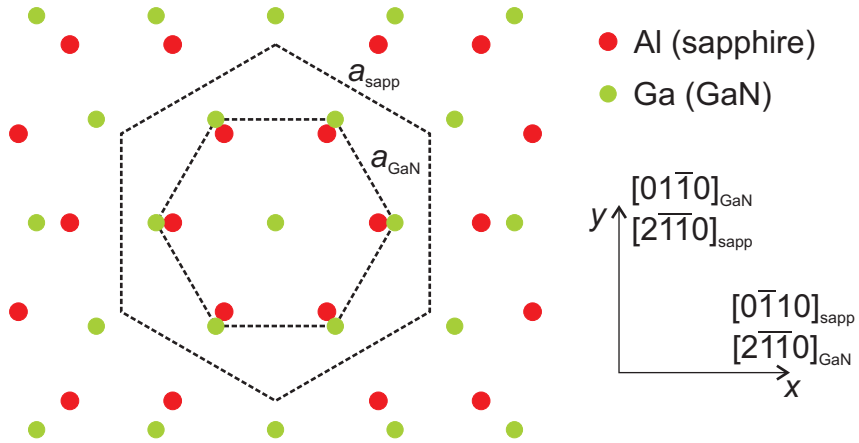


Figure 1.8: *c*-plane of GaN and sapphire. After Gil (1998).

The mismatch between sapphire and GaN lattice parameters leads to an isotropic in-plane compressive strain in GaN of

$$\varepsilon_{xx}^{m} = \varepsilon_{yy}^{m} = \frac{a_{\text{sapp}} - \sqrt{3}a_{\text{GaN}}}{\sqrt{3}a_{\text{GaN}}} = -0.139\,. \tag{1.1}$$

Another commonly used characteristic is the lattice mismatch which is the difference in lattice constants relative to the substrate lattice constants. In the case of *c*-plane GaN on *c*-plane sapphire, the lattice mismatch is 0.161.

The *c*-axis is a polar direction leading to spontaneous polarisation, and together with piezoelectric effects, this results in built-in charges causing electric fields parallel to the *c*-axis. This leads to a decreased overlap of the electron and hole wave functions and thus smaller recombinantion probabilities (Miller et al., 1984). Waltereit et al. (2000) have suggested that this problem may be overcome by the use of *non-polar* or *semi-polar* orientations where the *c*-axis is not perpendicular to the QW interfaces.

The *a*-plane GaN orientation is grown on *r*-plane sapphire, and belongs to a family of non-polar materials with rapidly growing importance for device design. The arrangement of atoms in the interface is shown in Fig. 1.9a, and GaN is in compression in all directions. However, the absolute values of compressive strains along the $[1\bar{1}00]_{\text{GaN}}$

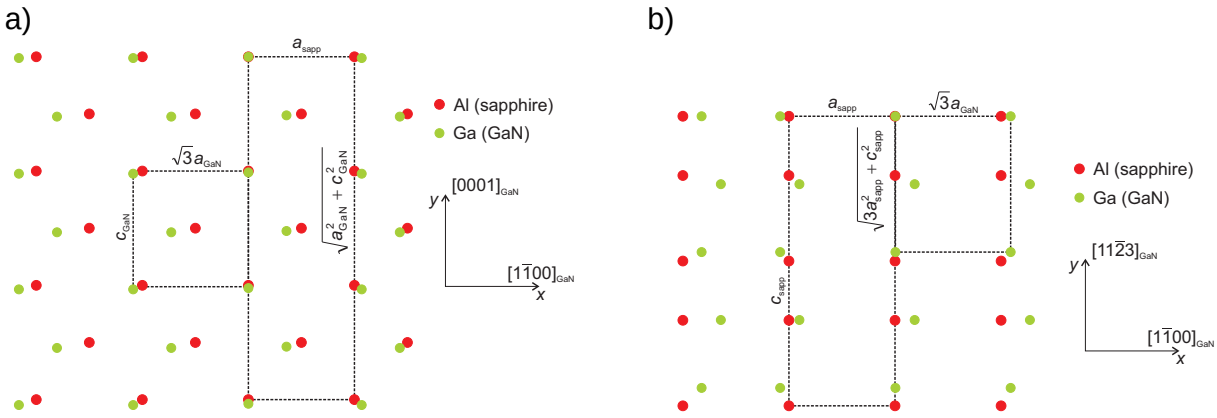


Figure 1.9: a) *a*-plane GaN on *r*-plane sapphire, b) $(11\bar{2}2)$-plane GaN and *m*-plane sapphire. After Gil (1998).

and $[0001]_{\mathrm{GaN}}$ directions differ

$$\varepsilon_{xx}^{m} = \frac{a_{\mathrm{sapp}} - \sqrt{3}a_{\mathrm{GaN}}}{\sqrt{3}a_{\mathrm{GaN}}} = -0.139\,, \tag{1.2a}$$

$$\varepsilon_{yy}^{m} = \frac{\sqrt{3a_{\mathrm{sapp}}^2 + c_{\mathrm{sapp}}^2} - 3c_{\mathrm{GaN}}}{3c_{\mathrm{GaN}}} = -0.012\,. \tag{1.2b}$$

Semi-polar $(11\bar{2}2)$-plane GaN is grown on *m*-plane sapphire. The lattice sites for those planes in GaN and sapphire which form the GaN–sapphire interface are shown in Fig. 1.9b. The material is in not in a unique strain state: it is in compression along the $[1\bar{1}00]_{\mathrm{GaN}}$ direction while it is in tension in the $[11\bar{2}3]_{\mathrm{GaN}}$ direction

$$\varepsilon_{xx}^{m} = \frac{a_{\mathrm{sapp}} - \sqrt{3}a_{\mathrm{GaN}}}{\sqrt{3}a_{\mathrm{GaN}}} = -0.139\,, \tag{1.3a}$$

$$\varepsilon_{yy}^{m} = \frac{c_{\mathrm{sapp}} - 2\sqrt{a_{\mathrm{GaN}}^2 + c_{\mathrm{GaN}}^2}}{2\sqrt{a_{\mathrm{GaN}}^2 + c_{\mathrm{GaN}}^2}} = 0.066\,. \tag{1.3b}$$

1.3.3 Mechanical properties

The lattice parameters of the III-nitrides were summarised in Table 1.1; elastic constants are given in Table 1.2. Whenever a material property of an alloy is needed, for example the *a* lattice constant of $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N}$, Vegard's law in a form

$$a(\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N}) = x\,a(\mathrm{AlN}) + (1-x)\,a(\mathrm{GaN}) \tag{1.4}$$

is used. Although its strict validity is under constant questioning[2], Dridi et al. (2003) concluded using first principle calculations that Vegard's law holds for AlGaN (both

[2]For example, the band gap dependence seems to be better expressed by adding the bowing parameter to Vegard's law

$$E_g(\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N}) = x\,E_g(\mathrm{AlN}) + (1-x)\,E_g(\mathrm{GaN}) + b\,x(1-x)\,.$$

lattice constants) and the a-lattice constant of InGaN and AlInN. The authors predicted some deviations from Vegard's law for the c-lattice constant of the latter two alloys. However, since no better estimate of lattice and elastic constants is available for the whole range of compositions, Vegard's law is used throughout this work.

	c_{11} [GPa]	c_{12} [GPa]	c_{13} [GPa]	c_{33} [GPa]	c_{44} [GPa]
AlN	411	149	99	389	125
GaN	377	160	114	209	81
InN	190	104	121	182	10

Table 1.2: Elastic constants used in this work (Madelung, 2004).

1.3.4 Substrates

In contrast to other semiconductors, such as silicon or GaAs, bulk single crystals of III-nitrides are not yet widely available. In the absence of bulk GaN crystals the use of foreign substrates is necessary for crystal growth. ZnO and SiC were identified as suitable substrates due to their similar characteristics to nitrides (Ponce, 1998; Jain et al., 2000). Some more exotic substrates such as GaAs, $MgAl_2O_4$, glass or quartz glass were also tried (Ponce, 1998; Jain et al., 2000). By far the most widely used substrate is sapphire Al_2O_3. This is surprising bearing in mind the large difference in thermal expansion and in lattice parameters between sapphire and the nitrides (up to 16 %). As a consequence, standard values of dislocation density[3] in GaN films grown on sapphire are $\rho \cong 10^8$–$10^{10}\,\mathrm{cm}^{-2}$ which may be further reduced to values $\rho = 5 \times 10^6$–$10^7\,\mathrm{cm}^{-2}$ using for example an epitaxial lateral overgrowth (ELOG) technique (Gibart, 2004). However, these values are still many orders higher than the required quality of silicon or GaAs for optoelectronic applications. Densities of about $10^6\,\mathrm{cm}^{-2}$ are fatal in conventional III-V semiconductor devices (Jain et al., 2000) since dislocations act as non-radiative recombination centres. As pointed out by Sasaoka et al. (1998) dislocations act as non-radiative recombination centres in GaN-based devices as well. A possible explanation for the high efficiency of GaN may be in the localisation of excitons and thus reduced interaction with dislocations. The mechanism of such a localisation has not yet been unambiguously identified although several models exist (Graham et al., 2003; Hitzel et al., 2005; Chichibu et al., 2006).

[3]The values mentioned here correspond to the threading dislocation density, see section 1.4.1.

Growth of *c*-plane GaN on the {111}-plane of silicon wafers has recently become of great interest for several reasons: high quality and large scale silicon wafers are available for low prices and device processing lines for silicon are already around. However, there are several difficulties which need to be overcome. Due to lattice and thermal mismatch, tensile strains as large as $\varepsilon \sim 0.2$ are generated in GaN layers which lead to cracking (Dadgar et al., 2007). For large size wafers the effects of bowing due to different thermal expansion coefficients of Si and GaN are more problematic. However, this is certainly a promising direction of the III-nitride research and commercialisation.

1.4 Defects in III-nitrides

1.4.1 Dislocations

A dislocation is a line defect in a crystal, which can greatly affect its material properties. A dislocation is usually represented by an oriented dislocation line, $\boldsymbol{l}$, and characterised by its *Burgers vector*, $\boldsymbol{b}$, describing displacements introduced in the crystal by the dislocation.

Based on the relationship of $\boldsymbol{l}$ and $\boldsymbol{b}$, three dislocation types are distinguished (see Fig. 1.10): an *edge*-type dislocations, where $\boldsymbol{l} \perp \boldsymbol{b}$, corresponding to an extra half-plane in the crystal, a *screw*-type dislocation with $\boldsymbol{l} \parallel \boldsymbol{b}$, and a *mixed*-type dislocations ($\boldsymbol{b} \not\perp \boldsymbol{l} \not\parallel \boldsymbol{b}$). The dislocation type can change as the dislocation line changes its direction in the crystal, since the Burgers vector $\boldsymbol{b}$ is a constant characteristic of a dislocation which remains unchanged along the whole single dislocation line. A dislocation can only make either a closed loop or end at the crystal surfaces. It cannot just end in the crystal. It is worth noting that the orientation of $\boldsymbol{b}$ in the standard definition (Burgers–Frank or a continuous elasticity definition, (see Cottrell, 1964)) depends on the actual orientation of the dislocation line $\boldsymbol{l}$: an opposite dislocation line direction $-\boldsymbol{l}$ yields an opposite orientation of the Burgers vector $-\boldsymbol{b}$. However, the uncertainty in sign of $\boldsymbol{b}$ and $\boldsymbol{l}$ does not affect their mutual relation in terms of defining an edge, screw or mixed type dislocation.

The Burgers vector of a *perfect dislocation* is a lattice vector. As such, there are three basic types of dislocations in GaN according to the Burgers vector: *a*-type dislocations with $\boldsymbol{b} = \frac{1}{3}\langle 11\bar{2}0\rangle$, *c*-type dislocations with $\boldsymbol{b} = \langle 0001\rangle$, and $(a+c)$-type dislocations with $\boldsymbol{b} = \frac{1}{3}\langle 11\bar{2}3\rangle$. It is important to distinguish between these two nomenclatures: an *a*-type dislocation can be either, an edge, screw or mixed dislocation depending on the actual dislocation line direction. Since the Burgers vector cannot change along the

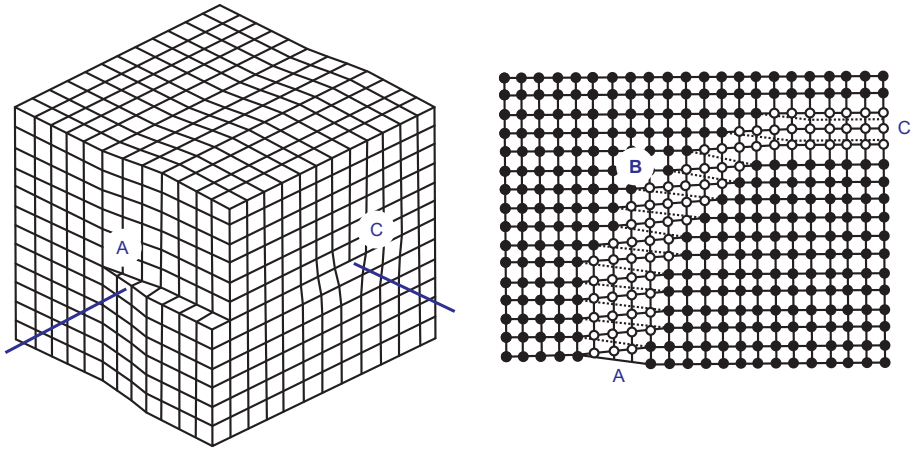


Figure 1.10: Changing dislocation type along a single dislocation line: screw-type (A), mixed-type (B) and edge-type (C). After http://oregonstate.edu/instruct/engr322/Exams/Previous/S98/ENGR322MT2.html.

dislocation line, the latter terminology is used in this text for describing dislocation types.

Glide of a dislocation is a conservative motion where no material is added or removed from the crystal as opposed to a nonconservative motion, *climb*, where diffusion of material to or from a dislocation takes place (Cottrell, 1964). During glide, a dislocation line remains in a *glide plane* which is generated by the vectors $\boldsymbol{l}$ and $\boldsymbol{b}$. A *slip system* is a convenient way to describe a dislocation in greater detail as it is given by the Burgers vector and the glide plane.

The slip systems in wurtzite nitrides were investigated by Jahnen et al. (1998) and Srinivasan et al. (2003). They found that the only operable slip systems, i.e. slip systems with a non-zero component of the resolved shear stress, are pyramidal slip systems. After considering also a Peierls force as an obstacle to the dislocation glide, they concluded that only the $\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$ and $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ slip systems are operable (see Fig. 1.11). Slip systems in material other than *c*-plane are discussed in section 3.4.1.

According to the dislocation line direction, yet another terminology is often used: *misfit dislocations* (MDs) lie in hetero-epitaxial interfaces and partially or fully relieve misfit strain due to lattice mismatch. Dislocations running through the material (mostly following the growth direction) are called *threading dislocations* (TD).

1.4.2 Other defects

A wide spectrum of other defects can be found in III-nitrides. Point defects (including vacancies and interstitial) are necessary for any diffusion driven process, for example

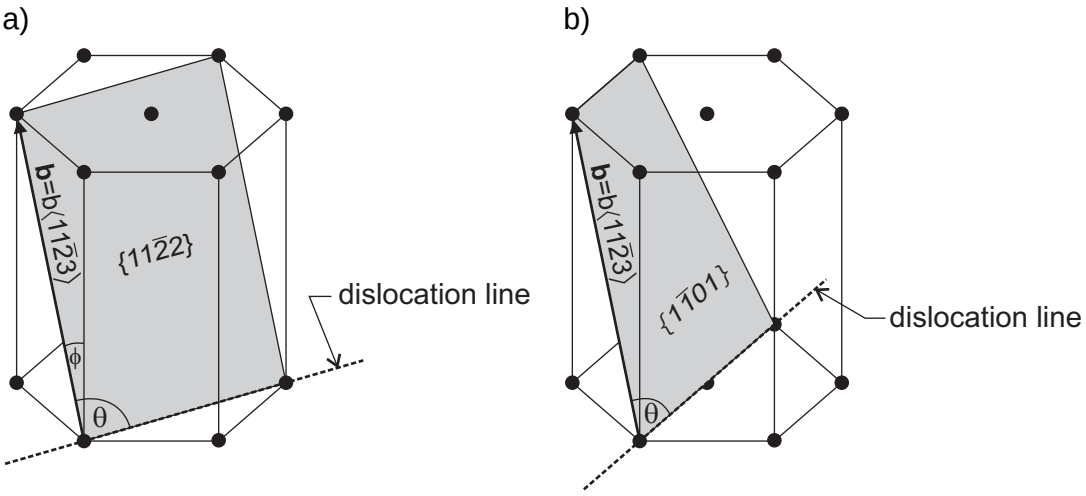


Figure 1.11: The two most favourable slip systems in the wurtzite nitrides: a) $\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$, and b) $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$.

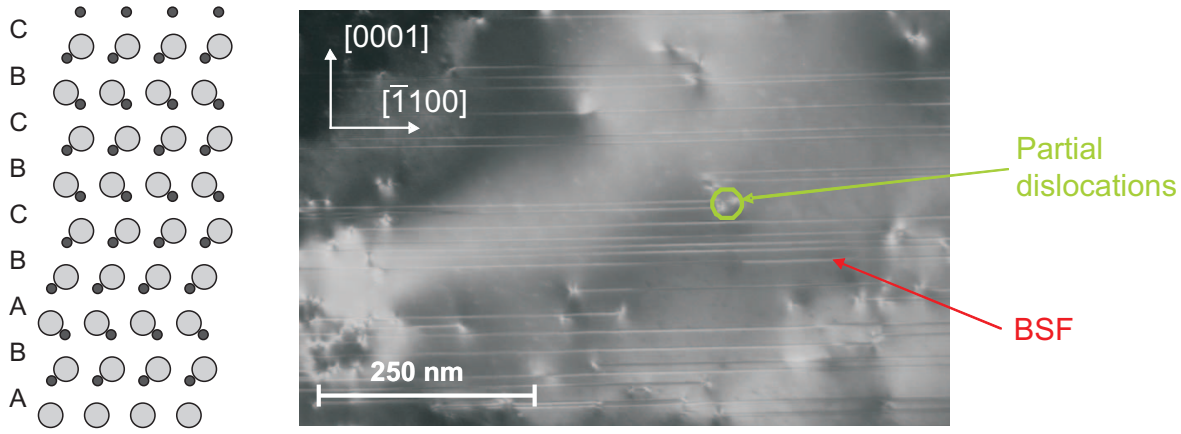


Figure 1.12: Schematic drawing of I_1 type basal stacking fault and a dark field plan view TEM image using $\boldsymbol{g} = 1\bar{1}02$ close to the $[11\bar{2}0]$ zone axis showing partial dislocations and basal plane stacking faults in GaN. (Courtesy of C. Johnston, University of Cambridge.)

dislocation climb. Dopants, which make III-N layers either *p*- or *n*-type semiconductors, can be also regarded as point defects.

Stacking faults are planar defects mainly observed in non-polar and semi-polar material. They are created by a fault in stacking of the hexagonal *c*-planes. The perfect stacking $\ldots ABABAB\ldots$ as discussed in section 1.3.1 changes to $\ldots ABABCBC\ldots$ where the ABC section corresponds to the stacking fault of type I_1. Similarly to dislocations, no stacking fault can spontaneously terminate inside the material; they have to either terminate at the crystal surface or be bounded with partial dislocations. Stacking faults in GaN are discussed e.g. in the paper by Zakharov et al. (2005); an example of stacking faults bounded with partial dislocations is shown in Fig. 1.12.

V-shape defects (or *V-defects*), hexagonal shape surface pits, are often observed in the InGaN/GaN systems where the InGaN QWs are under compressive strain. The apex of a V-defect is, in the majority of cases, connected to a threading dislocation. It has been suggested by Jahnen et al. (1998) that V-defects can act as effective centres for releasing compressive strain. An example of V-defects in the InGaN epilayer is shown in Fig. 1.13.

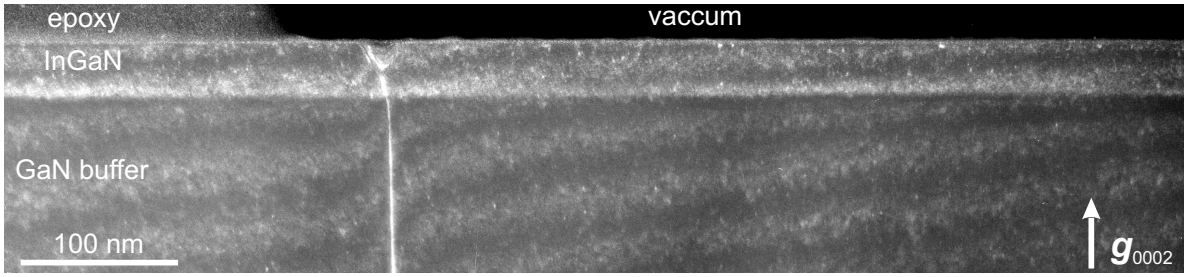


Figure 1.13: Dark field image using a reflection $\boldsymbol{g} = 0002$ of the cross-section of a sample with an $In_{0.133}Ga_{0.867}N$ epilayer grown on GaN. A V-defect connected to a threading dislocation can be observed. (Courtesy of Y. Zhang, University of Cambridge.)

Other commonly observed defects are cracks in structures under tensile strain (for example AlGaN epilayers on GaN or GaN on silicon).

1.5 III-nitrides from the experimental point of view

This section aims to give a very brief overview of experimental techniques which are used for growth and characterisation of III-nitrides. Although this work is theoretical, it has a close relation with experiments and a lot of the motivation originated from experimental work. Therefore, it is important and extremely useful also to understand, in basic terms, the experimental reality.

1.5.1 Growth techniques

There are a number of growth techniques and special treatments designed to improve a particular property of a material (for example to lower the number of dislocations). Generally, there are three main growth modes in epitaxy (see Fig. 1.14): Volmer-Weber mode where 3D islands are formed due to adatom-adatom interactions being stronger than adatom-growth surface interactions, Frank-van der Merwe mode (layer-by-layer growth mode) where an adatom sits preferentially on an atomically smooth surface, and Stranski-Krastanov where a transition from layer-by-layer to 3D growth mode occurs (e.g. Daruka and Barabási, 1997).

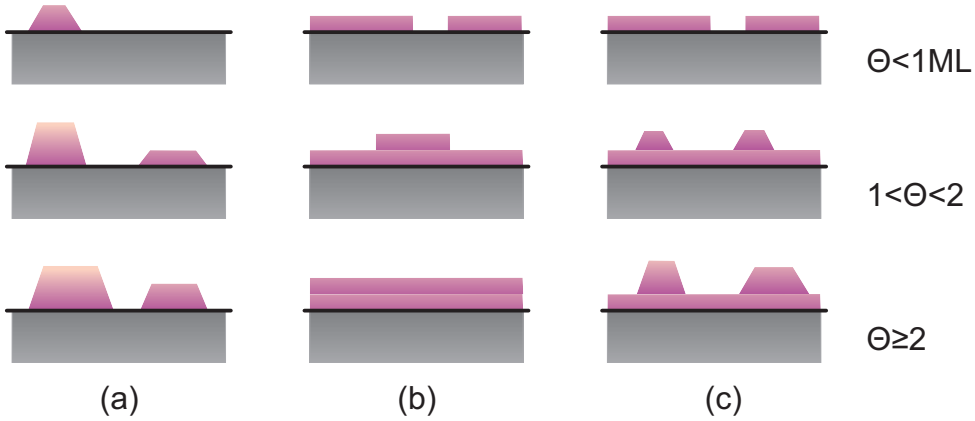


Figure 1.14: Cross-section views of the three primary modes of thin film growth including (a) Volmer-Weber (3D growth), (b) Frank-van der Merwe (layer-by-layer), and (c) Stranski-Krastanov mode. Each mode is shown for several different amounts of surface coverage, Θ (in monolayers). Adopted from http://en.wikipedia.org/wiki/Stranski-Krastanov_growth.

The three main methods for growth of III-nitrides are metalorganic chemical vapour deposition (MOCVD) or metalorganic vapour phase epitaxy (MOVPE), a specific subset of MOCVD for epitaxial growth, hydride vapour phase epitaxy (HVPE) and molecular beam epitaxy (MBE). In MOCVD and HVPE, the crystals are grown by chemical reactions between injected gases with designed components. It does not require high vacuum but pressures as low as 10^0–10^2 kPa are used. MOCVD and HVPE differ in the type of precursor gases and growth rate (HVPE is about $10\times$ faster) (Stringfellow, 1999). On the other hand, MBE requires ultra-high vacuum (10^{-8} Pa) and is based on a physical deposition of atoms. Consequently, the growth rate of MBE is often much slower when compared to MOCVD or HVPE.

1.5.2 Characterisation techniques

Several techniques for the characterisation of the epitaxial layers are widely used. The crystal structure is investigated using X-ray diffraction (XRD) methods. In general, XRD provides information about the space group and lattice parameters. By fitting simulated scans it is also possible to obtain information about thickness and composition of an epitaxial layer.

The surface topography is usually checked by means of atomic force microscopy (AFM) and related techniques (see an example in Fig. 1.15). A surface treatment with silane facilitates imaging of dislocations which terminate on the surface (Oliver et al., 2006). With some add-ons AFM is capable of probing many other surface properties, for example to study dopant profiles using scanning capacitance microscopy (SCM). Al-

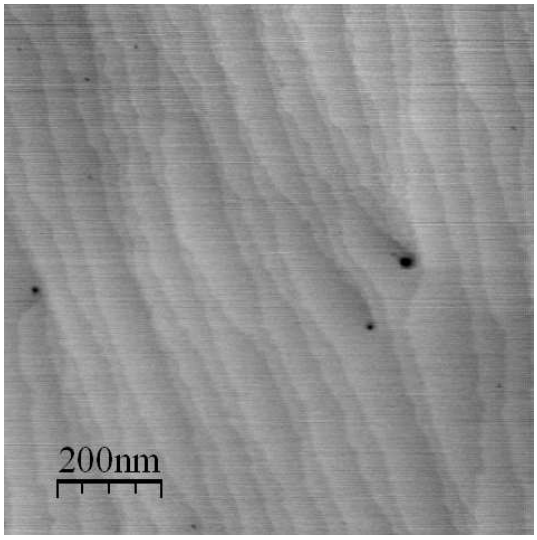


Figure 1.15: Intermittent contact mode AFM image of a GaN surface showing terraces and small pits related to dislocations. Screw- or mixed-type dislocations are assocated with step-edge terminations, whereas pits associated with edge-type dislocations are smaller and may be in the middle of a terrace. (Courtesy of R. Oliver, University of Cambridge.)

though AFM is in principle a surface method, cross-sectional views can be obtained by cleaving a sample (Sumner et al., 2008). The spatial resolution of AFM is given by the tip quality and is usually of the order of 10^1 nm; the vertical resolution can be as good as 0.3 Å.

Electron microscopy, and in particular, transmission electron microscopy (TEM) is probably one of the most universal techniques for characterising materials on the nanoscale. The two basic imaging modes are bright field and dark field, allowing studies of various defects such as dislocations or stacking faults. The final image contrast is built up by interference of the directly transmitted (bright field) or scattered (dark field) electron wave functions. The contrast and visibility of specific defects is determined by the selected diffraction condition, $\boldsymbol{g}$. In particular, the basic criterion for dislocation visibility in TEM images is $\boldsymbol{g} \cdot \boldsymbol{b} \neq 0$ where $\boldsymbol{b}$ is the dislocation Burgers vector (Williams and Carter, 1996). Many TEM micrographs are shown in this work to demonstrate various phenomena, a typical example being Fig. 1.13. The advantage of TEM is its high spatial resolution, often below 1 nm. The high resolution TEM (HRTEM) is capable of imaging single atomic columns; however, comparison with computer simulations is needed for correct interpretation of HRTEM images. The disadvantage of this technique is that it requires a very thin film ($\lessapprox 100$ nm) which requires careful specimen preparation. One also needs to be aware of the fact that the conventional TEM provides a *projection* of the studied structure.

Apart from direct imaging, several advanced techniques allowing chemical analysis of samples can be performed in the TEM. An example is high angle annular dark-field (HAADF) imaging in a scanning TEM where the contrast is sensitive to the atomic number Z (also called Z-contrast). Another family of techniques is called electron energy loss spectroscopy (EELS) and is discussed in greater detail in chapter 6. Energy filtered TEM (EFTEM) is based on EELS; the final image is made up of electrons from only a narrow energy window. Thus it allows simultaneous imaging and element anal-

ysis. A subset of EELS, named electron energy loss near edge structure (ELNES), gives information about the electronic structure of materials. The theoretical simulations needed for the interpretation of experimental spectra are the subject of chapters 7–9.

Electron diffraction revealing in the first instance local crystallographic information can be also done in TEM. There are many other sophisticated techniques such as convergent beam electron diffraction (CBED) which is capable (after thorough data processing) of providing experimental information about the charge density in a material.

The range of experimental techniques is, of course, much wider and contains scanning electron microscopy (SEM), 3D atom probe (3DAP) measurements which can give 3D information about the composition on the atomic scale, photoluminescence and electroluminescence techniques enabling detailed studies of optical properties etc.

Part I

Dislocation theory applied to wurtzite III-nitrides

"There are only two kinds of models: those that describe the working of things really well, and those that really work."

unknown author

Chapter 2

Dislocation energy

Elastic energy in a media caused by the addition of a dislocation is called *dislocation energy* which is traditionally derived within a framework of linear elasticity theory. This theory breaks down in the dislocation core region. As a consequence, an inner cut-off radius r_0, which separates the dislocation core, must be introduced. r_0 is usually in units of the the Burgers vector's length b and it is discussed in greater detail in sections 2.4 and 3.3.

In the case of a dislocation in an infinite continuum the outer cut-off radius R must also be employed in order to get a finite value of the dislocation energy. For dislocations inside a finite crystal, the smallest distance to the crystal surface is used as the value of R. For dislocations organised in an array, half of the next-nearest-dislocation distance is the suitable value for R (see for example Hirth and Lothe, 1982).

It is worth noting that in the case of infinite straight dislocations, only the *energy per unit length of the dislocation* $\mathrm{d}\mathcal{E}/\mathrm{d}l$ is usually evaluated[1]. The total energy of the dislocation can be divided into two parts

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}l} = \frac{\mathrm{d}\mathcal{E}_{\mathrm{core}}}{\mathrm{d}l} + \frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} \tag{2.1}$$

where $\mathrm{d}\mathcal{E}_{\mathrm{core}}/\mathrm{d}l$ and $\mathrm{d}\mathcal{E}_d/\mathrm{d}l$ account for energy inside and outside the dislocation core region, respectively. As argued by Hull and Bacon (2002), the core energy usually constitutes only a minor contribution to the total dislocation energy for large outer radii R, and thus will be neglected in the majority of the following calculations. Recent atomistic calculations provided estimates for the dislocation core energies of various dislocation/materials systems (see sections 2.4 and 3.4.2).

Let $\hat{\boldsymbol{\sigma}}^d$ be the stress tensor associated with the strain field $\hat{\boldsymbol{\epsilon}}^d$ caused by a straight dis-

[1] for simplicity often called just the dislocation energy

location. Then the strain energy of the dislocation is given by

$$\begin{aligned}\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} &= \frac{1}{2}\sum_{i,j}\iint_V \varepsilon_{ij}^d \sigma_{ij}^d \,\mathrm{d}V \\ &= \frac{1}{2}\sum_{i,j}\iint_V \frac{\partial}{\partial x_j}(u_i\sigma_{ij}^d)\,\mathrm{d}V - \frac{1}{2}\sum_{i,j}\iint_V u_i \frac{\partial \sigma_{ij}^d}{\partial x_j}\,\mathrm{d}V \end{aligned} \tag{2.2}$$

where u_i are corresponding displacements. Recalling the equilibrium conditions for elastic media

$$\sum_j \frac{\partial \sigma_{ij}}{\partial x_j} = 0 \tag{2.3}$$

implies that the last integral in Eq. 2.2 vanishes. The Gauss-Ostrogradski theorem transforms the first integral in Eq. 2.2 into a surface integral over the surface S enclosing the volume V

$$\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} = \frac{1}{2}\sum_{i,j}\int_S u_i \sigma_{ij}^d n_j \,\mathrm{d}S \tag{2.4}$$

where n_j refers to the outer normal of the surface S (see Fig. 2.2). The volume V has to be chosen in such a way that all the functions u_i and σ_{ij}^d are continuous inside it. Therefore, the surface S must also contain the dislocation slip plane on which the displacements undergo a step change given by the Burgers vector $\boldsymbol{b}$.

2.1 Isotropic approximation

Using the explicit analytical expressions for the stress tensor of straight edge and screw dislocations one gets using Eq. 2.4 (for more details see e.g. Hirth and Lothe, 1982; Hull and Bacon, 2002)

$$\frac{\mathrm{d}\mathcal{E}_{d,\,\mathrm{screw}}}{\mathrm{d}l} = \frac{\mu b_s^2}{4\pi}\ln\left(\frac{R}{r_0}\right)\,, \tag{2.5a}$$

$$\frac{\mathrm{d}\mathcal{E}_{d,\,\mathrm{edge}}}{\mathrm{d}l} = \frac{\mu b_e^2}{4\pi(1-\nu)}\ln\left(\frac{R}{r_0}\right) \tag{2.5b}$$

where μ and ν are the shear modulus and Poisson's ratio, respectively. b_s and b_e are the screw and edge components of the Burgers vector $\boldsymbol{b}$, respectively. If θ is the angle between the dislocation line and the Burgers vector then $b_e = b\sin\theta$ and $b_s = b\cos\theta$. Therefore, the energy of a mixed dislocation in the isotropic continuum is

$$\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} = \frac{\mu b^2(1-\nu\cos^2\theta)}{4\pi(1-\nu)}\ln\left(\frac{R}{r_0}\right)\,. \tag{2.6}$$

2.2 General approach for anisotropic materials

Steeds (1973) (following earlier accounts, for example, of Foreman (1955) or Chou and Eshelby (1962)) has described a general procedure for treating dislocations inside an anisotropic infinite continuum. A detailed derivation of equations for dislocation energy in material with hexagonal symmetry was given elsewhere (see Holec, 2006a,b). Only a brief overview of the derivation is given here and the results are summarised.

In many cases, it is convenient to treat the stress fields of the edge and the screw components separately and only then to combine them using Eq. 2.4 to get the dislocation energy. This can be done for dislocation lines lying along the high-symmetry directions, as is shown in the following sections. In a general case there are "cross-terms" due to which it is not possible to separate the edge and screw components. In such cases, a fully numerical solution is used (see e.g. section 2.3).

2.2.1 Dislocations in the hexagonal *c*-plane ((0001) plane)

The solution is demonstrated on an example of a dislocation lying in the hexagonal c-plane. A coordinate frame depicted in Fig. 2.1 is used. The considered problem is a plane strain problem. As a consequence, no quantity depends on the y-coordinate, i.e. $\partial/\partial y \equiv 0$. Displacements u_x and u_z correspond to the edge component of the considered dislocation whereas u_y corresponds to the screw component. Strain components are

$$\begin{aligned} \varepsilon_{xx} &= \frac{\partial u_x}{\partial x}\,, & \varepsilon_{xy} &= \frac{1}{2}\frac{\partial u_y}{\partial x}\,, \\ \varepsilon_{yy} &= 0\,, & \varepsilon_{xz} &= \frac{1}{2}\left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x}\right)\,, \\ \varepsilon_{zz} &= \frac{\partial u_z}{\partial z}\,, & \varepsilon_{yz} &= \frac{1}{2}\frac{\partial u_y}{\partial z}\,. \end{aligned} \tag{2.7}$$

The compatibility equations in this case provide two relations

$$\frac{\partial^2 \varepsilon_{xx}}{\partial z^2} + \frac{\partial^2 \varepsilon_{zz}}{\partial x^2} = 2\frac{\partial^2 \varepsilon_{xz}}{\partial x \partial z}\,, \tag{2.8a}$$

$$\frac{\partial \varepsilon_{yz}}{\partial x} - \frac{\partial \varepsilon_{xy}}{\partial z} = 0\,. \tag{2.8b}$$

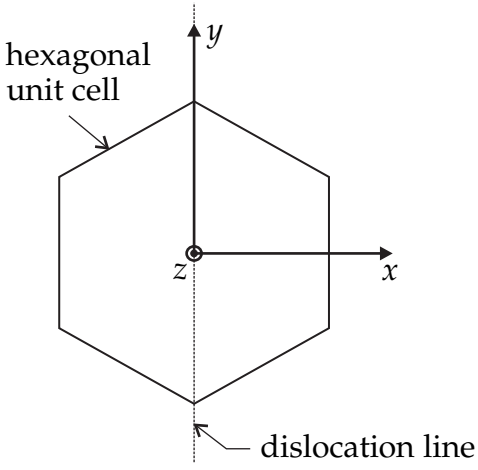


Figure 2.1: The coordinate system consists of three perpendicular axes x, y and z. The z-axis is perpendicular to the c-plane, the dislocation line lies along the y-axis.

The fact that $\varepsilon_{yy} \equiv 0$ yields an additional condition

$$0 = \varepsilon_{yy} = s_{12}\sigma_{xx} + s_{11}\sigma_{yy} + s_{13}\sigma_{zz} \quad \rightsquigarrow \quad \sigma_{yy} = -\frac{s_{12}}{s_{11}}\sigma_{xx} - \frac{s_{13}}{s_{11}}\sigma_{zz} \,. \tag{2.9}$$

Here, the compliances reflecting the hexagonal symmetry have been used

$$\begin{pmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ 2\varepsilon_{yz} \\ 2\varepsilon_{xz} \\ 2\varepsilon_{xy} \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & s_{13} & 0 & 0 & 0 \\ s_{12} & s_{11} & s_{13} & 0 & 0 & 0 \\ s_{13} & s_{13} & s_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{66} \end{pmatrix} \begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{pmatrix} \tag{2.10}$$

where $s_{66} = 2(s_{11} - s_{12})$. Steeds (1973) suggested employing two functions, F and ϕ, as potentials for expressing stresses

$$\begin{aligned} \sigma_{xx} &= \frac{\partial^2 F}{\partial z^2} \,, & \sigma_{xy} &= \frac{\partial \phi}{\partial z} \,, \\ \sigma_{xz} &= -\frac{\partial^2 F}{\partial x \partial z} \,, & \sigma_{yz} &= -\frac{\partial \phi}{\partial x} \,. \\ \sigma_{zz} &= \frac{\partial^2 F}{\partial x^2} \,, \end{aligned} \tag{2.11}$$

F and ϕ are, in fact, the Airy stress functions. Putting all together into the compatibility equations 2.8a and 2.8b yields

$$\left(s_{33} - \frac{s_{13}^2}{s_{11}}\right)\frac{\partial^4 F}{\partial x^4} + \left(2s_{13} + s_{44} - 2\frac{s_{12}s_{13}}{s_{11}}\right)\frac{\partial^4 F}{\partial x^2 \partial z^2} + \left(s_{11} - \frac{s_{12}^2}{s_{11}}\right)\frac{\partial^4 F}{\partial z^4} = 0 \,, \tag{2.12a}$$

$$s_{44}\frac{\partial^2 \phi}{\partial x^2} + s_{66}\frac{\partial^2 \phi}{\partial z^2} = 0 \,. \tag{2.12b}$$

It is clear from the last equations that in this case, the problem splits into two independent parts concerning the edge and the screw dislocation component separately.

Solution for the edge component – function *F*

The general form of the function F is (for more details see Steeds, 1973)

$$F = \sum_{n=1}^{2} \left(B_n g_n(\xi_n) + B_n^* g_n(\xi_n^*) \right) \tag{2.13}$$

where $\xi_n = x + p_n z$, $n = 1, 2$. Here, p_n and p_n^* are pairs of complex conjugate numbers and they are solutions of the characteristic equation

$$\left(s_{33} - \frac{s_{13}^2}{s_{11}}\right) + \left(2s_{13} + s_{44} - 2\frac{s_{12}s_{13}}{s_{11}}\right) p^2 + \left(s_{11} - \frac{s_{12}^2}{s_{11}}\right) p^4 = 0\,. \tag{2.14}$$

Following the discussion of Steeds (1973) and similar by Hirth and Lothe (1982) on physical requirements, the functional form of stress functions is limited to $\sigma \propto 1/r$:

$$\frac{\partial^2}{\partial \xi_n^2} g_n(\xi_n) = \frac{1}{\xi_n}\,. \tag{2.15}$$

The general form of the stress components corresponding to the edge component of the dislocation is thus

$$\sigma_{xx} = \sum_{n=1}^{2} \left(\frac{B_n p_n^2}{\xi_n} + \frac{B_n^* p_n^{*2}}{\xi_n^*} \right)\,, \tag{2.16a}$$

$$\sigma_{zz} = \sum_{n=1}^{2} \left(\frac{B_n}{\xi_n} + \frac{B_n^*}{\xi_n^*} \right)\,, \tag{2.16b}$$

$$\sigma_{xz} = -\sum_{n=1}^{2} \left(\frac{B_n p_n}{\xi_n} + \frac{B_n^* p_n^*}{\xi_n^*} \right)\,. \tag{2.16c}$$

Eq. 2.10 together with the definition of the strain components provide an expression for the displacement u_x

$$\begin{aligned} u_x &= \int \varepsilon_{xx}\, \mathrm{d}x \\ &= \sum_{n=1}^{2} \left\{ \left[\left(s_{11} - \frac{s_{12}^2}{s_{11}} \right) p_n^2 + \left(s_{13} - \frac{s_{12}s_{13}}{s_{11}} \right) \right] B_n \ln(\xi_n) + \right. \\ &\qquad \left. + \left[\left(s_{11} - \frac{s_{12}^2}{s_{11}} \right) p_n^{*2} + \left(s_{13} - \frac{s_{12}s_{13}}{s_{11}} \right) \right] B_n^* \ln(\xi_n^*) \right\}\,. \end{aligned} \tag{2.17}$$

Similarly, one obtains for u_z

$$\begin{aligned} u_z &= \int \varepsilon_{zz}\, \mathrm{d}z \\ &= \sum_{n=1}^{2} \left\{ \left[\left(s_{13} - \frac{s_{12}s_{13}}{s_{11}} \right) p_n + \left(s_{33} - \frac{s_{13}^2}{s_{11}} \right) \frac{1}{p_n} \right] B_n \ln(\xi_n) + \right. \\ &\qquad \left. + \left[\left(s_{13} - \frac{s_{12}s_{13}}{s_{11}} \right) p_n^* + \left(s_{33} - \frac{s_{13}^2}{s_{11}} \right) \frac{1}{p_n^*} \right] B_n^* \ln(\xi_n^*) \right\}\,. \end{aligned} \tag{2.18}$$

It is not difficult to show that u_x and u_z given by Eqs. 2.17 and 2.18, respectively, provide a consistent formula for σ_{xz}.

The first set of boundary conditions arises from the force equilibrium state of the media expressed as

$$\iint_S \sum_j \sigma_{ij} n_j \, \mathrm{d}S = 0 \tag{2.19}$$

for an arbitrary surface S enclosing the dislocation line. n_j denotes components of the outer normal to the integration surface S. This condition leads to

$$\sum_{n=1}^{2} (B_n p_n - B_n^* p_n^*) = 0 \qquad \text{for } i = x \,, \tag{2.20a}$$

$$\sum_{n=1}^{2} (B_n - B_n^*) = 0 \qquad \text{for } i = z \,. \tag{2.20b}$$

The second set of boundary equations is provided by the displacement relations. The integral of the displacement acquisitions along a closed loop encircling the dislocation line is equal to the Burgers vector

$$\oint \mathrm{d}\boldsymbol{u} = \boldsymbol{b} \,. \tag{2.21}$$

Substituting u_x from Eq. 2.17 into the above formula yields

$$b_x = 2\pi i \left(s_{11} - \frac{s_{12}^2}{s_{11}} \right) \sum_{n=1}^{2} \left(p_n^2 B_n - p_n^{*2} B_n^* \right) \,. \tag{2.22a}$$

Similarly, for the z component one obtains

$$b_z = 2\pi i \left(s_{33} - \frac{s_{13}^2}{s_{11}} \right) \sum_{n=1}^{2} \left(\frac{B_n}{p_n} - \frac{B_n^*}{p_n^*} \right) \,. \tag{2.22b}$$

The stress field of the edge component of the dislocation is fully described by Eqs. 2.16a–2.16c where constants p_n are roots of Eq. 2.14, and constants B_n are solutions of the system of linear equations 2.20a, 2.20b, 2.22a and 2.22b.

Solution for the screw component of a dislocation – function ϕ

A solution for the screw component can be found in exactly the same way as shown above for the edge component. The remaining stress components are

$$\sigma_{xy} = \frac{C p_3}{\xi} + \frac{C^* p_3^*}{\xi^*} \,, \tag{2.23a}$$

$$\sigma_{yz} = -\frac{C}{\xi} - \frac{C^*}{\xi^*} \,, \tag{2.23b}$$

where $p_3 = i\sqrt{s_{44}/s_{66}}$. The boundary conditions provide additional relations for the constants C and C^*. Eq. 2.19 for $i = y$ gives

$$C - C^* = 0\,. \tag{2.24a}$$

The condition for the displacement u_y yields

$$-2\pi i s_{44}\left(\frac{C}{p_3} - \frac{C^*}{p_3^*}\right) = b_y\,. \tag{2.24b}$$

The stress field associated with the screw component of the dislocation is given by Eqs. 2.23a and 2.23b, where $p_3 = i\sqrt{s_{44}/s_{66}}$ and the coefficients C and C^* are given by solving the system of linear equations 2.24a and 2.24b as

$$C = C^* = -\frac{b_y}{4\pi\sqrt{s_{44}s_{66}}}\,. \tag{2.25}$$

Energy of a dislocation

A dislocation may be produced by making a cut along the intended slip plane, displacing one cut surface relative to the other by the Burgers vector $\boldsymbol{b}$ and welding the material back together. This is shown schematically in Fig. 2.2. The work done by this procedure transforms to (and is equal to) the energy of the dislocation. If the dislocation is inside an infinite medium, then the surface S from Eq. 2.4 consists of parts S_1 and S_2 (corresponding to the cut along the slip plane), and a part enclosing the dislocation core. The difference between the displacements on the surfaces S_1 and S_2 equals to the Burgers vector $\boldsymbol{b}$. As a consequence, Eq. 2.4 simplifies to

$$\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} = \frac{1}{2}\int_{S_2}\sum_{i,j} b_i\sigma_{ij}^d n_j\,\mathrm{d}S + \frac{1}{2}\int_{\text{around core}}\sum_{i,j} b_i\sigma_{ij}^d n_j\,\mathrm{d}S \tag{2.26}$$

where n_j now denotes the outer normal to the surface S_2. The latter part of Eq. 2.26 is neglected in the present calculation. The outer cut-off radius is R, the dislocation core radius is denoted as r_0. Then one obtains

$$\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} = \frac{1}{2}\sum_{n=1}^{3}\left[B_n(-b_x p_n + b_z) - C_n b_y + B_n^*(-b_x p_n^* + b_z) - C_n^* b_y\right]\ln\frac{R}{r_0} \tag{2.27}$$

where a simplifying notation $B_3 = 0$, $C_1 = 0$, $C_2 = 0$ and $C_3 = C$ was used.

2.2.2 Dislocations along the hexagonal *c*-axis ([0001] direction)

The energy of a dislocation lying in the hexagonal c-plane was derived in the previous section. Exactly the same approach can now be applied to a dislocation with an

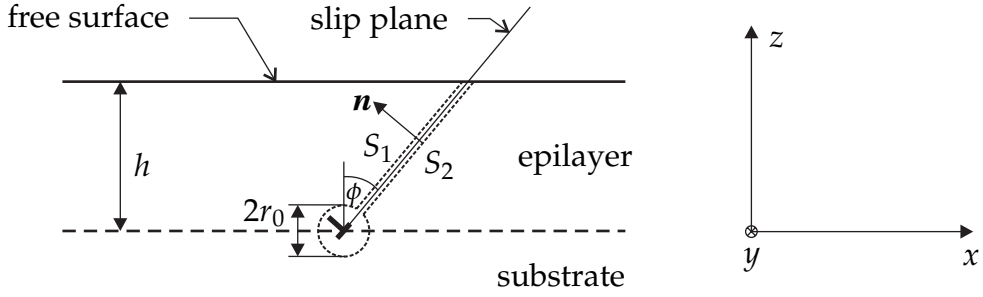


Figure 2.2: The coordinate system and the cut plane for the cut-displace-weld procedure used for the estimation of the dislocation energy as described in the text. Burgers vector $\boldsymbol{b}$ lies in the slip plane.

arbitrary dislocation line direction. The only difference is that the solutions for the edge and screw components do not separate in the general case. As a consequence, a more complex system of equations for coefficients B_n and C_n is obtained. The general expression for the dislocation energy can be written as

$$\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} = \sum_{i,j} K_{ij} b_i b_j \ln\frac{R}{r_0} = A_{\mathrm{cont}} \ln\frac{R}{r_0} \tag{2.28}$$

where K_{ij} are numerical parameters depending on the elastic constants and the particular direction of the dislocation line with respect to the hexagonal axes. A_{cont} is the so called *prelogarithmic term* (see section 2.4).

An exception is the case of a dislocation along the hexagonal c-axis, i.e. along the [0001] direction (along the z-axis in the chosen cartesian coordinate frame). Due to the isotropic symmetry of the hexagonal c-plane, Eqs. 2.12a and 2.12b take a different form and thus prevent using exactly the same procedure as described in section 2.2.1.

Energy of the threading dislocation (along the [0001] direction)

Compatibility equations 2.8a and 2.8b now form a simpler system than in section 2.2.1 and take the same form as in the isotropic case:

$$\frac{\partial^4 F}{\partial x^4} + 2\frac{\partial^4 F}{\partial x^2 \partial y^2} + \frac{\partial^4 F}{\partial y^4} = 0\,, \tag{2.29}$$

$$\frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial x^2} = 0\,. \tag{2.30}$$

The different symmetry (hexagonal instead of isotropic) arises through the boundary conditions.

Solution for the edge component can be found by following a similar derivation of Hirth and Lothe (1982) for the isotropic media. Without loss of generality, the coordinate system can be rotated in the c-plane so that the edge component $\boldsymbol{b}_e$ of the Burgers

vector $\boldsymbol{b}$ lies along the x-axis. The solution is

$$F = \frac{\beta}{2} y \ln\left(x^2 + y^2\right) \,. \tag{2.31}$$

The boundary condition expresses the idea of Volterra dislocation formation. This yields β:

$$-b_e = \lim_{y\to 0, y>0} \int_{-\infty}^{\infty} \left(\varepsilon_{xx}(x,y) - \varepsilon_{xx}(x,-y)\right) \mathrm{d}x = \beta \left(s_{11} - \frac{s_{13}^2}{s_{33}}\right) 2\pi \quad \rightsquigarrow$$

$$\rightsquigarrow \quad \beta = -\frac{b_e}{2\pi} \frac{1}{s_{11} - s_{13}^2/s_{33}} \,. \tag{2.32}$$

The strain energy for the edge component of the dislocation becomes

$$\frac{\mathrm{d}\mathcal{E}_{d,\mathrm{edge}}}{\mathrm{d}l} = \frac{b_e^2}{8\pi} \frac{s_{33}}{s_{11}s_{33} - s_{13}^2} \ln\frac{R}{r_0} \,, \quad b_e^2 = b_x^2 + b_y^2 \,. \tag{2.33}$$

The solution of Eq. 2.30 is a simplified version of the one presented in section 2.2.1. The dislocation energy per unit length for the screw component $\boldsymbol{b}_z$ is thus given by

$$\frac{\mathrm{d}\mathcal{E}_{d,\mathrm{screw}}}{\mathrm{d}l} = \frac{b_s^2}{4\pi s_{44}} \ln\frac{R}{r_0} \,, \quad b_s = b_z \,. \tag{2.34}$$

All together, the final formula for the dislocation energy is

$$\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} = \frac{1}{4\pi} \left(\frac{s_{33}}{2(s_{11}s_{33} - s_{13}^2)} (b_x^2 + b_y^2) + \frac{1}{s_{44}} b_z^2 \right) \ln\frac{R}{r_0} \,. \tag{2.35}$$

2.2.3 Numerical comparison

To give a better idea of the difference between the above derived formulae compared with the isotropic ones, some numerical values for the $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}$ system are listed below. A dislocation lying in the c-plane of $\mathrm{In}_{0.20}\mathrm{Ga}_{0.80}\mathrm{N}$ has energy

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}l} = (9.05 b_x^2 + 8.06 b_y^2 + 9.22 b_z^2) \ln\frac{R}{r_0} \times 10^9\,\mathrm{GPa} \,, \tag{2.36}$$

whereas a dislocation lying perpendicular to the c-plane (i.e. along the $[0001]$ direction) has energy

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}l} = (11.12(b_x^2 + b_y^2) + 6.47 b_z^2) \ln\frac{R}{r_0} \times 10^9\,\mathrm{GPa} \,. \tag{2.37}$$

The isotropic equivalents of elastic constants for GaN were taken from Madelung (2004): Young's modulus $E = 324\,\mathrm{GPa}$ (obtained from Brillouin scattering) and Bulk modulus

$B = 195\,\mathrm{GPa}$ (powder X-ray diffraction data). The shear modulus and Poisson's ratio were calculated as (Brdička et al., 2000)

$$\mu = \frac{3BE}{9B - E} = 132\,\mathrm{GPa}\,, \tag{2.38a}$$

$$\nu = \frac{3B - E}{6B} = 0.22\,. \tag{2.38b}$$

As for InN, Madelung (2004) gives only a value for the bulk modulus $B = 147\,\mathrm{GPa}$ (calculated). Using Poisson's ratio $\nu = 0.29$ (Vickers et al., 2003), the shear modulus was calculated as (Brdička et al., 2000)

$$\mu = 3B\frac{1 - 2\nu}{2 + 2\nu} = 71\,\mathrm{GPa}\,. \tag{2.39}$$

The isotropic formula for $In_{0.20}Ga_{0.80}N$ then gives

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}l} = (12.46b_e^2 + 9.57b_s^2)\ln\frac{R}{r_0} \times 10^9\,\mathrm{GPa}\,. \tag{2.40}$$

All the above Burgers components are in nanometres, results are in J/m. It is clear from these examples and from Table 2.1 that when the hexagonal symmetry is considered, lower energies that in the isotropic case are obtained. Moreover, it demonstrates that the dislocation energy depends differently on the Burgers vector's components for various dislocation line orientations as shown also in Table 2.1.

	A_{cont} [eV/Å]			
	$\boldsymbol{l} \perp [0001]$	$\boldsymbol{l} \perp [0001]$	$\boldsymbol{l} \parallel [0001]$	$\boldsymbol{l} \parallel [0001]$
	hexagonal	isotropic	hexagonal	isotropic
$b = a,\ \boldsymbol{b} \parallel \boldsymbol{x}$	1.54, screw	1.63, screw	1.89, edge	2.12, edge
$b = a,\ \boldsymbol{b} \parallel \boldsymbol{y}$	1.37, edge	2.12, edge	1.89, edge	2.12, edge
$b = c,\ \boldsymbol{b} \parallel \boldsymbol{z}$	4.14, edge	5.59, edge	2.90, screw	4.29, screw

Table 2.1: Prelogarithmic terms A_{cont} (Eq. 2.28) for two different dislocation line directions $\boldsymbol{l}$ (in the c-plane ($\boldsymbol{l} \perp [0001]$) and along the c-direction) and three Burgers vectors oriented along the three Cartesian axis evaluated with (Eqs. 2.36 and 2.37) and without (Eq. 2.40) hexagonal symmetry.

2.2.4 Effect of the free surface for dislocations in the *c*-plane

A dislocation in an infinite material was assumed in all the above derivations. However, real structures contain well confined layers (at least in one direction) and thus a question about the effect of hetero-interfaces and free surfaces arises.

Willis et al. (1990) presented a mathematical derivation of the dislocation energy in isotropic material within the vicinity of a free surface. The same approach but with hexagonal symmetry taken into account was used previously by Holec (2006b). Moreover, a system of two different materials was investigated in the latter report. As a consequence, a solution corresponding to a dislocation lying in a hetero-interface of a thick substrate and a thin epilayer was obtained. The numerical results summarised by Holec (2006b) suggest that the effect of the free surface more or less cancels the effect of the hexagonal symmetry in the case of an $In_xGa_{1-x}N$ epilayer on a GaN substrate (for $x \lessapprox 0.3$).

This solution needs a tedious numerical treatment and is computationally very complex. In addition, these results do not bring any new insight and some approximations used are debatable. Therefore, results using this approach are not further exploited in this work. Readers interested in a more detailed discussion are referred to the report by Holec (2006b).

2.3 Geometrical properties of dislocations in wurtzite GaN

Many considerations about dislocations are based on the minimum energy criterion, which is one of the basic physical principles. Sections 2.1 and 2.2 gave recipes for the calculation of the dislocation energy in isotropic and hexagonal materials. Two applications of the above will be given in this section, both showing significant differences between isotropic and hexagonal symmetries.

2.3.1 Notation used for description of geometric relations

All directions are equivalent in the isotropic case. As a consequence, only one angle, ϑ, is needed to fully describe a particular configuration of a dislocation line direction and its Burgers vector (see Fig. 2.3a). Hexagonal symmetry requires two angles, θ and ϕ, for the description of the dislocation line direction (see Fig. 2.3b); the Burgers vector is described separately in this case.

2.3.2 Dislocation types in bulk

For a given Burgers vector $\boldsymbol{b}$ one may question which orientation of the dislocation line is energetically the most favourable. The minimum value of $\mathrm{d}\mathcal{E}_d^{\mathrm{iso}}/\mathrm{d}l$ with respect to ϑ is for $\vartheta = 0°$ regardless of the dislocation line orientation. Therefore, the screw-type

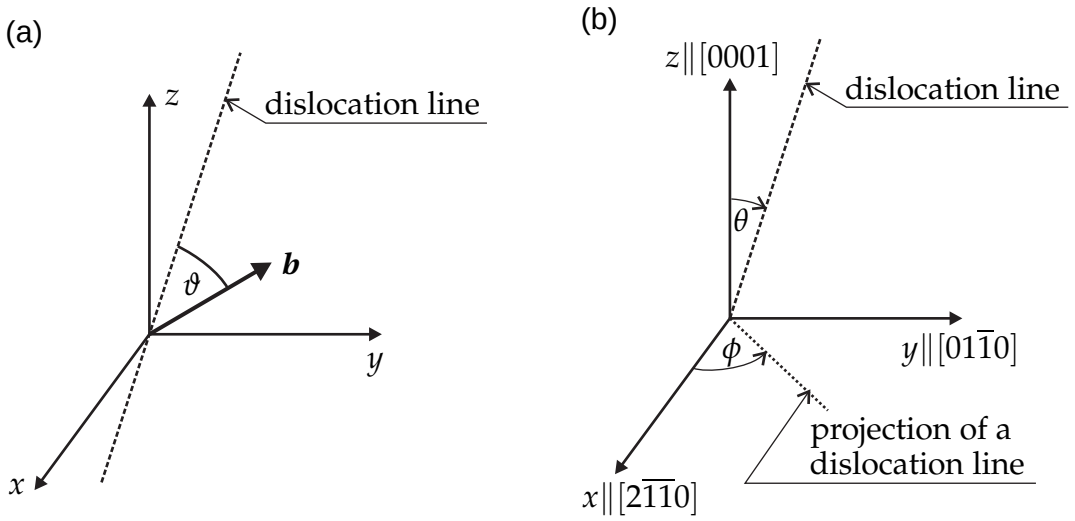


Figure 2.3: (a) Only the angle ϑ between the dislocation line and the Burgers vector is needed to fully describe the geometrical configuration in the isotropic case. (b) In the hexagonal (and more generally, in the non-isotropic case) two angles, θ and ϕ, are needed for description of the dislocation line direction.

dislocation[2] has always the lowest energy within the isotropic approximation.

The hexagonal dislocation energy of the c-type dislocation ($\boldsymbol{b} = [0001]$) in GaN as a function of the two angles θ and ϕ is plotted in Fig. 2.4. One may notice that the dislocation energy does not depend on the angle ϕ. This is because the hexagonal c-plane is isotropic and therefore for this Burgers vector all dislocation lines with different angles ϕ are equivalent. It follows from Fig. 2.4 that the lowest energy configuration is obtained for $\theta = 0°$ or $\theta = 180°$ which both correspond to the screw-type dislocation.

The situation is different for the a-type ($\boldsymbol{b} = \frac{1}{3}\langle\bar{2}110\rangle$) and the $(a+c)$-type ($\boldsymbol{b} = \frac{1}{3}\langle\bar{2}113\rangle$) dislocations. The dislocation energies as functions of the angles θ and ϕ are shown in Figs. 2.5 and 2.6.

Dislocation line directions with the lowest energy of the a-type dislocation are $\phi = 0°$ or $\phi = 180°$ and $\theta \approx 65°$ or $\theta \approx 115°$. As the Burgers vector $\boldsymbol{b} = \frac{1}{3}\langle\bar{2}110\rangle$ is characterised by $\phi_{\boldsymbol{b}} = 0°$ and $\theta_{\boldsymbol{b}} = 90°$, the Burgers vector and the dislocation line are coplanar and the angle between them is approximately 25°. Therefore, the energetically most favourable configuration for the a-type dislocation is a mixed-type dislocation, unlike the screw-type dislocation predicted by the isotropic approximation.

The lowest dislocation energy of the $(a+c)$-type dislocation is obtained for the dislocation line along the $[0001]$ direction ($\theta = 0°$ or $\theta = 180°$) whereas the Burgers vector $\boldsymbol{b} = \frac{1}{3}\langle\bar{2}113\rangle$ is characterised by $\phi_{\boldsymbol{b}} = 0°$ and $\theta_{\boldsymbol{b}} \approx 31.6°$. Therefore, the angle between the Burgers vector and the dislocation line is approximately 31.6°. As a consequence,

[2] A screw-type dislocation has its Burgers vector parallel to the dislocation line direction.

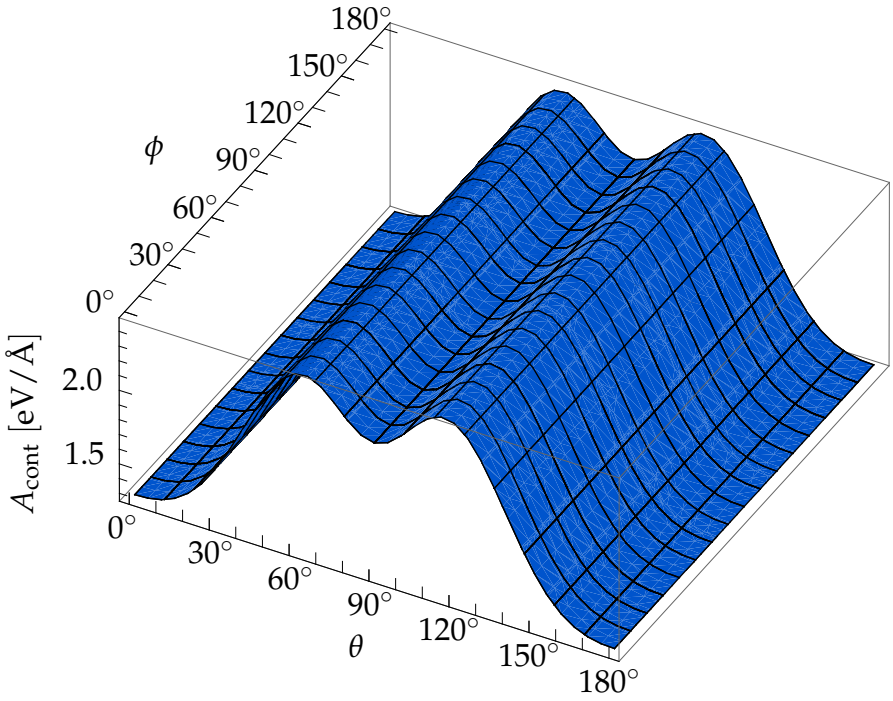


Figure 2.4: Dislocation energy per unit length for *c*-type dislocations (Burgers vector $\boldsymbol{b} = [0001]$) in GaN for various dislocation line orientations given by different angles θ and ϕ.

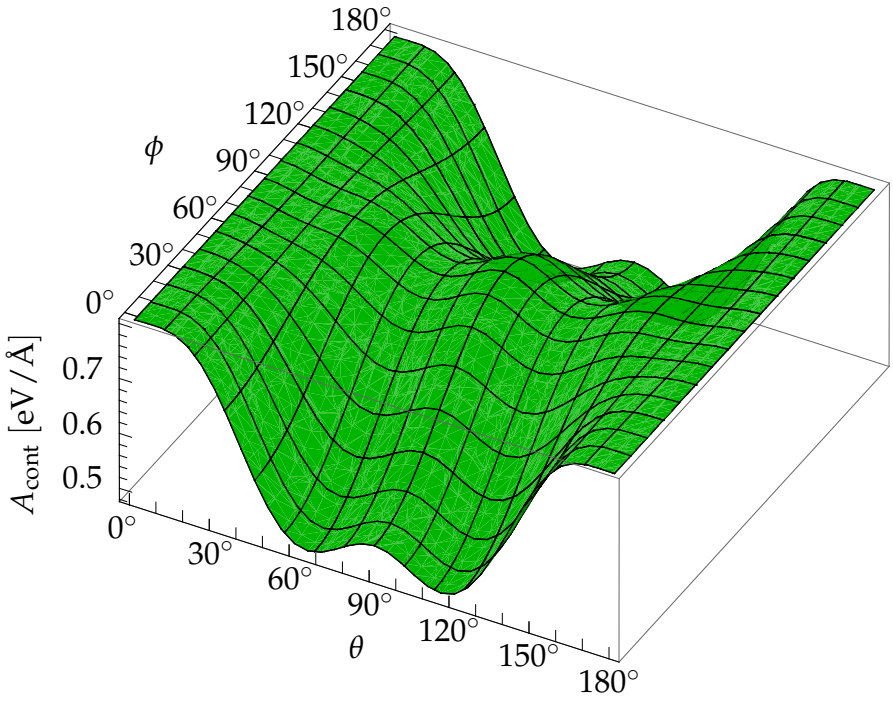


Figure 2.5: Dislocation energy per unit length for *a*-type dislocations (Burgers vector $\boldsymbol{b} = \frac{1}{3}\langle 2\overline{1}\overline{1}0\rangle$) in GaN for various dislocation line orientations given by different angles θ and ϕ.

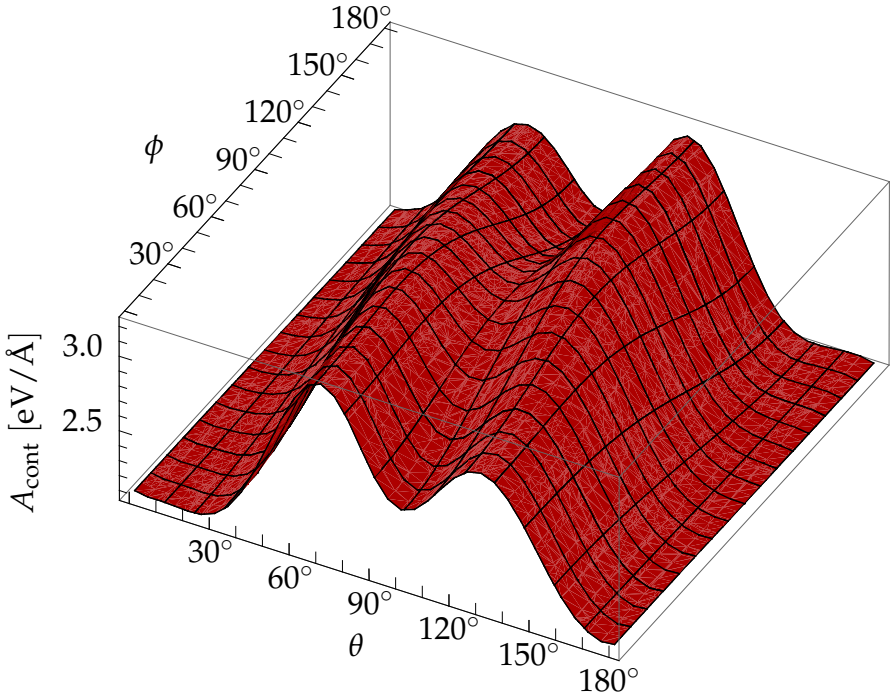


Figure 2.6: Dislocation energy per unit length for $(a + c)$-type dislocations (Burgers vector $\boldsymbol{b} = \frac{1}{3}\langle 2\bar{1}\bar{1}3\rangle$) in GaN for various dislocation line orientations given by different angles θ and ϕ.

the energetically most favourable dislocation type is again the mixed-type dislocation which contradicts the prediction of the isotropic theory.

2.3.3 Dislocation types near the free surface

In real finite crystals, dislocations which are not closed loops have to terminate at the free surfaces of crystals and thus have to have finite length. As a consequence, the dislocation direction is not only given by the minimum of the dislocation energy per unit length $\mathrm{d}\mathcal{E}_d/\mathrm{d}l$. Also the actual length of the dislocation must be taken into account.

How does a dislocation know which direction is potentially the most favourable for it? This comes as a result of growth. When a thin layer is grown on top of a surface where a dislocation terminates, the dislocation extends in order to terminate at the (new) free surface again. In this very thin layer the dislocation is mobile enough to adopt the energetically most convenient direction.

Isotropic approximation

In addition to the angle ϑ between the Burgers vector and the dislocation line, φ describes the angle between the dislocation line and the $[0001]$ direction. If the film thick-

ness grows by Δh, then the dislocation length increases by $\Delta h/\cos\varphi$. The dislocation energy increases by

$$\Delta\mathcal{E}_d = \frac{\mu b^2(1-\nu\cos^2\vartheta)}{4\pi(1-\nu)}\frac{\Delta h}{\cos\varphi}\ln\frac{R}{r_0}\,. \tag{2.41}$$

Minimisation of Eq. 2.41 with respect to φ for different kinds of Burgers vectors (different ϑs) gives the following predictions: for the c-type dislocations ($\vartheta+\varphi=0°$) the optimal value is $\varphi=0°$, for the a-type dislocations ($\vartheta+\varphi=90°$) the optimal value is $\varphi=0°$ (dislocation line along $[0001]$ direction) and for the $(a+c)$-type dislocations ($\vartheta+\varphi\approx 32°$) the optimal value is $\varphi\approx 15.6°$. The same results were obtained and published by Mathis et al. (2001).

One should note that this approach is not absolutely correct as the energy does not include the effect of a free surface (see section 2.2.4). However, it is supposed to be good enough to give some feeling about what are the important differences between isotropic and hexagonal theories.

Dislocations in the wurtzite material with (0001) surface

The same approach is now used for the wurtzite material with the (0001) free surface. Because of the complexity of the dislocation energy formula, all results were obtained numerically. The calculation reveals that for the (0001) surface, dislocation lines along the $[0001]$ direction constitute the lowest energy configuration for all three dislocation types (a-, $(a+c)$- and c-type). This contradicts the results obtained within the isotropic approximation. It is also in good agreement with the experimental TEM observations where dislocations (regardless of their type) are usually vertical rather than inclined straight lines in thick GaN.

Different surface facets

Hiramatsu et al. (1999) discussed changes in growth modes and the preferred growth facets at different temperatures and III/V ratios. They concluded that for the 3D growth mode, the most preferable growth facets are $\{11\bar{2}2\}$ and $\{1\bar{1}01\}$. Therefore, analysis similar to the above was performed also for these facets.

All together, there are 20 different Burgers vectors: $\pm\boldsymbol{c}$, $\pm\boldsymbol{a}_i$ and $\pm\boldsymbol{c}\pm\boldsymbol{a}_i$ ($i=1,2,3$). The results are presented graphically in Figs 2.7 and 2.8. It can be concluded that the a-type dislocations (green lines) tend to propagate almost perpendicularly to the facets whereas the $(a+c)$-type and the c-type dislocations are inclined towards the $[0001]$

direction. A TEM image showing such behaviour has been reported by Lang et al. (2007).

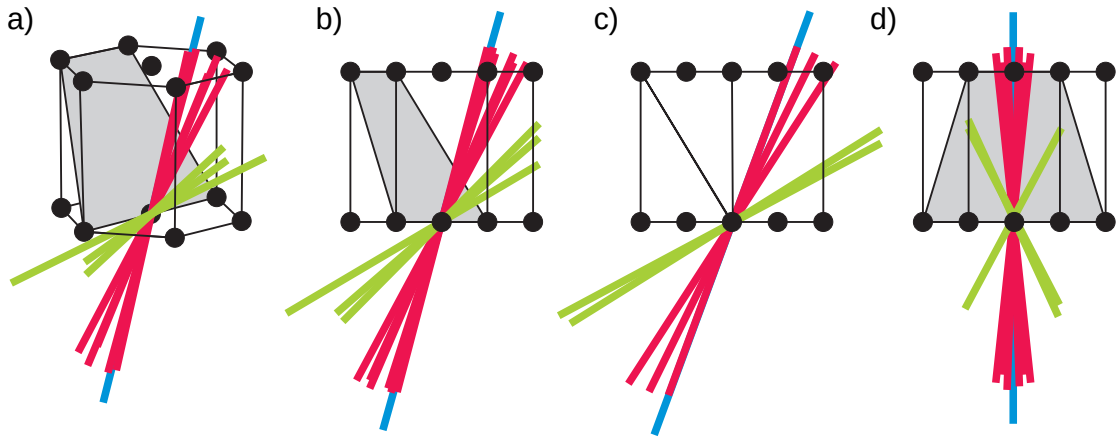


Figure 2.7: Dislocation lines' directions near the $(1\bar{1}01)$ facet for several projections: (a) a 3D view and a view with the direction of projection being (b) $[01\bar{1}0]$, (c) $[11\bar{2}0]$ and (d) $[\bar{1}100]$. The green, red and blue lines represent the a-, $(a+c)$- and c-type dislocations, respectively.

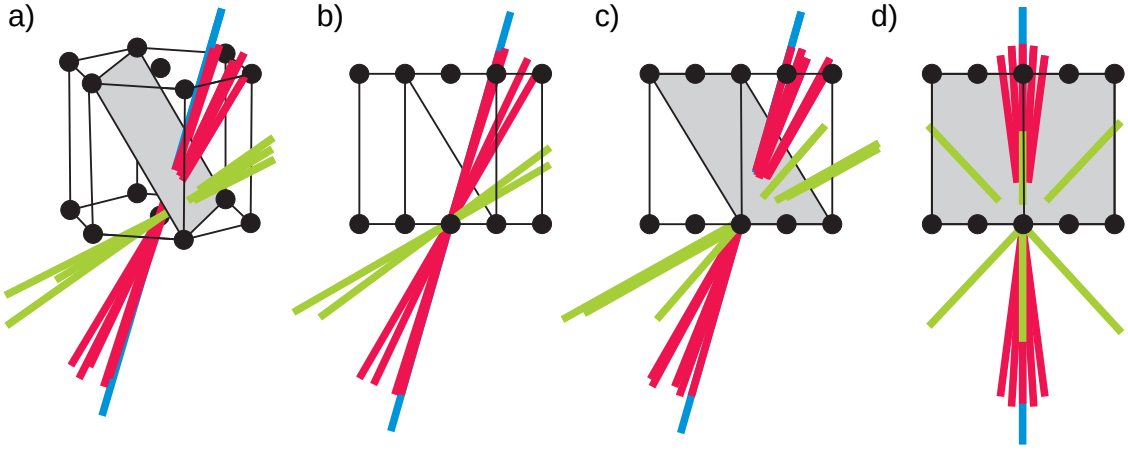


Figure 2.8: Dislocation lines' directions near the $(2\bar{1}\bar{1}2)$ facet for several projections: (a) a 3D view and a view with the direction of projection being (b) $[01\bar{1}0]$, (c) $[11\bar{2}0]$ and (d) $[\bar{1}100]$. The green, red and blue lines represent the a-, $(a+c)$- and c-type dislocations, respectively.

2.4 Beyond linear elasticity – energy of the dislocation core

Recent atomistic calculations provided some details about dislocation core structures in III–nitrides. In each case, a comparison of several dislocation core configurations yielded the energetically most favourable structure. It turned out that the 5/7-atom ring core possesses the lowest energy configuration in all investigated cases. Despite this, experimental observations of some other configurations were reported in the

material	dislocation	r_c [nm]	E_{core} [eV/Å]	A [eV/Å]	reference	A_{cont} [eV/Å]
GaN	$(a+c)$-type	0.72	3.12	2.15	[1]	2.12
GaN	a-type	0.7	1.57	0.91	[2]	0.81
GaN	a-type	0.60	1.61	0.80	[3]	0.81
AlN	a-type	0.83	1.71	0.83	[3]	0.90
InN	a-type	0.54	1.66	0.56	[3]	0.41

[1] Belabbas et al. (2007)

[2] Lymperakis (2005)

[3] Kioseoglou et al.

Table 2.2: Parameters of the 5/7-atom ring dislocation cores resulting from the atomistic calculations. A_{cont} is the prelogarithmic term from Eq. 2.35 obtained within the linear theory of elasticity. A is an estimation of the same quantity, this time based on the atomistic calculations.

literature too. Xin et al. (1998) observed an 8-atom ring whereas Lymperakis (2005) reported on an 4-atom ring configuration.

Belabbas et al. (2007) estimated that the dislocation core radius of the $(a+c)$-type dislocation in GaN is $\approx b$. Lymperakis (2005) and Kioseoglou et al. calculated $r_0 \approx 2b$ for the a-type dislocations in III-nitrides. All these calculations were done for dislocation lines along the $[0001]$ direction. Belabbas et al. (2007), Lymperakis (2005) and Kioseoglou et al. also provided estimates of the core energies E_{core} for these dislocations. Their results, as will be used later in this work, are summarised in Table 2.2.

A very good agreement of the prelogarithmic factors based on the continuum elasticity theory, A_{cont}, and resulting from the atomistic calculations, A, was obtained. This suggests both approaches yielded comparable results and thus can be combined into a kind of a multiscale approach (see section 3.4.2).

2.5 Summary

A detailed derivation of the dislocation energy with hexagonal symmetry being taken into account was given in this chapter. The results were subsequently applied to two case studies to demonstrate some significant differences between isotropic and hexagonal approximations. In the first one, geometrical properties of straight dislocations in an infinite medium were investigated. It has been shown that the often used ar-

gument "the screw-type dislocation has lower energy than the mixed- and edge-type dislocations" does not hold in the case of wurtzite GaN. A similar conclusion is expected also for other wurtzite III-nitrides. The second study focused on an estimate of the energetically optimal dislocation line directions near a free surface of a finite crystal. Again, different conclusions result from the isotropic and hexagonal theories. The hexagonal theory predicts (in agreement with TEM observations) that the dislocation line along the $[0001]$ direction possesses the lowest energy configuration regardless of the dislocation type. a-type dislocations tend to propagate almost perpendicularly to the free surface whereas the $(a+c)$- and c-type dislocations are slightly inclined towards the $[0001]$ direction in the case of inclined $\{11\bar{2}2\}$ and $\{1\bar{1}01\}$ facets. In the end of the chapter, a summary of results from the literature on atomistic calculations of dislocation core structures of III-nitrides was given.

Chapter 3

Critical thickness

The generation of misfit strain during hetero-epitaxial growth is a well known phenomenon. Misfit strain arises from the difference between the in-plane lattice parameter(s) of the substrate and the epilayer. The basic idea of epitaxial growth theory is that for very thin layers it is energetically favourable to accommodate the misfit strain elastically while for thicker epilayers the strain is accommodated by introducing defects. These are *misfit dislocations* (see e.g. Srinivasan et al., 2003), *V-defects* (see e.g. Jahnen et al., 1998), *surface roughening* (see e.g. Tersoff and Legoues, 1994), *stacking faults* (see e.g. Johnston et al., 2008) or *cracking* (see e.g. Vennéguès et al., 2005). V-defects are commonly observed in $In_xGa_{1-x}N/GaN$ systems whereas cracking is typical for $Al_xGa_{1-x}N/GaN$ systems. Stacking faults have been reported in non-polar and semi-polar GaN grown on sapphire.

Probably the most commonly observed defects are misfit dislocations (MDs) which can be found in silicon semiconductors, GaAs-based systems and also in III-nitrides. MDs lie in the epilayer–substrate interface. The minimum layer thickness at which MDs are introduced is called the *critical thickness* (CT) and is denoted by h_c throughout this text.

A summary of various CT models, several modifications to the energy balance model and a critical comparison with available experimental data are presented in this chapter.

3.1 Experimental critical thickness

3.1.1 Methods of measuring the critical thickness

Several experimental techniques exist to determine the critical thickness. One of the most straightforward is to use *transmission electron microscopy* (TEM) to observe directly the misfit dislocations. For example, a plot of the critical thickness against the compo-

sition can be obtained from TEM by investigating a series of samples that either have varying layer thicknesses with constant composition, or varying composition with constant thickness. However, this method is time demanding and needs a lot of samples. On the other hand, it provides a direct view of the strained hetero-interface and, in addition, constitutes a simple method of how to measure the layer thickness. TEM and other electron microscopy techniques are often used in combination with other methods (described below) to validate the observations. However, the detection limit of the dislocation generation onset in TEM is poor (typically for the dislocation density of the order of $10^4\,\mathrm{cm}^{-2}$) due to a very small area of TEM samples.

Photoluminescence microscopy (PLM) is another technique for a direct observation of the dislocations in the interface (see e.g. Gourley et al. (1988)). To obtain PLM images, the sample is irradiated with laser light of an appropriate wavelength (depending on the investigated material) which is focused to a fine spot. The light is absorbed near the incident surface and the generated electrons and holes are then transported to the quantum well(s) where they recombine with high efficiency. As dislocations act as non-radiative recombination centres they appear as dark line defects in the final PLM micrograph. The determination of the critical thickness is then similar to that by TEM.

Other methods are also *ex situ* approaches and therefore need a series of samples but, compared with TEM or PLM, misfit dislocations are not directly observed. *X-ray diffraction* provides accurate data for determining the in-plane (perpendicular to the growth direction) lattice constants by analysing the peaks of X-ray rocking curves (used for example by Bean et al. (1984) or Park et al. (2003)). For a strained layer the in-plane lattice constant is the same as that of the substrate. After the critical thickness is reached, relaxation of the film occurs and separate peaks corresponding to the layer and the substrate can be observed. Generally, it seems that X-ray diffraction provides higher values of the critical thickness than PLM measurements which points to the fact that X-ray diffraction is not capable of picking up the absolute onset of strain relaxation (see Gourley et al., 1988, and references therein).

Reed et al. (2000) pointed out that broad X-ray rocking curve widths and the high density of threading dislocations make X-ray analysis more difficult in the case of III-nitrides. Therefore, Parker et al. (1999) and Reed et al. (2000) used *photoluminescence* (PL). The method is based on the red-shift of the PL spectra as the layer thickness increases. Further investigation showed that this red-shift is closely connected with the onset of the relaxation process and therefore may be used as an indicator for determination of the critical thickness.

Reed et al. (2000) suggested another possibility of measuring the critical thickness. They used data from measurements of the electrical properties (mobility and conductivity). They found these measurements in good agreement with the values determined by the PL technique although depending on the chosen method the results are systematically shifted.

Floro et al. (2004) measured stress *in situ* during deposition. They used a multibeam optical stress sensor (MOSS), a laser deflectometry technique. It measures the curvature of a wafer imposed by the stress in an epilayer. The curvature can be directly related to the stress in the epilayer and thus change in the curvature marks the onset of strain relief.

3.1.2 Overview of published experimental results for GaN-based systems

There are very few papers reporting on critical thickness of III-nitride systems, namely $In_xGa_{1-x}N/GaN$ and $Al_xGa_{1-x}N/GaN$ systems. This is probably due to obstacles and uncertainty in the experimental techniques described in the previous section. Despite this, a collection of published results focused on the presence of MDs was assessed for a critical comparison with the herein presented theoretical calculations.

InGaN/GaN systems

InGaN epilayers on GaN substrates are under compressive stress. The most commonly observed defects are V-defects which are connected to dislocations threading from the GaN substrate. However, MDs are quite often observed too. Jahnen et al. (1998) suggested that V-defects may be a source for MDs. Experimental evidence for such a mechanism was given by Costa et al. (2006) and it was subsequently observed by Liu et al. (2006b). These results were obtained on structures grown on conventional GaN substrates with a threading dislocation density (TDD) of about $10^9\,cm^{-2}$. Observations tagged in Table 3.1 as "low TDD substrates" used GaN substrates with TDD of about $10^7\,cm^{-2}$. Absence of TDs which may serve either as MD nucleation centres themselves or as nucleation sites for V-pits is probably the reason for thicker dislocation free epilayers (see e.g. reports by Liu et al. (2006b) or Srinivasan et al. (2003)).

The available experimental data for the $In_xGa_{1-x}N/GaN$ system are summarised in Table 3.1.

x	h [nm]	MDs?	ref.	notes
0.05	100	no	[1]	low TDD substrate
0.11	100	yes	[1]	half-loops generated at apexes of V-pits, a-type dislocations ($\boldsymbol{b} = \boldsymbol{a}$), low TDD substrate
0.17	100	yes	[1]	in addition to half-loops, straight lines are generated, a-type dislocations ($\boldsymbol{b} = 2\boldsymbol{a}$), low TDD substrate
0.10	100	h_c	[2]	$(a+c)$-type MDs arranged in arrays, low TDD substrate
0.18	6.0	yes	[3]	a-type MDs, MQW sample
0.11	100	h_c	[4]	punch-out mechanism of MDs generation ($(\boldsymbol{a}+\boldsymbol{c}) + (\boldsymbol{a}-\boldsymbol{c}) \rightarrow 2\boldsymbol{a}$)
0.17	100	yes	[4]	a-type MDs ($\boldsymbol{b} = 2\boldsymbol{a}$), long straight lines, low TDD substrate
0.33	112	yes	[5]	MDs present in the heterointerface *and* also in the QW itself
0.14	100	h_c	[6]	MDs generated at apexes of V-pits
0.20–0.25	3.0	yes	[7]	MQW samples, observed MD loops
0.06	130	h_c	[8]	CL imaging
0.08	102	h_c	[8]	
0.15	74	h_c	[8]	
0.20	60	h_c	[8]	
0.05	100	h_c	[9]	averaged CL imaging, conductivity and mobility measurements
0.10	33	h_c	[9]	
0.16	14	h_c	[9]	
0.13	39.7	no	[10]	single layer, based on TEM
0.08	36.6	no	[10]	
0.03	33.5	no	[10]	
0.08	5.7	no	[11]	5 layers with different compositions in one sample (separation about 100 nm)
0.12	5.7	no	[11]	
0.18	6.4	no	[11]	
0.22	6.4	yes	[11]	
0.23	6.4	yes	[11]	

[1] Liu et al. (2006b)
[2] Srinivasan et al. (2003)
[3] Lü et al. (2004)
[4] Liu et al. (2006a)
[5] Cho and Yang (2002)
[6] Jahnen et al. (1998)
[7] Costa et al. (2006)
[8] Parker et al. (1999)
[9] Reed et al. (2000)
[10] Unpublished results. Courtesy of M. Kappers and Y. Zhang, University of Cambridge.
[11] Unpublished results. Courtesy of M. Kappers and S. Duggi, University of Cambridge.

Table 3.1: Experimental observations of MDs in $In_xGa_{1-x}N/GaN$ systems. h denotes thickness of the investigated epilayer. Column "MDs?" indicates whether MDs were or were not observed. In some papers, the actual critical thickness was estimated. Such cases correspond to h_c in the "MDs?" column.

AlGaN/GaN systems

The lattice constants of AlN are smaller than those of GaN and therefore $Al_xGa_{1-x}N$ grown on GaN is under tensile strain. As a consequence, AlGaN layers often crack (see e.g. Vennéguès et al., 2005). Despite that, MDs are often observed. Vennéguès et al. (2005) carried out a detailed study of the onset of MDs in AlGaN/GaN layers in the whole composition range. Bethoux and Vennéguès (2005) concluded that in their samples all cracks were accompanied by MDs that could be divided into two groups: (i) long straight MDs and (ii) bowed MDs connected at both ends to the cracks. Floro et al. (2004) used the MOSS technique that is capable of immediate registration of the strain state in the epilayer. As a consequence, an *in situ* measurement of the CT was possible.

Experimental data from the literature are summarised in Table 3.2.

3.2 Critical thickness models

3.2.1 Historical overview

Following earlier work of Frank and van der Merwe (1949a,b), van der Merwe (1963a) published a theory for the calculation of stresses at an interface between two adjacent crystals that have different lattice and/or elastic constants. Subsequently, van der Merwe (1963b) applied these results to the problem of a thin crystalline epilayer on an infinite substrate. This model assumes isotropic materials and the calculated properties correspond to an epilayer *fully relaxed* by an array of evenly-spaced edge dislocations (see Fig. 3.1).

According to a later discussion by Willis et al. (1990), some assumptions in the original van der Merwe's calculation are not obvious although they provide correct results. This theory is mathematically rigorous but rather complex, which may explain why it has not been widely used.

Matthews and Blakeslee (1974) introduced a *force balance model* (MB model). They investigated alternating layers of GaAs and $GaAs_{0.5}P_{0.5}$ and revealed that some threading dislocations bent back and forth in the layers' interfaces, which gave rise to misfit dislocations. Based on these observations they identified that two forces (misfit force and tension of the dislocation line) were acting on the dislocations and the critical thickness was determined by their equilibrium. The CT values obtained by this calculation were compared with their observations and they were found to be in good agreement.

x	h [nm]	MDs?	ref.	notes
0.14	72.1	no	[1]	experiments are sensitive only to the presence of MDs, not
0.14	106.1	yes	[1]	to the cracks; therefore, these results show the onset in
0.24	44.2	no	[1]	MDs; multilayered samples
0.24	57.7	yes	[1]	
0.28	25.9	no	[1]	
0.28	34.8	yes	[1]	
0.39	17.3	no	[1]	
0.39	24.0	yes	[1]	
0.46	10.6	no	[1]	
0.46	15.4	yes	[1]	
0.59	5.4	no	[1]	
0.59	6.9	yes	[1]	
0.71	4.4	no	[1]	
0.71	7.0	yes	[1]	
0.88	3.0	no	[1]	
0.88	3.9	yes	[1]	
1.00	2.6	no	[1]	
1.00	4.8	yes	[1]	
1.00	3.5	no	[2]	multilayered sample, surface undulations are observed be-
1.00	10	yes	[2]	fore MDs, 60° dislocations (a-type, along $[11\bar{2}0]$)
1.00	40	yes	[2]	single layer, cracks present
0.70	25	no	[2]	single layer, cracks do not reach heterointerface
0.43	23	no	[2]	single layer, cracks do not reach heterointerface
0.20	200	yes	[3]	cracked AlGaN layers on GaN
0.20	350	yes	[3]	all cracks are accompanied by presence of MDs
0.20	500	yes	[3]	two types of MDs: long straight MDs and bowed MDs
0.20	6000	yes	[3]	bonded to the crack
1.00	56	yes	[4]	TEM evidence, otherwise AFM based paper
0.17	140	h_c	[5]	laser deflectometry by MOSS. Direct measurement of CT;
0.29	92.0	h_c	[5]	however, authors propose that crack are generated first.
0.37	30.0	h_c	[5]	

[1] Lee et al. (2004)
[2] Vennéguès et al. (2005)
[3] Bethoux and Vennéguès (2005)
[4] Gherasimova et al. (2004)
[5] Floro et al. (2004)

Table 3.2: Experimental observations of MDs in $Al_xGa_{1-x}N/GaN$ systems. h denotes thickness of the investigated epilayer. Column "MDs?" indicates whether MDs were or were not observed. In some papers, the actual critical thickness was estimated. Such cases correspond to h_c in the "MDs?" column.

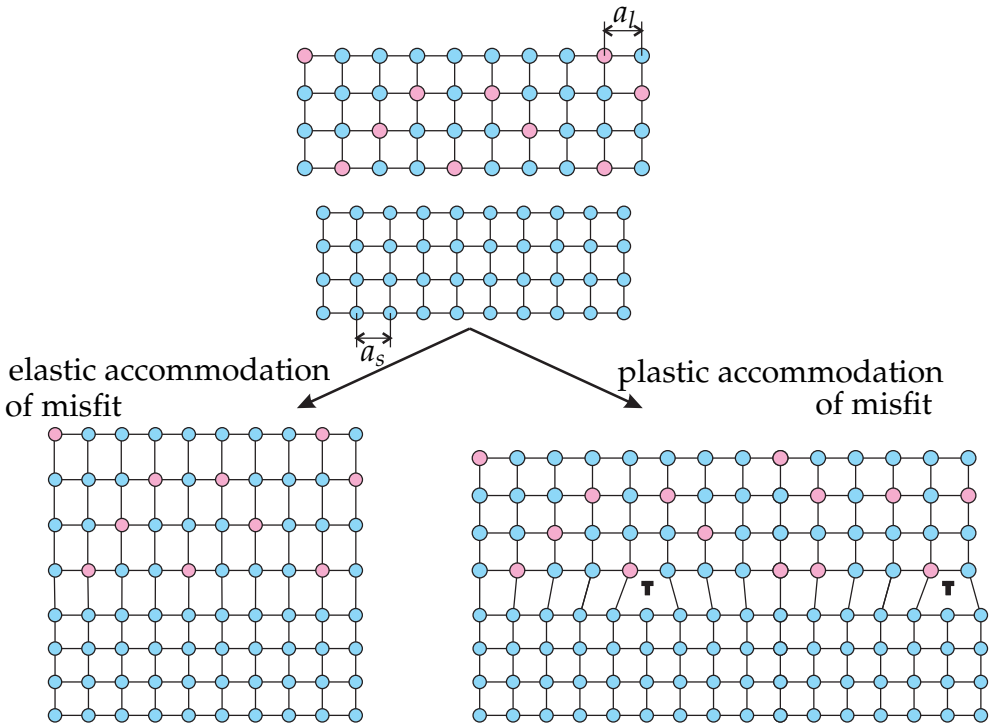


Figure 3.1: The lattice constants of the epilayer (bigger lattice constant) and the substrate do not match. When the epitaxial layer is grown, the misfit strain may be accommodated either elastically and a homo-epitaxial layer is obtained or plastically by introducing misfit dislocations in the layer–substrate interface.

Despite this success, some objections were raised. First, a probably more transparent approach would be using energies (the forces acting on a dislocation are defined as energy gradients anyway). Second, it is not clear what role is played by the dislocation bending and how big an obstacle this constitutes. Last, the effect of interfaces and a free surface of a sample is not taken into account. More details about this model are given in section 3.2.2.

All these objections were a driving force for further investigations. People and Bean (1985) published a paper on the CT calculation based purely on an energy balance of the dislocation self-energy and the elastic energy (discussed in a greater detail by Holec (2006b)). Their calculation was performed for a strained $Si_{1-x}Ge_x$ layer on a silicon substrate. They compared the experimental data published by Bean et al. (1984) with values predicted by their model and found them to be in very good agreement. People and Bean (1985) also compared the same data with the force balance model and with van der Merwe's calculations where they found discrepancies for the calculated CT up to one order of magnitude[1]. However, some drawbacks and uncertainties of this model were pointed out by Hu (1991).

[1]This is due to the dislocation energy being restricted only to the area limited by so called the dislocation width (Holec, 2006b).

The story then continued with a semi-empirical model by Dodson and Tsao (1987) and studies of the excess stress for glide of a dislocation in a strained epitaxial layer by Freund and Hull (1992) and Fischer and Richter (1994). They resulted in a well-known model by Fischer et al. (1994) which has been recently revised (see Fischer, 2006). It is one of the "force models" and it was again discussed in detail by Holec (2006b).

Almost at the same time, calculations done by Willis et al. (1990), Jain et al. (1997) and Freund and Suresh (2003) based on the minimisation of the overall energy appeared in the literature (energy balanced (EB) model). This approach is described in section 3.2.3. When formulated coherently, the MB and EB models give the same predictions which are in agreement with basic physical principles. One of the advantages of the EB model is that additional considerations such as hexagonal symmetry or more complex geometry can be incorporated in a straightforward manner. Therefore, mainly the EB model is used in this work.

The development of a good critical thickness theory was obtained in the late 1990s. Reviews such as those published since then by Jain et al. (1997) or Freund and Suresh (2003) show good agreement between the EB model and the experimental data for conventional SiGe/Si and GaAs-based systems. However, some questions have arisen recently while using these CT models for GaN-based semiconductors.

The real situation is very complex, which leads to a very wide spread of experimental data (as mentioned in previous section 3.1.2). Different authors used different CT models in order to fit their experimental findings without any discussion of their actual validity. Basic predictions and several extensions of CT models (especially the EB model) will be discussed in the rest of this chapter together with a critical comparison of theory with available experimental data.

3.2.2 Model of Matthews and Blakeslee

This model, originally proposed by Matthews and Blakeslee (1974), is often called the *force balance model*. Balance of two forces[2], the line tension $\boldsymbol{F}_l$ and the misfit force $\boldsymbol{F}_\varepsilon$, is sought.

The line tension $\boldsymbol{F}_l$ which acts against the dislocation line prolongation, is given by the dislocation energy per unit length

$$F_l = -\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l}\,. \tag{3.1}$$

[2]Sometimes, also the Peierls force (see e.g. Jahnen et al., 1998) or image forces are taken into consideration. However, this work limits itself to the original concept of only two forces.

Dislocation energy was discussed in detail in chapter 2. The misfit force is estimated using the Peach-Koehler formula for a force acting on a dislocation with Burgers vector $\boldsymbol{b}$ lying along a unit vector $\boldsymbol{l}$ in an external uniform stress field $\hat{\boldsymbol{\sigma}}$ (see for example Hirth and Lothe, 1982)

$$\frac{\mathrm{d}\boldsymbol{F}}{\mathrm{d}l} = (\boldsymbol{b}\cdot\hat{\boldsymbol{\sigma}})\times\boldsymbol{l}\,. \tag{3.2}$$

In coordinates, this equation reads

$$\frac{\mathrm{d}F_k}{\mathrm{d}l} = -\sum_{i,j}\varepsilon_{kij}l_i\sum_{\ell}\sigma_{j\ell}b_\ell \tag{3.3}$$

where ε_{kij} is the permutation operator. The stress field caused by mismatch of lattice parameters between an epilayer and a substrate is generally a bi-axial stress with the only non-zero components

$$\sigma_{xx} = \sigma_{yy} = \sigma_m \tag{3.4}$$

where σ_m is the misfit stress. According to Fig. 3.2a, the direction of the threading dislocation is $\boldsymbol{l} = (-\sin\phi, 0, \cos\phi)$. The misfit stress tensor is

$$\hat{\boldsymbol{\sigma}}_m = \begin{pmatrix} \sigma_m & 0 & 0 \\ 0 & \sigma_m & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.5}$$

and the Burgers vector has components $\boldsymbol{b} = (-b\sin\theta\sin\phi, b\cos\theta, b\sin\theta\cos\phi)$. As a consequence, Eq. 3.3 gives for the y-component of the misfit force acting on the dislocation

$$\frac{\mathrm{d}F_y}{\mathrm{d}l} = b\sigma_m\sin\theta\sin\phi\cos\phi\,. \tag{3.6}$$

If h denotes the thickness of an epilayer then the length of the threading dislocation is $h/\cos\phi$. Therefore, the total misfit force acting in the y-direction, i.e. "competing" with the line tension of the misfit dislocation, is

$$F_\varepsilon = \frac{\mathrm{d}F_y}{\mathrm{d}l}\frac{h}{\cos\phi} = b\sigma_m h\sin\theta\sin\phi\,. \tag{3.7}$$

The total force acting on the dislocation in the y-direction is $F_{\text{total}} = F_l + F_\varepsilon$. For very small epilayer thicknesses h, the total force F_{total} is negative and therefore it acts against extension (and possibly also against creation) of the misfit dislocation. On the contrary,

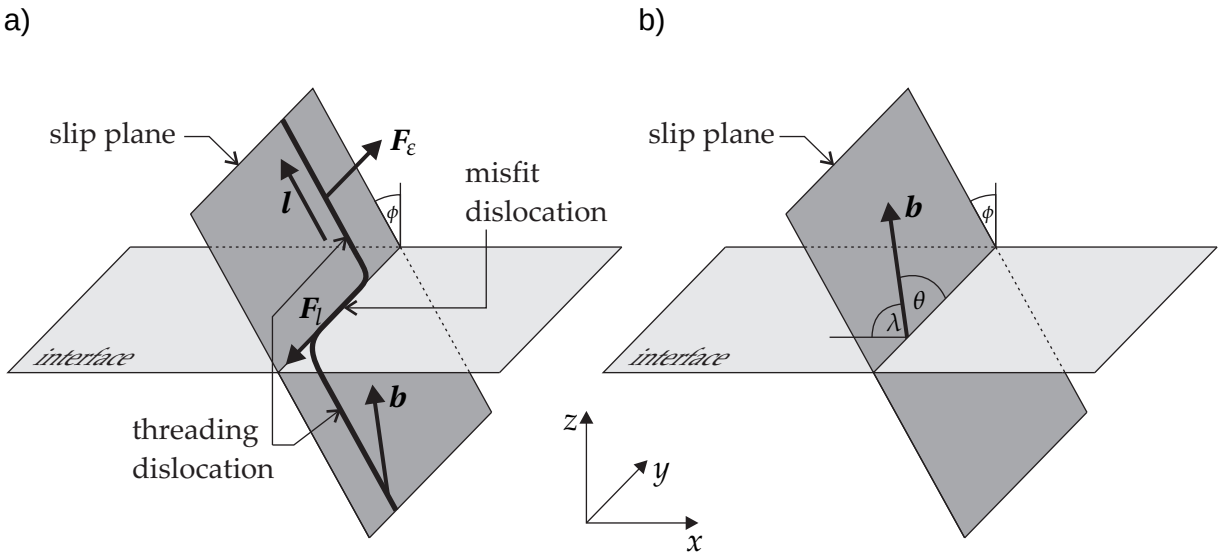


Figure 3.2: A mechanism of creation of the misfit dislocation proposed by Matthews and Blakeslee (1974): a) bending of a pre-existing threading dislocation. The two "competing" forces taken into account are the line tension $\boldsymbol{F}_l$ and the misfit force $\boldsymbol{F}_\varepsilon$. b) Different angles describing the geometry.

F_{total} is positive for thick epilayers and thus extends the misfit dislocation. The critical thickness is therefore determined by $F_{\text{total}}(h_c) = 0$. Eqs. 3.1 and 3.7 yield

$$h_c = \frac{1}{b\sigma_m \sin\theta \sin\phi} \frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l}(h_c) \,. \tag{3.8}$$

Taking the isotropic dislocation energy Eq. 2.6 one obtains the most often used equation for the CT within the MB model

$$h_c = \frac{b(1-\nu\cos^2\theta)}{8\pi(1+\nu)|\varepsilon_m|\sin\theta\sin\phi} \ln\left(\frac{h_c}{r_0}\right) \tag{3.9}$$

where the isotropic relation between the bi-axial misfit strain and the misfit stress was used

$$\sigma_m = 2\varepsilon_m \mu \frac{1+\nu}{1-\nu} \,. \tag{3.10}$$

3.2.3 Energy balance model

Balancing total energy in the system is a concept used already by van der Merwe (1963a) and often since then. It has been recently reviewed by Freund and Suresh (2003). For a very thin epilayer, the elastic strain energy $\mathrm{d}\mathcal{E}_m/\mathrm{d}l$ relieved by the introduction of a MD is smaller than the dislocation energy $\mathrm{d}\mathcal{E}_d/\mathrm{d}l$. Therefore, it is energetically unfavourable for the system to introduce any MDs. On the other hand, the energy $\mathrm{d}\mathcal{E}_m/\mathrm{d}l$ exceeds the dislocation energy $\mathrm{d}\mathcal{E}_d/\mathrm{d}l$ for thicker epilayers and thus the relaxation of misfit strain by introducing MDs is preferred. The particular epilayer thickness h_c which separates these two states is the *critical thickness*.

Energy relieved by a misfit dislocation

The geometry from Fig. 2.2 is assumed. The top surface of a specimen is stress free and therefore $\sigma^m_{zz} = 0$. The only non-zero components of the uniform bi-axial misfit stress are $\sigma^m_{xx} = \sigma^m_{yy} = \sigma_m$ within the thin layer. All directions within the xy-plane (the hexagonal c-plane) are equivalent and thus the misfit strain components are $\varepsilon^m_{xx} = \varepsilon^m_{yy} = \varepsilon_m$. From Hooke's law it follows

$$0 = \sigma^m_{zz} = c_{13}\varepsilon^m_{xx} + c_{13}\varepsilon^m_{yy} + c_{33}\varepsilon^m_{zz} \quad \rightsquigarrow \quad \varepsilon^m_{zz} = -2\frac{c_{13}}{c_{33}}\varepsilon_m \,. \tag{3.11}$$

Using again Hooke's law, the misfit stress takes the form

$$\sigma_m = \sigma^m_{xx} = \frac{(c_{11}+c_{12})c_{33} - 2c_{13}^2}{c_{33}}\varepsilon_m \,. \tag{3.12}$$

Following Freund and Suresh (2003), the work $\partial\mathcal{W}/\partial l$ done by the mismatch stress while bringing the unit length of the dislocation from the free surface to its position in the interface is

$$\frac{\mathrm{d}\mathcal{W}}{\mathrm{d}l} = -\frac{\mathrm{d}\mathcal{E}_m}{\mathrm{d}l} = \int_{\frac{h}{\cos\phi}}^{0} \sum_{i,j} b_i \sigma_{ij} n_j \,\mathrm{d}\ell \tag{3.13}$$

where ℓ is the coordinate along the slip plane and perpendicular to the dislocation line. The integrand is constant along the integration path. As a consequence, Eq. 3.13 results (in agreement with Eq. 3.6) in

$$\frac{\mathrm{d}\mathcal{E}_m}{\mathrm{d}l} = b\sigma_m h \sin\phi \sin\theta \tag{3.14}$$

where the following relations were used:

$$\boldsymbol{b} = \begin{pmatrix} -b\sin\phi\sin\theta \\ b\cos\theta \\ b\cos\phi\sin\theta \end{pmatrix}, \quad \boldsymbol{\sigma}^m = \begin{pmatrix} \sigma_m & 0 & 0 \\ 0 & \sigma_m & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{n} = \begin{pmatrix} -\cos\phi \\ 0 \\ \sin\phi \end{pmatrix}. \tag{3.15}$$

Critical thickness criterion

The critical thickness criterion is

$$\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l}(h_c) = \frac{\mathrm{d}\mathcal{E}_m}{\mathrm{d}l}(h_c) \,. \tag{3.16}$$

Furthermore, Freund and Suresh (2003) have shown that this approach gives the same results as the force balance model. This is because the concept presented here is the so called *equilibrium critical thickness* which does not take into consideration any dislocation sources.

3.3 Discussion or results

Some basic results on the CT are summarised in this section. They are an extension of a similar analysis in previous work (Holec, 2006b, only for InGaN/GaN systems).

Graphs in Fig. 3.3 represent CT predictions for $Al_xGa_{1-x}N/GaN$ and $In_xGa_{1-x}N/GaN$ systems as calculated using models by Matthews and Blakeslee (1974), People and Bean (1985) and Fischer et al. (1994). Here, the MB model corresponds *exactly* to the EB model. The calculations were done for the $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ slip system (see Fig. 1.11) and thus using the inner cut-off radius $r_0 = b$ (see Table 2.2).

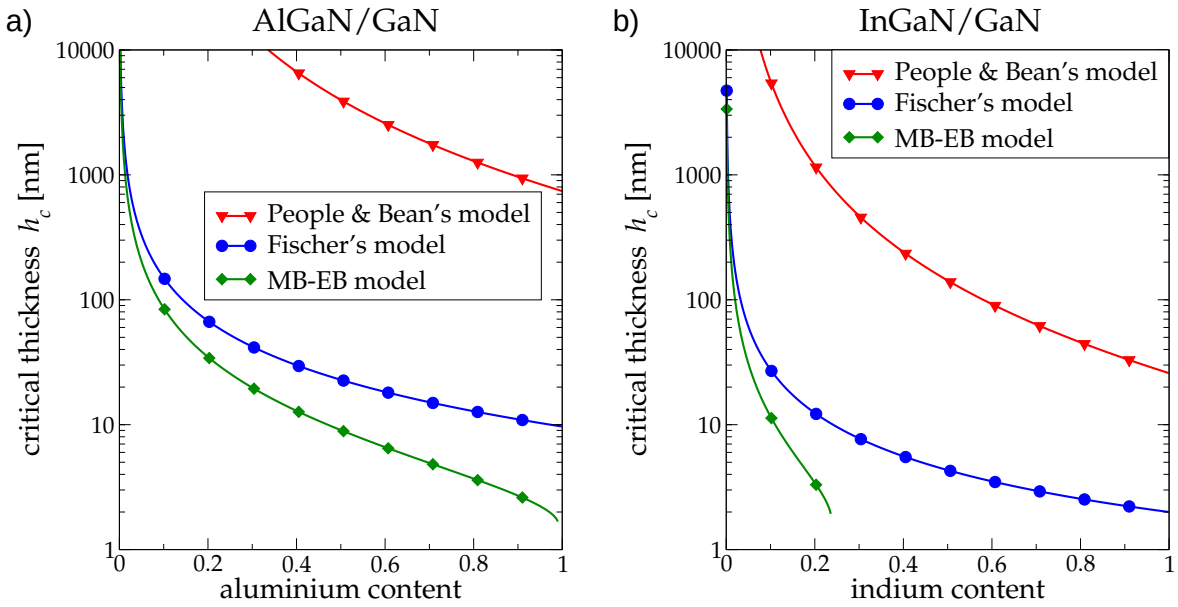


Figure 3.3: Comparison of CT predictions by various models for (a) AlGaN/GaN and (b) InGaN/GaN systems. These curves correspond to the $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ slip system.

One can observe that the model of People and Bean gives extremely high estimates of the CT and thus it can be immediately ruled out as inappropriate for III-nitrides. The Fisher's model gives the CT in the whole range of compositions for both systems. Also looking at the experimental data (Table 3.1, Table 3.2 and Fig. 3.6) this model gives reasonable estimates for the CT. These results were obtained within the isotropic approximation. As discussed by Holec (2006b), there are some aspects of this model that are not well understood and which prevent its correction for hexagonal symmetry and other geometrical concerns. The last curve in these graphs corresponds to the simplest isotropic formulation of the EB model. The EB model in this form strongly underestimates the possible values of the CT. In the case of the InGaN/GaN system it does not predict any CT values for $x \gtrapprox 0.25$ (this means that in such cases there is no solution to Eq. 3.16). Therefore, there is a need for corrections and extension of the EB model in order to use it for III-nitrides.

Holec et al. (2007) reported on the contribution of hexagonal symmetry for InGaN/GaN systems. The CT curves corresponding to the EB model for the isotropic and hexagonal approximations as presented in sections 2.1 and 2.2.1 are shown in Fig. 3.4a. The differences are clearer in Fig. 3.4b where the hexagonal-to-isotropic ratio of the CT is plotted. In this particular case, the hexagonal symmetry lowers the predicted CT values by 10–20 % for $x \leq 0.3$. This small correction may become a significant difference in the region around $x \approx 0.2$ where the isotropic value is $h_c = 4.3\,\text{nm}$ whereas the hexagonal value is $h_c = 3.6\,\text{nm}$. These values are very close to the thickness used for quantum wells in InGaN based LEDs.

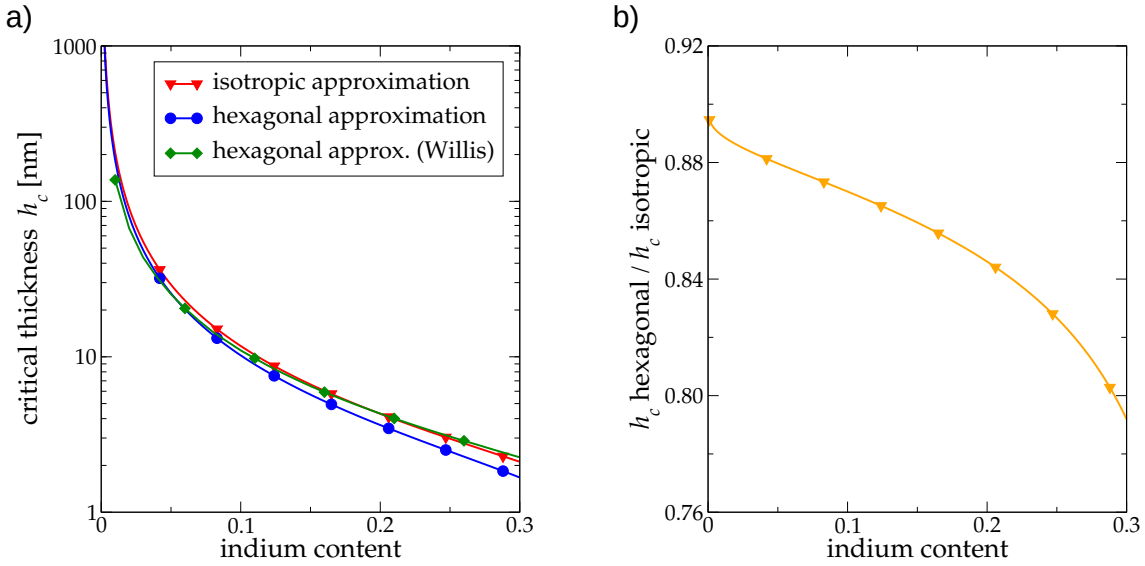


Figure 3.4: (a) Various approximations to the dislocation energy, (b) hexagonal-to-isotropic ratio of critical thickness. The CT values correspond to InGaN/GaN system with the $\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$ slip system and the inner cut-off radius $r_0 = 2b$.

The green curve in Fig. 3.4a corresponds to what is supposed to be the most accurate estimation of dislocation energy (see Willis et al. (1990) and its correction for hexagonal symmetry by Holec (2006b)). It takes into account the hexagonal symmetry, the effect of a free surface of an epilayer, and different elastic constants in both an epilayer and a substrate. This result suggests that these three effects approximately cancel out and that the final critical thickness curve is very close to the isotropic case, in particular in the region around $x \approx 0.2$.

Srinivasan et al. (2003) concluded that only the prismatic slip systems $\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$ and $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ are operable for glide in wurtzite III-nitrides (see Fig. 1.11). However, there are also reports of pure a-type MDs (see e.g. Vennéguès et al., 2005) which are theoretically the most efficient MD types for relieving the misfit strain and thus the basal plane slip system $\frac{1}{3}\langle 11\bar{2}0\rangle\{0001\}$ is included for theoretical considerations here. Fig. 3.5 shows CT of AlGaN/GaN and InGaN/GaN systems for these three slip

systems. In both cases, the basal plane slip system $\frac{1}{3}\langle 11\bar{2}0\rangle\{0001\}$ yields the lowest CT values. However, neither the $\frac{1}{3}\langle 11\bar{2}0\rangle\{0001\}$ not the prismatic slip systems predict any CT values for high x in both material systems. Differences between the two prismatic slip systems $\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$ and $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ are not very big in the case of the AlGaN/GaN system. The highest CT values are obtained for the $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ slip system in both material cases. For this slip system the MDs lie along the $\langle 11\bar{2}0\rangle$ directions which corresponds with the observations of Vennéguès et al. (2005).

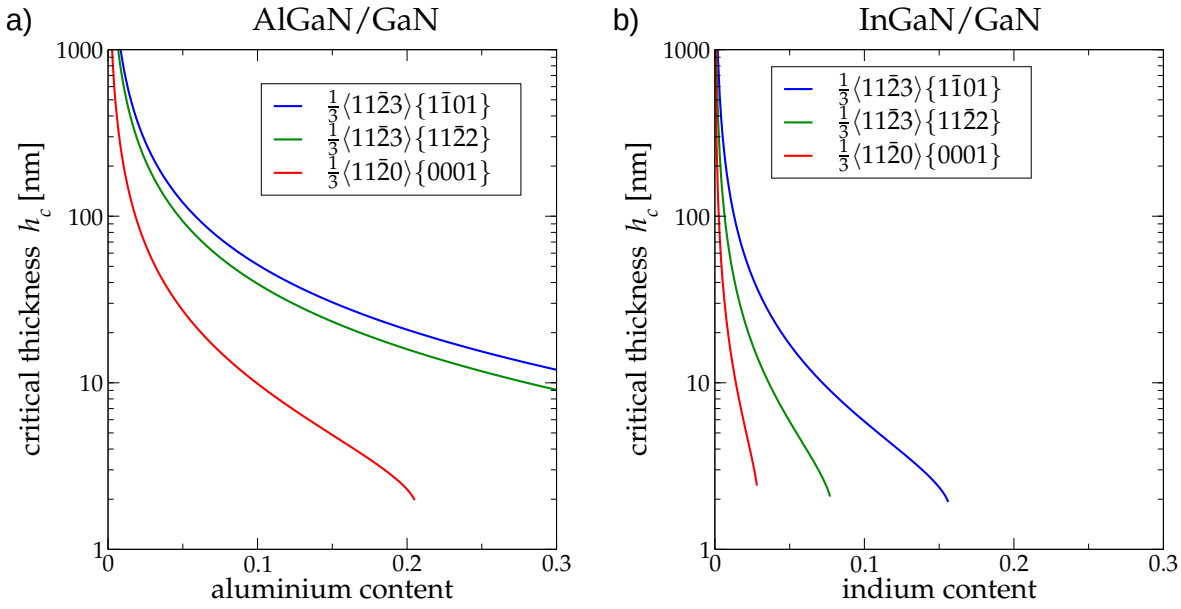


Figure 3.5: Variations in critical thickness of (a) AlGaN/GaN and (b) InGaN/GaN with slip systems. The cut-off radii used for prismatic slip systems were $r_0 = b$, for the basal slip system was used $r_0 = 2b$ (see Table 2.2).

Finally, Fig. 3.6 shows the results of calculations together with available experimental data from section 3.1.2. These calculations were done for the basal $\frac{1}{3}\langle 11\bar{2}0\rangle\{0001\}$ and prismatic $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ slip systems with the core radius $r_0 = 2b$ and $r_0 = b$, respectively. The dislocation core energy is also included as it is discussed later in section 3.4.2. It is evident that the theoretical curves from the EB model underestimate slightly the CT with respect to the experimental values. In the case of the AlGaN/GaN system (Fig. 3.6a) it is usually cracks that are observed first. This is expected to push the experimental onset of MDs to higher epilayer thicknesses when compared with theoretical calculations where MDs are the only strain relieving mechanism. The same argument can be used for the InGaN/GaN system (Fig. 3.6b) where V-defects are often observed experimentally. However, the data points corresponding to the AlGaN/GaN system seem to correlate quite well with the prismatic $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ slip system. On

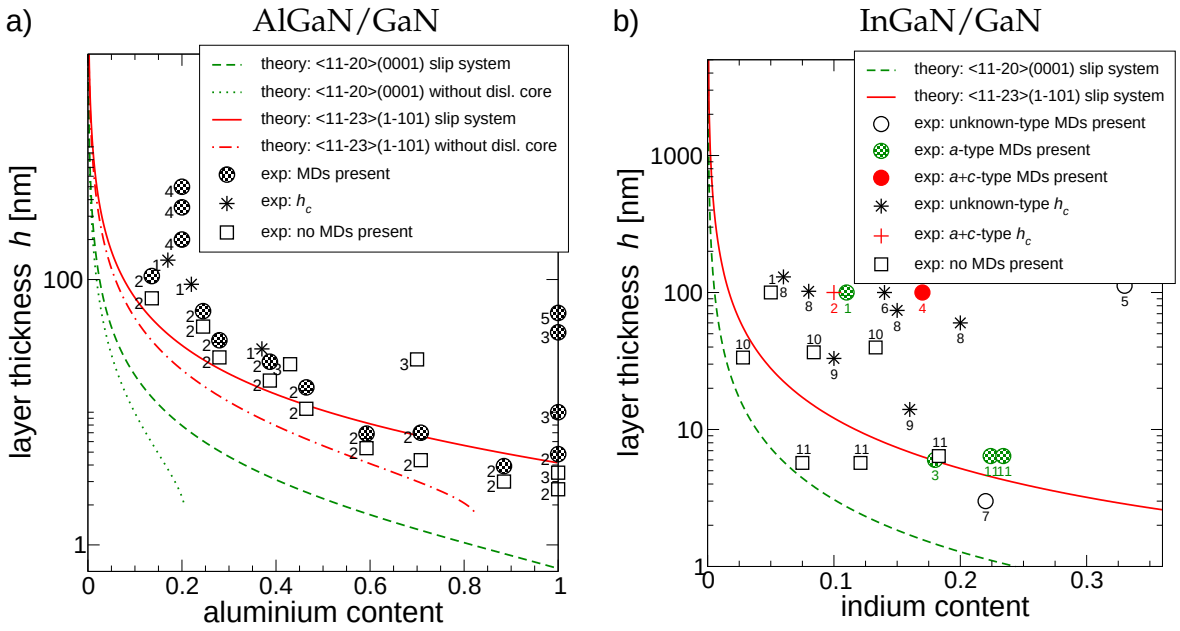


Figure 3.6: Comparison of theoretical predictions of CT with experimental observations for (a) AlGaN/GaN and (b) InGaN/GaN systems. Calculations were done for the basal $\frac{1}{3}\langle 11\bar{2}0\rangle\{0001\}$ and prismatic $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ slip systems with the core radius $r_0 = 2b$ and $r_0 = b$, respectively, and with the dislocation core energy included. Experimental data and corresponding citations refer to Tables 3.1 and 3.2.

the other hand, much larger scatter of the experimental data points is observed for the InGaN/GaN system.

The present theoretical model predicts the *equilibrium critical thickness* which does not account for the actual MDs generation mechanism. Therefore, theoretical CT values for both the basal $\frac{1}{3}\langle 11\bar{2}0\rangle\{0001\}$ and prismatic $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ slip systems were shown in the above graphs while noting that the experimental onset of a-type dislocations (unlike the $(a+c)$-type MDs) is expected to be significantly higher than the predicted CT values, due to the inoperability of the basal slip system.

3.4 Advanced topics

Holec (2006b) used the EB model to investigate several "advanced" topics. Those included (a) the effect of raised temperature during growth on the critical thickness, (b) the CT for dislocation half-loops, (c) density of misfit dislocations after the CT is reached, and (d) a simple model for the CT of multiple quantum wells. In this section are summarised some additional recently investigated applications based on the EB model for the CT.

3.4.1 Critical thickness of non-polar and semi-polar InGaN/GaN

In all previous calculations only the *c*-plane systems (i.e. where the epilayer–substrate interface was the hexagonal *c*-plane) were assumed. That orientation of GaN (and its related alloys) is called *polar* because the surface plane always terminates with either Ga or N. Qualitatively different properties of so called *non-polar* GaN (and related alloys), i.e. grown on the *a*-plane or the *m*-plane of the wurtzite substrates, have led to an increased interest in these structures. Holec and Humphreys (2007) calculated CT values for a single $In_xGa_{1-x}N$ epilayer grown on the *a*- and *m*-plane GaN substrates and compared them with the values obtained for the *c*-plane GaN.

The critical thickness criterion (Eq. 3.16) for differently orientated substrates differs in both terms, the dislocation energy $\mathrm{d}\mathcal{E}_d/\mathrm{d}l$ and the misfit relief $\mathrm{d}\mathcal{E}_m/\mathrm{d}l$. These terms need to be recalculated for all orientations of the considered systems. The dislocation energy $\mathrm{d}\mathcal{E}_d/\mathrm{d}l$ can be obtained using the procedure from sections 2.2.1 and 2.2.2. The misfit relief $\mathrm{d}\mathcal{E}_m/\mathrm{d}l$ is a straightforward generalisation of the procedure from section 3.2.3.

Slip systems of individual orientations

Srinivasan et al. (2003) concluded that the active slip system for the *c*-plane GaN was $\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$. On the other hand, Jahnen et al. (1998) identified $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ as the active slip system. Therefore, both these slip systems are considered here.

Possible slip systems were derived in a similar manner for the *m*-plane and the *a*-plane material. The glide planes used were planes intersecting the hexagonal cell unit. Four slip systems were found for the *m*-plane material and three slip systems in a case of the *a*-plane material. These are shown in Figs. 3.7 and 3.8, respectively, and are listed together with their geometrical properties in Table 3.3.

Peierls force

Several possible slip systems for each material variant were identified. However, dislocations do not operate with the same ease on all of them. As Srinivasan et al. (2003) pointed out, the Peierls force acts as a frictional force against the dislocation motion. The Peierls force used in their paper was derived within the framework of linear isotropic elasticity. Despite the fact that hexagonal symmetry of wurtzite materials

		b	ϕ	θ
c-plane	$\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$	$\sqrt{a^2+c^2}$	$\arctan\frac{a}{c}$	90°
	$\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$	$\sqrt{a^2+c^2}$	$\arctan\frac{\sqrt{3}a}{2c}$	$\arccos\frac{a}{2b}$
m-plane	$\frac{1}{3}\langle 11\bar{2}0\rangle\{1\bar{1}00\}$	a	30°	90°
	$\langle 1\bar{1}00\rangle\{11\bar{2}0\}$	$\sqrt{3}a$	60°	90°
	$\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$	$\sqrt{a^2+c^2}$	$\arctan\frac{2c}{a\sqrt{3}}$	$180^\circ-\arccos\frac{a}{2b}$
	$\langle 1\bar{1}01\rangle\{1\bar{1}02\}$	$\sqrt{3a^2+c^2}$	$\arctan\frac{c}{a\sqrt{3}}$	90°
a-plane	$\frac{1}{3}\langle 11\bar{2}0\rangle\{1\bar{1}00\}$	a	60°	90°
	$\langle 1\bar{1}00\rangle\{11\bar{2}0\}$	$\sqrt{3}a$	30°	90°
	$\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$	$\sqrt{a^2+c^2}$	$\arctan\frac{c}{a}$	90°

Table 3.3: Geometrical properties of the slip systems considered in this study: b is the length of the Burgers vector, ϕ is the angle between slip plane and the normal to the interface, and θ is the angle between the Burgers vector and the dislocation line.

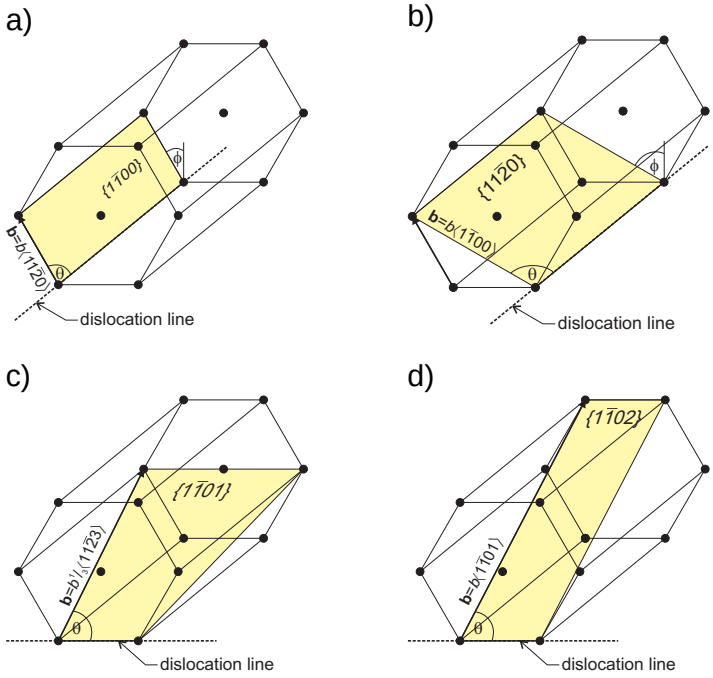


Figure 3.7: Four possible slips systems for the m-plane material: a) $\frac{1}{3}\langle 11\bar{2}0\rangle\{1\bar{1}00\}$, b) $\langle 1\bar{1}00\rangle\{11\bar{2}0\}$, c) $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$, and d) $\langle 1\bar{1}01\rangle\{1\bar{1}02\}$.

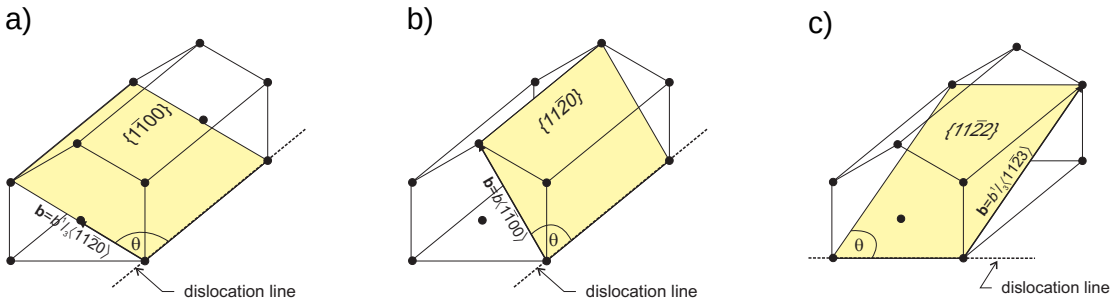


Figure 3.8: Three possible slips systems for the a-plane material: a) $\frac{1}{3}\langle 11\bar{2}0\rangle\{1\bar{1}00\}$, b) $\langle 1\bar{1}00\rangle\{11\bar{2}0\}$, and c) $\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$.

is otherwise used here, their formula is used to provide some guidelines as to how big this effect is for various slip systems. The Peierls force takes the form

$$F_P = 2Gbh\sec\phi\left(\frac{1-\nu\cos^2\theta}{1-\nu}\right)\omega\exp\left(\frac{-2\pi d(1-\nu\cos^2\theta)\omega}{(1-\nu)b}\right) \tag{3.17a}$$

where

$$\omega = \exp\left(\frac{4\pi^2 nkT}{5GV}\right)\,. \tag{3.17b}$$

Apparently, the only parameters that change from one slip system to another are the length b of the Burgers vector, the angles θ and ϕ, and the inter-planar distance d for planes parallel with the slip plane. By substituting $n = 4$ (number of atoms in the unit cell), $T = 1000\,\mathrm{K}$ (approximate growth temperature), $a = 3.189\,\text{Å}$, $c = 5.185\,\text{Å}$,

c-plane

	F_P/h [N/m]
$\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$	55.12
$\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$	11.42

m-plane

	F_P/h [N/m]
$\frac{1}{3}\langle 11\bar{2}0\rangle\{1\bar{1}00\}$	0.09
$\langle 1\bar{1}00\rangle\{11\bar{2}0\}$	49.18
$\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$	21.44
$\langle 1\bar{1}01\rangle\{1\bar{1}02\}$	63.24

a-plane

	F_P/h [N/m]
$\frac{1}{3}\langle 11\bar{2}0\rangle\{1\bar{1}00\}$	0.15
$\langle 1\bar{1}00\rangle\{11\bar{2}0\}$	27.31
$\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$	89.61

Table 3.4: Peierls force F_P/h per unit thickness of the layer calculated from Eq. 3.17a. The tables show F_P/h for the *c*-plane, *m*-plane, and *a*-plane materials' slip systems.

$\gamma = 120^\circ$ (unit cell parameters for volume V) and $G = 200\,\text{GPa}$ (the shear modulus) into Eq. 3.17b, one obtains $\omega \approx 1$.

With $\omega \approx 1$, Poisson's ratio $\nu \approx 0.3$, and the inter-planar distance d substituted Eq. 3.17a, the values of F_P/h are obtained. These may serve as a comparison of the effect of the Peierls force for different systems. The results are summarised in Table 3.4. As a consequence, only the the most favourable slip-systems with the smallest Peierls (friction) force are taken as the final result: $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$ for the *c*-plane material and $\frac{1}{3}\langle 11\bar{2}0\rangle\{1\bar{1}00\}$ for both the *m*- and the *a*-plane materials.

Results

Fig. 3.9 shows the critical thickness curves for all slip systems considered here. The most favourable slip systems for each type of material (the *a*-, the *m*-, and the *c*-plane) are shown separately also in Fig. 3.10a. In the whole range $0 < x < 0.3$ of indium compositions, the *m*-plane material has the lowest CT whereas the *a*-plane material has the highest CT values. This is even more apparent from the graph in Fig. 3.10b where ratios of the *a*- and the *m*-plane material CT to the *c*-plane material CT are shown. The CT of the *m*-plane material is about 0.7–0.9 of the *c*-plane CT for the indium compositions $0 < x < 0.3$; the *m*-plane CT approaches the *c*-plane CT values for high indium concentrations.

On the other hand, values of the *a*-plane CT are about 1.3–2.1 times higher than those of the *c*-plane system for indium compositions ranging $0 < x < 0.3$. The higher the indium content is the higher this ratio becomes.

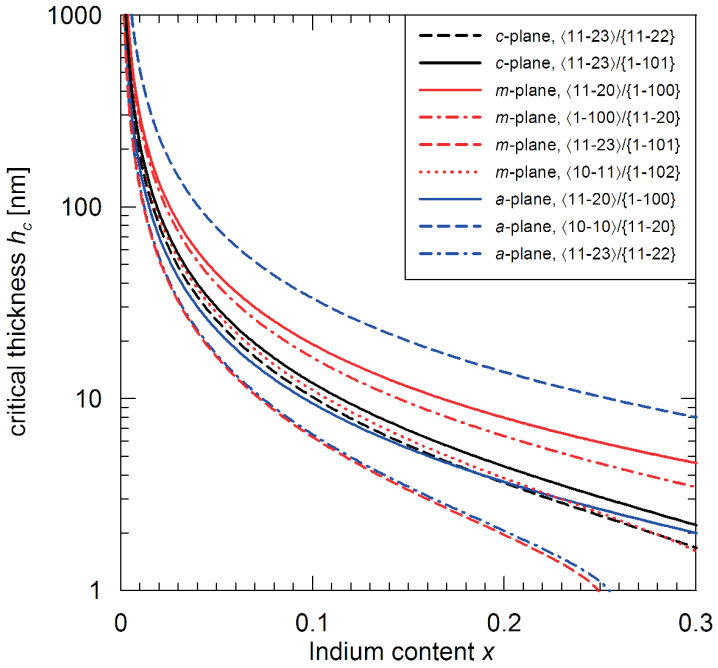


Figure 3.9: Critical thickness versus indium composition curves for the nine slip systems investigated in this study.

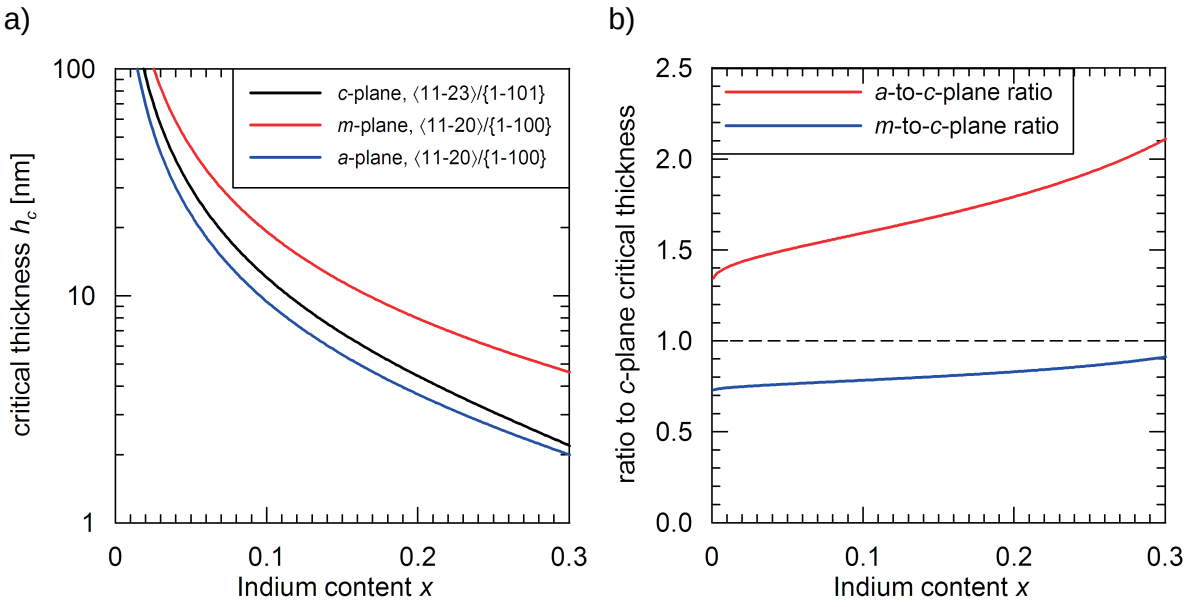


Figure 3.10: Critical thickness calculations for different substrate orientations: a) the h_c curves for the *c*-, the *m*-, and *a*-plane materials, b) ratio of the *a*-to-*c*-plane material and the *m*-to-*c*-plane material CT.

Conclusions

This study showed that the CT for the *m*-plane $In_xGa_{1-x}N$/GaN system is very close (but lower) than the CT for the *c*-plane material. On the contrary, in the case of *a*-plane material the CT is approximately 1.5–2 times higher than the CT for the *c*-plane system.

3.4.2 Effect of the dislocation core energy on the equilibrium critical thickness

Up to now, the dislocation core energy has been neglected. Section 2.4 summarised results of recent atomistic calculations that provided some parameters of dislocation cores in III-nitrides. The question now arises: What is the effect of the dislocation core energy on the CT values?

If it is assumed that $\mathrm{d}\mathcal{E}_{\mathrm{core}}/\mathrm{d}l$ does not change too much with the indium content, it can be added into Eq. 3.16 as a constant contribution to $\mathrm{d}\mathcal{E}_d/\mathrm{d}l$ and subsequently the critical thickness curves can be recalculated. The results are shown in Fig. 3.11 (solid lines). The EB model now does not "break" at all in the case of hexagonal symmetry as it did without the dislocation core energy being taken into account (see section 3.3). Moreover, compared with the results without the core energy term, the CT values are significantly higher. The model remains still quite simple but now retains much more physics. The same physics can now be applied to the AlGaN/GaN system. Results are shown in Fig. 3.12.

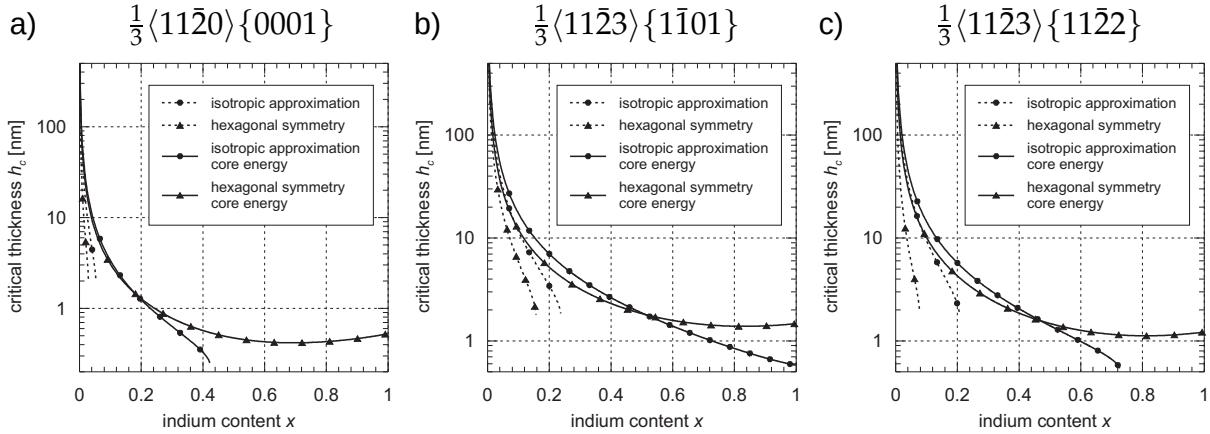


Figure 3.11: Critical thickness curves for a single layer of InGaN on GaN for three different slip systems: a) $\frac{1}{3}\langle 11\bar{2}0\rangle\{0001\}$, b) $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$, and c) $\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$. Dashed and solid lines correspond to the result without and with the dislocation core term, respectively.

One should be aware that all the atomistic calculations dealt with dislocation lines along the $[0001]$ direction whereas MDs in this case lie in the (0001) plane, i.e. along either the $\langle 11\bar{2}0\rangle$ or the $\langle 1\bar{1}00\rangle$ directions. So far there has been no report in the literature of similar calculations for dislocations along the latter two directions. As a consequence, the atomistic results *do not* correspond to the dislocations used for the CT calculations. Nevertheless, these results are being used while assuming that the physically correct parameters of MDs would not differ too much from the only currently available atomistic results.

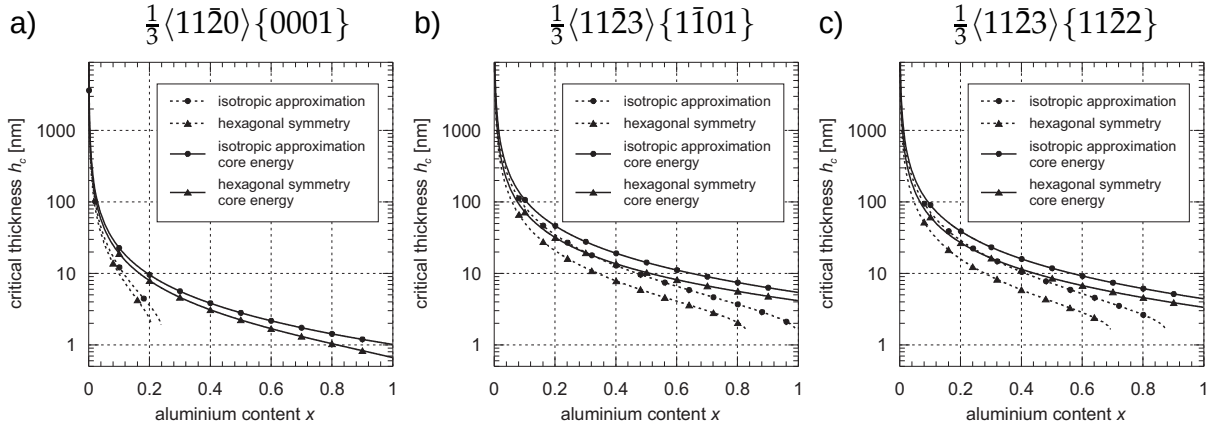


Figure 3.12: Critical thickness curves for a single layer of AlGaN on GaN for three different slip systems: a) $\frac{1}{3}\langle 11\bar{2}0\rangle\{0001\}$, b) $\frac{1}{3}\langle 11\bar{2}3\rangle\{1\bar{1}01\}$, and c) $\frac{1}{3}\langle 11\bar{2}3\rangle\{11\bar{2}2\}$. Dashed and solid lines correspond to the result without and with the dislocation core term, respectively.

As reported by Belabbas et al. (2007), Lymperakis (2005) and Kioseoglou et al., the energetically lowest configuration is the 5/7-atom ring for all Burgers vectors. Therefore, these parameters were used for the calculations. Table 2.2 also includes prelogarithmic terms as obtained from Eq. 2.28 for comparison with values that result from the atomistic calculations. Although the dislocation lines of the MDs do not correspond to those from the atomistic calculations, the prelogarithmic terms correspond well, which suggests that the assumptions made here are reasonable.

Kioseoglou et al. provided a very detailed study of the *a*-type [0001] dislocations in III-nitrides. Using their results it is also possible to estimate the effect of neglecting the

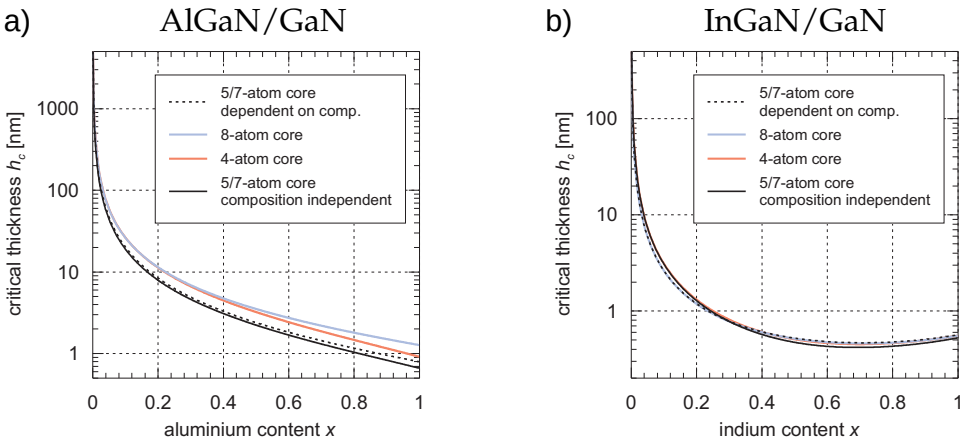


Figure 3.13: Critical thickness curves for a single layer of InGaN or AlGaN on GaN the slip system $\frac{1}{3}\langle 11\bar{2}0\rangle\{0001\}$. Black dashed and solid lines correspond to the energetically lowest dislocation core configuration with the 5/7-atoms ring with and without taking into account the composition dependence of the core energy, respectively. Pink and blue lines correspond to the 4- and 8-atoms dislocation core configurations, respectively.

dependence of $\mathcal{E}_{\text{core}}$ on the composition x. According to Kioseoglou et al., the 5/7-atom ring core is the lowest energy configuration for AlN, GaN and InN. Applying Vegard's law yields

$$\frac{\mathrm{d}\mathcal{E}_{\text{core}}}{\mathrm{d}l}(\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N}) = x\frac{\mathrm{d}\mathcal{E}_{\text{core}}}{\mathrm{d}l}(\mathrm{AlN}) + (1-x)\frac{\mathrm{d}\mathcal{E}_{\text{core}}}{\mathrm{d}l}(\mathrm{GaN}) \tag{3.18}$$

and a similar expression for InGaN. Also other dislocation core configurations (4- and 8-atom rings) as discussed by Kioseoglou et al. were included. The results are shown in Fig. 3.13. From these figures it is clear that the assumption of a composition independent core energy does not effect the results significantly (compare black dashed and solid lines in Fig. 3.13). Different core configurations give the same CT in the case of the InGaN/GaN system and slightly higher CT in the case of AlGaN/GaN than the CT for the 5/7-atom core. One should also note that in the case of InGaN/GaN the CT ranges down to 0.4 nm for $x \approx 0.6$. These values are smaller than the dislocation core radius ($r_{\text{core}} \approx 0.7$ nm) and therefore they are hard to interpret. As a consequence, one should not take seriously CT values below ≈ 1 nm. That is, epilayers with high indium or aluminium concentration are so strained that the growth of substrate lattice matched layers becomes impossible.

3.4.3 Staircase structures

Cherns et al. (2007) reported on staircase structures of threading dislocations observed in AlGaN multilayer samples. A series of alternating $\mathrm{Al}_{0.5}\mathrm{Ga}_{0.5}\mathrm{N}$ QWs with GaN barrier layers was grown on a $\mathrm{Al}_{0.45}\mathrm{Ga}_{0.65}\mathrm{N}$ buffer layer. A series of three samples with different QW thicknesses (1.5 nm, 2.5 nm and 3.5 nm) was grown and investigated. The thickness of barrier layers was 10.5 nm in all cases. TEM images with low magnification showed dislocations inclined in the QWs/barriers region, with the inclination of the TDs from the vertical direction increasing with QW thickness. Higher magnification TEM images then revealed that the inclination is in fact caused by a staircase structure such as shown in Fig. 3.14. Cherns (2007) concluded that the observed dislocations are *a*-type.

In this section some initial theoretical approaches to rationalise these observations are discussed.

Misfit strain, misfit stress and misfit force

Only two types of layers are distinguished: an unstrained substrate (under the lowest quantum well (QW)) or barrier (in-between QWs and above the uppermost one QW)

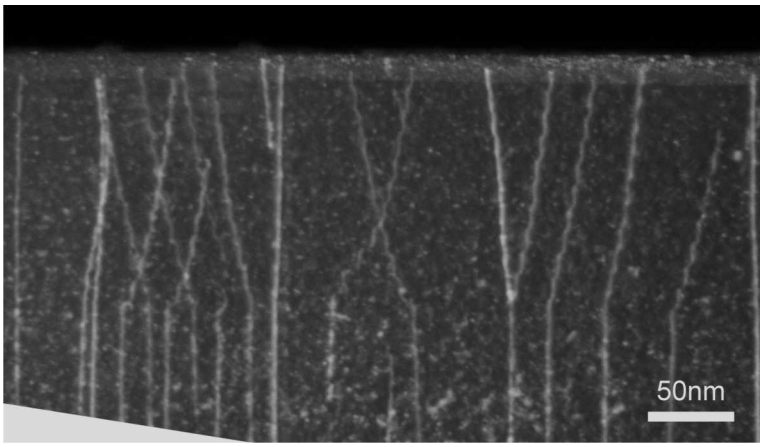


Figure 3.14: A weak-beam dark-field TEM image ($\boldsymbol{g}$–$4\boldsymbol{g}$ condition, $\boldsymbol{g} = 11\bar{2}0$) of a sample with 3.5 nm thick $Al_{0.5}Ga_{0.5}N$ QWs and 10.5 nm thick GaN barriers. (Courtesy of P. Cherns.)

and the QWs themselves. The force acting on a dislocation can be calculated using the Peach-Koehler formula (Eq. 3.2). In this case, $\boldsymbol{b} = (0, b, 0)^T$ (a-type dislocations), $\mathrm{d}\boldsymbol{l} = (0, 0, \mathrm{d}l)^T$ (threading dislocation segment) and $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}^m$. Therefore the misfit force is

$$\frac{\mathrm{d}\boldsymbol{F}^m}{\mathrm{d}l} = (b\sigma^m_{yy}, 0, 0)^T . \tag{3.19}$$

The Peach-Koehler misfit force acting on a TD segment has the only non-zero component in the x-direction.

Dislocation energy and dislocation line tension

In the assumed configuration, dislocations always remain edge-type (a TD is initially edge-type and the misfit force acts in a direction perpendicular to TD segments and the Burgers vector). Energy of an edge-type, infinitely long dislocation per unit length (within isotropic approximation) is

$$\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} = \frac{\mu b^2}{4\pi(1-\nu)} \ln\frac{h}{r_{\mathrm{core}}} + \frac{\mathrm{d}\mathcal{E}_{\mathrm{core}}}{\mathrm{d}l} . \tag{3.20}$$

Eq. 3.20 presents also the line tension of a long dislocation; it is a force that acts *against* prolongation of a dislocation.

As pointed out by Kroupa and Brown (1961), a very short misfit dislocation (MD) segment should be treated as a kink rather than as a dislocation: the interaction between adjacent TDs is important and cannot be neglected. It can be shown (see Hirth and Lothe, 1982 and Kroupa and Brown, 1961) that the energy of a kink of length l (and a

geometrical configuration such that the whole kinked dislocation is edge-type) is

$$\mathcal{E}_k = \frac{\mu b^2 l}{4\pi(1-\nu)}\left(\ln\frac{2l}{r_{\text{core}}} - 3\right) + l\frac{\mathrm{d}\mathcal{E}_{\text{core}}}{\mathrm{d}l}\,. \tag{3.21}$$

The elastic part of the above equation is negative for

$$l < l_{\text{core}} = \frac{\mathrm{e}^3}{2} r_{\text{core}} \approx 10 r_{\text{core}}\,. \tag{3.22}$$

This is a region where the dislocation core effects are significant. In the following, the kink energy is set to $\mathcal{E}_k = l\,\mathrm{d}\mathcal{E}_{\text{core}}/\mathrm{d}l$ for $l < l_{\text{core}}$. The line tension acting against prolongation of a kink is then

$$T_k = \frac{\mathrm{d}\mathcal{E}_k}{\mathrm{d}l} = \frac{\mu b^2}{4\pi(1-\nu)}\left(\ln\frac{2l}{r_{\text{core}}} - 2\right) + \frac{\mathrm{d}\mathcal{E}_{\text{core}}}{\mathrm{d}l}\,. \tag{3.23}$$

The elastic part of the tension is negative for $l < r_{\text{core}}\mathrm{e}^2/2 \approx 3.7 r_{\text{core}}$ and diverges for $l \to 0$. Similarly to the kink energy formula, it is set to $T_k = \mathrm{d}\mathcal{E}_{\text{core}}/\mathrm{d}l$ for $l < l_{\text{core}}$. l_{core} is now smaller than it was in the case of the kink energy formula (by a factor e); however, all these adjustments are only estimates of orders and thus in all the following calculations is kept $l_{\text{core}} \approx 10 r_{\text{core}}$ for both cases.

For a given h, the MD segment is treated as a kink until its length exceeds $l > h\mathrm{e}^3/2 \approx 10h$ at which length the dislocation energy term becomes smaller than the kink energy term. Therefore, switching between these two modes of MDs (a kink or a long dislocation) is driven by the minimum energy principle.

Substrate and a single QW

A misfit dislocation will be created in the epilayer–substrate interface if the total energy in the system is lowered by this process. The strain energy relief by the introduction of a MD is

$$\mathcal{W}(h,l) = F_x^m h l = 2\mu\frac{1+\nu}{1-\nu}\varepsilon_{yy}^m b h l\,. \tag{3.24}$$

The energy introduced into the system by the MD segment is

$$\mathcal{E}(h,l) = \begin{cases} l\frac{\mathrm{d}\mathcal{E}_{\text{core}}}{\mathrm{d}l} & \text{for } l \lessapprox l_{\text{core}}\,, \\ \frac{\mu b^2}{4\pi(1-\nu)} l\left(\ln\frac{2l}{r_{\text{core}}} - 3\right) + l\frac{\mathrm{d}\mathcal{E}_{\text{core}}}{\mathrm{d}l} & \text{for } l_{\text{core}} \lessapprox l \lessapprox 10h\,, \\ \frac{\mu b^2}{4\pi(1-\nu)} l \ln\frac{h}{r_{\text{core}}} + l\frac{\mathrm{d}\mathcal{E}_{\text{core}}}{\mathrm{d}l} & \text{otherwise}\,. \end{cases} \tag{3.25}$$

For a particular epilayer thickness h, a MD segment of length l is energetically favourable if

$$\mathcal{E}(h,l) - \mathcal{W}(h,l) < 0\,. \tag{3.26}$$

The h–l-parameter space can be divided into four regions according to this criterion (see Fig. 3.15). In area "1", the dislocation energy is bigger than the energy relief and thus MD segments of corresponding lengths and QW thicknesses are energetically unfavourable. On the other hand, for QW thicknesses h lower than a certain value, h_k, and reasonably short MD segment length (area "2") the criterion 3.26 holds. As a consequence, for QW thicknesses lower than h_k there exists a maximum length (as a function of h) of MD segment (which is regarded as a kink rather than a dislocation) that reduces the total energy in the system. For thicker QWs, $h > h_k$ (area "3"), any length of a MD segment lowers the total energy and thus once the MD segment is created it extends to infinity. The last area, "4", corresponds to very small h or l, i.e. to the dislocation core region where this continuum picture does not apply.

Note: h_k is, in fact, the ordinary CT for the energy balance model where the dislocation energy is assumed to be in the form of Eq. 3.20 regardless of the MD segment length.

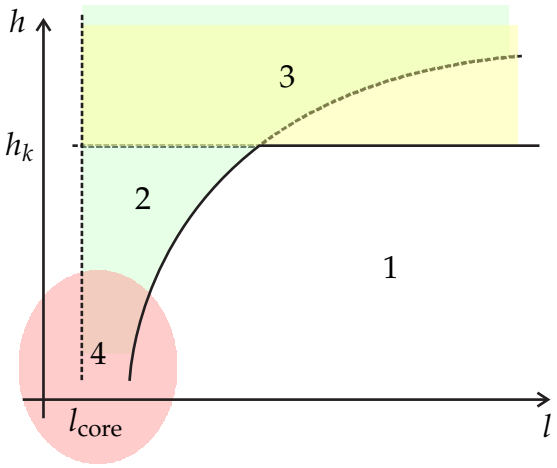


Figure 3.15: h–l-parameter space. Different regions discussed in the text are marked "1"–"4".

Comparison with experimental observations of staircase structures

For comparison with experimental observations (see Cherns, 2007, p. 126), average lengths of the MD segments corresponding to the observed dislocation inclinations were calculated for several QW thicknesses. All these layer thicknesses lie above the h_k which is just below 1.5 nm for this system. Therefore, MD segments in all these three cases are expected to extend to infinity (in theory) and not to find any equilibrium MD segment length[3].

[3] with an exception of the 1.5 nm QWs which are very close to h_k and thus may be affected by inaccuracies in numerical values of material constants and/or not working with hexagonal symmetry etc.

The different MD segment lengths may be explained by the different growth times needed for growth of the QWs. Thicker layers need longer time which corresponds to longer MD segments. If the growth rate is independent of the QW thickness, h, then one would expect a linear dependence of the MD segment length on h. And indeed, the three experimental data points seem to follow a linear trend.

To support this idea it would be interesting to see the same structures but with a pause after the growth of each QW (but keeping all other conditions unchanged). Longer MD segments are expected in such structures. In fact, the linear dependence in Fig. 3.16 should only move to the right.

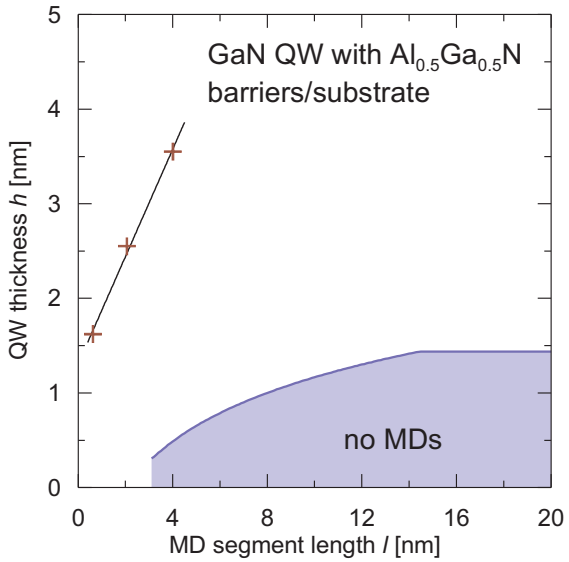


Figure 3.16: A GaN quantum well on top of an $Al_{0.5}Ga_{0.5}N$ substrate. MDs are expected in the region with white background. Crosses correspond to the experimentally observed configurations.

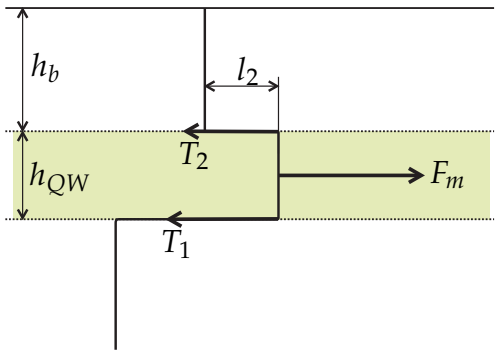


Figure 3.17: Forces acting on a TD segment in a QW capped with a barrier.

Single QW with a capping layer (barrier on top)

The situation changes when a layer on the top of the QW starts to grow. The termination top point of a TD in the QW is no longer free and the overgrown barrier/capping layer acts as an additional obstacle to the motion of a TD segment in a QW. In the following, the problem of whether the overgrown layer can stabilise a moving TD segment in a QW is investigated.

For simplicity it is assumed that the TD segment in the barrier is immobile. The forces acting on the TD segment in the QW are shown in Fig. 3.17. The line tensions $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ of the MD segments between the QW and the substrate and between the QW and the

capping layer, respectively, act against the misfit force $\boldsymbol{F}_m$. The magnitude of the total force is

$$F_{\text{total}} = F_m - T_1 - T_2 \,. \tag{3.27}$$

Furthermore, it is assumed that $F_m - T_1 > 0$, i.e. that the TD segment in the QW was moving before the capping layer was introduced. The condition when $F_{\text{total}} = 0$ is now sought. The misfit force is given by Eq. 3.19. As the MD segment was already in the regime of "a long dislocation", the line tension T_1 is given by Eq. 3.20. For the last contribution, T_2, it is assumed that the MD segment (of length l_2) in the upper interface is small enough to be considered as a kink. Therefore, Eq. 3.27 translates to

$$F_{\text{total}} = b2\varepsilon_{yy}^m \mu \frac{1+\nu}{1-\nu} h_{\text{QW}} - \frac{\mu b^2}{4\pi(1-\nu)} \left(\ln \frac{h_{\text{QW}} + h_b}{r_{\text{core}}} + \ln \frac{2l_2}{r_{\text{core}}} - 2 \right) - 2l \frac{\mathrm{d}\mathcal{E}_{\text{core}}}{\mathrm{d}l} \,. \tag{3.28}$$

The TD segment moves if and only if $F_{\text{total}} > 0$. Inspection of Eq. 3.28 reveals that the only positive term is the misfit force; this term is independent of any barrier parameters. On the other hand, the negative line tension terms become more negative with increasing barrier thickness h_b.

In all the above consideration it has been assumed that the upper TD segment was immobile. Obviously, T_2 acts on this segment and thus there is a driving force on it. The theory of dislocations provides an important result that dislocation velocity is proportional to the applied stress $v \propto \tau^n$ where n is some constant. However, if there exists a minimum applied stress, τ_0, for which the dislocation moves then the dislocation velocity law changes to

$$v = \begin{cases} A(\tau - \tau_0)^n & \text{for } \tau > \tau_0 \,, \\ 0 & \text{otherwise} \,. \end{cases} \tag{3.29}$$

This mechanism can be imagined as some kind of a "friction" which one needs to overcome in order to move a dislocation.

Comparison with experimental findings

Let's assume that the stress threshold τ_0 is such that when the barrier reaches a thickness $h_{b,\text{stab}}$ when the TD segment in the QW stops moving, the length of the upper TD segment satisfies $l_2 < l_{\text{core}}$. This corresponds to either high stress τ_0 or a very quickly growing barrier. In this case, $T_2 = l\mathcal{E}_{\text{core}}$ and the stabilising barrier thickness from Eq. 3.28 is

$$h_{b,\text{stab}} = r_{\text{core}} \exp\left[8\pi(1+\nu) \frac{h_{\text{QW}}}{b} \varepsilon_{yy}^m - \frac{\mathrm{d}\mathcal{E}_{\text{core}}}{\mathrm{d}l} \frac{4\pi(1-\nu)}{\mu b^2} \right] - h_{\text{QW}} \,. \tag{3.30}$$

Results are summarised in Table 3.5. The experimental thickness of barriers was 10.5 nm. Therefore, the samples with 1.5 nm and 2.5 nm QWs are well below this limit which suggests that the TD segments in the QWs became immobile. On the other hand, for the 3.5 nm QW, the barrier thickness needed for TD segment stabilisation was not reached and the TD segments in the QWs are expected to move further.

h_{QW} [nm]	1.5	2.5	3.5
$h_{b,stab}$ [nm]	≈0	3.55	19.84

Table 3.5: Calculated barrier thicknesses that prevent a TD segment in the QW from further motion.

If the TD segments are not stabilised, it seems reasonable to deduce that the lowest MD segment (in the lowest interface) is the longest one and MD segments are getting shorter in the upper interfaces of a MQW structure. In such a case the overall dislocation shape would be bowed having the biggest inclination from the vertical direction in the lowest interface. Cherns (2007, p. 127) reported that some dislocations were "bowed" for the 3.5 nm QWs whereas in the case of 1.5 nm and 2.5 nm QWs all dislocations appeared to be straight (that is with a constant inclination).

3.5 Conclusions

The critical thickness of AlGaN/GaN and InGaN/GaN systems was calculated using the energy balance model. This model was corrected for the hexagonal symmetry of wurtzite III-nitrides. The best agreement with experimental data was obtained when the dislocation core energy was included. Despite the satisfying agreement with experimental observations one has to bear in mind some limiting factors as well. Firstly, the model presents an equilibrium critical thickness which does not concider any dislocation generation mechanism. Further, the model does not include differences in material properties in the underlying substrate and the epilayer itself (even though an attempt to account for this was reported as well in Fig. 3.4). Also the free surface which affects the dislocation strain/stress field is not included.

Obstacles to the experimental determination of the CT are also significant. Directly imaging techniques such as TEM are extremely time-consuming and inaccurate due to low MD densities; on the other hand, measurements like XRD or MOSS may not be sensitive enough to detect the actual onset of strain relief. As discussed, for example by Eaglesham et al. (1988), this is also a problem in TEM where one can observe only

MDs *present* in the sample (and thus it is not guaranteed that the very first MD will be observed).

Another big question concerns the origin of MDs. Nucleation centres for new MDs at the free surface may be surface roughness or surface steps. Other possibility may be nucleation of MDs from apexes of V-defects in the InGaN/GaN system (as suggested by Jahnen et al. (1998)) or from cracks in the AlGaN/GaN system (as reported by Bethoux and Vennéguès (2005)). A lot of experimental effort has been put into clarifying of the origin of MDs in III-nitrides. However, there is still no unique answer as to how the MDs are being introduced. The situation is much more complex because of other strain relieving mechanisms (such as V-pits or cracks) competing with MDs.

Chapter 4

Reduction of the threading dislocation density

4.1 Introduction

Reduction of the threading dislocation density (TDD) remains one of the main challenges in GaN-based materials research. Several sophisticated methods of growing low-dislocation density layers such as epitaxial lateral overgrowth (see review by Beaumont et al., 2001), use of SiN_x interlayers (Kappers et al., 2007) or using ScN masking layer (Moram et al., 2007) interlayers have been reported in the literature.

There are also several reports on experimental measurements of TDD reduction. Two examples of dislocation density ρ as a function of layer thickness h of HVPE[1] grown GaN are shown in Fig. 4.1. The striking difference in these reports is in the slope of these curves, i.e. in the rate of reduction. Another interesting observation concerns the difference in the curves' shapes: Jasinski and Liliental-Weber (2002) reported an almost linear dependence in the log–log graph whereas Albrecht et al. (1999) observed an "S"-shape dependence. This is even more interesting considering that the range of dislocation densities ρ and layer thicknesses h investigated by Albrecht et al. (1999) is covered by the report of Jasinski and Liliental-Weber (2002). Albrecht et al. (1999) seem to get to a point of *saturation* where any further TDD reduction becomes very slow or impossible (see the point at $h \approx 100\,\mu$m). Similar experience is also seen with MOCVD[2] growth (Kappers et al., 2007, and private communication). Such observations suggest that the actual growth conditions, the growth technique and substrate quality prior to the growth are important in TDD reduction.

[1] hydride vapour phase epitaxy
[2] metal-organic chemical vapour deposition

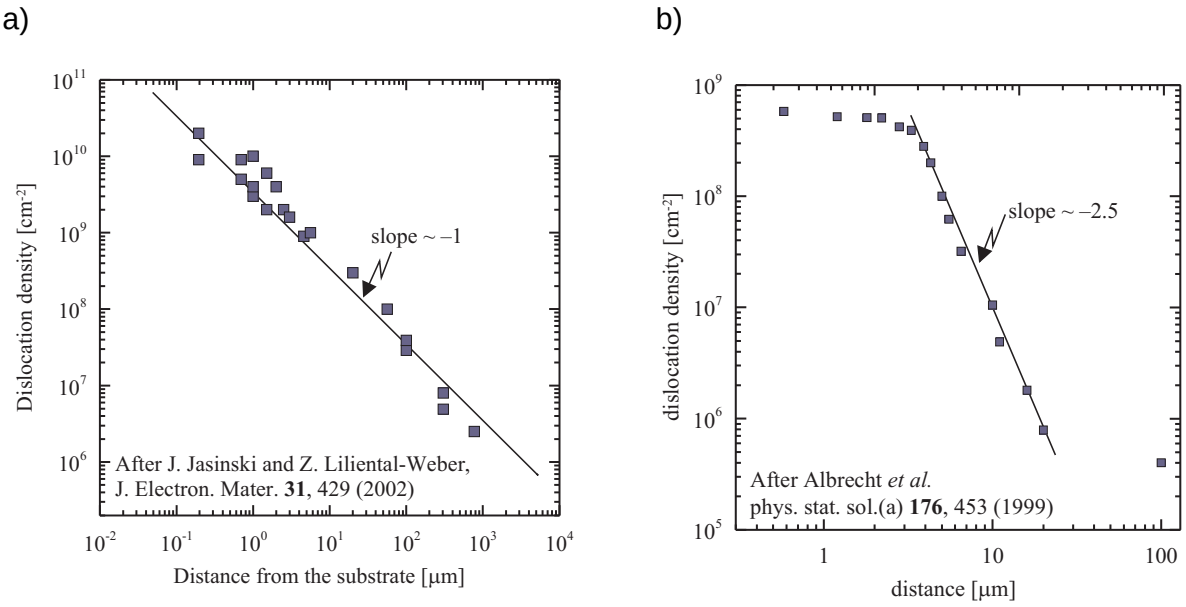


Figure 4.1: Threading dislocation density reduction during thickening of GaN layers by HVPE reported by (a) Jasinski and Liliental-Weber (2002) and (b) Albrecht et al. (1999).

Modelling the mechanisms of TDD during growth aims (i) to increase understanding of the basic physics associated with TDD reduction and (ii) to help to identify effective ways for TDD reduction. The ultimate long term goal is to build a model with predictive capabilities so much fewer experimental work would be needed.

This chapter discusses how the TDD reduction is effected by islands formed during the initial stage of the standard hetero-epitaxial growth.

4.1.1 TDD reduction model by Romanov (layer-by-layer growth)

Romanov et al. (1996, 1997) introduced a model based on "reactions" of inclined dislocations. Two dislocations with Burgers vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ can react when they, by some process, become contiguous. If $\boldsymbol{b}_1 = -\boldsymbol{b}_2$, i.e. two dislocations with opposite Burgers vector meet they *annihilate*. Other possible reaction types are *scattering* ($\boldsymbol{b}_1 + \boldsymbol{b}_2 \rightarrow \boldsymbol{b}_3 + \boldsymbol{b}_4$), *reduction* ($\boldsymbol{b}_1 + \boldsymbol{b}_2 \rightarrow \boldsymbol{b}_3$) or no reaction at all. In a first approximation, energies of reacting and resulting dislocations are compared in order to determine which reaction type is favourable.

The energy per unit length of a dislocation with the Burgers vector $\boldsymbol{b}$ is, within the isotropic approximation, (cf. Eq. 2.6)

$$\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} = \frac{\mu|\boldsymbol{b}|^2(1-\nu\cos^2\theta)}{4\pi(1-\nu)}\ln\left(\frac{R}{r_0}\right) . \tag{4.1}$$

The above expression can be very roughly approximated as

$$\frac{\mathrm{d}\mathcal{E}_d}{\mathrm{d}l} \approx \text{const.}|\boldsymbol{b}|^2 \,. \tag{4.2}$$

The energy criterion determining which reaction takes place is called *Frank's rule* and for the reaction $\boldsymbol{b}_1 + \boldsymbol{b}_2 \rightarrow \boldsymbol{b}_3$ (see Fig. 4.2) has form

$$|\boldsymbol{b}_1|^2 + |\boldsymbol{b}_2|^2 > |\boldsymbol{b}_3|^2 \qquad \text{(Frank's rule)} \,. \tag{4.3}$$

If this condition is satisfied then the reaction occurs. Frank's rule for scattering looks similar.

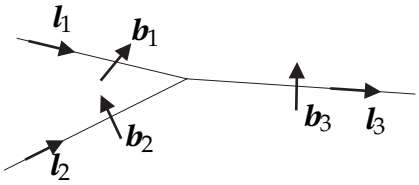


Figure 4.2: An example of a dislocation reaction (reduction) $\boldsymbol{b}_1 + \boldsymbol{b}_2 \rightarrow \boldsymbol{b}_3$. Vectors $\boldsymbol{l}_1$, $\boldsymbol{l}_2$ and $\boldsymbol{l}_3$ are the respective dislocation line directions.

In Romanov's model, threading dislocations do not move physically (i.e. do not glide or climb) but the "movement" results from growth. Suppose the situation sketched in Fig. 4.3. The TD segments are inclined with respect to the growth direction (here vertical) and as a consequence, the terminating points at the upper surface "move" laterally and can eventually become contiguous. Of course, in the full 3D model, it is very unlikely that two random TDs will intersect each other. Therefore, Romanov et al. (1996) introduced a *reaction radius* r_I (generally a different value for each reaction type). If the distance between terminating points of two dislocations is shorter than r_I, then the reaction can take place. The growth surface is assumed to be planar which corresponds with the Frank-van der Merwe's layer-by-layer growth mode.

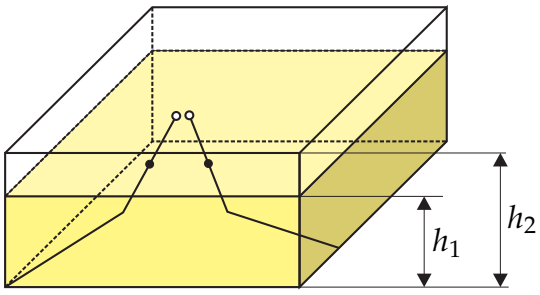


Figure 4.3: Lateral movement of the terminating points of dislocation lines on the top surface. As the layer gets thicker ($h_1 \rightarrow h_2$), some dislocations may converge and get closer.

Subsequently, Romanov et al. (1996, 1997) applied this model to cubic InGaAs/GaAs and SiGe/Si (001) systems where TDs are "60° dislocations" ($\langle 110 \rangle \{111\}$ slip system).

Such dislocations lie on planes that are inclined with respect to the [001] growth direction and thus during the growth may become close enough to react. Mathis et al. (2001) used this model for hexagonal *c*-plane ((0001) plane) GaN grown on (0001) sapphire. The inclination of (some) dislocations was obtained by considering the dislocation energy (see section 2.3.3). Within the isotropic theory, the optimal dislocation line direction for *c*- and *a*-type dislocations in GaN is vertical whereas for $(a + c)$-type dislocations it is approximately 15.6° inclined from the vertical direction. Mathis et al. (2001) also presented TEM evidence for some dislocations being inclined and claimed agreement with the above mentioned inclination angle.

4.1.2 Experimental observations

However, as it is shown in section 2.3.3, the same concept determines that all dislocation types are vertical, providing the hexagonal symmetry of wurtzite GaN is used. Moreover, most TEM micrographs are not conclusive about the inclination[3] of TDs and many of them suggest that the crucial stage in the TDD reduction is the formation of faceted GaN islands during the initial stage of growth (see Fig. 4.4 and Fig. 4.5). During this stage, dislocations change their direction due to the presence of inclined island facets. This is a completely new mechanism that was not accounted for in the original model by Romanov et al., and subsequently it was not used by Mathis et al. (2001). Therefore, an immediate question arises: How does the initial islanded stage affect the model?

4.1.3 Bougrov's model

Bougrov et al. (2006) published work on a phenomenological model for GaN. They included intentionally inclined dislocations using a redirection layer to the Romanov's model, as an analogue for a two-step growth process (Lang et al., 2006, 2007). The advantage of this treatment is that their equations remain very simple while retaining the 3D character of the investigated situation. However, various dislocation types remain unresolved in this model and its phenomenological character inhibits a deeper understanding of the effect of dislocation bending. Therefore, a new model is presented here which accounts for all dislocation types (various Burgers vectors) separately and takes the actual geometry of islands into consideration.

[3]This seems to depend also on the actual growth technique used. Since Cambridge-grown samples are obtained using the MOCVD, the statements herein are based on the observations of those samples.

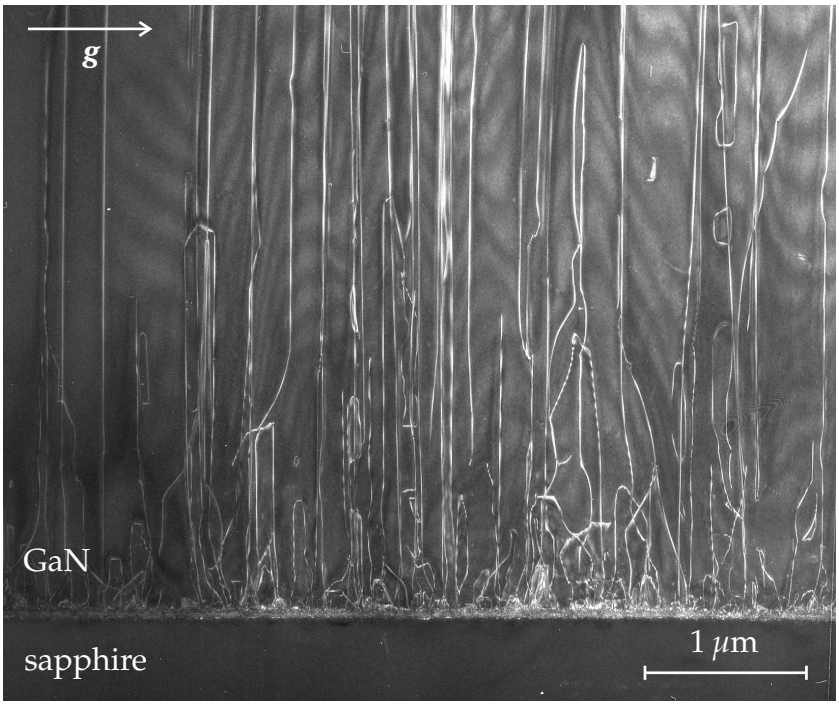


Figure 4.4: TEM dark field image ($\boldsymbol{g} = 1\bar{2}10$) of a typical thick gallium nitride specimen. a- and $(a+c)$-type dislocation are visible under this condition. A high density of threading dislocations generated in the interface between the sapphire substrate and the thick GaN layer can be observed. However, the TDD does not change significantly further away from the interface. (Courtesy of J. Barnard.)

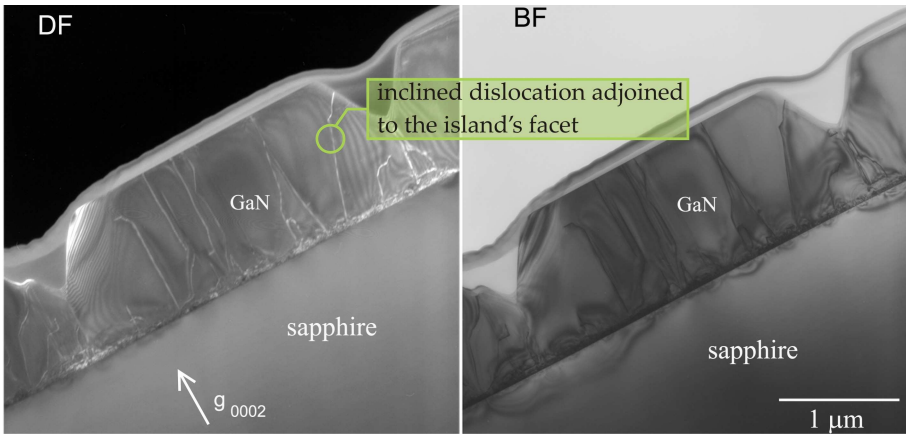


Figure 4.5: TEM dark field and bright field images with ($\boldsymbol{g} = 0002$) of gallium nitride islands grown on sapphire. The effect of faceted islands resulting in inclined TDs is clearly demonstrated. (Courtesy of D.V. Sridhara Rao.)

4.1.4 An improved model with islands

A major assumption of the present model is a simplification of the geometry to 2D. The description of the islands becomes much more difficult than for a flat surface so a numerical instead of an analytical treatment is needed. Despite this simplification, the 2D model brings some insight and, although not quantitatively, at least qualitatively

accounts for the contribution of the islands. In order to keep results comparable with the original Romanov/Mathis's model, the isotropic approximation is used.

Several nucleation centres are assumed on the substrate surface (see Fig. 4.6a). On each of them an island nucleates (Fig. 4.6b). As the islands grow, misfit dislocations are generated at the substrate–epilayer interface, which subsequently turn upwards and propagate through the film as TDs (Fig. 4.6c). The facet on which a particular dislocation terminates determines its direction. A dislocation can change its direction either by becoming adjacent to a different surface (after coalescence of two islands or simply due to its movement) or by changing its type after a reaction with another TD. The reaction can be either *annihilation* (which reduces the total number of dislocations by 2), *reduction* (reduction by 1 dislocation), or *scattering* (where the total number of dislocations remains constant but the dislocation energy is reduced). After the islands have coalesced, growth continues in the 2D (layer-by-layer) mode (Fig. 4.6d) where some minor reduction of the TDD still takes place.

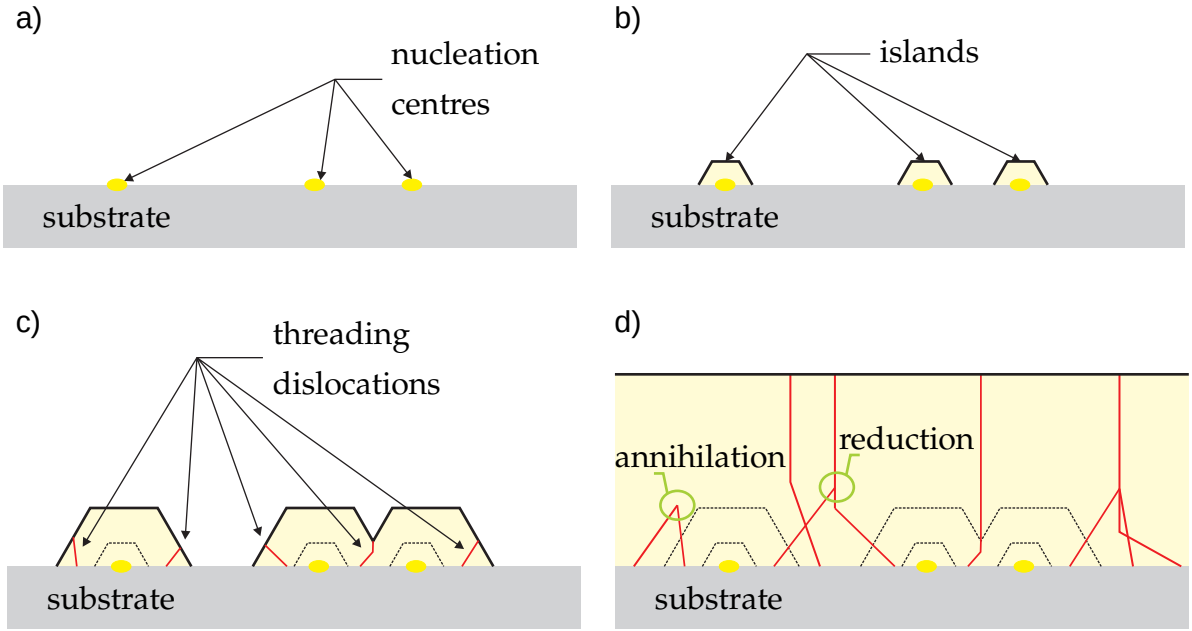


Figure 4.6: Stages of the growth: a) substrate with nucleation centres; b) small islands start growing at the nuclei; c) islands get bigger and some of them coalesce; dislocations are generated in the substrate–epilayer interface; d) the fully coalesced thick layer after dislocation reactions (reduction and annihilation).

Dislocations in the substrate–layer interface are generated because of the lattice mismatch of the two materials. Their purpose is to relax the misfit strain. Only dislocations with an edge component of the Burgers vector lying in the interface are capable of relaxing the misfit strain. Therefore, only the a- and $(a+c)$-type dislocations are assumed to be generated at the interface; the c-type dislocations are introduced only later on during growth as a result of dislocation reactions. The a- and $(a+c)$-type dis-

locations are assumed to be equally populated in the interface and they are randomly[4] distributed along the interface.

There are 20 dislocation types in the real 3D situation: 2 c-type ($\pm\boldsymbol{c}$), 6 a-type ($\pm\boldsymbol{a}_i$, $i = 1,2,3$), and 12 $(a+c)$-type ($\pm\boldsymbol{a}_i \pm \boldsymbol{c}$, $i = 1,2,3$). In the 2D case, their number is reduced to only 8 in total: 2 c-type ($\pm\boldsymbol{c}$), 2 a-type ($\pm\boldsymbol{a}$), and 4 $(a+c)$-type ($\pm\boldsymbol{a} \pm \boldsymbol{c}$). The dislocation reactions are listed in Table 4.1.

	$\boldsymbol{c}$	$-\boldsymbol{c}$	$\boldsymbol{a}$	$-\boldsymbol{a}$	$\boldsymbol{a}+\boldsymbol{c}$	$\boldsymbol{a}-\boldsymbol{c}$	$-\boldsymbol{a}+\boldsymbol{c}$	$-\boldsymbol{a}-\boldsymbol{c}$
$\boldsymbol{c}$	–	A	–	–	–	$\boldsymbol{a}$	–	$-\boldsymbol{a}$
$-\boldsymbol{c}$	A	–	–	–	$\boldsymbol{a}$	–	$-\boldsymbol{a}$	–
$\boldsymbol{a}$	–	–	–	A	–	–	$\boldsymbol{c}$	$-\boldsymbol{c}$
$-\boldsymbol{a}$	–	–	A	–	$\boldsymbol{c}$	$-\boldsymbol{c}$	–	–
$-\boldsymbol{a}+\boldsymbol{c}$	–	$\boldsymbol{a}$	–	$\boldsymbol{c}$	–	$\{\boldsymbol{a},\boldsymbol{a}\}$	–	A
$-\boldsymbol{a}-\boldsymbol{c}$	$\boldsymbol{a}$	–	–	$-\boldsymbol{c}$	$\{\boldsymbol{a},\boldsymbol{a}\}$	–	A	–
$-\boldsymbol{a}+\boldsymbol{c}$	–	$-\boldsymbol{a}$	$\boldsymbol{c}$	–	–	A	–	$\{-\boldsymbol{a},-\boldsymbol{a}\}$
$-\boldsymbol{a}-\boldsymbol{c}$	$-\boldsymbol{a}$	–	$-\boldsymbol{c}$	–	A	–	$\{-\boldsymbol{a},-\boldsymbol{a}\}$	–

Table 4.1: Table of reactions for the 2D case based on Frank's rule (Eq. 4.3). 'A' is annihilation, '–' means that the two dislocations do not react. If the result is a pair of dislocations, the reaction is called scattering. During scattering, the number of dislocations is not reduced, however the total energy of dislocations is lowered.

The model is not dependent on chosen units, thus it is a *scalable* model (see section 4.2.5). The length units are labelled "d.u." (distance units) and time is measured in "t.u." (time units). The modelled width of the substrate is 800 d.u.; this can be considered as the periodicity length of the model. *Standard parameters* of the model (given in Table 4.2) constitute a base for comparisons between various parameter studies.

4.2 Results and discussion

The dislocation types are randomly distributed along the substrate–epilayer interface. Therefore, a series of 50 runs was carried out for every study; these were subsequently

[4] Randomly in terms of their type, not position. See below.

parameter	value
periodicity of the model	800 d.u.
initial number of dislocations	80
initial dislocation spacing	10 d.u.
number of island nuclei	4
distance between neighbouring nuclei	200 d.u.
growth rates	1 d.u./t.u. (horizontal [top] facet) 2 d.u./t.u. (side [inclined] facets)
inclination of facets	$\approx 30^\circ$ from the vertical direction

Table 4.2: A set of the standard parameters of the model used for various comparisons.

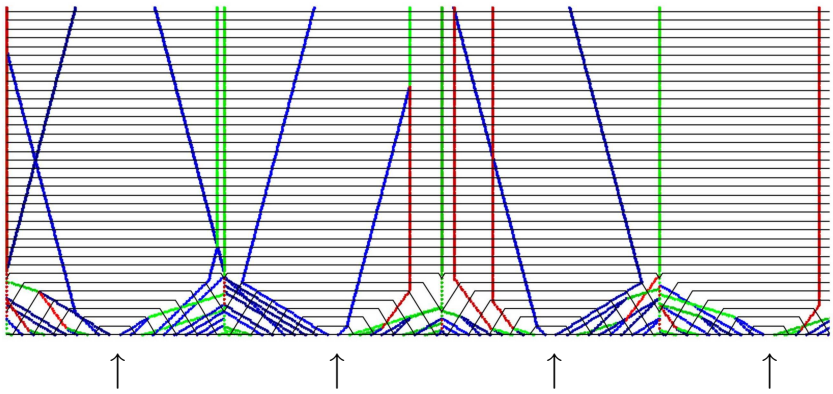


Figure 4.7: A typical output of the simulation for the standard parameters and total time of 300 t.u.. The black lines show the evolution of a layer surface every 8 t.u.. Red, green and blue colours corresponds to the c-, a- and $(a+c)$-type dislocations, respectively. The arrows show centres of the islands.

averaged. The graphs presented below show the averaged values of the number of dislocations as a function of various parameters, often time or layer thickness. In the case of a sample with islands, layer thickness is defined as a thickness of an equivalent uniform layer made of the same volume of material. The model was implemented in the system *Mathematica*.

A typical picture of dislocations threading through an epilayer as obtained from the simulations is shown in Fig. 4.7.

4.2.1 Contribution of the islands

Contribution of the islands to the TDD reduction is revealed in a basic comparison summarised in Fig. 4.8a. Results of studies of the TD density reduction with (i.e. the 3D model) and without (i.e. the layer-by-layer growth model) islands being taken into account are compared. Several general trends can be observed:

- The major reduction in the number of dislocations occurs at the very beginning of the growth. The speed of reduction decreases as the thickness increases because there are fewer TDs in the material. As a consequence, the dislocation density eventually saturates and no further reduction occurs. Both these statements are intuitively clear and can be expected. It is not clear from the literature whether this saturation appears experimentally: many results suggest that there is no saturation (see for example Jasinski and Liliental-Weber, 2002, and Fig. 4.1). However, Albrecht et al. (1999) published results where saturation was observed. Kappers et al. (2007) also observed this saturation in their samples once the *coalescence thickness* was reached. This may be due to differences in growth technique or different experimental conditions. It is also possible that saturation could have occurred in the other examples if the growth had continued or that their experimental findings are misinterpreted here.

- The TDD reduction is much faster in the case of islands: a near-to-saturation stage is reached at the thickness of $\approx$ 50 d.u. (the coalescence thickness is $\approx$ 60 d.u.) in the presence of the islands in contrast to $\approx$ 150–200 d.u in the case of the layer-by-layer growth.

- The final number of dislocations is approximately 27 (a 66% reduction from the initial value of 80 dislocations) in the presence of islands whereas it is approximately 48 (a 40% reduction from the initial number of 80 dislocations) in the case of the layer-by-layer growth. Therefore, islands in this case account for additional reductions of the TDD by 26%.

- The variation of possible final numbers of TDs due to different starting configurations of dislocations is slightly larger in the case of the islands than without them. However, there is no or very small overlap of the possible results in these two growth modes. This underlines the important contribution of islands.

Fig. 4.8b shows results based on a less strict version of Frank's rule: Eq. 4.3 modified to

$$|\boldsymbol{b}_1|^2 + |\boldsymbol{b}_2|^2 \geq |\boldsymbol{b}_3|^2 \quad . \tag{4.4}$$

While the layer-by-layer growth mode seems to be unaffected by this change, in the case with the islands a further reduction by approximately 7% is observed, that is the final average number of dislocations is $\approx$ 22. This result suggests that the case with islands is more sensitive to the actual form of the reaction criterion, which leaves space for further investigations on a more accurate form of Frank's rule.

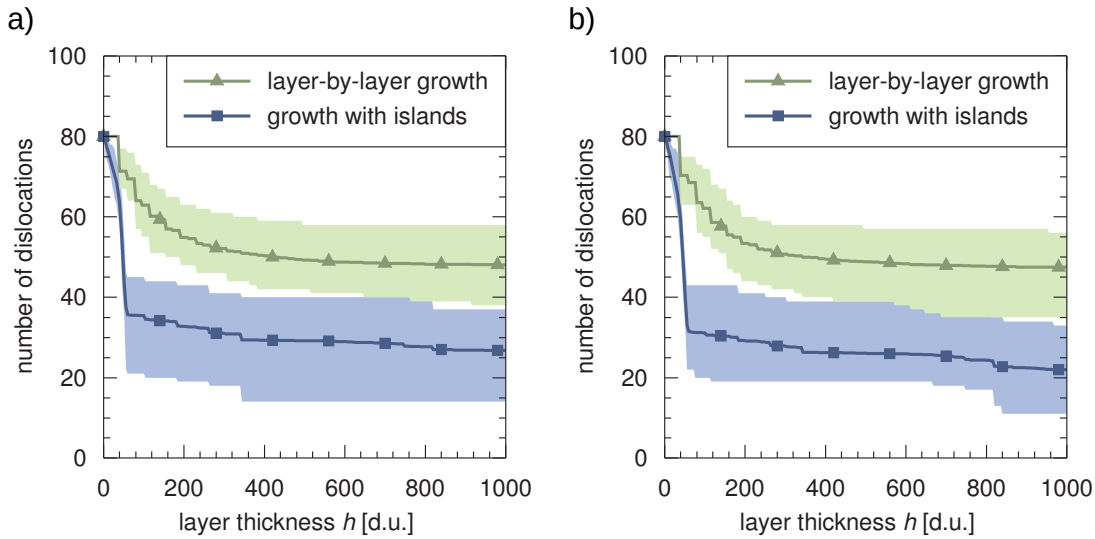


Figure 4.8: Effect of the islands on the total TD density: blue and green curves are cases with and without islands taken into account, respectively. Simulations were carried out with the standard parameters with (a) Frank's rule (Eq. 4.3) or (b) the weaker Frank's rule (Eq. 4.4) applied.

4.2.2 Size effects: distance between islands

Further studies concentrate on the dependence of the TDD reduction on various parameters of the proposed model, in particular, on the geometry. As a first case study, dependence on island separation is discussed (see Fig. 4.9): while keeping all but one parameter constant and equal to the standard values (see Table 4.2), the distance λ between neighbouring islands was varied. As a consequence, since the "base" of the model is constant (and equal to 800 d.u.), the total number of islands in the model varies.

Numerical results are summarised in Table 4.3. Evidently, there is a trend towards lower numbers of dislocations with increasing λ. The last row corresponding to $\lambda =$ 800 d.u. is probably affected by the periodicity error of the model. Despite the fact that the ranges given by the standard deviation (again based on 50 runs for each λ) overlap, the trend is clearly evident. What is even more encouraging is that this conclusion is in good agreement with the experimental observations: Kappers et al. (2007) observed increased TDD reduction with higher SiN_x coverage prior to the regrowth which corre-

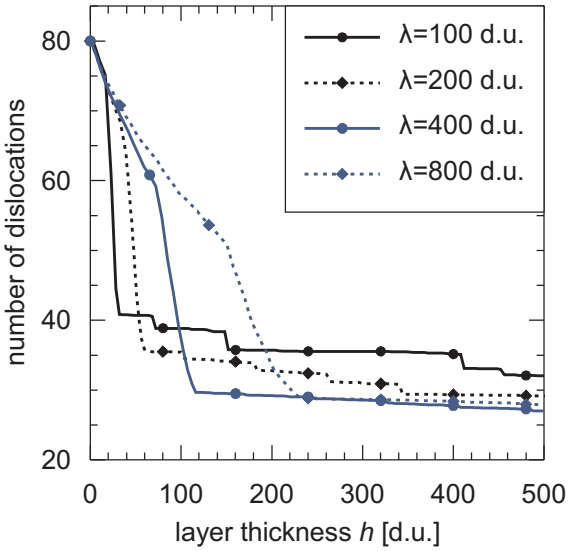


Figure 4.9: Results of simulations with various distances λ between neighbouring islands.

	$n(500\,\text{d.u.})$
$\lambda = 100\,\text{d.u.}$	$n = (32.0 \pm 4.6)$
$\lambda = 200\,\text{d.u.}$	$n = (29.2 \pm 5.5)$
$\lambda = 400\,\text{d.u.}$	$n = (27.0 \pm 6.6)$
$\lambda = 800\,\text{d.u.}$	$n = (27.9 \pm 5.7)$

Table 4.3: Total numbers of dislocations for various distances λ between neighbouring islands and the layer being 500 d.u. thick.

sponds to sparser islands. The effect shown here is quite small, but the initial reduction varies strongly.

4.2.3 Random nucleation

Results of a small extension of the model are summarised in this section. Instead of having evenly spaced islands which are all nucleated at the beginning of the simulation, nucleation centres that appear at random positions in the substrate–epilayer interface every Δt time units are assumed.

Fig. 4.10a shows the final numbers of dislocations after 1000 t.u. of the growth corresponding to $\approx$ 1000 d.u. thick layers, for values of Δt ranging from 0.1 t.u. to 50 t.u.. There is no dependence of the final number of dislocations on the actual value of Δt.

However, by inspecting the graph in Fig. 4.10b one can identify different rates of reduction: for larger Δt the reduction is slower. Bigger Δt corresponds to sparser islands, though not exactly in the sense of section 4.2.2. In both cases of sparse islands, large inclined facets of the islands, which promote the dislocation reduction, are obtained.

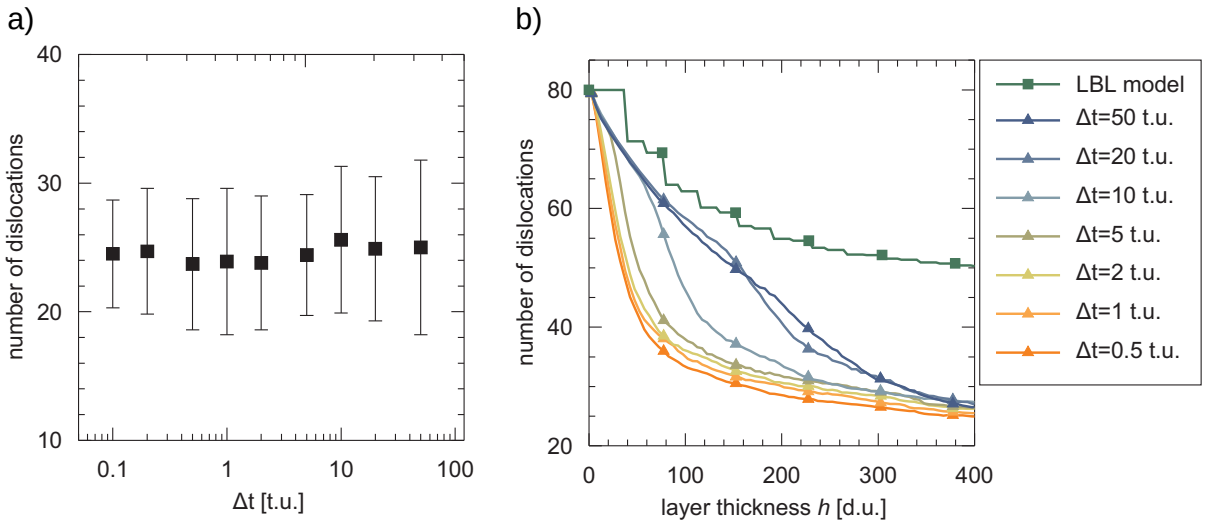


Figure 4.10: (a) Total numbers of dislocations after 1000 t.u. and (b) the number of dislocations as a function of the layer thickness for various intervals Δt between subsequent nucleation of islands. 'LBL' stands for layer-by-layer growth, the situation with no islands.

4.2.4 Coalescence thickness

In order to compare simulations with some experimental data, a study of TD density dependence on coalescence thickness was carried out. Variation of the number of nuclei within one period of 800 d.u. effectively translates into changes of the coalescence thickness. Statistically evaluated results of simulations together with experimental data from the literature are shown in Fig. 4.11. Although it is not possible to correlate the absolute values of dislocation density, qualitatively correct trends are observed. This suggests that the model extended to 3D would be capable of quantitative comparisons.

4.2.5 Scalability of the model

This rather technical section discusses an important property of the model: scalability. As was mentioned in the description of the model (section 4.1.4) it is independent of length and time units. A multiplication of distances by factor α does not change the physics and as a consequence, the same results are expected. To demonstrate that the

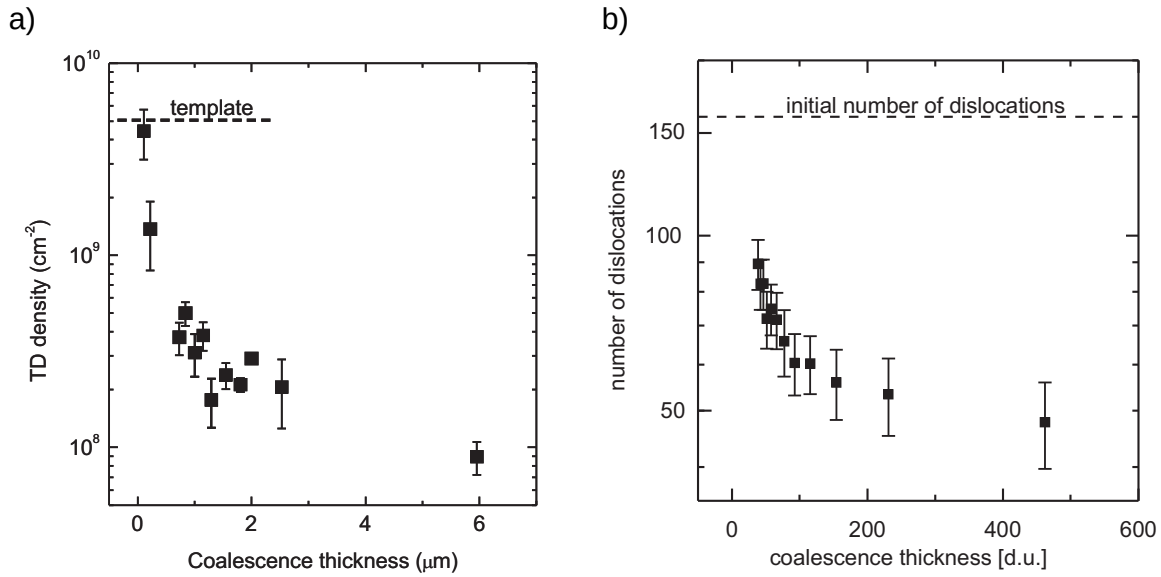


Figure 4.11: TD density as a function of coalescence thickness: (a) experimental data taken from Kappers et al. (2007), and (b) results of simulations.

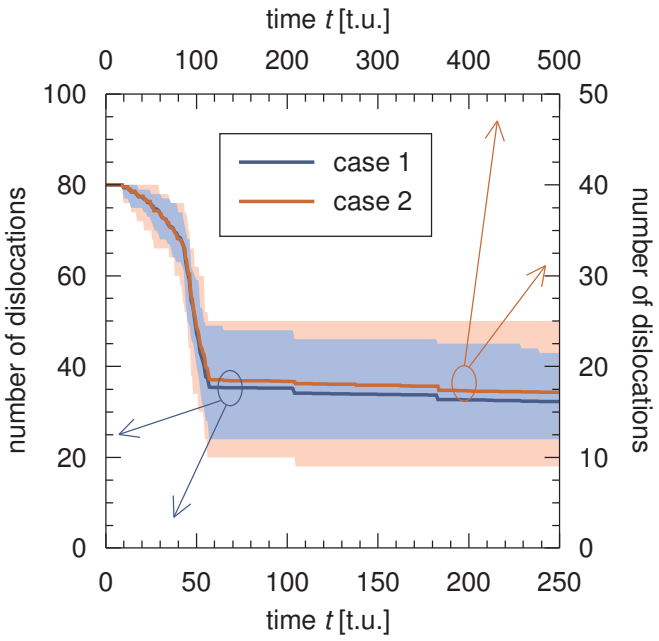


Figure 4.12: A proof of scalability of the model: blue results (case 1) correspond to the standard parameters, red results (case 2) correspond to twice as long distances and twice faster time.

numerical implementation of the model preserves this physically important property, a case study was carried out with results shown in Fig. 4.12. "Case 1" corresponds to the set of standard parameters. In "Case 2", all lengths were multiplied by a factor of 2, i.e. 40 evenly distributed dislocations with the spacing of 20 d.u. and 2 islands separated by 400 d.u. In order to preserve the growth rates of different facets and thus the overall geometry of the model, time goes twice as fast in case 2 than in case 1. The numerical results correspond very well. Small discrepancies between the two cases

can be ascribed to the boundary conditions, i.e. a periodicity error. The results are reasonable since the wider spread of data is in case 2 which is more affected by the boundary conditions.

This result is considered as a proof of the physical correctness of the numerical implementation of the model.

4.3 Conclusions and outlook

Results obtained so far using an improved model for threading dislocation density reduction including islands were summarised in this chapter. To start with, the model was simplified to 2D. The main purpose of this study was to estimate qualitatively the influence of islands on the TDD reduction and to investigate trends linked with various parameters of the proposed model. A good qualitative agreement with experiments was obtained as well. Therefore it is believed that the model as it stands is capable of *qualitative* predictions of various trends.

Future work may include further studies on the effects of different parameters of the model as well as new geometries (including different facets, different growth rates of the facets). In addition, using the knowledge gained with the 2D model, a significant improvement can be achieved by extending this model in 3D. Another significant improvement would be to take into consideration the actual dislocation–dislocation interactions. This would involve calculation of dislocation stress fields and, in fact, employing some methods of dislocation dynamics. Such work would constitute a major effort in the development and employment of new analytical and numerical methods and high optimisation of the code used for simulations. However, a model capable of *quantitative* predictions is expected as an outcome.

Part II

First principle calculations of electron energy loss near edge structures of III-nitrides

"A theory is something nobody believes, except the person who made it. An experiment is something everybody believes, except the person who made it."

Albert Einstein

Chapter 5

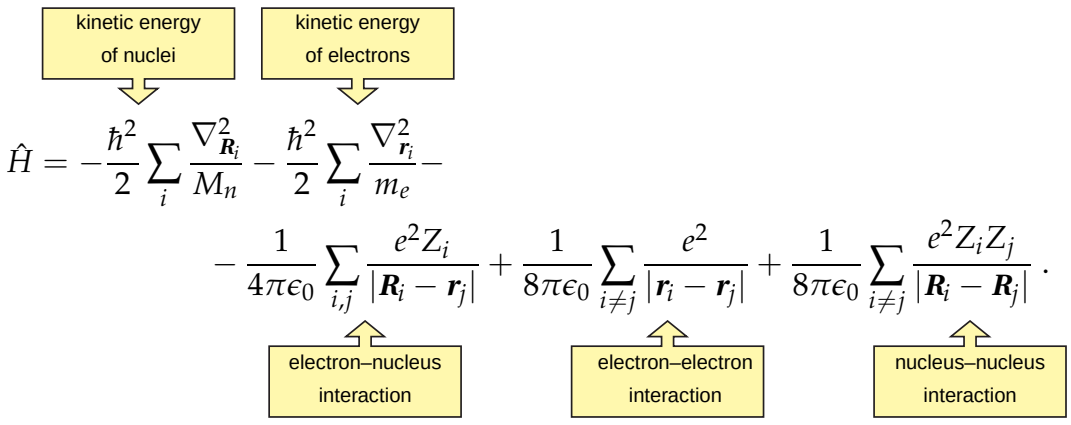
First principle calculations

This chapter aims to give a basic overview of the density functional theory (DFT) and related concepts. It is mostly based on the text by Cottenier (2002) and the manuals of `Wien2k` (Blaha et al., 2000) and `S/PHI/nX` (Boeck et al., 2003).

5.1 Density functional theory in brief

5.1.1 Formulation of a many-body problem

A solid state matter can be viewed as a collection of nuclei, positively charged ions at positions $\boldsymbol{R}_i$, and much lighter electrons at positions $\boldsymbol{r}_i$. Assuming that the atomic number is Z and that there are N atoms, the problem translates into solving a system of $N + ZN$ electromagnetically interacting objects. This is called a *many-body problem*. The corresponding exact Hamiltonian is



All nuclei in the above equation can be, in principle, of different types (Z_i). Different terms correspond to various contributions as described in the equation. This Hamiltonian takes into account only electrostatic Coulombic interactions. Despite that, it is practically impossible to solve such a problem exactly and thus several approximations must be made.

Nuclei are several orders heavier than electrons (mass of electron $m_e = 9.1 \times 10^{-31}$ kg vs. mass of proton $m_p = 1.7 \times 10^{-27}$ kg). This implies that nuclei move significantly slower than electrons. As a consequence, it is electrons that in the first instance maintains the inner equilibrium and respond to all external impulses. Therefore, nuclei can be viewed as "frozen", and only the electronic part remains to be solved. Nuclei are included as a source of fixed background positive charge. This is known as the *Born-Oppenheimer approximation*.

Under this approximation, the kinetic energy of (now static) nuclei is zero and the nucleus–nucleus interaction reduces to a constant. Equation 5.1 is left with a kinetic energy of electrons, $\hat{T}$, the potential energy due to electron–electron interaction, $\hat{V}$. The electron–nucleus interaction can be understood as an interaction of electrons with a positive external potential, $\hat{V}_{\text{ext}}$. The many-body Hamiltonian becomes

$$\hat{H} = \hat{T} + \hat{V} + \hat{V}_{\text{ext}} \,. \tag{5.2}$$

It is interesting to note that the kinetic energy, $\hat{T}$, and electron–electron potential, $\hat{V}$, are universal terms not depending on a particular system. The material-specific information is completely contained in the $\hat{V}_{\text{ext}}$ term.

5.1.2 Theorems of Hohenberg and Kohn

Although the problem is significantly reduced by the Born-Oppenheimer approximation, it is still far too complex to be solved. There are several methods of how to further simplify the solution of Eq. 5.2. A modern and powerful method, which is nowadays widely used for solid state calculations, is the *density functional theory* (DFT) firstly published by Hohenberg and Kohn (1964). The DFT works with the *electron density*, $\rho(\boldsymbol{r})$, as the main variable.

The central ideas of DFT were formulated in several theorems. In what follows, formulations given by Cottenier (2002) are used as they are easier to understand in today's scientific language. The first theorem was published by Hohenberg and Kohn (1964) and it states:

> THEOREM 1: There is a *one-to-one* correspondence between the ground state density $\rho(\boldsymbol{r})$ of a many-electron system and the external potential $\hat{V}_{\text{ext}}$.

An immediate consequence is that a ground state expectation value of any observable $\hat{O}$ is a *unique* functional of the ground state electron density

$$\langle \Psi | \hat{O} | \Psi \rangle = O[\rho] \,. \tag{5.3}$$

The most important part of the above Theorem 1 is that it claims a one-to-one correspondence. The electron density $\rho(\boldsymbol{r})$ can be calculated from the many-electron wave function as

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{N} \int \Psi^*(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_i \equiv \boldsymbol{r}, \ldots, \boldsymbol{r}_N) \cdot \\ \cdot \Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_i \equiv \boldsymbol{r}, \ldots, \boldsymbol{r}_N) \mathrm{d}\boldsymbol{r}_1 \mathrm{d}\boldsymbol{r}_2 \ldots \cancel{\mathrm{d}\boldsymbol{r}_i} \ldots \mathrm{d}\boldsymbol{r}_N \,. \quad (5.4)$$

Therefore, the implication that $\rho(\boldsymbol{r})$ is fully given by the external potential $\hat{V}_{\text{ext}}$ (through the many-electron wave function Ψ) is straightforward. However, the reverse implication is not so straightforward and in fact that is where the strength of the Theorem 1 originates. Intuitively it may seem that the electron density contains less information than the wave function, but Theorem 1 guarantees that they do contain the same amount of information (in terms of what knowledge can be extracted about atoms, molecules, or solids).

> THEOREM 2: For $\hat{O}$ being the Hamiltonian $\hat{H}$, the ground state total energy functional $H[\rho] = E_{V_{\text{ext}}}[\rho]$ is of the form
>
> $$E_{V_{\text{ext}}}[\rho] = \underbrace{\langle \Psi | \hat{T} + \hat{V} | \Psi \rangle}_{F_{HK}[\rho]} + \langle \Psi | \hat{V}_{\text{ext}} | \Psi \rangle \\ = F_{HK}[\rho] + \int \rho(\boldsymbol{r}) V_{\text{ext}}(\boldsymbol{r}) \, \mathrm{d}^3\boldsymbol{r} \quad (5.5)$$
>
> where the Hohenberg-Kohn density functional $F_{HK}[\rho]$ is *universal* for any many-electron system. $E_{V_{\text{ext}}}[\rho]$ reaches its *minimal* value (equal to the ground state total energy) for the ground state density corresponding to V_{ext}.

This second theorem makes it possible to use the variational principle in order to find the ground state electron density. Out of the infinite number of possible electron densities, only the one (up to an additive constant) which minimises $E_{V_{\text{ext}}}[\rho]$ is the ground state density.

5.1.3 Self-consistent scheme for solving the Kohn-Sham equation

The Hohenberg-Kohn functional can be further divided into

$$F_{HK}[\rho] = T_0[\rho] + V_H[\rho] + V_{xc}[\rho] \quad (5.6)$$

where T_0 is the kinetic energy of a non-interacting electron gas, V_H is the Hartree potential and V_{xc} the exchange-correlation part. This accounts for any self-interactions

within the electron gas and constitutes the difficult part which needs additional approximations (see below).

THEOREM 3: The exact ground state electron density $\rho(\boldsymbol{r})$ of an N-electron system is

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{N} \phi_i^*(\boldsymbol{r})\phi_i(\boldsymbol{r}) \tag{5.7}$$

where the single-particle wave functions $\phi(\boldsymbol{r})$ are the N-lowest energy solutions of the Kohn-Sham equation

$$\hat{H}_{KS}\phi = \epsilon\phi \tag{5.8}$$

where

$$\hat{H}_{KS} = \underbrace{-\frac{\hbar^2}{2m_e}\nabla^2}_{\hat{T}_0} + \underbrace{\frac{e^2}{4\pi\epsilon_0}\int \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r}'-\boldsymbol{r}|}\,\mathrm{d}^3\boldsymbol{r}'}_{\hat{V}_H} + \hat{V}_{xc} + \hat{V}_{\text{ext}}\,. \tag{5.9}$$

It is again important to note how strong this theorem is. In fact it says that the problem of solving a system of many coupled partial differential equations corresponding to the Hamiltonian in Eq. 5.2 can be transformed to solving a Schrödinger-like equation of non-interacting particles (Eq. 5.8). The single particle wave functions $\phi(\boldsymbol{r})$ do not have any physical meaning on their own; it is only the electron density $\rho(\boldsymbol{r})$ they generate which has a meaningful physical interpretation.

Also it is important to note that both, the Hartree potential, V_H, and the exchange-correlation potential, V_{xc}, depend on the electron density ρ, which in turn depends on them through the Kohn-Sham equation (Eq. 5.8). Therefore, the Kohn-Sham equation constitutes a *self-consistent problem* which can be solved by an iterative procedure. The initial electron density $\rho_0(\boldsymbol{r})$ needs to be guessed; after that it is used to construct the corresponding potentials. Subsequently, the Kohn-Sham equation is solved and a new electron density, $\rho_n(\boldsymbol{r})$, is constructed from the solution. This is repeated until convergency within a prescribed accuracy is reached.

5.1.4 The exchange-correlation potential

Apart from the Born-Oppenheimer approximation, the above described DFT scheme is exact. However, for any practical use of it, an approximation to the yet unknown exchange-correlation potential must be made.

A widely used approximation is the *local density approximation* (LDA) proposed already by Kohn and Sham (1965). The exchange-correlation energy functional takes the form

$$V_{xc}^{\mathrm{LDA}}[\rho] = \int \rho(\boldsymbol{r})\epsilon_{xc}(\rho(\boldsymbol{r}))\,\mathrm{d}^3\boldsymbol{r} \tag{5.10}$$

where the function ϵ_{xc} is the density of the exchange-correlation energy for the homogeneous electron gas (so called jellium model). The name for this approximation reflects the fact that the exchange-correlation energy density depends only on the value of electron density at a particular point. A more sophisticated approach is the *generalised gradient approximation* (GGA) which has the form

$$V_{xc}^{\mathrm{GGA}}[\rho] = \int \rho(\boldsymbol{r})\epsilon_{xc}(\rho(\boldsymbol{r}), |\nabla\rho(\boldsymbol{r})|)\,\mathrm{d}^3\boldsymbol{r}\,. \tag{5.11}$$

There are several parametrisations for the functional form of ϵ_{xc} in which various parameters are fixed against a comparison with a large set of experimental data. The often used parametrisation is the one of Perdew et al. (1996) usually termed as PBE employed also for all calculations in this work. In addition, there exists many other advanced approximations (e.g. hybrid potentials).

5.1.5 Solving the one step of Kohn-Sham equation

The above described scheme leads to a single particle Kohn-Sham equation (Eq. 5.8). Rather than looking for the eigenfunctions $\phi_m(\boldsymbol{r})$ directly the solution is expressed in a chosen basis set $\{\phi_p^{\mathrm{basis}}(\boldsymbol{r})\}$ as

$$\phi_m(\boldsymbol{r}) = \sum_{p=1}^{P} c_p^m \phi_p^{\mathrm{basis}}(\boldsymbol{r}) \tag{5.12}$$

where c_p^m are constant coefficients that need to be determined. The function space where the eigenfunctions $\phi_m(\boldsymbol{r})$ belong to, has an infinite dimension and thus P should be in principle infinite. However, for any practical use P needs to be cut to a finite value which will obviously introduce some deviation from the exact solution. Consequently, the "quality" of a solution depends on a chosen basis set and the actual value of P.

For a chosen basis set, the problem of the Kohn-Sham equation reduces to solving a system of linear equations of a form

$$\begin{bmatrix} \cdots & \cdots & \cdots \\ \vdots & \langle\phi_i^{\mathrm{basis}}|\hat{H}_{KS}|\phi_j^{\mathrm{basis}}\rangle - \epsilon_m\langle\phi_i^{\mathrm{basis}}|\phi_j^{\mathrm{basis}}\rangle & \vdots \\ \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} c_1^m \\ \vdots \\ c_P^m \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{5.13}$$

Diagonalisation of the above Hamiltonian matrix leads to P eigenvalues and P sets of coefficients c_p^m. Although this may seem trivial, in real problems, where the matrix sizes are of order 10^4 and larger, this step is the most time consuming part of the whole calculation.

A particular choice of a basis set can significantly reduce the value of P needed for a prescribed precision. Therefore, some basis sets are more efficient than others. The two most often used basis sets are plane waves (in this work used in the pseudopotential code `S/PHI/nX`, see section 5.3.1) and (linearised) augmented plane waves (in this work utilised in the `Wien2k` code, see section 5.2.1).

5.2 Short introduction to Wien2k

`Wien2k` is a full-potential all-electron code developed by Blaha et al. (2000) at the Institut für Materialchemie, Technische Universität Wien, Austria. It utilises the (linearised) augmented plane waves (LAPW) approach which is described in the following section.

5.2.1 The family of (L)APW basis sets

Generally speaking, weakly bonded electrons (for example valence electrons) are well described by plane waves which are solutions to the Hamiltonian with a zero potential. On the other hand, a core electron (deep in energy) "feels" practically only the nucleus to which it is bonded and thus it is well described by spherical harmonics (solutions for a single free atom). The (L)APW approach combines these two basis sets by setting up a *muffin tin sphere* (MTS), S_α, on each atom (labelled α). The rest of the space is the *interstitial region* (denoted I). A schematic division of a unit cell is shown in Fig. 5.1.

The problem with the original APW method is that the basis set depends on the eigenvalues $\epsilon_{\boldsymbol{k}}^n$ which are sought using this basis set. Thus, apart from diagonalisation of Eq. 5.13, one needs also to guess roots of the secular equation Eq. 5.13. Andersen (1975) proposed to use a Taylor expansion for estimation of the basis set functions at yet unknown energies $\epsilon_{\boldsymbol{k}}^n$

$$u_\ell^\alpha(r', \epsilon_{\boldsymbol{k}}^n) = u_\ell^\alpha(r', E_0) + (E_0 - \epsilon_{\boldsymbol{k}}^n) \underbrace{\left.\frac{\partial u_\ell^\alpha(r', E)}{\partial E}\right|_{E=E0}}_{\dot{u}_\ell^\alpha(r', E_0)} + \mathcal{O}(E_0 - \epsilon_{\boldsymbol{k}}^m)^2 \,. \tag{5.14}$$

Here, $u_\ell^\alpha(r, E)$ are solutions to the radial part of the Schrödinger equation for a free single atom α, $r' = |r - R_\alpha|$ is the distance from the centre of an atom α. E_0 is some trial

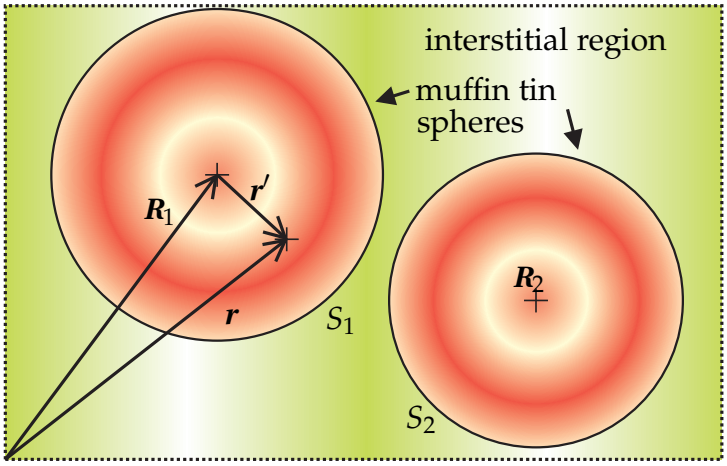


Figure 5.1: A schematic division of a unit cell (dotted area) into muffin tin spheres, S_α, (red areas) where an expansion into spherical harmonics is made, a the interstitial region, I, (green area) where the wave function is expanded into plane waves. Notation used in the text is also shown.

energy, close to the expected value of $\epsilon^n_{\boldsymbol{k}}$. The LAPW basis set takes the form

$$\phi^{\boldsymbol{k}}_{\boldsymbol{K}}(\boldsymbol{r}) = \begin{cases} \frac{1}{\sqrt{V}} e^{i(\boldsymbol{k}+\boldsymbol{K})\cdot\boldsymbol{r}} & \boldsymbol{r} \in I \\ \sum_{\ell,m} \left(A^{\alpha,\boldsymbol{k}+\boldsymbol{K}}_{\ell m} u^\alpha_\ell(r', E^\alpha_{1,\ell}) + B^{\alpha,\boldsymbol{k}+\boldsymbol{K}}_{\ell m} \dot{u}^\alpha_\ell(r', E^\alpha_{1,\ell}) \right) Y^\ell_m(\hat{r}') & \boldsymbol{r} \in S_\alpha \end{cases} \tag{5.15}$$

Y^ℓ_m are the spherical harmonics, $\hat{r}$ represents a pair (φ, ϑ) of polar angles specifying the direction $\boldsymbol{r} - \boldsymbol{R}_\alpha$. A requirement of a continuous wave function across the MTS boundary sets a system of equations for the coefficients $A^{\alpha,\boldsymbol{k}+\boldsymbol{K}}_{\ell m}$ and $B^{\alpha,\boldsymbol{k}+\boldsymbol{K}}_{\ell m}$. In order to make this method work efficiently, E_0 from Eq. 5.14 is different for every physically important ℓ (i.e. s-states, p-states, etc.) on each atom α rather than being a universal single value constant. This is reflected by the notation $E^\alpha_{1,\ell}$.

The LAPW basis set is appropriate for delocalised valence electrons. The deep core electrons, fully localised on a particular atom (i.e. inside a particular MTS) are expanded only to spherical harmonics. Despite this, there are still problems with description of (deeper) valence states of the same character. These are called semi-core states and are described as local orbitals (LO) in a following way

$$\phi^{lm}_{\alpha,LO}(\boldsymbol{r}) = \begin{cases} 0 & \boldsymbol{r} \notin S_\alpha \\ \left(A^{\alpha,LO}_{\ell m} u^\alpha_\ell(r', E^\alpha_{1,\ell}) + B^{\alpha,LO}_{\ell m} \dot{u}^\alpha_\ell(r', E^\alpha_{1,\ell}) + \right. & \\ \qquad \left. + C^{\alpha,LO}_{\ell m} u^\alpha_\ell(r', E^\alpha_{2,\ell}) \right) Y^\ell_m(\hat{r}') & \boldsymbol{r} \in S_\alpha \end{cases} \tag{5.16}$$

where $E^\alpha_{2,\ell}$ corresponds to the higher energy of the two states. Although this increases the size of a basis set and thus increases the computational time, much better accuracy

can be obtained including LOs. Therefore, they are implicitly used in `Wien2k` calculations.

The performance of the original APW basis set can be greatly improved by using a concept named APW+lo, where a fixed energy $E^{\alpha}_{1,\ell}$ is used for an APW base function with a subsequent LO-like correction (termed here lo). This method was developed by Sjöstedt et al. (2000). The valence d- and f- states need many basis functions (i.e. high $\boldsymbol{K}_{\max}$) in order to be described accurately; however, a much lower number of basis functions is needed for their accurate description within the APW+lo scheme (Cottenier, 2002). Such a mixed basis set of LAPW/APW+lo orbitals seems to be the best approach and, in combination with LOs for semi-core states, is routinely used in `Wien2k`.

5.2.2 Practicalities of running Wien2k

There are plenty of input files which need to be set up correctly in order to make the calculation run. Fortunately, it is only the master input file, the *struct-file* that needs to be set up manually (or using a user-friendly interface `w2web`, a part of the `Wien2k` distribution). The rest of the input files is generated automatically, and only some minor adjustments in specific cases are needed. The struct-file contains information about the structure itself: numbers, types and positions of the atoms, the space group and symmetry operations. Crutial parameters are the MTS radii which can be different for each inequivalent atom; they are also given in the struct-file.

Having the struct-file prepared, one has to initialise the calculation (see the top part of the scheme in Fig. 5.2). In particular, several checks of the structure are done, followed by a calculation of atomic densities by the `lstart` routine which are later used by the `dstart` routine for calculation of the starting electron density ρ_0. The energy separating core states on one side and semi-core and valence states on the other side is required by `lstart` in order to set up basis sets for various orbitals correctly (see section 5.2.1). The value of -8 Ry was used for all calculations in this work as this value minimised "leakage" of the core electrons out of MTS (i.e. prevented states extending beyond MTS from being treated as core rather than semi-core states).

The other important characteristics of the numerical calculation are $\ell_{\max}$, a quantity defining the number of terms being used for an expansion into spherical harmonics, and $K_{\max}$, the magnitude of maximum $\boldsymbol{K}$ directly determining the basis set size. Based on the convergence tests done (see section 7.2 and appendix A), the value of $K_{\max} \approx 2.5$ was used. This corresponds roughly to 85 eV energy cut-off of the equivalent plane wave basis set.

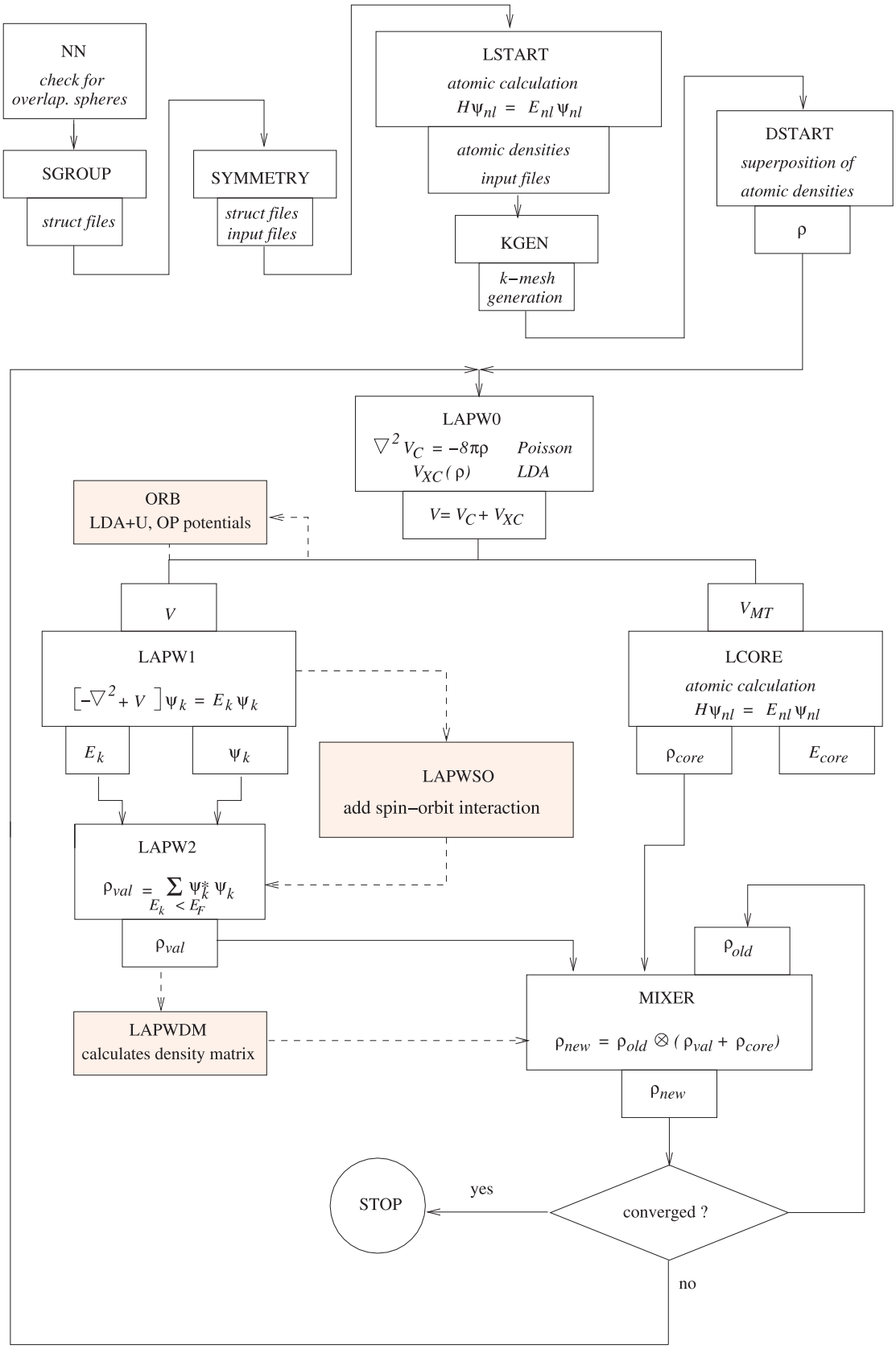


Figure 5.2: A flowchart of a standard `Wien2k` calculation. Taken from the userguide (Blaha et al., 2000).

The last important parameter is the number of *k*-points, i.e. a sampling density of the reciprocal space. All calculations need to be converged with respect to this number; more specific details are given in corresponding sections of following chapters.

After the initialisation, the iterative process as described in section 5.1 proceeds. This is depicted in the main part of the scheme in Fig. 5.2. This scheme represents a general calculation in `Wien2k`. Calculations in this work used only `lapw0` (calculation of potential corresponding to the electron density ρ_{n-1} and its decomposition into spherical harmonics), `lapw1` (diagonalisation of the Hamiltonian), `lapw2` (calculation of the new electron density corresponding to the semi-core and valence states), `lcore` (calculation of the electron density corresponding to the core electrons) and `mixer` (combining the old, valence and core electron densities into a new one).

Convergence criteria used in all present calculations were (i) total energy changes between two subsequent cycles by less than 0.0001 Ry, and (ii) integrated electron charge does not fluctuate by more than 0.001 e.

Once a converged electron density (in terms of energy, electron density and a number of *k*-points) was obtained, analysis of the material properties proceeded. In particular, routines `spaghetti` and `tetra` were used for bandstructure and density of states calculations. These results are (in principle) readily available from outputs of `lapw1` and `lapw2`. An example of such a result for AlN is shown in Fig. 5.3.

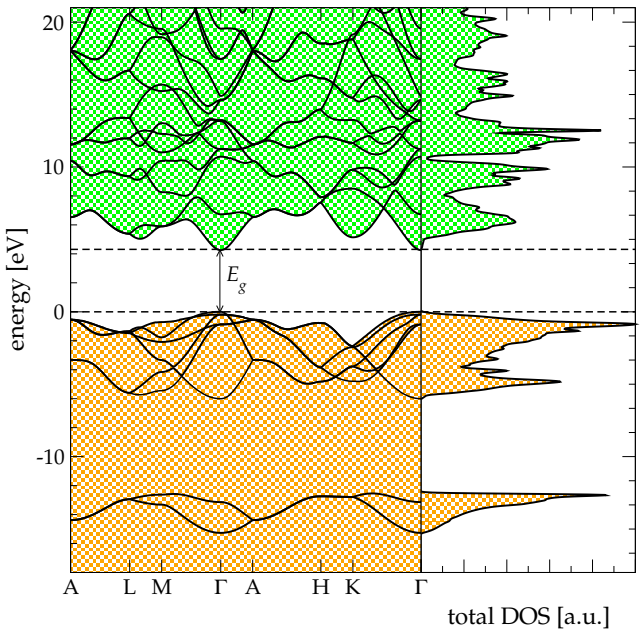


Figure 5.3: A band structure and total density of states of a ground state of AlN. Calculation details are given in section 7.2. Note the underestimated value of the band gap, E_g, (4.2 eV instead of the experimental value of 6.2 eV) which is a common outcome a basic DFT calculations.

The `lapw5` routine was used for calculations of the electron density as presented in section 7.3.4. The majority of the following work concerns electron energy loss near edge structure (ELNES) for which the `Telnes.2` program and the `broadening` routine were used. They are described in greater detail in chapter 6.

All the above mentioned programs and routines are parts of the `Wien2k` package.

5.3 Short introduction to S/PHI/nX

`S/PHI/nX` is a pseudopotential DFT code developed by Boeck et al. (2003) at the Max-Planck Institut für Eisenforschung in Düsseldorf, Germany. The majority of results presented in chapter 9 were obtained using this program and thus in this section a short overview of the basic principles of a pseudopotential method and of this code are given.

5.3.1 Principles of pseudopotential method

The core electrons are tightly bonded to nuclei, as a result, they are localised on a particular atom nucleus and are rapidly oscillating functions (see blue curves in Fig. 5.4). This makes it difficult to construct a basis set which describes both the localised core and extended valence electrons in a satisfactory way, as was discussed in section 5.2.1. However, for material properties associated with bonding, the core electrons are redundant as only the valence electrons participate in bonding. Therefore, the potential of the atomic nucleus and the core electrons is replaced with a *pseudopotential* which (i) generates smooth wave functions and (ii) outside a given radius, r_c reproduces those wave functions and the potential of a full electron system.

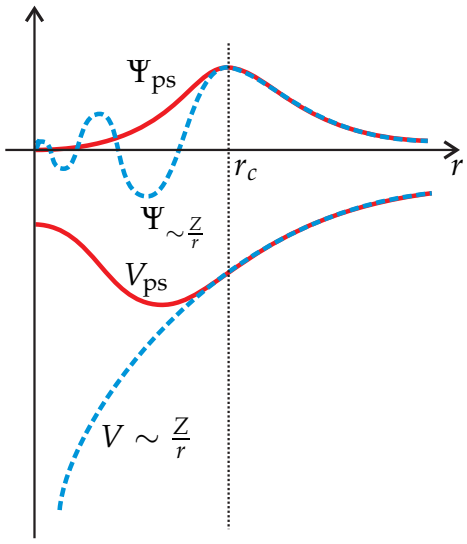


Figure 5.4: Schematic relationship between true (blue dashed lines) and pseudo (red solid lines) wave functions and potential. Perfect agreement is required for $r > r_c$.

Howsoever simple this concept may look, it brings in fact several difficulties. It is not easy to generate a good quality pseudopotential in the first instance. In order to be beneficial, a pseudopotential must demand as small a basis set as possible (*hardness* of the pseudopotential) while keeping the accuracy. Furthermore,

it must describe different bonding configurations accurately (*transferability* of the pseudopotential). It is a piece of art to construct a reliable and high quality pseudopotential (Boeck et al., 2003) and thus it is usually better to rely on well established and tested pseudopotentials from literature rather than generate new ones. This was also the case of calculations presented in chapter 9.

An important feature of the pseudopotential approach is that it works only with valence electrons. Core electrons are included in the pseudopotential. This is usually a great advantage which results in a huge speed-up of the calculations. However, it becomes of a problem for the ELNES calculations (see chapter 6) where matrix elements of the core electron states are needed. Evidently, also the core hole approximation is inaccessible with pseudopotentials. Fortunately, since the fine features in ELNES are governed mainly by a fluctuation of the projected density of states, ground state pseudopotential calculations can be used as rough estimates of what features can be expected in ELNES spectra.

The great advantage of pseudopotentials is that the plane wave basis set is suitable for smooth pseudo wave functions. Plane waves are easy to handle and, because of the smoothness of wave functions, the basis set size can be reasonably small. As was mentioned in section 5.2.1, the diagonalisation of the Hamiltonian is the most time consuming part of a calculation, and thus the relatively small plane wave basis set reduces the calculation time significantly. The plane wave basis set, as implemented in `S/PHI/nX`, is

$$\phi_{\boldsymbol{K}}^{\boldsymbol{k}}(\boldsymbol{r}) = \mathrm{e}^{\mathrm{i}(\boldsymbol{k}+\boldsymbol{K})\cdot\boldsymbol{r}}\,. \tag{5.17}$$

The Kohn-Sham equation (Eq. 5.13) takes a much simpler form in this representation than in the case of (L)APWs (see Boeck et al., 2003).

Another advantage of the plane wave basis set is that they are well suited for calculation of forces on the ions (Boeck et al., 2003). Therefore, the structure relaxation is easier in this case than in the case of `Wien2k` calculations. As a consequence, all `S/PHI/nX` calculations presented in chapter 9 correspond to relaxed structures.

5.3.2 Practicalities of running S/PHI/nX

`S/PHI/nX`, being a newer piece of software, has a different philosophy from the `Wien2k` code. All details about the structure, parameters for initialisation, pseudopotentials,

k-point sampling, etc. are given in a single input file[1] and the calculation is afterwards performed by a single program.

Apart from pseudopotentials, it is mainly the k-point sampling, plane wave cut-off energy E_{cut} (determining the largest $\boldsymbol{K}$ in the basis set) and a minimisation scheme that influence the quality of results. Convergence tests should always be made for the number of k-points and energy cut-off E_{cut}. Each of the minimisation schemes is suitable for a different situation; details about them can be found in the manual (Boeck et al., 2003).

There are several "add-ons" which allow calculation of additional quantities once a converged electron density was obtained. One of them, called `sxpdos`, calculates the density of states projected on different atoms and states. This add-on was used for the calculations in chapter 9.

5.4 Summary

This chapter introduced the basic ideas and concepts of a widely used and well established density functional theory which allows calculations of material properties from first principles. Its advantages as well as drawbacks and limitations were discussed. In the second part of the chapter, basic information about two codes used in this work was given. The majority of the calculations in the chapters 7 and 8 was done using the full-potential all-electron linear augmented plane wave code called `Wien2k`. The studies of strain-dependence presented in chapter 9 were done using the pseudopotential plane wave code `S/PHI/nX`. A few most important practicalities of running calculations with these programs were mentioned with references to particular software manuals.

[1] In fact, it can be several files nested in a single input file and thus making a single unit.

Chapter 6

Electron Energy Loss Near Edge Structure

Electron energy loss spectroscopy (EELS) is an analytical technique performed in a transmission electron microscope in which fast electrons (typically about 100 keV and more) impart fractions of their energy to a thin specimen. There are many energy loss processes corresponding to different interactions of the incident electrons with the specimen.

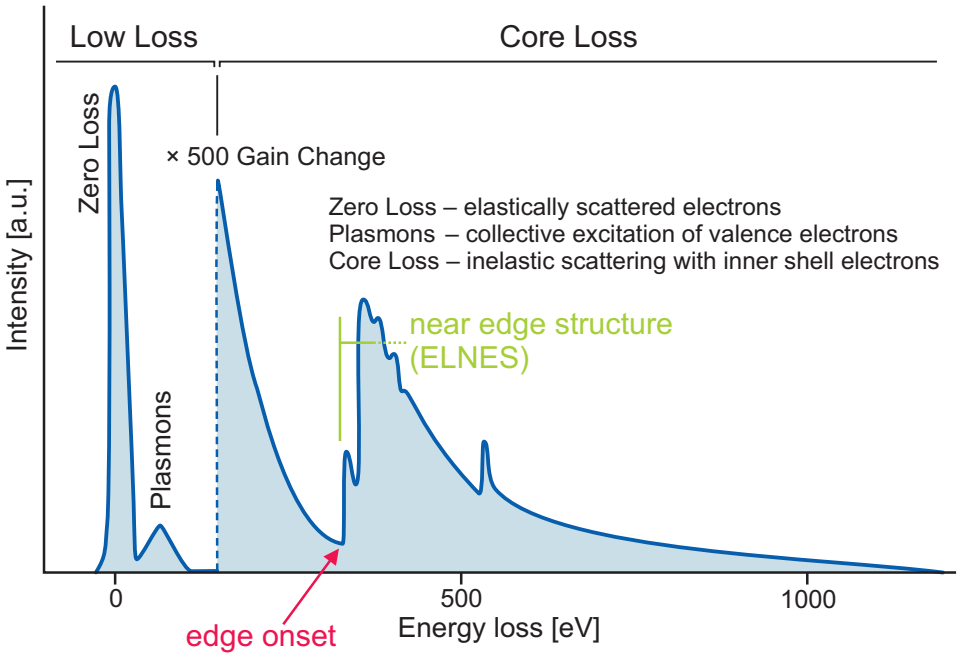


Figure 6.1: A drawing of a typical EELS spectrum.
(Adopted from http://www7430.nrlssc.navy.mil/facilities/emf/gif.htm.)

A typical EELS spectrum is shown in Fig. 6.1. The low-loss region corresponds to energy losses below roughly 50 eV and the dominant feature is the zero-loss peak. The zero loss peak represents electrons that suffered no inelastic scattering apart from phonon scattering (i.e. practically no energy loss). The low-loss spectrum also con-

tains plasmon excitations, i.e. peaks corresponding to electron energy losses due to interaction with collective oscillations of the electron density. The major features in the remainder of the spectrum, denoted as the core loss regime, correspond to characteristic core loss edges related to the allowed electronic excitations between energy levels in an atom. Since each element has a unique electron configuration with specific energy levels, the edge energies are characteristic for each element. The intensity of the ionisation edge is directly related to the amount of element present in the specimen, therefore allowing a quantitative analysis (see e.g. Egerton, 1996).

In addition to the quantitative elemental analysis, the ionisation edges have a small intensity fluctuation just above the edge onset. These fine features are called electron energy loss near edge structure (ELNES). The ELNES has been shown to reflect the local environments of atoms such as coordination or bonding type. Having the advantage of small probe sizes (typically about 1 nm or less) and high energy resolution (in modern microscopes with monochromators, energy resolution about 0.1 eV), this technique becomes an excellent tool for investigating of the electronic structure of materials.

There are some other experimental techniques closely related to EELS and ELNES. X-ray absorption near edge structure (XANES) works on the same principle as ELNES but the core electrons are excited by X-ray instead of incident electrons. As the cross-section (see section 6.1.2) for X-rays is much smaller than that for electrons, XANES is performed on bulk samples using synchrotron radiation. Other techniques are Auger electron spectroscopy (AES), X-ray photoelectron spectroscopy (XPS), both probing the core states. The excited electrons escape the material and are subsequently detected; thus these techniques are surface sensitive. Similar to XPS is ultraviolet photoelectron spectroscopy (UPS) which probes the valence band. A more detailed description with further references can be found in the review by Keast et al. (2001).

This chapter gives a short overview of the theory of ELNES which is essential for understanding of the results presented in chapters 7–9. It is beyond the scope of this work to give an exhaustive description of the theory. An excellent overview of ELNES of both, the theory and experiment, was published by Keast et al. (2001). The classic book of Egerton (1996) concentrates mostly on the practical issues of experimental EELS but also gives a summary of the theory behind it. All ELNES spectra in this work were calculated using the `Wien2k` code (Blaha et al., 2000) described in greater detail in chapter 5. Hébert (2007) gave an excellent review on practicalities of running `Wien2k` for ELNES calculations. This chapter is mostly based on the three above mentioned literature sources.

6.1 Theory of ELNES in short

6.1.1 Basics of the excitation process

During the excitation process, the incident electron interacts with a core electron and imparts a fraction of its energy to the core electron. This enables the core electron to overcome the attraction of the nucleus and, in the case of ELNES, to be excited to an unoccupied state (energy level). This is shown schematically in Fig. 6.2.

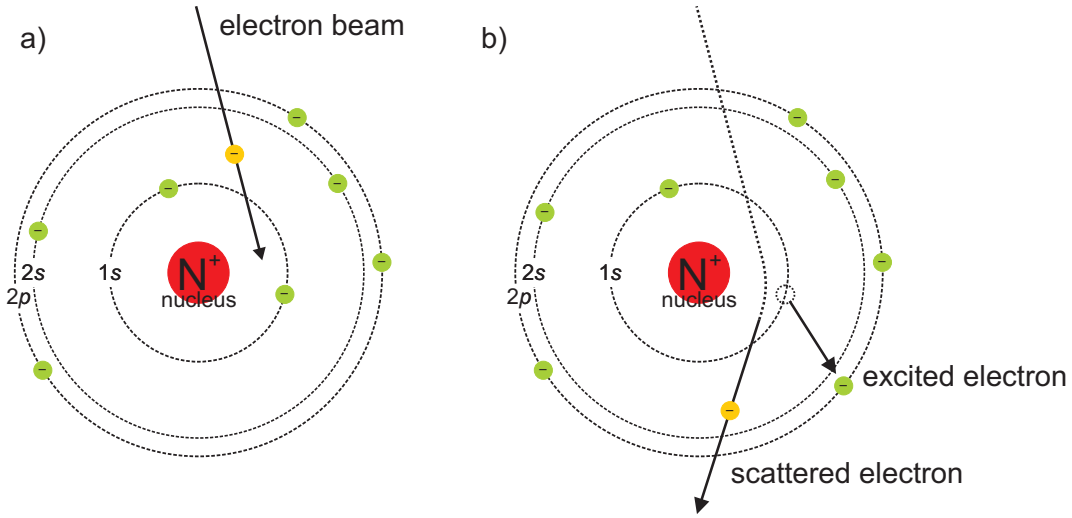


Figure 6.2: Excitation process in EELS: a) incident electron "hits" a particular atom and b) gets scattered while exciting a core electron to previously unoccupied state.

Upon this ionisation, the atom itself is transformed from a ground state configuration to an excited state with an empty state in the core level, called a *core hole*. The possible energies imparted to the excited electron are controlled by a distribution of available empty states. Therefore, fluctuations of the intensity just above the edge onset reflect directly the density of (unoccupied) states.

By a simple calculation, Keast et al. (2001) showed that the typical time scale of excitation is about 10^{-19}–10^{-20} s. Since the lifetime of the excitation is much longer (typically about 10^{-14}–10^{-15} s), Keast et al. (2001) argued that the excited state may be influenced by the presence of the core hole. This is discussed in greater detail in section 6.2.1. On the other hand, the excitation time scale is relatively short compared with the typical oscillation time in crystals of about 10^{-13} s, and thus ion vibrations may be neglected during the excitation. As a result, Born-Oppenheimer approximation holds (see section 5.1.1) which means that the energy levels available for the excited electron are simply those of a fixed atomic geometry. This simplifies both the simulation and interpretation of the ELNES.

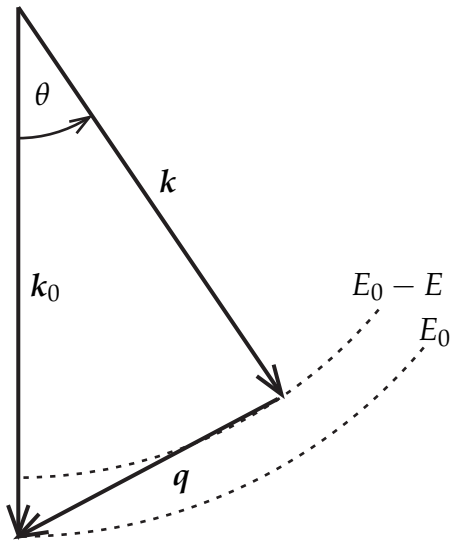


Figure 6.3: Excitation process in EELS: wave vectors and energy levels.

Assume an incident electron of energy E_0 and represented by a plane wave with wave vector $\boldsymbol{k}_0$. During the interaction with a core electron, it loses energy E and transfers momentum $\boldsymbol{q} = \boldsymbol{k}_0 - \boldsymbol{k}$ to the specimen. The scattering angle, θ, and momentum transfer are related as follows

$$q^2 = k_0^2(\theta^2 + \theta_E^2) \tag{6.1}$$

where θ_E is a characteristic scattering angle for the energy loss, E, given by

$$\theta_E = \frac{E}{\gamma m_0 v^2} \,. \tag{6.2}$$

Here, $\gamma = 1/\sqrt{1 - v^2/c^2}$ is the relativistic factor, m_0 is the rest energy of an electron and v is the velocity of an incident electron. For non-relativistic cases, Eq. 6.2 can be simplified to $\theta_E \approx E/2E_0$.

6.1.2 Double differential scattering cross-section

Using Fermi's Golden Rule from time-dependent perturbation theory, Bethe (1930) obtained the double differential scattering cross-section (DDSCS) in a form (see e.g. Hébert, 2007)

$$\frac{\partial^2\sigma}{\partial\Omega\partial E} = \frac{4\gamma^2}{a_0^2 q^4}\frac{k_f}{k_i}\sum_{i,f}|\langle f|\exp(\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r})|i\rangle|^2\,\delta(E - E_f + E_i)\,. \tag{6.3}$$

Here, the sum is done over all empty final, $|f\rangle$, and initial, $|i\rangle$, one-electron states, $a_0 = 0.053$nm is the Bohr radius. DDSCS expresses the probability of scattering an incoming electron with energy loss from $(E, E + \partial E)$ and a momentum transfer $\boldsymbol{q}$ going through the solid angle $\partial\Omega$.

Nelhiebel et al. (1999), Schattschneider et al. (2001) and some other authors pointed out that Eq. 6.3 is strictly correct only for isotropic samples. For the case of anisotropic material, Schattschneider et al. (2005) derived the following formula

$$\frac{\partial^2\sigma}{\partial\Omega\partial E} = \frac{4\gamma^2}{a_0^2 q^4}\frac{1}{(q^2 - q_z^2\beta^2)^2}\frac{k_f}{k_i}\sum_{i,f}\left|\left\langle f\left|\exp(\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r})\left(1 - \frac{\boldsymbol{p}\cdot\boldsymbol{v}_0}{m_0c^2}\right)\right|i\right\rangle\right|^2\delta(E - E_f + E_i) \tag{6.4}$$

where $\boldsymbol{p}$ is the momentum operator and $\boldsymbol{v}_0$ is the velocity of the incoming fast electron. $\boldsymbol{q}_z$ is the momentum transfer projected onto the direction of $\boldsymbol{v}_0$. This can be geometrically interpreted in a following way (Jorissen, 2007): in the dipole limit, a relativistic Hamiltonian shrinks the impulse transfer in the direction of propagation. The general case is much more complex.

The energy losses are small compared with the incident electron energy, therefore the momentum transfer $\boldsymbol{q}$ is also small. The magnitude of $\boldsymbol{q}$ can be reduced experimentally even more by the collection aperture which limits the maximum scattering angle of the collected signal. The assumption of small $\boldsymbol{q}$ is known as the *dipole approximation*. The exponential in Eqs. 6.3 and 6.4 can be expanded as

$$\exp(\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r}) = 1 + \boldsymbol{q}\cdot\boldsymbol{r} - \frac{1}{2!}(\boldsymbol{q}\cdot\boldsymbol{r})^2 + \ldots \tag{6.5}$$

and approximated only by the first two terms of this expansion. The first term drops out due to the orthogonality of the initial and final state ($\langle f|i\rangle = 0$). Therefore, DDSCS under the dipole approximation takes the form of Eqs. 6.3 and 6.4 for the isotropic and anisotropic cases, respectively, with $\boldsymbol{q}\cdot\boldsymbol{r}$ being substituted for $\exp(\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r})$.

6.1.3 ELNES intensity

The ELNES intensity, I, is proportional to the DDSCS. In the latest version of the `Telnes` program[1] (`Telnes.2`) both are implemented, the isotropic and anisotropic version of DDSCS according to Eqs. 6.3 and 6.4. Nelhiebel et al. (1999) showed that the isotropic DDSCS (Eq. 6.4) expresses the average over all directions of $\boldsymbol{q}$. Therefore, these two cases in the subsequent text are referred to as directionally averaged and directionally resolved calculations, respectively.

In either case, there are two important parts in DDSCS formulae: the matrix element $|\ldots|^2$ and the term $\delta(E - E_f + E_i)$ reflecting the joined density of states.

The matrix element expresses the radial overlap of the initial and final wave functions. It varies slowly with energy and determines the basic shape of the edge (Keast et al., 2001), e.g. the saw-tooth shape for K-edges. Moreover, the matrix element is non-zero only for certain combinations of initial and final states. This is known as *dipole selection rules* and in this case it implies that the matrix element is non-zero only when $\Delta l = \pm 1$, where Δl is the change in the angular-momentum quantum number upon excitation. The comonly used nomenclature for EELS edges together with their characteristics are listed in Table 6.1.

[1] A part of the `Wien2k` package, see chapter 5.

Edge	Initial state	Initial state quantum numbers n	l	j	Final state symmetry
K	$1s^{1/2}$	1	0	1/2	p
L_1	$2s^{1/2}$	2	0	1/2	p
L_2	$2p^{1/2}$	2	1	1/2	s or d
L_3	$2p^{3/2}$	2	1	3/2	s or d
M_1	$3s^{1/2}$	3	0	1/2	p
M_2	$3p^{1/2}$	3	1	1/2	s or d
M_3	$3p^{3/2}$	3	1	3/2	s or d
M_4	$3d^{3/2}$	3	2	3/2	p or f
M_5	$3d^{5/2}$	3	2	5/2	p or f

Table 6.1: Nomenclature of EELS and ELNES edges and their characteristics (after Keast et al., 2001).

Superimposed on the basic shape of the edge given by the matrix element is a fine structure due to the joint density of states which is varying much faster with energy than the matrix element. Generally speaking, density of states (DOS) expresses the number of states available at a certain energy and it was discussed more rigorously in chapter 5. Since the initial core state is highly localised on a particular atom, it does not undergo the energy broadening due to interaction with orbitals of other atoms in the solid state matter. As a consequence, the initial energy E_i is has a very small spread or is almost a single value and therefore joined density of states can be approximated by partial density of states (or projected density of states, PDOS) corresponding to the final state. The PDOS is a subset of the total DOS projected on the appropriate angular-momentum symmetry. Consequently, the ELNES is said to measure the PDOS. Moreover, due to very small probe sizes, the ELNES in fact measures PDOS locally.

6.1.4 Broadening of spectra

The energy resolution of ELNES spectra provided by calculations is far beyond what is accessible experimentally. There are several origins for broadening of the theoretical ELNES spectra to the final shape observed in a microscope: the lifetime of the excited state, lifetime of the core hole and instrumental broadening.

The latter two are easier to model and are better understood. Hébert (2007) suggested to model the instrumental broadening by a Gaussian distribution

$$I_{\text{broadened}}(E) = \int \frac{I(E')}{\sigma\sqrt{2\pi}} \exp\left(\frac{(E-E')^2}{2\sigma^2}\right) \mathrm{d}E' \,. \tag{6.6}$$

The full width at half maximum (FWHM), σ, is usually about 0.5–1 eV. With some most recent instruments and special techniques it is possible to go down to about 0.1 eV. The broadening due to core hole lifetime is generally low for edges accessible by EELS (Hébert, 2007). Fuggle and Inglesfield (1992) showed that this can be modelled by a convolution with a Lorentzian profile

$$I_{\text{broadened}}(E) = \int I(E') \frac{\gamma}{\pi\left((E-E')^2+\gamma^2\right)} \mathrm{d}E' \,. \tag{6.7}$$

The broadening parameters, γ, for various elements and edges have been tabulated (see Fuggle and Inglesfield, 1992, and references therein) and are implemented in the `Telnes.2` code.

The broadening due to the excited state lifetime is more complicated. At the edge onset, the broadening is zero and it increases with increasing energy (Hébert, 2007). Several schemes have been proposed in the literature. The `broadening` module[2] implements broadening by a Lorentzian profile with an energy-dependent broadening parameter: a constant broadening independent on energy, an empirical linear function of energy $\gamma(\Delta E) = 0.1\Delta E$ which was proposed by Weijs et al. (1990), a quadratic broadening proposed by Muller et al. (1998) and the Heisenberg's uncertainty principle based formula derived by Moreau et al. (2006). All these formulae constitute approximate estimates and a particular choice influences the resulting theoretical spectrum as demonstrated in Fig. 6.4.

Hébert (2007) concluded that the linear function proposed by Weijs et al. (1990) had proved to be a good compromise and gave reasonable results in most cases. Therefore, if not stated otherwise, the linear excited state lifetime broadening is used for the calculations reported in this text.

6.2 Modelling of ELNES

There are several codes allowing simulation of the ELNES. They can be divided into two groups: multiple scattering methods which work in the real space (for example the `FEFF` program by Ankudinov et al. (1998)) and band structure (BS) based approaches.

[2] A part of the `Wien2k` package, see chapter 5.

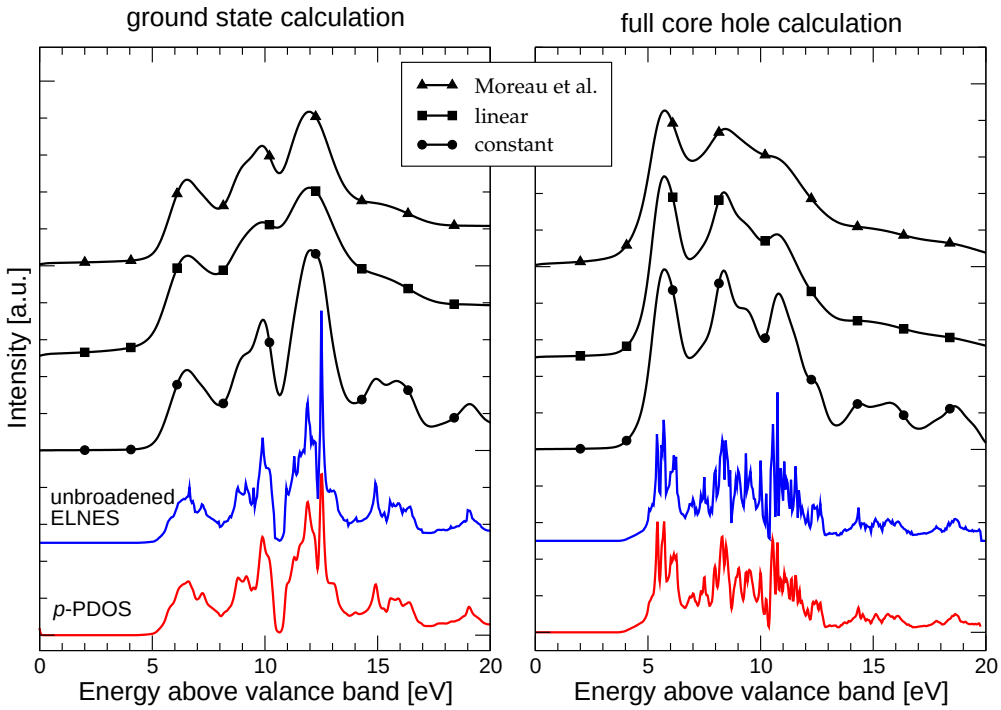


Figure 6.4: Effect of various models for excited state lifetime broadening. The bottom (red) curve is corresponding PDOS, the second bottom (blue) curve is unbroadened ELNES. Calculations correspond to N K-edge of AlN in ground state and with a full core hole of 1 e.

The BS methods are based on density functional theory (DFT) (Hohenberg and Kohn, 1964; Kohn and Sham, 1965, and chapter 5) and are applied in reciprocal space. The advantage of BS-based calculations is that once the charge density is obtained, a wide variety of physical "coherent" properties can be derived. The drawback is that DFT is a *ground state theory* and thus, strictly speaking, calculations of excited states are not guaranteed to work. Nevertheless, the calculation of ELNES with DFT works quite well (Hébert, 2007).

All ELNES spectra in this work were obtained using a full potential linear augmented plane wave DFT code called `Wien2k` (Blaha et al., 2000) which was introduced in greater detail in chapter 5.

6.2.1 The core hole approach

Results published in the literature suggest that better agreement with experimental data can be obtained for simulations using core holes (Hébert et al., 2003; Lazar et al., 2004; Luitz, 2006). In a core hole calculation of N K-edge ELNES, one takes an electron (or fraction of it) from its ground state position at the 1*s* core state of N and puts it in the lowest unoccupied state above the Fermi level or adds it as a background charge.

This can be done very easily in `Wien2k` as it is a full potential code with all electrons (including the core electrons) treated explicitly. Most core hole calculations in the literature take into account only a full core hole (i.e. a whole electron being taken from the core to the unoccupied states), some consider also a partial core hole of 0.5 e (so called *Slater's transition state*). But in theory, one can excite any fraction of an electron in order to create a core hole (Luitz, 2006).

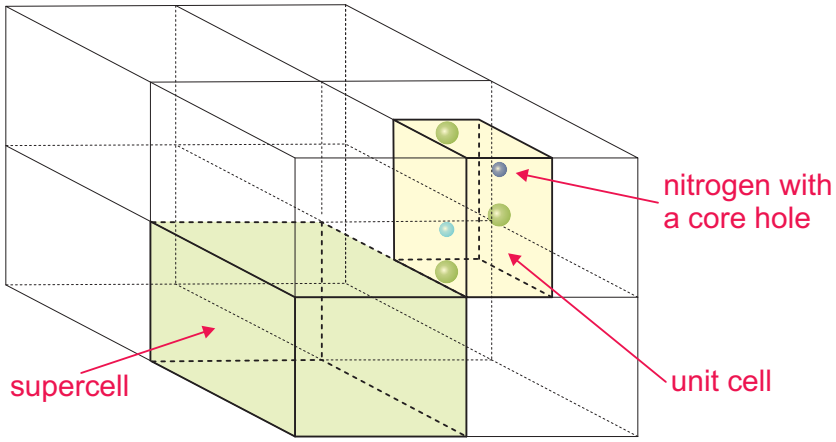


Figure 6.5: A periodic pattern of $2 \times 2 \times 1$ supercells, each containing one unit cell with one core hole.

In the core hole calculations, the basic unit cell (yellow box in Fig. 6.5) is first copied to create a supercell (denoted by the green box). The supercell is described by the number of unit cells in each direction; for example a $2 \times 2 \times 1$ supercell means that there are two unit cells in x- and y-directions and one in the z-direction, in total the supercell contains four unit cells. A core hole is subsequently created on one of the atoms in the supercell. The supercell size should be converged in order to be large enough to screen interactions of core holes in neighbouring supercells.

The reason why core hole calculations give better agreement than ground state calculations is that the presence of a core hole can cause a shift of the energy levels and thus modify the final DOS, similar to what happens in reality. However, one must be aware that the core hole calculations are only another approximation in the ground state DFT, and that there is no guarantee that they describe the physics in a meaningful way. Nevertheless, practise shows that the core hole simulations usually work quite well and therefore, being aware of all the facts above, they can be used in a beneficial way.

Another commonly used approximation is the so called $Z + 1$ approximation (see e.g. Muller, 2006) in which an atom on which the excitation takes place is replaced by its neighbour in the periodic table of elements with the atomic number increased by one (e.g. replace C by N). This approximation is not used in the present work.

6.2.2 ELNES of III-nitrides

The edge onsets for most elements have been tabulated and could be found for example in the book by Egerton (1996). These values are however only approximative. Depending on a concrete local environment of an atom, they can vary by several eVs. The relevant values for III-nitrides are listed in Table 6.2. This work focuses on the N K-edge for two reasons: (i) experimental data for this edge was readily available at the beginning of the study, and (ii) N is present in all binary and ternary alloys investigated here and thus allows a comparison over a variety of compounds.

Element	Edge	Initial stage	Energy [eV]
N	K	$1s^{1/2}$	400
Al	K	$1s^{1/2}$	1560
Al	L_1	$2s^{1/2}$	118
Al	L_3	$2p^{3/2}$	73
Ga	L_1	$2s^{1/2}$	1298
Ga	L_2	$2p^{1/2}$	1142
Ga	L_3	$2p^{3/2}$	1115
In	M_4	$3d^{2/2}$	451
In	M_5	$3d^{5/2}$	443

Table 6.2: Energies of EELS edges available for III-nitrides (after Egerton, 1996).

Chapter 7

Electronic structure of III-nitride binary compounds

7.1 Introduction

Several theoretical studies of ELNES on III-nitrides have been reported on in the literature previously. Lazar et al. (2004) studied theoretically and experimentally the nitrogen K-edge of cubic and hexagonal GaN. They also discussed the effect of a core hole on the spectra. Radtke et al. (2004) studied experimentally the evolution of the N K-edge with aluminium composition in $Al_xGa_{1-x}N$ alloys. Radtke et al. (2003) discussed the anisotropy of the ELNES N K-edge in AlN together with a detailed modelling of the N K-edge using a core hole. Using ground state calculations Keast et al. (2002) gave a detailed experimental and theoretical report of the N K-edge evolution in $In_xGa_{1-x}N$. Mizoguchi et al. (2003, 2004) discussed the N K-edge of AlN, GaN and InN with respect to different crystal structures and orientation. Gao et al. (2004) calculated N K-edges of wurtzite AlN, GaN and InN using pseudopotentials.

Despite the above mentioned high quality papers, a unifying theoretical study needed for understanding of the EELS spectra of III-nitrides is missing. For example, there is lack of a detailed study of the core hole effect on the AlN and InN N K-edge similar to the results of Lazar et al. (2004) for GaN. This is an important starting point for accurate simulations of the ELNES N K-edge of the III-nitride ternary alloy.

This lack of information was a driving force for a systematic study and to build a methodology for simulations of ELNES of III-nitrides. Besides the analysis of ELNES spectra, the relevant electronic structure of III-nitride binary compounds is discussed in detail. This chapter provides the necessary background which is subsequently used in the next chapter for a theoretical study of III-nitride alloys. Some of these results have been published already (see Holec et al., 2008a,b).

7.2 Computational details

The `Wien2k` code (Blaha et al., 2000) was used for all calculations presented in this chapter. The electronic structure (in particular, the site and symmetry projected density of states (PDOS)) of several III–V semiconductor materials was studied within the framework of density functional theory (DFT) (Hohenberg and Kohn, 1964; Kohn and Sham, 1965). The PDOS, which defines the number of electron states per unit energy range on a particular atomic orbital, is an essential input for the `Telnes.2` program to enable the calculation of ELNES spectra.

The lattice parameters of the binary compounds AlN, GaN and InN used in the calculations are summarised in Table 7.1. Wyckhoff 2b atomic positions for the atoms in a wurtzite unit cell were used. In the core hole calculations, this unit cell was copied in order to build a $2 \times 2 \times 2$ supercell, followed by an excitation of one core electron to an unoccupied state of one of the nitrogen atoms.

	a [nm]	c [nm]	u	R_{MT}(N)	R_{MT}(group III)
AlN	0.311	0.498	0.382	1.65	1.85
GaN	0.319	0.519	0.377	1.65	1.92
InN	0.354	0.570	0.380	1.65	2.00

Table 7.1: Lattice constants and muffin-tin radii used in the calculations in this chapter. The values are taken from papers by Dridi et al. (2003) and Wright and Nelson (1995).

The parametrisation of Perdew et al. (1996) was applied to the generalised gradient approximation (GGA) of the exchange-correlation potential. Basis functions were expanded simultaneously as spherical harmonic functions (inside non-overlapping muffin-tin spheres (MTS) centred at atomic sites) and as plane waves in the interstitial region. The values adopted for radii of MTS, R_{MT}, are given in Table 7.1. The l-expansion (azimuthal quantum number) of the non-spherical potential and charge density inside MTS was carried out up to $l_{\mathrm{max}} = 10$. The plane waves were expanded up to a cut-off parameter, K_{max}, fulfilling the relation $R_{\mathrm{MT}} K_{\mathrm{max}} = 5$ where R_{MT} is the average radius of MTS. The ground state calculations were repeated also for a higher cut-off parameter with $R_{\mathrm{MT}} K_{\mathrm{max}} = 6$. Since no significant difference between these two cases was found (see appendix A), the lower value was used for the rest of the calculations since it saves computational time.

The self-consistent iteration process was repeated until the charge density and the total energy were converged within the precision of 0.001 e and 0.0001 Ry, respectively.

Subsequently, the convergence in terms of the density of states (DOS) was sought by increasing the number of *k*-points (describing the discretisation of reciprocal space) as suggested by Hébert (2007). Using these convergence criteria it was noticed that the calculations were usually well converged for about 5 000–10 000 *k*-points in the whole first Brillouin zone (translating to about 528–980 *k*-points, respectively, in the irreducible wedge of the first Brillouin zone). A proportionally smaller number of the *k*-points was used for the supercell calculations. The convergence results are shown in appendix A.

It is worth mentioning that such a high number of *k*-points is not necessary if one is interested *only* in broadened ELNES. The very fine features in the (P)DOS as well as in the unbroadened ELNES obtained by this convergence test are experimentally inaccessible. In the theoretical spectra they are smeared by the broadening process and one finds the same broadened ELNES spectra with approximately one order lower number of *k*-points. However, the unbroadened ELNES as well as the (P)DOS are, of course, affected.

If not stated otherwise, the theoretical ELNES in this chapter were calculated employing following parameters: electron beam energy 200 keV, convergence semi-angle of 1.0 mrad, collection semi-angle of 2.5 mrad. A value of 0.6 eV was used for spectrometer broadening of the calculated spectra.

7.3 Discussion of results

7.3.1 ELNES N K-edge of AlN, GaN and InN

In Fig. 7.1 are shown variations of the N K-edge of AlN, GaN and InN with the core hole charge increasing from 0 e (ground state) to 1 e (full core hole). An immediate observation is that increasing the core hole charge raises significantly the intensity of the first peak (marked "A"). At the same time, intensity of the third peak (marked "C") decreases. This can be intuitively understood in terms of decreasing shielding of the nucleus, and thus increased attraction of the unoccupied electronic states, which results in their shift towards the lower energies. This effect is discussed in section 7.3.2 in greater detail.

In order to estimate which core hole charge gives the best agreement with experimental data and therefore what core hole charge to use in subsequent time-expensive simulations of alloys, the peak positions and intensities of the simulated data of GaN and AlN (Fig. 7.1) were compared with several experimental spectra. In particular, results

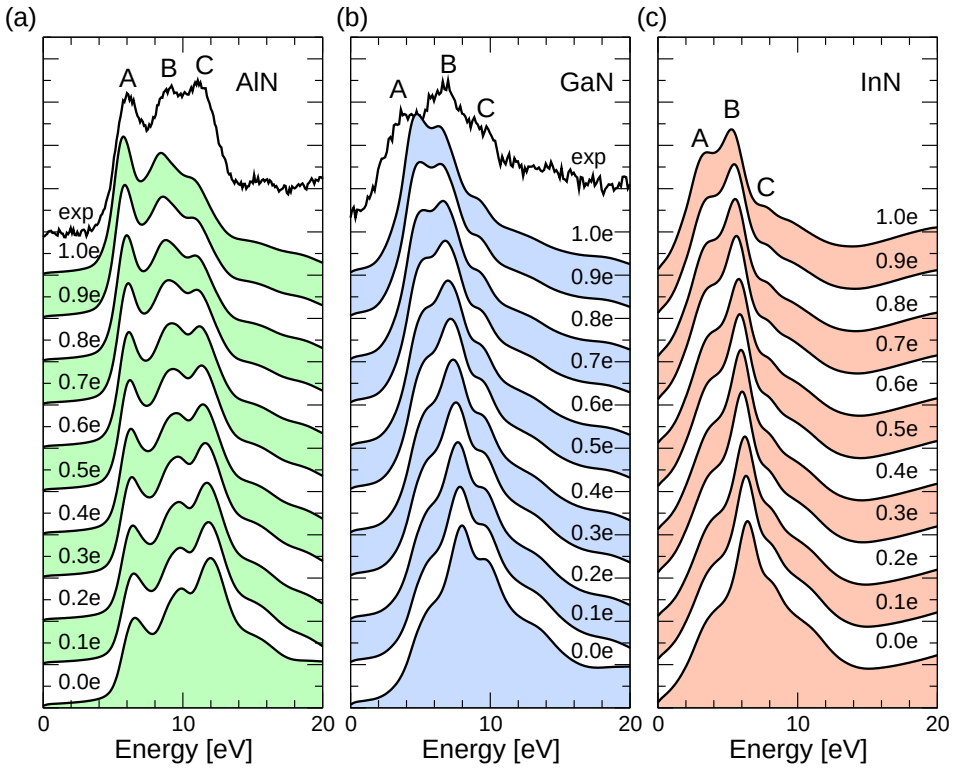


Figure 7.1: Nitrogen K-edge of a) AlN, b) GaN, and c) InN as calculated in the ground state (0.0 e), using partial (0.1–0.9 e) and full (1.0 e) core hole approach. Experimental N K-edge of AlN and GaN is also shown (after Cherns, 2007).

of Radtke et al. (2004), Lazar et al. (2004) (only GaN), Cherns (2007) and Holec et al. (2008a) were considered. The results discussed below are summarised in a graphical form in Fig. 7.2 and 7.3; the data itself can be found in Table B.1.

Two criteria were used. The first one concerned relative intensities of the three peaks "A"–"C". GaN spectra of Radtke et al. (2004) and Cherns (2007) agree very well; opposite trends are observed for the monochromated results given by Lazar et al. (2004) and Holec et al. (2008a). Despite the scatter of results (see Fig. 7.2a) which is caused by a data noise and inability to locate the peak positions precisely (especially in the case of the first and third peaks), it can be concluded that the best agreement in terms of intensity ratios in the majority of cases for GaN is obtained for the core hole charge of 0.4–0.5 e. The second criterion considered the peak positions (absolute distances between peaks). The scatter of the results is much worse in this case (see Fig. 7.3a), and a simple comparison of peak distances for experimental and simulated data is not conclusive at all. However, there might have been some misalignment in the calibration of experimental set up which resulted in a differently "stretched" energy axis in various experiments. Thus, ratios of distances between the three main peaks were calculated. After that, almost a perfect agreement for all experimental results was obtained, pointing to the optimal core hole charge 0.4 e for GaN (see Table B.1).

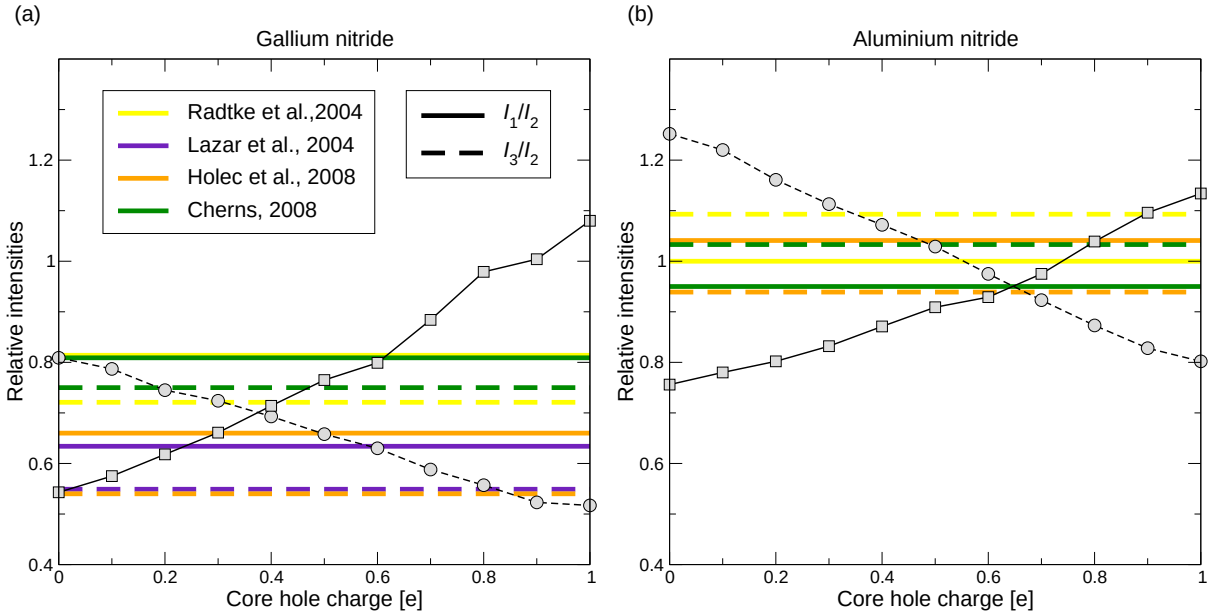


Figure 7.2: Comparison of the relative peak intensities of the calculated N K-edge ELNES spectra (Fig. 7.1) with experimental spectra from the liteature for a) GaN and b) AlN. The main three peaks were considered. Solid line corresponds to the I_1/I_2 ratio, dashed line to the I_3/I_2 ratio.

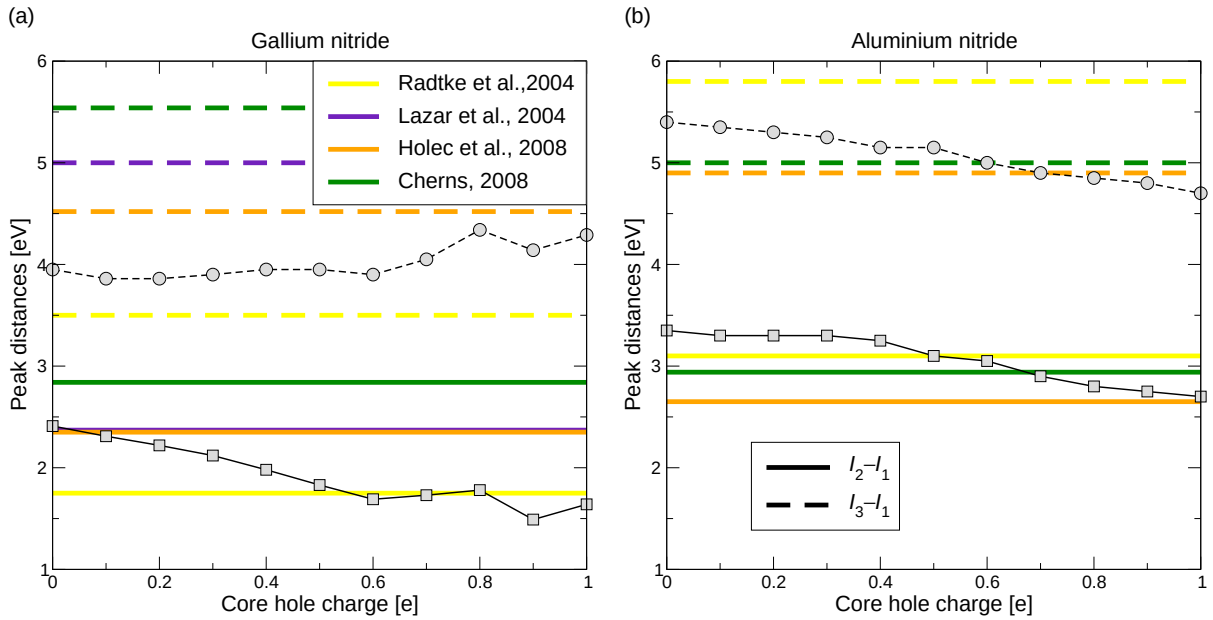


Figure 7.3: Comparison of the relative peak positions of the calculated N K-edge ELNES spectra (Fig. 7.1) with experimental spectra from the liteature for a) GaN and b) AlN. The main three peaks were considered. Solid line corresponds to the $I_2 - I_1$ distance, dashed line to the $I_3 - I_1$ distance.

The same procedure was repeated for AlN. The scatter of data is smaller in this case (see Fig. 7.2b and 7.3b) due to better defined peaks and a smaller signal-to-noise ratio. This allows more precise location of the peaks. The intensity-based comparison favours a core hole charge of about 0.7 e. In the case of comparison based on peak positions, 0.7 e is obtained again as the best value. On the other hand, the peak distance ratios give somewhat higher values (even over 1 e; see Table B.2).

As has been mentioned before, the comparison with experimental data is not absolutely conclusive since there are big variations between individual experimental spectra. These may be attributed to the different experimental conditions or samples might be of different quality. The comparisons favour partial core holes rather than ground state (0.0 e) or full core hole (1.0 e) calculations. Moreover, GaN tends to favour partial core hole charge just below 0.5 e, whereas in the case of AlN, core hole charge above 0.5 e seems to be the best. The same analysis for InN and experimental spectra of Gao et al. (2004) and Mizoguchi et al. (2004) favours 0.4 e core hole charge. Therefore, the partial core hole of 0.5 e is used for all calculations of alloys in chapter 8.

7.3.2 *p*-PDOS of nitrogen atom in AlN

This section attempts to rationalise observations from the previous section 7.3.1 by a detailed investigation of the density of states available to a core electron excited to the conduction band state. In particular, AlN is discussed here. The study focuses on the N K-edge; therefore, the corresponding final states are unoccupied p-states in the conduction band of nitrogen atoms. Moreover, in the case of (partial) core hole calculations, the site projected DOS on nitrogen with the core hole is needed.

Figure 7.4 shows p-PDOS, $(p_x + p_y)$-PDOS and p_z-PDOS as obtained for the ground state, partial core hole (0.5 e) and full core hole calculations. In correspondence with the ELNES spectra the energy range is divided into four regions: the first ELNES peak ($5\,\text{eV} \lessapprox E - E_F \lessapprox 8\,\text{eV}$) covering the feature "A" in Fig. 7.4, the second ELNES peak ($8\,\text{eV} \lessapprox E - E_F \lessapprox 10.5\,\text{eV}$) covering the features "B" and "C" in Fig. 7.4, the third ELNES peak ($10.5\,\text{eV} \lessapprox E - E_F \lessapprox 13.5\,\text{eV}$) covering the features "D" and "E" in Fig. 7.4, and the tail ($E - E_F \gtrapprox 14\,\text{eV}$) covering the feature "F" in Fig. 7.4.

The shape and intensity of the first ELNES peak is clearly given by the number of unoccupied $p_x + p_y$ states; the intensity of the peak increases with increasing charge of the core hole (see Fig. 7.4a and 7.4b). p_z-PDOS provides a broad basis for this peak and it is almost independent of the core hole charge.

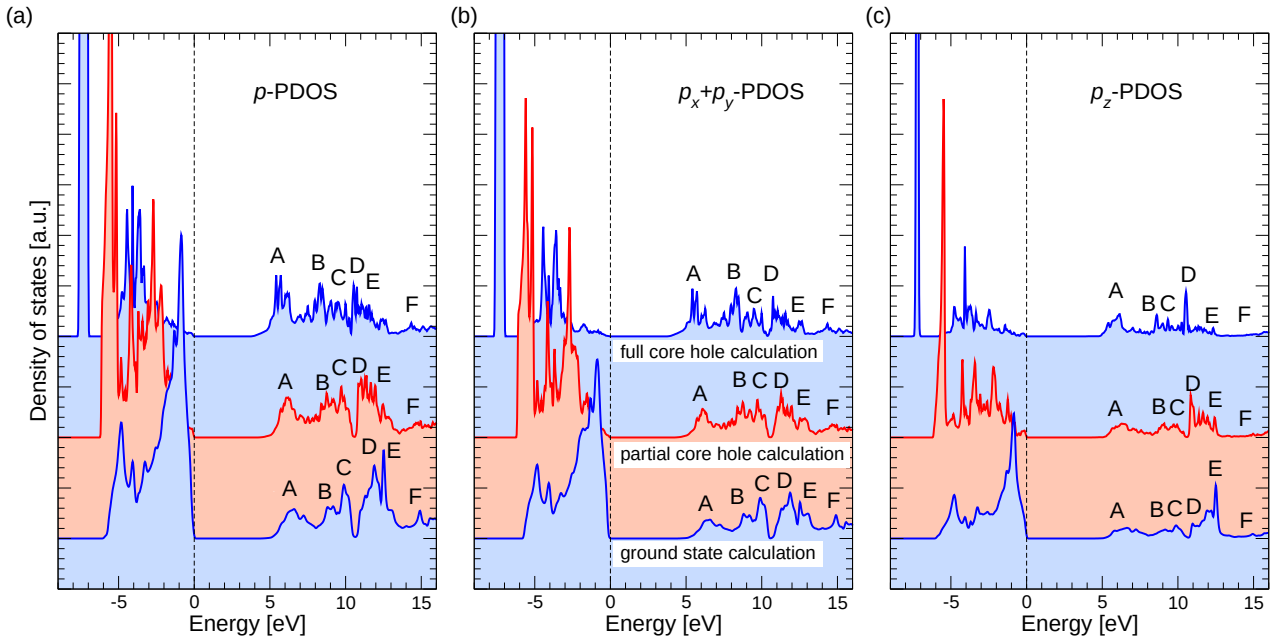


Figure 7.4: Density of final states for electrons contributing to the N K-edge: a) the total p-PDOS, the PDOS projected on b) $p_x + p_y$ states and c) p_z states. The bottom spectra correspond to the ground state calculation, the middle ones to the partial core hole of 0.5 e and the top ones to the full core hole case. Important features discussed in the text are labelled "A"–"F".

The second peak has the form of a twin-peak (features "B" and "C" in Fig. 7.4). The peak "B" is weak in the ground state calculation and thus the overall shape of the second peak is dictated by the peak "C". However, with the increasing core hole charge, the peak "B" intensifies at the expense of the peak "C". As a consequence, the second peak moves to lower energies with increasing core hole charge as can be observed in Fig. 7.1. Once again, the main contribution in this region comes from $(p_x + p_y)$-PDOS (compare Fig. 7.4a with 7.4b). The two peaks "B" and "C" can be also recognised in the p_z-PDOS (see Fig. 7.4c) with similar trends in their respective heights as observed for $(p_x + p_y)$-PDOS, though their peak values are much smaller than the corresponding maxima for $(p_x + p_y)$-PDOS.

The third peak exhibits very similar tendency to the second one: it consists of two very close peaks (the features "D" and "E"). Their maxima are almost the same in the ground state case of $(p_x + p_y)$-PDOS; with increasing core hole charge the peak "E" decreases and becomes only a very weak shoulder on the high-energy side of the third peak. On the other hand, the p_z-PDOS constitutes in this case the major contribution to p-PDOS and, in the ground state case, completely rules over the $(p_x + p_y)$-PDOS in the "E" peak.

Finally, the tail region with the feature "F" seems to be independent of the core hole charge. Its low-energy side is contributed by the $p_x + p_y$ states whereas the high-

energy side comes from p_z states. Both contributions are individually independent of the core hole charge.

The general observation that the peaks "move" towards lower energy with the higher core hole charge (or that in the case of the twin-peaks the low-energy peak intensifies whereas the high-energy peak weakens) can be intuitively understood in terms of shielding. With the increasing core hole charge (from zero to one electron) the shield of the core electrons weakens and thus the originally unoccupied states in the conduction band are bonded more strongly to the atom nucleus. This results in their shift to lower energies.

An interesting observation can also be made by looking at the valence p-PDOS, i.e. just below the 0 eV energy. The integrated value of the ground state p-PDOS from -6.5 eV to 0 eV gives a value of 2.89; $(p_x + p_y)$-PDOS in the same region yields 1.93 whereas p_z-PDOS gives 0.96. This agrees well with the fact that there are three valence p electrons provided by nitrogen atoms[1]. The reason why one does not get the exact number 3 as might be expected is that the concept of "an electron belonging to a particular atom" does not make sense in the case of a crystalline solid anymore: some electrons (in particular, the valence electrons) become delocalised. Therefore, so called *interstitial DOS* is introducted as the difference between the total DOS and that part of the charge that can be unambiguously associated with particular atoms and states. As a consequence PDOS does not give absolute numbers but rather how much of some electronic state can be identified with a particular atom.

In the full core hole case, the total p-PDOS significantly changes its character. There is a sharp narrow peak between -8 eV and -7 eV and a broad spectrum between -5.5 eV and 0 eV. Similar behaviour is observed also in the $(p_x + p_y)$-PDOS and p_z-PDOS. A detailed analysis reveals that these sharp peaks correspond to the valence $2p$ states: the integrated value of the total p-PDOS peak is 3.08 with 2.10 of $(p_x + p_y)$-character and 0.97 of p_z-character. The "tail" between -5.5 eV and 0 eV constitutes approximately one electron state (0.84). This is (in the calculation) filled by "leaking" electrons from the neighbouring atoms to the nitrogen with the core hole. The partial core hole of 0.5 e lies somewhere in between the ground state and the full core hole case.

The observation of the valence states shifting to lower energies is consistent with the concept of decreasing screening effect with increasing core hole charge. Moreover, a profound narrowing of the band corresponding to the valence p electrons can be observed.

[1] Electron configuration of nitrogen is $1s^2\,2s^2\,2p^3$.

7.3.3 Origins of the N K-edge variations of III-nitrides

Figure 7.1 shows very similar behaviour (reproduced in Fig. 7.5) of InN and GaN in terms of the shape of the N K-edge ELNES spectrum: both have a very strong central peak (labelled "B" in the figure) surrounded with "shoulders" on both sides (labelled "A" and "C"). In contrast, the ground state of AlN shows three distinct peaks "A", "B", and "C". This section addresses these variations in terms of PDOS.

Density of unoccupied states projected on both atoms (N and either Al, or Ga, or In) and *s*, *p* and *d* states for the ground state calculations is shown in Fig. 7.5. Nitrogen *p* states, the final states of N K-edge ELNES excitation, are shown in purple colour. The three peaks of AlN ELNES N K-edge can be clearly identified in Fig. 7.5a, including their increasing intensity with higher energy losses. Similarly, one very sharp peak at ≈7–8 eV and at ≈6–7 eV can be seen in Fig. 7.5b and 7.5c, respectively.

In all three cases, a strong correlation between nitrogen *p* states on one hand and group III element *s* and *p* states on the other hand is clearly seen in the region of several eVs above the valence band edge (0 eV). This suggests quite a strong interaction between N *p* states and Al/Ga/In *s* and *p* states. Since valence bands of the individual elements are only partially filled, these unoccupied states are solid-states analogues to molecular anti-bonding or-

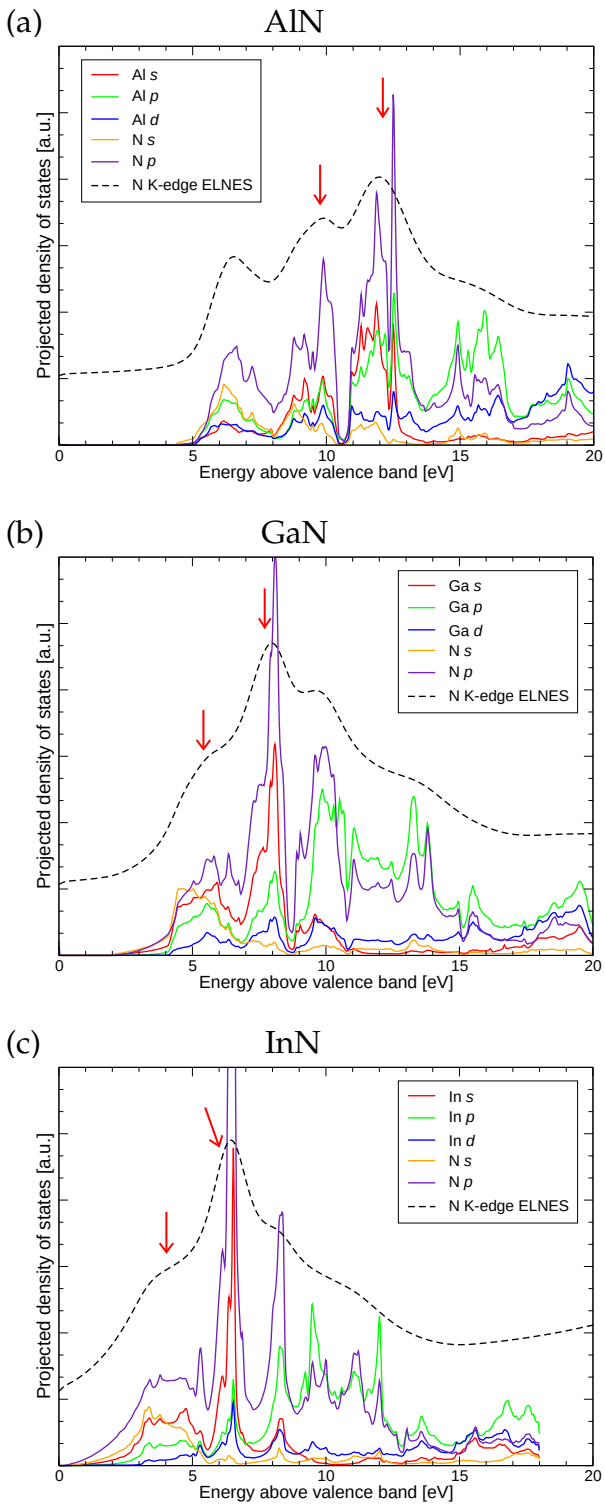


Figure 7.5: Projected density of states of a) AlN, b) GaN, and c) InN. 0 eV on the energy axis corresponds to the top of the valence band.

bitals. One can see that group III element p states to N p states interactions are particularly strong in all three ELNES peaks as well as in the tail region at about 4 eV above the third peak. The striking difference between AlN on one side and GaN and InN on the other side is the interaction of N p states with Al/Ga/In s states. These contribute mainly to the second and third peaks for AlN but to the first and second peaks in the cases of GaN and InN (emphasised by red arrows in Fig. 7.5). The different interaction between the unoccupied states modifies the profile of N p states which is reflected in the N K-edge ELNES.

7.3.4 Bonding of III-nitrides

A significant difference between AlN on the one side and GaN and InN on the other side is the electronic configuration of Al, Ga and In. While Al has in the valence shell[2] only three electrons (having a configuration $3s^2\,3p^1$), Ga and In have thirteen electrons in their valence shells (having respective configurations $3d^{10}\,4s^2\,4p^1$ and $4d^{10}\,5s^2\,5p^1$). Differences in bonding are thus expected; this section focuses AlN and GaN.

Table 7.2 gives energy levels of valence states[3]. Considering that the necessary condition for a formation of a bond is the overlap of atomic energy levels of participating states, bonding of Al $3p$ and $3s$ with N $2p$ states is predicted for AlN. In the case of GaN, additional bonds may be created between Ga $3d$ and Al $2s$ states. As a consequence, the GaN bond is predicted to be stronger than the AlN bond in the sense that more electrons per atom contribute to the bonding.

Al	N	Ga
-3.01 eV $3p$		$4p$ -2.91 eV
-7.28 eV, -8.04 eV } $3s$	$2p$ { -8.29 eV, -8.31 eV } $2p$	$4s$ { -8.48 eV, -9.17 eV
	-15.31 eV, -15.52 eV } $2s$	$3d$ { -18.79 eV, -18.86 eV, -19.25 eV, -19.31 eV

Table 7.2: Atomic energy levels of valence orbitals of Al, N and Ga.

[2] What is here called the valence electrons are, in fact, valence and semi-core electrons in the terminology of the Wien2k (see discussion in section 5.2.1): these occupy shallow energy levels ($\lessapprox 20$ eV below the top of the valence band). As a consequence, the term valence electrons in this text is not limited to only the unfilled shells.

[3] Occupied states with the highest energy. Below them is usually an energy gap of several tens of eVs.

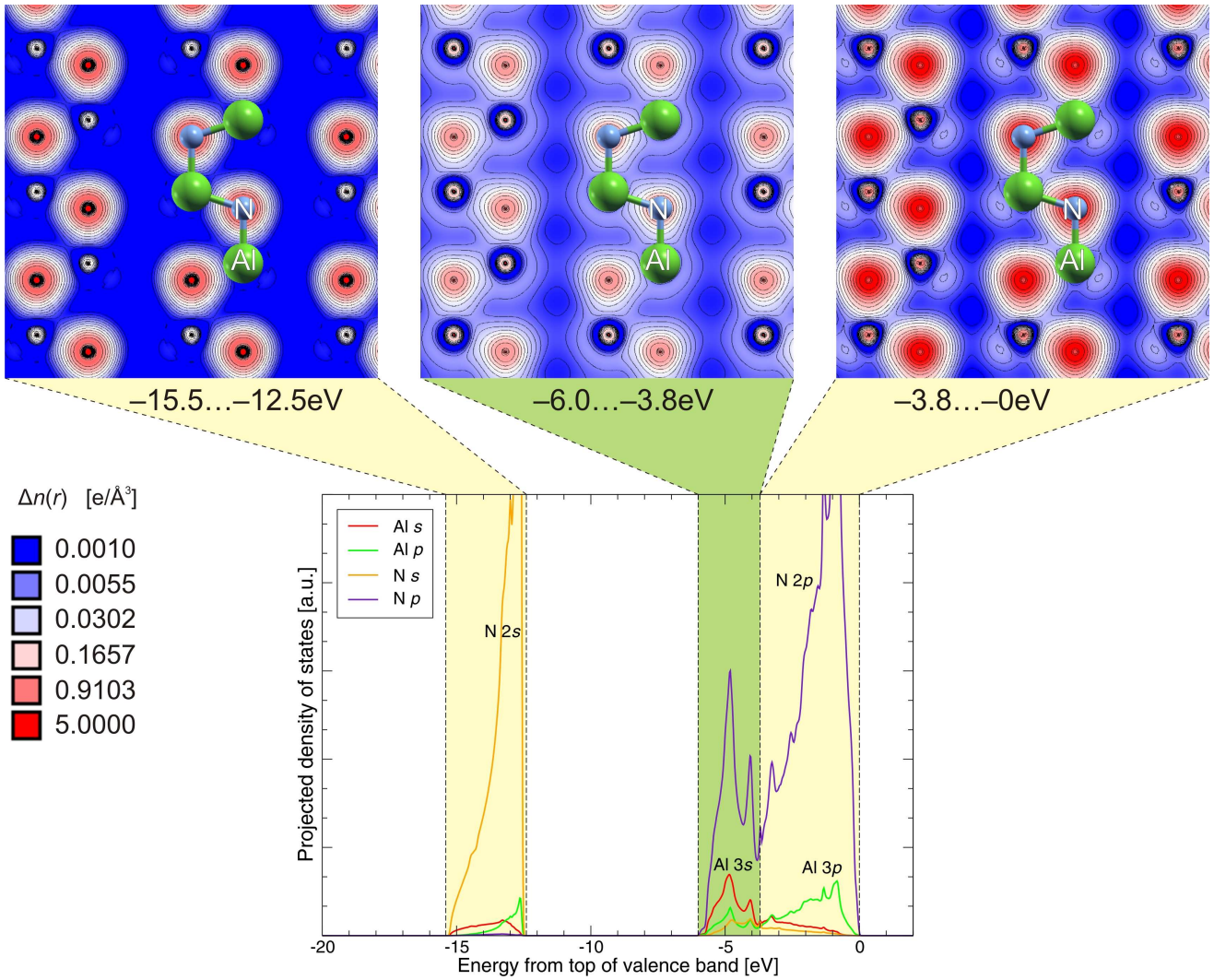


Figure 7.6: Projected density of states of AlN corresponding to the valence band. Electron density maps of the valence electrons for several energy windows as discussed in the text are also shown. Maps are plotted on the $\{11\bar{2}0\}$ planes.

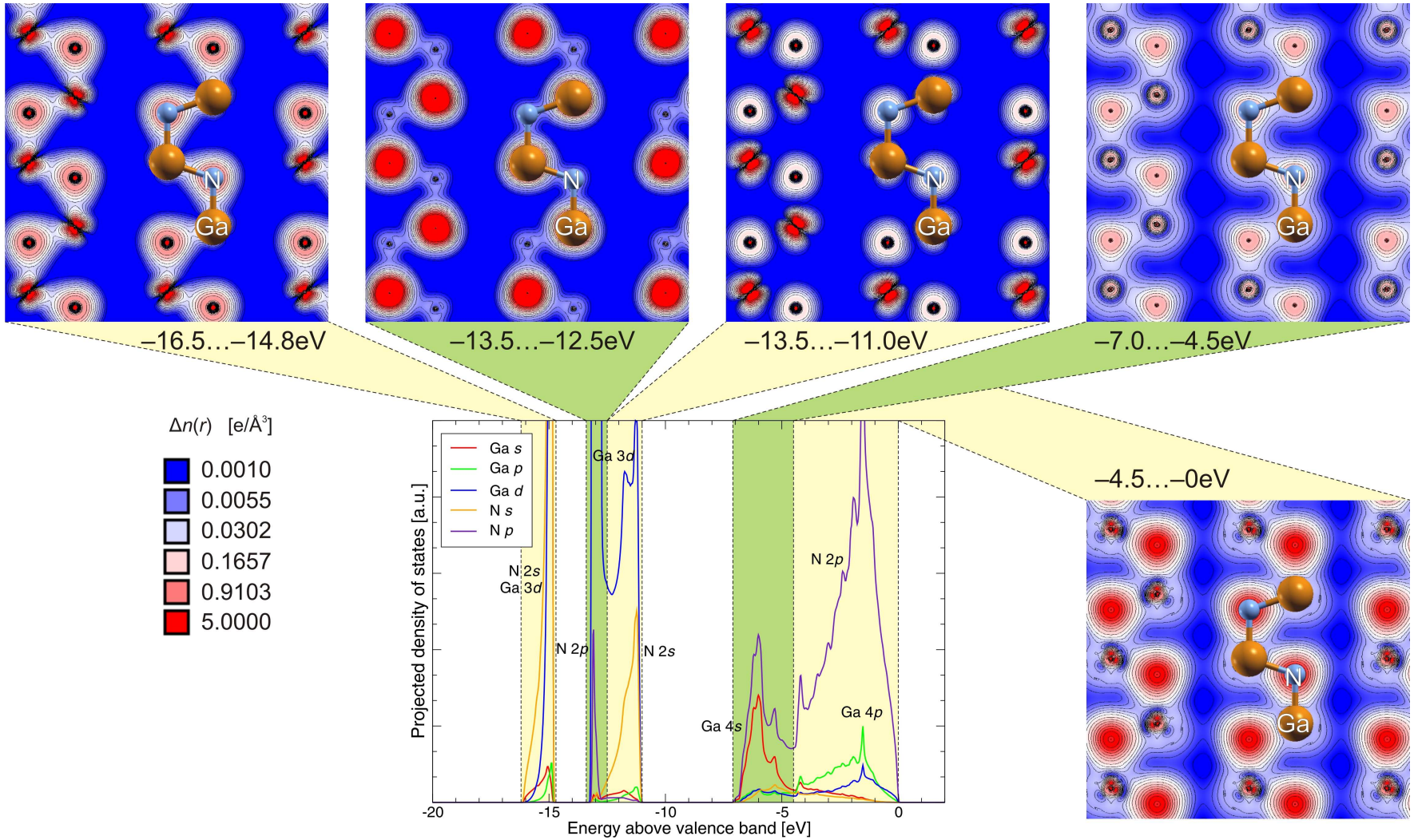


Figure 7.7: Projected density of states of GaN corresponding to the valence band. Electron density maps of the valence electrons for several energy windows as discussed in the text are also shown. Maps are plotted on the $\{11\bar{2}0\}$ planes.

Corresponding PDOSes together with electron density maps representing spatial distribution of valence electrons in various energy windows are shown in Fig. 7.6 for AlN and in Fig. 7.7 for GaN.

N 2*s* states make up a band at about −15.5 eV to −12.5 eV in AlN (Fig. 7.6). A very small contribution of Al 3*s* and 3*p* states is present in this range as well. However, the interaction is too weak to contribute significantly to bonding. This is also illustrated in the corresponding energy density map where contours on the N sites remain almost perfectly circular (corresponding to the spherical symmetry of the *s* states). On the contrary, in the case of the higher energy windows at −6.0 eV to −3.8 eV (N 2*p*–Al 3*s* interaction) and −3.8 eV to 0 eV (N 2*p*–Al 3*p* interaction), an accumulation of charge in between two neighbouring atoms is observed.

A similar analysis for GaN unveils that N 2*s* participate in bonding in GaN by interacting with Ga 3*d* states (Fig. 7.7). This bond has been discussed by Dudešek et al. (1998); the present results are in good agreement with their results. Deepest in the energy (the window from −16.5 eV to −14.8 eV) lie bonding N 2*s*–Ga 3*d* states. These are followed by non-bonding Ga 3*d* states between −13.5 eV and −12.5 eV. The N 2*p* states forming the peak seen in PDOS in this energy window do not interact strongly with Ga 3*d* states and thus does not contribute to bonding. The next energy window at −13.5 eV to −11.0 eV corresponds to N 2*s*–Ga 3*d* anti-bonding states as suggested by Dudešek et al. (1998). The different distribution of electrons in the bonding and anti-bonding states is clearly demonstrated by the electron density maps in Fig. 7.7. The structure of the top of valence band of GaN is similar to that of AlN apart from a much more complex structure of the electron density on the Ga sites in the energy window from −3.8 eV to 0 eV. The PDOS shows that in this region are present both Ga 4*p* and 3*d* states possibly suggesting some form of *pd* hybridisation as discussed by Mizoguchi et al. (2003).

Differences in electron density distribution in AlN, GaN and InN are presented in Fig. 7.8 where difference maps between the crystal valence electron density and the superimposed atomic valence densities are shown. Red regions correspond to the electron accumulation in a crystal with respect to the atomic electron distribution while blue regions correspond to the electron depletion in a crystal. In white regions, no electron redistribution occurs. Electron accumulation between two neighbouring atomic sites can be clearly observed in all three cases (AlN, GaN and InN) reflecting the covalent character of the bonds. At the same time, electrons accumulate closer to the N than to the group III element sites suggesting also a partly ionic character of the bonds. The complex shape of Ga orbitals results in a directionally better interaction than in AlN.

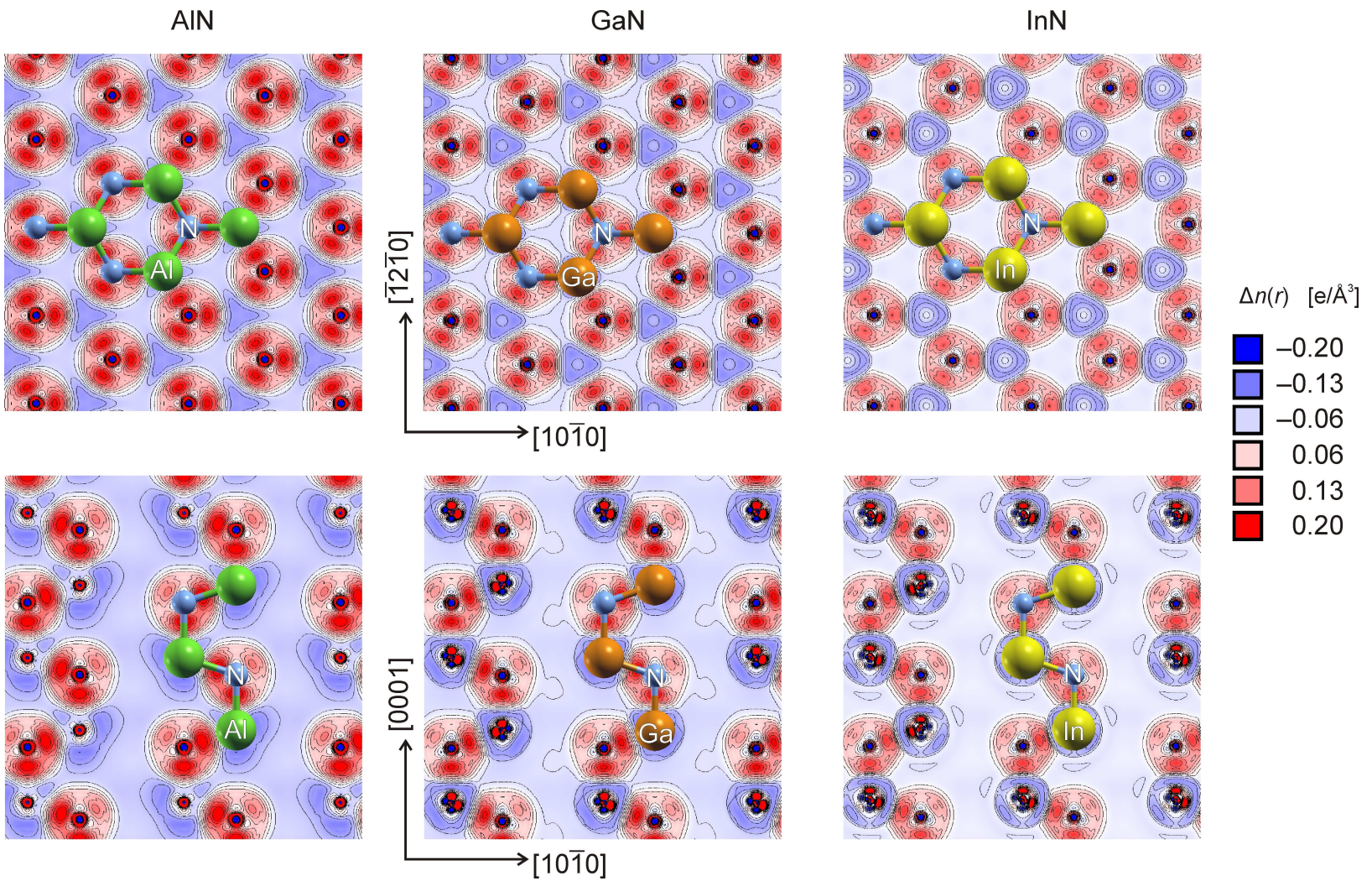


Figure 7.8: Difference maps of AlN, GaN and InN on two different planes. The difference maps show variations between the crystal and the superimposed atomic valence electron densities

The difference maps also show that InN has very similar behaviour to GaN (showing the same differences from AlN). This is consistent with the predicted similarities between GaN and InN, such as previously shown for PDOS or N K-edge.

To conclude this section, the most striking difference between AlN on the one hand and GaN and InN on the other hand is the presence of the valence *d* states in GaN and InN which interact with the N 2*s* states. As a consequence, more electrons contribute to the bonds in GaN and InN than in AlN; the bonding in GaN and InN is also more directional than in AlN.

7.3.5 Directionally resolved N K-edge ELNES of AlN

An improved relativistic model for ELNES simulations has been implemented in the recent version of the `Telnes` program (`Telnes.2`). This brings the possibility of calculating correctly the directionally resolved ELNES, e.g. ELNES depending on the mutual orientation of a single-crystal sample and the electron beam direction. If changes of major features in the spectra with different electron beam–sample misorientations are big enough, one would, in principle, be able to use ELNES for determining the crystal orientation of a sample.

Fig. 7.9 shows N K-edge ELNES spectra of AlN corresponding to three different samples: *m*-plane and *a*-plane (cross-sectional) samples (i.e. with the incident electron beam oriented along the hexagonal *m*-axis ($\langle 1\bar{1}00\rangle$ direction) and *a*-axis ($\langle 11\bar{2}0\rangle$ direction), respectively, and a *c*-plane (plan view) sample with the electron beam along the $\langle 0001\rangle$ direction. Corresponding directionally averaged ELNES spectra are also given in Fig. 7.9. All spectra were calculated from the AlN full core hole using constant broadening. It is worth noting that here are discussed the total ELNES spectra as they would be obtained for different sample-to-beam orientations. The decomposition of the signal into the "parallel" and "perpendicular" components (Schattschneider et al., 2001) corresponding to p_z and $p_x + p_y$ states, respectively, for the *c*-plane samples was done by Radtke et al. (2003) for AlN.

The variations between individual spectra are not big, however a closer inspection reveals some differences. There are significant variations between the orientation resolved spectra for the small collection semi-angle $\beta = 1.0\,\mathrm{mrad}$: the second peak is significantly higher than the two others for the *m*-plane sample; a difference in intensities of the second and third peak is notably smaller for the *c*-plane sample than for the other two samples. On the contrary, for $\beta = 2.5\,\mathrm{mrad}$ all three directionally resolved spectra are the same as the directionally averaged one. This suggests that $\beta = 2.5\,\mathrm{mrad}$

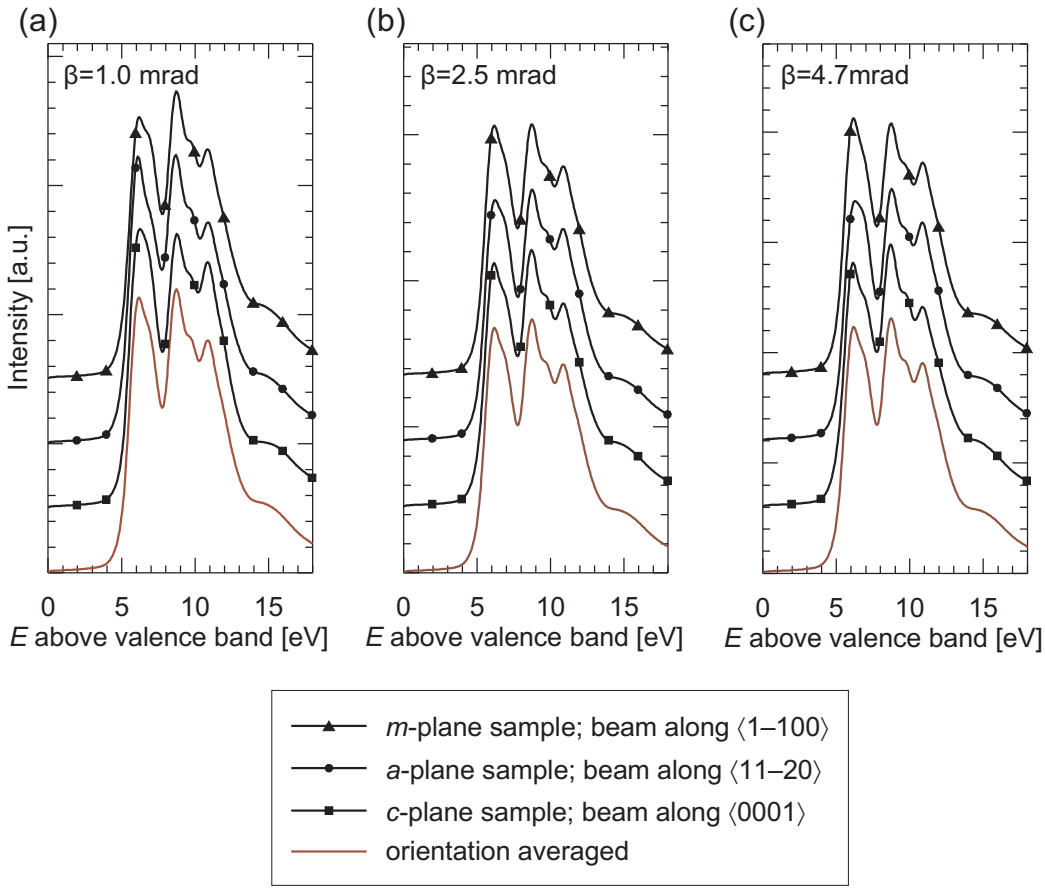


Figure 7.9: N K-edge ELNES spectra of AlN calculated for the full core hole case: three different mutual orientations of the single crystal sample and the electron beam direction and directionally averaged spectra are shown for the collection semi-angle a) $\beta = 1.0$ mrad, b) $\beta = 2.5$ mrad, and c) $\beta = 4.7$ mrad.

is very close to the so called *magic angle* defined by Jouffrey et al. (2004), for which the anisotropy disappears.

For incident electrons with energy $E_0 = 200$ keV and energy losses around $E \approx 400$ eV the characteristic scattering angle is $\theta_E \approx E/2E_0 \approx 1$ mrad. Jouffrey et al. (2004) calculated theoretically the magic angle to be $1.46\theta_E$ for 200 keV (≈ 1.5 mrad in this case); the experimental work done by Daniels et al. (2003) showed that the magic angle is around $2\theta_E$ (≈ 2 mrad in this case). These values are in good agreement with the value $\beta = 2.5$ mrad for which no variations with the mutual beam-to-sample orientation in the simulated ELNES spectra are observed.

Lazar et al. (2004) estimated theoretically (by comparing orientation resolved and unresolved ELNES) that $\beta = 4.7$ mrad can be taken as a good approximation to the magic angle for the N K-edge and 200 keV beam energy. Looking at the presented results slight anisotropy is observed for those conditions (see Fig. 7.9c). The first peak has a different shape for the *a*-plane sample; the first peak is the most intense one in the case

of the *a*-plane sample unlike for the other orientations. This corresponds to a lower value of the magic angle as was discussed in the previous paragraphs.

7.4 Summary of the chapter

Nitrogen K-edge electron energy loss near edge structures of AlN, GaN and InN were studied in detail in this chapter. In particular, N K-edge spectra corresponding to the ground states (0 e), partial (0.1–0.9 e) and full core hole (1 e) calculations were performed. By comparing them with several available experimental spectra from the literature was concluded that the best agreement is obtained for a core hole charge of about 0.6–0.7 e for AlN, about 0.4–0.5 e for GaN and about 0.4 e for InN. As all these values lie close to a core hole charge of 0.5 e and since the changes in the range 0.4–0.6 e are not big in either of the studied cases, a core hole charge of 0.5 e will be used in the next chapter for studies of ternary alloys.

The variations in N K-edge shape with core hole charge were traced to rearrangement of the final *p* states of nitrogen. Generally speaking, they move towards lower energies with increasing core hole charge. This effect was ascribed to a weaker shielding effect in the presence of a core hole (i.e. in the absence of a (fraction of a) core electron) which results in stronger bonding of the unoccupied states to the core. The discussion of ground state valence and unoccupied PDOSs of AlN, GaN and InN revealed a strong coupling between the *s* states of group III elements and the N *p*. Since the distribution of the *s* states of group III elements is different for Al, Ga and In (but similar for Ga and In due to the presence of *d* electrons), the N K-edge ELNES varies accordingly.

Chapter 8

ELNES of ternary alloys

8.1 Special quasi-random structures

The crucial question when dealing with *ab initio* calculations of alloys is how to arrange atoms within the periodically repeated supercell. Obviously, if the supercell contains N atom sites which can be occupied by atoms of either type A or type B, only the compositions $x = m/N$ can be obtained, $0 \leq m \leq N$ and m is an integer. The problem is which m of the N sites to populate with atoms A.

Early works approached the problem of alloys by developing the *virtual crystal approximation* (VCA) (Nordheim, 1931) and the *site-coherent potential approximation* (SCPA) (Soven, 1967; Velický et al., 1968). These are so-called *non-structural* theories since they approach alloys by "average occupation" of atom sites. They have been applied to a wide variety of alloys and are able to capture effects with symmetry-preserving uniform volume changes. However, the principal simplification in these methods lies in the association of the average alloy properties with those of "effective atoms" on sites, bonds etc. A longer discussion was given by Lymperakis (2005) and Wei et al. (1990) and references cited therein.

On the contrary, *structural* theories consider the actual configuration of atoms in the structure. For a binary (or pseudo-binary[1]) alloy with N sites, there are in total 2^N possible configurations (including those which are equivalent due to crystal symmetry). A measurable property of the alloy, for example the total energy E, is then the average over the ensemble of all these configurations

$$E = \langle E \rangle = \sum_{\sigma \in \Omega_m} \rho(\sigma) E(\sigma) \ . \tag{8.1}$$

[1]Strictly speaking, as a binary alloy is called A_xB_{1-x}, as a pseudobinary is called $A_xB_{1-x}C = (AC)_x(BC)_{1-x}$. Within the context of this text, pseudobinary alloys are called ternary alloys.

Here, $\rho(\sigma)$ denotes the probability of finding the configuration σ in the ensemble and Ω_m is the subset of all possible configurations containing these with a fixed composition given by m. Although this is in principle a solution to the "alloy problem", it still constitutes a major obstacle: it is not possible to probe all 2^N configurations[2]. Wei et al. (1990) described a methodology which justifies utilisation of only a very few specially designed structures (*special quasi-random structures*, SQS) which closely resemble a random alloy with a given composition x.

This chapter first describes in detail the process of identification of SQS for wurtzite structures. These are subsequently used for calculation of the ELNES of III-nitride ternary alloys.

8.1.1 Cluster expansion method

When trying to find a finite structure mimicking an infinite random alloy, artificial correlations beyond some distance, due to the periodic nature of the calculations, are introduced. However, many physical properties, for example projected density of states, depend strongly on the local environment. Therefore, the selection process of SQS is driven by the principle of a close reproduction of the structural correlation functions (defined below) of a random alloy for the first few nearest neighbour shells around a given site. In what follows, only a short introduction is given; more details can be found, for example, in the paper by Wei et al. (1990).

Each property $E(\sigma)$ of a configuration σ is discretised into *figures* (k, m). A figure is defined by the number k of atoms located on its vertices ($k = 1$ for sites themselves, $k = 2$ for pairs of atoms, etc.) and the order m of neighbour distances separating them ($m = 1$ for the nearest neighbours, $m = 2$ for the second nearest neighbours, etc.). The number of figures per site equivalent due to crystal symmetry is denoted by $D_{(k,m)}$. Each site i is assigned a "spin" variable $\hat{S}_i$ having a value 1 if the site i is occupied by the atom A, and -1 if it is occupied by the atom B. Defining as $\Pi_{(k,m)}(\boldsymbol{l}, \sigma)$ a product of the spin variables corresponding to a figure (k, m) located at an atom site $\boldsymbol{l}$ in a configuration σ, the lattice average over all locations $\boldsymbol{l}$ becomes

$$\overline{\Pi}_{(k,m)}(\sigma) = \frac{1}{ND_{(k,m)}} \sum_{\boldsymbol{l}} \Pi_{(k,m)}(\boldsymbol{l}, \sigma) \,. \tag{8.2}$$

$\overline{\Pi}_{(k,m)}(\sigma)$ is called the *structural correlation coefficient* (SCC). Sanchez et al. (1984) proved that $\{\Pi_{(k,m)}(\boldsymbol{l}, \sigma)\}$ provides a complete set of orthogonal functions. This allows expressing a property $E(\sigma)$ as

[2] For example, the ensemble size for $2 \times 2 \times 2$ supercell with 2 atoms in the unit cell is 65 536 which is far beyond any practical utilisation.

$$E(\sigma) = N \sum_{(k,m)} D_{(k,m)} \overline{\Pi}_{(k,m)}(\sigma) \epsilon_{(k,m)} \tag{8.3}$$

where $\epsilon_{(k,m)}$ denotes the contribution of the figure (k, m) to the physical property E. Note, that $\epsilon_{(k,m)}$ does not depend on the figure position $\boldsymbol{l}$. The ensemble average over all configurations σ now becomes

$$\langle E \rangle = N \sum_{(k,m)} D_{(k,m)} \langle \overline{\Pi}_{(k,m)} \rangle (\sigma) \epsilon_{(k,m)} \ . \tag{8.4}$$

Wei et al. (1990) showed that Eq. 8.3 is fairly rapidly convergent and thus that a very good approximation to an exact value given by Eq. 8.4 can be obtained by taking only a few configurations σ. In particular, in what follows only a single structure is considered which mimics best the random alloy. For a perfectly random binary alloy it can be easily shown that

$$\overline{\Pi}^R_{(k,m)} = \langle \overline{\Pi}^R_{(k,m)} \rangle = (2x - 1)^k \ . \tag{8.5}$$

Let D_m denote the number of neighbour sites in the m^{th} shell. Then in a random binary alloy of composition $A_x B_{1-x}$, the number of atoms of opposite type in the m^{th} shell is given by

$$O^R_m = 2 D_m x (1 - x) \ . \tag{8.6}$$

In what follows, figures up to the 4^{th} shell are checked for all possible configurations in a $2 \times 2 \times 2$ wurtzite supercell. Those corresponding best with a random alloy of the same composition in terms of both, SCC (Eq. 8.5) and the numbers of neighbours of opposite type (Eq. 8.6) are chosen for the subsequent studies.

8.1.2 SQS of 2×2×2 supercell of wurtzite material

As the topology of figures depends only on the geometrical configuration of atoms within the unit cell they may change with changing composition x: big changes in configuration of atoms may cause changes in the nearest-neighbour relations (some atoms are no longer the nearest neighbours whereas some others become the nearest neighbours of a given atom) which would lead to different figures. Fortunately, the c/a ratio of neither AlGaN, InGaN, nor AlInN changes that much with composition x to be able to cause changes in nearest-neighbour relations (i.e. the nearest-neighbour shells

up to the fourth order remain unchanged with x). Therefore, the estimated figures are independent on composition x in the studied alloys.

The $2 \times 2 \times 2$ supercell of wurtzite III-N structures is considered in the following. The binary and pseudo-binary alloys are treated in the same way within the cluster expansion theory. Therefore, the evaluation of figures is simplified by omitting the nitrogen atoms. The figures are described using numbered atom sites as shown in Fig. 8.1. Atoms in dark red colour correspond to the actual atoms in the supercell, the pink ones are their periodic expansion.

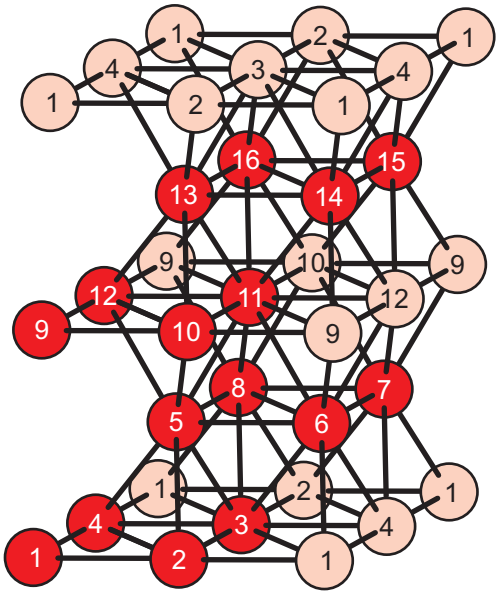


Figure 8.1: Numbered atom sites in a $2 \times 2 \times 2$ supercell of the wurtzite III-N structure.

(1,0) figures

This figure corresponds to single atom sites (the figure consists of only one vertex). Therefore, the average value of SCC is

$$\overline{\Pi}_{(1,0)}(\sigma) = \frac{1 \times (\text{number of Al atoms}) + (-1) \times (\text{number of Ga atoms})}{\text{total number of atoms in the supercell}} . \qquad (8.7)$$

(2,1) figures

This figure includes pairs of atom sites in the nearest neighbour distance. All inequivalent pairs contributing to this figure are listed in Table 8.1.

Each of these pairs corresponds to two identical figures (each time located on a different atom). Because SCC is an average value over all pairs and each pair has the same

$\{1,6\}$	$\{1,7\}$	$\{1,8\}$	$\{1,14\}$	$\{1,15\}$	$\{1,16\}$	$\{2,5\}$	$\{2,7\}$	$\{2,8\}$	$\{2,13\}$
$\{2,15\}$	$\{2,16\}$	$\{3,5\}$	$\{3,6\}$	$\{3,8\}$	$\{3,13\}$	$\{3,14\}$	$\{3,16\}$	$\{4,5\}$	$\{4,6\}$
$\{4,7\}$	$\{4,13\}$	$\{4,14\}$	$\{4,15\}$	$\{5,10\}$	$\{5,11\}$	$\{5,12\}$	$\{6,9\}$	$\{6,11\}$	$\{6,12\}$
$\{7,9\}$	$\{7,10\}$	$\{7,12\}$	$\{8,9\}$	$\{8,10\}$	$\{8,11\}$	$\{9,14\}$	$\{9,15\}$	$\{9,16\}$	$\{10,13\}$
$\{10,15\}$	$\{10,16\}$	$\{11,13\}$	$\{11,14\}$	$\{11,16\}$	$\{12,13\}$	$\{12,14\}$	$\{12,15\}$		

Table 8.1: All inequivalent pairs contributing to the $(2,1)$ figure.

weight, the formula for SCC of this figure is

$$\overline{\Pi}_{(2,1)}(\sigma) = \frac{1}{48} \sum_{\{i,j\}\in\text{Table 8.1}} \hat{S}_i \hat{S}_j \,. \tag{8.8}$$

As each atom site under consideration has six nearest neighbours, the degeneracy coefficient is $D_{(2,1)} = 6$.

(2,2) figure

Pairs of atoms corresponding to this figure are separated by the second nearest neighbour distance. These are the nearest atoms within the (0001) planes. Similarly to the case of the $(2,1)$ figure, the SCC is given by

$$\overline{\Pi}_{(2,2)}(\sigma) = \frac{1}{24} \sum_{\{i,j\}\in\text{Table 8.2}} \hat{S}_i \hat{S}_j \tag{8.9}$$

where all inequivalent pairs are listed in Table 8.2. Due to the periodic boundary conditions in the (0001) planes imposed by the $2 \times 2 \times 2$ supercell, each pair in Table 8.2 corresponds to four placements of this figure: two at each of the two vertices. Each atom site has six second nearest neighbours and thus $D_{(2,2)} = 6$.

$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$	$\{5,6\}$	$\{5,7\}$
$\{5,8\}$	$\{6,7\}$	$\{6,8\}$	$\{7,8\}$	$\{9,10\}$	$\{9,11\}$	$\{9,12\}$	$\{10,11\}$
$\{10,12\}$	$\{11,12\}$	$\{13,14\}$	$\{13,15\}$	$\{13,16\}$	$\{14,15\}$	$\{14,16\}$	$\{15,16\}$

Table 8.2: All inequivalent pairs contributing to the $(2,2)$ figure.

(2,3) figure

The neighbours in the third nearest neighbour shell are atoms in the (0001) planes immediately above and below the (0001) plane containing the atom under consideration. Due to the periodicity and size of the supercell, all three sites in the upper plane are of

the same kind. The same holds for the atoms in the bottom plane. Thus, each pair in Table 8.3 represents in fact six different placements of this figure: three for each of the two vertices. It follows the degeneracy coefficient in this case is $D_{(2,3)} = 6$. The SCC corresponding to this figure is given by

$$\overline{\Pi}_{(2,3)}(\sigma) = \frac{1}{16} \sum_{\{i,j\} \in \text{Table 8.3}} \hat{S}_i \hat{S}_j \,. \tag{8.10}$$

$\{1,5\}$	$\{1,13\}$	$\{2,6\}$	$\{2,14\}$	$\{3,7\}$	$\{3,15\}$	$\{4,8\}$	$\{4,16\}$
$\{5,9\}$	$\{6,10\}$	$\{7,11\}$	$\{8,12\}$	$\{9,13\}$	$\{10,14\}$	$\{11,15\}$	$\{12,16\}$

Table 8.3: All inequivalent pairs contributing to the $(2,3)$ figure.

(2,4) figure

The fourth nearest neighbours are the atoms straight above and below the considered atom site, i.e. in the $\langle 0001 \rangle$ direction. Again, due to the periodicity imposed by the size of the supercell, both sites are occupied by the same atom. Therefore, each of the pairs listed in Table 8.4 represents four placements of this figure: two for each of the two vertices. The SCC corresponding to this figure is

$$\overline{\Pi}_{(2,4)}(\sigma) = \frac{1}{8} \sum_{\{i,j\} \in \text{Table 8.4}} \hat{S}_i \hat{S}_j \tag{8.11}$$

and the degeneracy is $D_{(2,4)} = 2$.

$\{1,9\}$	$\{2,10\}$	$\{3,11\}$	$\{4,12\}$	$\{5,13\}$	$\{6,14\}$	$\{7,15\}$	$\{8,16\}$

Table 8.4: All inequivalent pairs contributing to the $(2,4)$ figure.

(3,1) figure

There are no such figures. Atoms in the nearest neighbour distance are the closest neighbours lying in two adjacent (0001) planes. Assume $\{\alpha, \beta\}$ is such a pair of atom sites. The nearest neighbour γ to β is either in the same plane as α, in which case the distance between α and γ is at least the second nearest neighbour distance, or lies in the other adjacent (0001) plane to the plane containing β. In that case, the distance between α and γ is at least two (0001) inter-planar distances, i.e. at least the fourth nearest neighbour distance.

(3,2) figure

There are three different figures with atom sites separated up to second neighbour distance as shown in Fig. 8.2. The second configuration (Fig. 8.2b) is not possible to obtain, though. It follows from the discussion of the figure $(2,2)$ that atom sites in the second nearest neighbour distance lie in the same (0001) plane whereas those separated by the first nearest neighbour distance lie in two different, adjacent (0001) planes. Figure 8.2b implies that α lies in the same (0001) plane as β which is in the same (0001) plane as γ. However, γ and α are in two different adjacent (0001) planes, which contradicts the earlier statement.

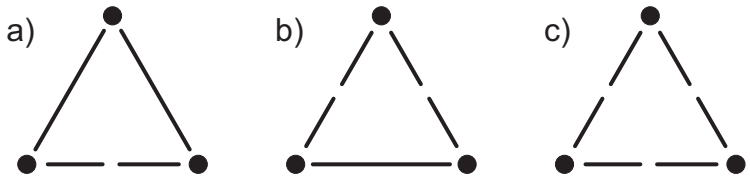


Figure 8.2: Three theoretically possible $(3,2)$ figures. However, only the a) and c) configurations are realisable for wurtzite materials.

The figure $(3,2)^a$ (see Fig. 8.2a) consists of two atoms being in the same (0001) plane and the third one being their common nearest neighbour lying in the adjacent (0001) plane. All inequivalent triplets of atomic sites corresponding to this figure are listed in Table 8.5. Each of the triplets corresponds to three different placements of this figure: each time located at one of its vertices. The degeneracy coefficient is $D^a_{(3,2)} = 18$.

$\{1,7,8\}$	$\{1,6,7\}$	$\{1,6,8\}$	$\{1,14,15\}$	$\{1,14,16\}$	$\{1,15,16\}$	$\{1,2,8\}$	$\{1,4,6\}$
$\{1,3,6\}$	$\{1,2,16\}$	$\{1,4,14\}$	$\{1,3,14\}$	$\{1,2,7\}$	$\{1,4,7\}$	$\{1,3,8\}$	$\{1,2,15\}$
$\{1,4,15\}$	$\{1,3,16\}$	$\{2,5,7\}$	$\{2,5,8\}$	$\{2,7,8\}$	$\{2,13,15\}$	$\{2,13,16\}$	$\{2,15,16\}$
$\{2,3,5\}$	$\{2,4,5\}$	$\{2,3,13\}$	$\{2,4,13\}$	$\{2,3,8\}$	$\{2,4,7\}$	$\{2,3,16\}$	$\{2,4,15\}$
$\{3,5,6\}$	$\{3,5,8\}$	$\{3,6,8\}$	$\{3,13,14\}$	$\{3,13,16\}$	$\{3,14,16\}$	$\{3,4,6\}$	$\{3,4,5\}$
$\{3,4,14\}$	$\{3,4,13\}$	$\{4,5,6\}$	$\{4,5,7\}$	$\{4,6,7\}$	$\{4,13,14\}$	$\{4,13,15\}$	$\{4,14,15\}$
$\{5,10,11\}$	$\{5,10,12\}$	$\{5,11,12\}$	$\{5,6,11\}$	$\{5,8,11\}$	$\{5,7,12\}$	$\{5,6,12\}$	$\{5,8,10\}$
$\{5,7,10\}$	$\{6,9,12\}$	$\{6,9,11\}$	$\{6,11,12\}$	$\{6,7,12\}$	$\{6,8,11\}$	$\{6,7,9\}$	$\{6,8,9\}$
$\{7,9,10\}$	$\{7,9,12\}$	$\{7,10,12\}$	$\{7,8,9\}$	$\{7,8,10\}$	$\{8,9,10\}$	$\{8,9,11\}$	$\{8,10,11\}$
$\{9,14,15\}$	$\{9,14,16\}$	$\{9,15,16\}$	$\{9,10,16\}$	$\{9,12,14\}$	$\{6,11,14\}$	$\{9,10,15\}$	$\{9,12,15\}$
$\{9,11,16\}$	$\{10,13,15\}$	$\{10,13,16\}$	$\{10,15,16\}$	$\{10,11,13\}$	$\{10,12,13\}$	$\{10,11,16\}$	$\{10,12,15\}$
$\{11,13,14\}$	$\{11,13,16\}$	$\{11,14,16\}$	$\{11,12,14\}$	$\{11,12,13\}$	$\{12,13,14\}$	$\{12,13,15\}$	$\{12,14,16\}$

Table 8.5: All inequivalent triples contributing to the $(3,2)^a$ subfigure.

$\{1,2,3\}$	$\{1,2,4\}$	$\{1,3,4\}$	$\{2,3,4\}$	$\{5,6,7\}$	$\{5,6,8\}$	$\{5,7,8\}$	$\{6,7,8\}$
$\{9,10,11\}$	$\{9,10,12\}$	$\{9,11,12\}$	$\{10,11,12\}$	$\{13,14,15\}$	$\{13,14,16\}$	$\{13,15,16\}$	$\{14,15,16\}$

Table 8.6: All inequivalent triplets contributing to the $(3,2)^b$ subfigure.

Three atom sites lying in the same (0001) plane forming the smallest possible equilateral triangle correspond to the figure $(3,2)^b$ depicted in Fig. 8.2c. Each atom site is a vertex belonging to six triangles; due to the periodicity and the size of the supercell, each triplet of vertices from Table 8.6 corresponds to six such triangles: two for each of its three vertices. The degeneracy for this subfigure is $D^b_{(3,2)} = 6$. The SCC corresponding to this figure is given by

$$\overline{\Pi}_{(2,3)}(\sigma) = \frac{1}{96+2\times 16}\left(\sum_{\{i,j,k\}\in\text{Table 8.5}} \hat{S}_i\hat{S}_j\hat{S}_k + 2\sum_{\{i,j,k\}\in\text{Table 8.6}} \hat{S}_i\hat{S}_j\hat{S}_k\right) . \quad (8.12)$$

(3,3) figure

There are six topologically inequivalent subfigures, all shown in Fig. 8.3. Nevertheless, only one of them can be realised in the wurtzite structures. The argument, why none of the subfigures in Fig. 8.3a, 8.3d and 8.3f can exist, is exactly the same as the one used for the figure $(3,1)$. The non-existence of the subfigure in Fig. 8.3c can be proved in the same way as for subfigure in Fig. 8.2b. Lastly, each position of any nearest neighbour atom pair within the (0001) plane (i.e. atoms within the second nearest neighbour distance) have different nearest neighbours within the third nearest neighbour distance. Therefore, neither the subfigure Fig. 8.2e exists. The only remaining subfigure contributing to the $(3,3)$ figure is thus the one in Fig. 8.3b.

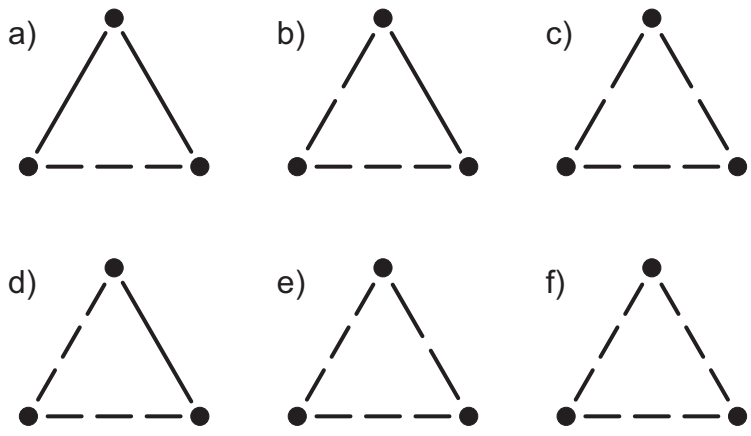


Figure 8.3: Six theoretically possible $(3,3)$ figures: only one of them is realisable.

Due to the periodicity in the (0001) planes imposed by the $2 \times 2 \times 2$ supercell, each of the triplets listed in Table 8.7 corresponds to 6 placements of this figure: for each of the three vertices. The total degeneracy of this figure is $D_{(3,3)} = 36$ and the corresponding SCC is

$$\overline{\Pi}_{(3,3)}(\sigma) = \frac{1}{96} \sum_{\{i,j,k\} \in \text{Table 8.7}} \hat{S}_i \hat{S}_j \hat{S}_k \, . \tag{8.13}$$

$\{1,2,5\}$	$\{1,4,5\}$	$\{1,2,13\}$	$\{1,4,13\}$	$\{1,2,6\}$	$\{1,4,8\}$	$\{1,2,14\}$	$\{1,4,16\}$
$\{1,3,5\}$	$\{1,3,13\}$	$\{1,3,7\}$	$\{1,3,15\}$	$\{2,4,6\}$	$\{2,4,14\}$	$\{2,3,6\}$	$\{2,3,14\}$
$\{2,4,8\}$	$\{2,4,16\}$	$\{2,3,7\}$	$\{2,3,15\}$	$\{3,4,7\}$	$\{3,4,15\}$	$\{3,4,8\}$	$\{3,4,16\}$
$\{1,8,5\}$	$\{1,7,5\}$	$\{1,6,5\}$	$\{1,16,13\}$	$\{1,15,13\}$	$\{1,14,13\}$	$\{2,5,6\}$	$\{2,6,8\}$
$\{2,6,7\}$	$\{2,13,14\}$	$\{2,14,16\}$	$\{2,14,15\}$	$\{3,7,8\}$	$\{3,5,7\}$	$\{3,6,7\}$	$\{3,15,16\}$
$\{3,13,15\}$	$\{3,14,15\}$	$\{4,6,8\}$	$\{4,7,8\}$	$\{4,5,8\}$	$\{4,14,16\}$	$\{4,15,16\}$	$\{4,13,16\}$
$\{5,6,9\}$	$\{5,7,9\}$	$\{5,8,9\}$	$\{5,6,10\}$	$\{5,7,11\}$	$\{5,8,12\}$	$\{5,9,10\}$	$\{5,9,12\}$
$\{5,9,11\}$	$\{6,7,11\}$	$\{6,7,10\}$	$\{6,8,10\}$	$\{6,8,12\}$	$\{6,9,10\}$	$\{6,10,11\}$	$\{6,10,12\}$
$\{7,8,11\}$	$\{7,8,12\}$	$\{7,10,11\}$	$\{7,11,12\}$	$\{7,9,11\}$	$\{8,11,12\}$	$\{8,9,12\}$	$\{8,10,12\}$
$\{9,10,13\}$	$\{9,10,14\}$	$\{9,11,13\}$	$\{9,11,15\}$	$\{9,12,13\}$	$\{9,12,16\}$	$\{9,13,16\}$	$\{9,13,14\}$
$\{9,13,15\}$	$\{10,11,14\}$	$\{10,11,15\}$	$\{10,12,14\}$	$\{10,12,16\}$	$\{10,13,14\}$	$\{10,14,15\}$	$\{10,14,16\}$
$\{11,12,15\}$	$\{11,12,16\}$	$\{11,14,15\}$	$\{11,16,15\}$	$\{11,13,15\}$	$\{12,15,16\}$	$\{12,13,16\}$	$\{12,14,16\}$

Table 8.7: All inequivalent triplets contributing to the $(3,3)$ figure.

(4,1) figure

This does not exist for similar reasons that the figure $(3,1)$ does not exist: if three vertices cannot be fit into the wurtzite structure in such a way that all distances between them are the nearest neighbour distances, there is no way of adding one more atom site into the consideration and to fulfil the same requirement.

(4,2) figure

The fact that the figures $(3,1)$ and $(3,2)^b$ have no real meaning for wurtzite structures reduces the number of possibilities for the $(4,2)$ figures significantly. In fact, the only remaining possibility is a pyramid with base in the (0001) plane and an apex in an adjacent (0001)

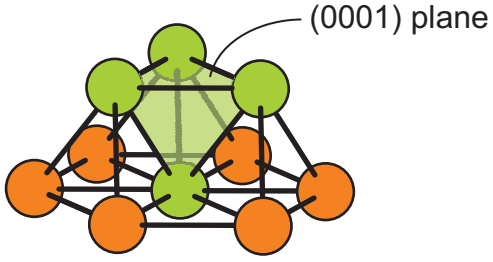


Figure 8.4: The pyramid constituting the $(4,2)$ figure.

plane as shown in Fig. 8.4. Each of the groups listed in Table 8.8 corresponds exactly to four placements of this figure, each time located on different apexes. The degeneracy of this figure is $D_{(4,2)} = 8$ and the SCC is given by

$$\overline{\Pi}_{(4,2)}(\sigma) = \frac{1}{32} \sum_{\{i,j,k,l\} \in \text{Table 8.8}} \hat{S}_i \hat{S}_j \hat{S}_k \hat{S}_l \,. \tag{8.14}$$

$\{1,6,7,8\}$	$\{1,14,15,16\}$	$\{2,5,7,8\}$	$\{2,13,15,16\}$	$\{3,5,6,8\}$	$\{3,13,14,16\}$
$\{4,5,6,7\}$	$\{4,13,14,15\}$	$\{9,6,7,8\}$	$\{9,14,15,16\}$	$\{10,5,7,8\}$	$\{10,13,15,16\}$
$\{11,5,6,8\}$	$\{11,13,14,16\}$	$\{12,5,6,7\}$	$\{12,13,14,15\}$	$\{5,2,3,4\}$	$\{6,1,3,4\}$
$\{7,1,2,4\}$	$\{8,1,2,3\}$	$\{5,10,11,12\}$	$\{6,9,11,12\}$	$\{7,9,10,12\}$	$\{8,9,10,11\}$
$\{13,2,3,4\}$	$\{14,1,3,4\}$	$\{15,1,2,4\}$	$\{16,1,2,3\}$	$\{13,10,11,12\}$	$\{14,9,11,12\}$
$\{15,9,10,12\}$	$\{16,9,10,11\}$				

Table 8.8: All inequivalent groups of four sites contributing to the $(4,2)$ figure.

8.1.3 Chosen configurations for ternary III-N alloys

All relevant configurations for each composition x were investigated. The set of SCC corresponding to the figures discussed above was evaluated for each configuration. These were subsequently compared with those of the random alloy of the same composition and the best matching structures were selected[3]. At the same time, average numbers of nearest neighbours of opposite type in the first four nearest neighbour shells were calculated and compared with the prediction (Eq. 8.6) for the random alloys[3]. It is worth noting that in this case all structures with the SCC best matching those of the random alloy for a given composition x, have the same set of the numbers of nearest neighbours. One representative for each composition was selected from the set of the best matching structures (which are in terms of the used model equivalent). The selected SQS for various compositions are listed in Table 8.9.

8.2 Nitrogen K-edge ELNES of III-nitride alloys

8.2.1 Computational details

`Wien2k` code (Blaha et al., 2000) was used to calculate the ground state and excited (core hole) charge densities of III-nitride alloys within the framework of the density functional theory (Hohenberg and Kohn, 1964; Kohn and Sham, 1965). The computational

[3]Detailed values are given in Appendix. See Table C.1 and C.2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$x = 0.0000$	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga
$x = 0.0625$	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Al
$x = 0.1250$	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Al	Ga	Ga	Al	Ga
$x = 0.1875$	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Al	Ga	Ga	Al	Al
$x = 0.2500$	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Al	Ga	Ga	Al	Al	Ga	Al	Ga	Ga
$x = 0.3125$	Ga	Ga	Ga	Ga	Ga	Ga	Ga	Al	Ga	Ga	Al	Al	Ga	Al	Ga	Al
$x = 0.3750$	Ga	Ga	Ga	Ga	Ga	Ga	Al	Al	Ga	Al	Ga	Al	Ga	Al	Al	Ga
$x = 0.4375$	Ga	Ga	Ga	Al	Ga	Ga	Al	Ga	Ga	Ga	Al	Al	Al	Al	Ga	Al
$x = 0.5000$	Ga	Ga	Ga	Al	Ga	Ga	Al	Ga	Ga	Al	Al	Al	Al	Ga	Al	Al
$x = 0.6250$	Ga	Ga	Ga	Al	Ga	Al	Al	Al	Al	Ga	Al	Al	Al	Al	Ga	Al
$x = 0.6875$	Ga	Ga	Al	Al	Ga	Al	Ga	Al	Ga	Al	Al	Al	Al	Al	Al	Al
$x = 0.7500$	Ga	Ga	Al	Al	Ga	Al	Al	Al	Al	Al	Al	Al	Al	Al	Ga	Al
$x = 0.8125$	Ga	Ga	Al	Al	Ga	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al
$x = 0.8750$	Ga	Al	Al	Al	Al	Ga	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al
$x = 0.9375$	Ga	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al
$x = 1.0000$	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al	Al

Table 8.9: Selected quasi-random structures best matching the SCC and number of opposite type neighbours to the random alloy. Numbers represent atom sites in the $2 \times 2 \times 2$ supercell as shown in Fig. 8.1.

parameters were the same as those for binary compounds summarised in section 7.2: GGA approximation to the exchange-correlation potential parametrised by Perdew et al. (1996), l-expansion up to $l_{\max} = 10$, MTS radii from Table 7.1 and unit cell parameters obtained using Vegard's law with the boundary values corresponding to those binary compounds given in Table 7.1. As the cut-off parameter was used $R_{\mathrm{MT}}K_{\max} = 5$ based on a similar test as discussed in section 7.2. The number of k-points in the whole first Brillouin zone was 250 which corresponds approximately to the same density of k-points as 4 000 k-points in the case of the unit cell.

The theoretical ELNES in this chapter were calculated employing parameters corresponding to the experimental conditions (Cherns, 2007): electron beam energy 200 keV, convergence semi-angle of 4.7 mrad, collection semi-angle of 0.2 mrad. A value of 0.4 eV was used for spectrometer broadening of the calculated spectra. All spectra are directionally averaged.

8.2.2 Ground state evolution of the N K-edge

ELNES is highly sensitive to the local environment of a particular atom on which the energy loss takes place. In the simple cases of AlN, GaN and InN, all nitrogen atoms have exactly the same local neighbourhood (i.e. they are equivalent) and thus the N K-edge ELNES is unique. A different situation occurs with alloys: the local environment differs from nitrogen to nitrogen. Therefore, ELNES N K-edge spectra were calculated for all individual atoms in the supercell and the final spectrum was obtained by averaging them. Apart from the $x = 0.125$ and $x = 0.875$ supercells, all nitrogen atoms in the supercells are inequivalent and thus their contribution to the final spectra is the same. In the cases of $x = 0.125$ and $x = 0.875$, there are four pairs of equivalent atoms. Each of the spectra corresponding to these pairs was thus counted twice into the final averaged spectrum.

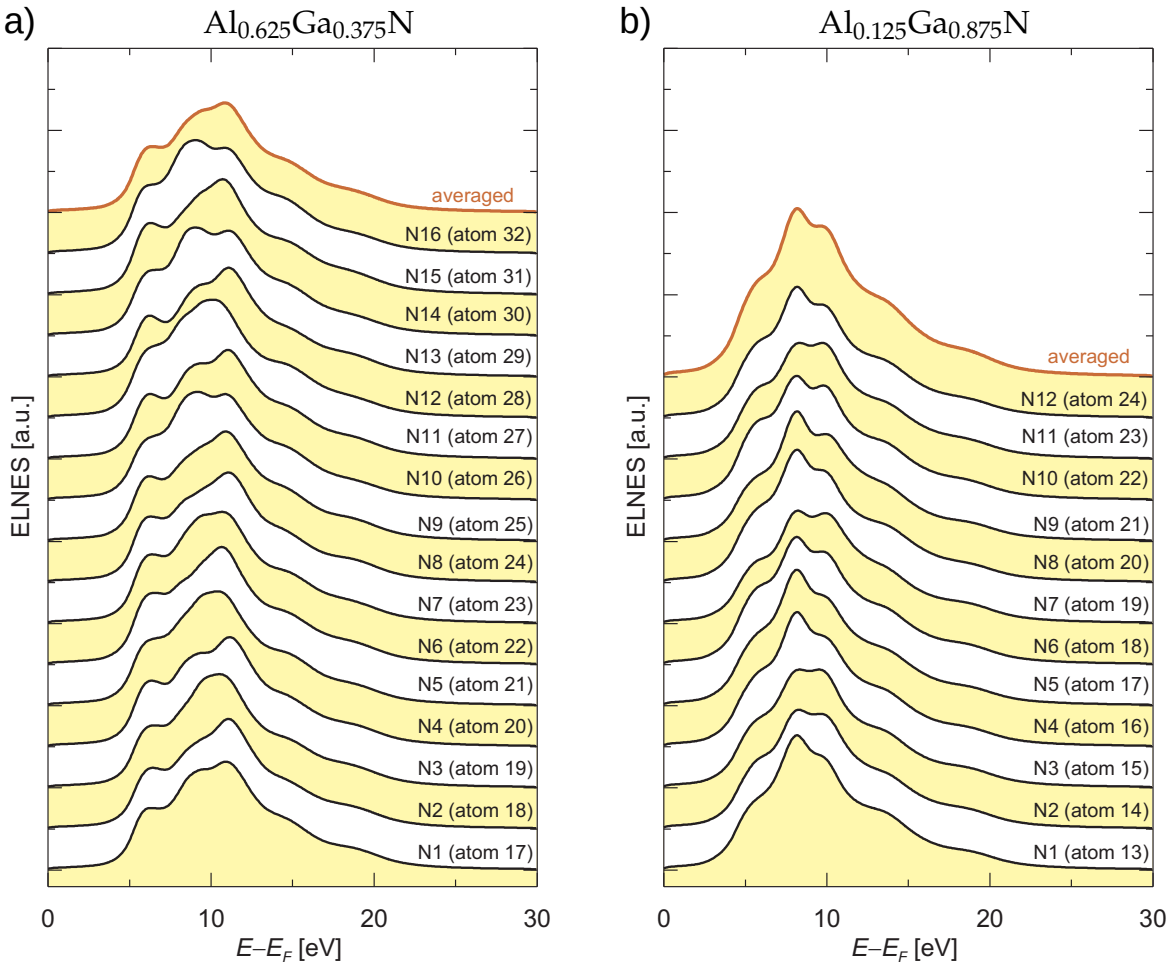


Figure 8.5: N K-edge ELNES for a) $Al_{0.625}Ga_{0.375}N$ and b) $Al_{0.125}Ga_{0.875}N$. The spectra were calculated for all non-equivalent atoms in the supercell and were subsequently averaged.

Examples of N K-edge ELNES are shown in Fig. 8.5a for $x = 0.625$ and in Fig. 8.5b for $x = 0.125$. Strong variations between graphs for different atoms can be observed which justifies the need for explicit calculation of all of them in order to obtain the final spectrum. The labels in Fig. 8.5 emphasise the fact the curves correspond to different N atoms (labelled 1 to 12 or 16) which make up half of the total number of inequivalent atoms in a supercell (24 or 32).

Figure 8.6 shows the evolution of the N K-edge of $Al_xGa_{1-x}N$, $In_xGa_{1-x}N$ and $Al_xIn_{1-x}N$ as obtained from the ground state calculations. It was concluded in the previous chapter (section 7.3.1) that GaN and InN behave very similarly in terms of the N K-edge. This is confirmed also in Fig. 8.6c with the evolution of the N K-edge of $In_xGa_{1-x}N$ alloy. The central peak at about 6–8 eV above the top of valence band (in these calculations) dominates for all values x. The most significant variation thus is a diminution of the "shoulder" on the high energy side of this peak, as x varies from 0 to 1.

A qualitatively different behaviour can be observed for $Al_xGa_{1-x}N$ and $Al_xIn_{1-x}N$ (Figs 8.6a and 8.6b): in both cases a transition from a single peak spectrum for $x = 0$ to a triple peak spectrum for $x = 1$ is observed. The reasons for this were discussed in section 7.3.3. Yet another observation is that the energy onset moves significantly towards lower energies with increasing (average) atomic number Z. This effect is most profound in the case of AlInN (Fig. 8.6b).

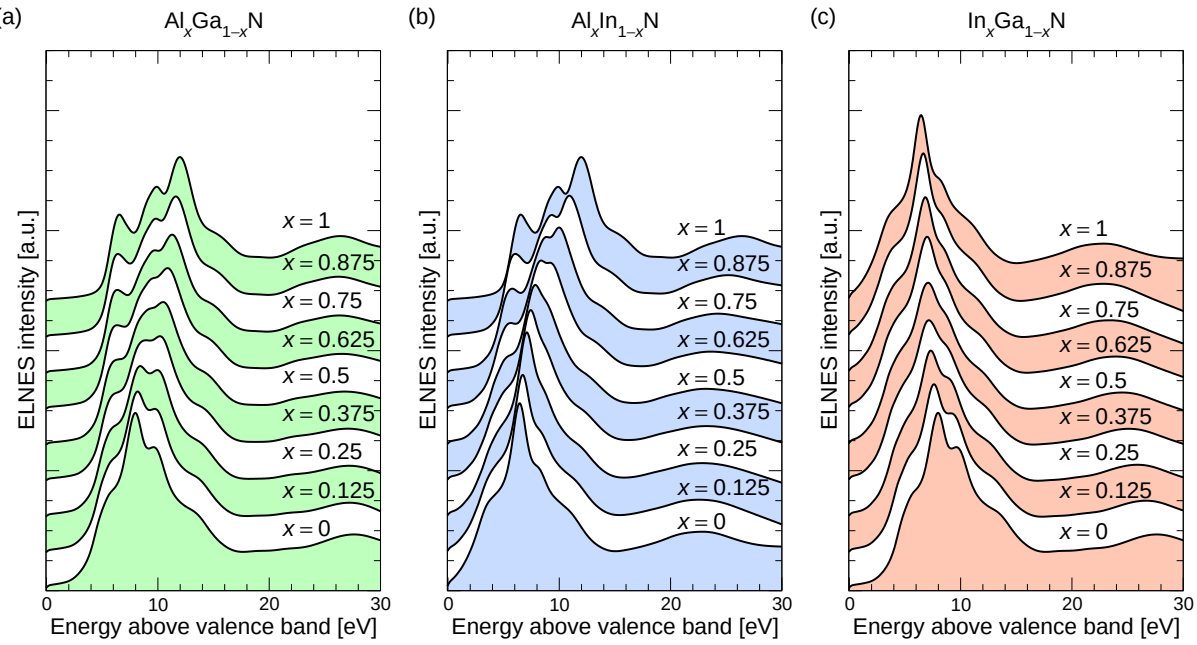


Figure 8.6: Evolution of N K-edge ELNES for a) $Al_xGa_{1-x}N$, b) $Al_xIn_{1-x}N$, and c) $In_xGa_{1-x}N$ as obtained from the ground state calculations.

8.2.3 Core hole calculations of the N K-edge

Figure 8.7 shows the evolution of $Al_xGa_{1-x}N$, $Al_xIn_{1-x}N$ and $In_xGa_{1-x}N$ N K-edge as calculated using the 0.5 e partial core hole. The binary compounds of pure AlN and GaN were calculated with different cut-off energies E_{max} (for searching the eigenvalues of the Hamiltonian, see section 5.2) and thus the "tail" region from about 20 eV above top of valence band is not present. Lower E_{max} speeds up calculations but, in turn it limits the range for ELNES calculations.

Similar trends to those discussed for the ground state calculation in section 8.2.2 can be observed. The transition from a single central peak to a triple peak character spectrum in the case of AlGaN (Fig. 8.7a) and AlInN (Fig. 8.7b) takes place only at a quite high Al fraction, at about $x \approx 0.75$. On the other hand, the single peak character persists with only small changes up to $x \approx 0.5$. N K-edge evolution of InGaN (Fig. 8.7c) becomes even less interesting than in the ground state case as the high energy shoulder at the GaN end is partly transferred into the central peak as a consequence of the presence of a core hole (see discussion in section 7.3.2).

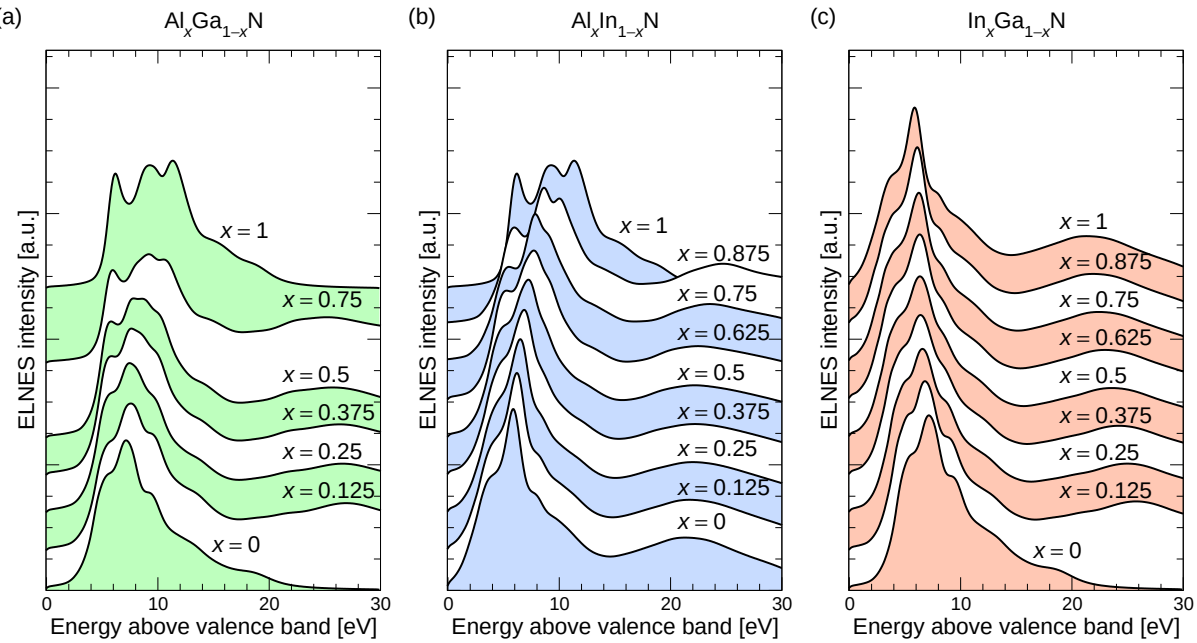


Figure 8.7: Evolution of N K-edge ELNES for a) $Al_xGa_{1-x}N$, b) $Al_xIn_{1-x}N$, and c) $In_xGa_{1-x}N$ as obtained from the partial core hole (0.5 e) calculations.

Figure 8.8 shows evolution of ELNES N K-edge of wurtzite $Al_xGa_{1-x}N$ for different aluminium fractions x using the ground state, partial (0.5 e) and full core hole calculations. Two experimental data sets (Cherns, 2007; Holec et al., 2008a) from two different microscopes are also shown for comparison. Better agreement of the calculated with experimental spectra is obtained for the Cambridge data set (Cherns, 2007) which was

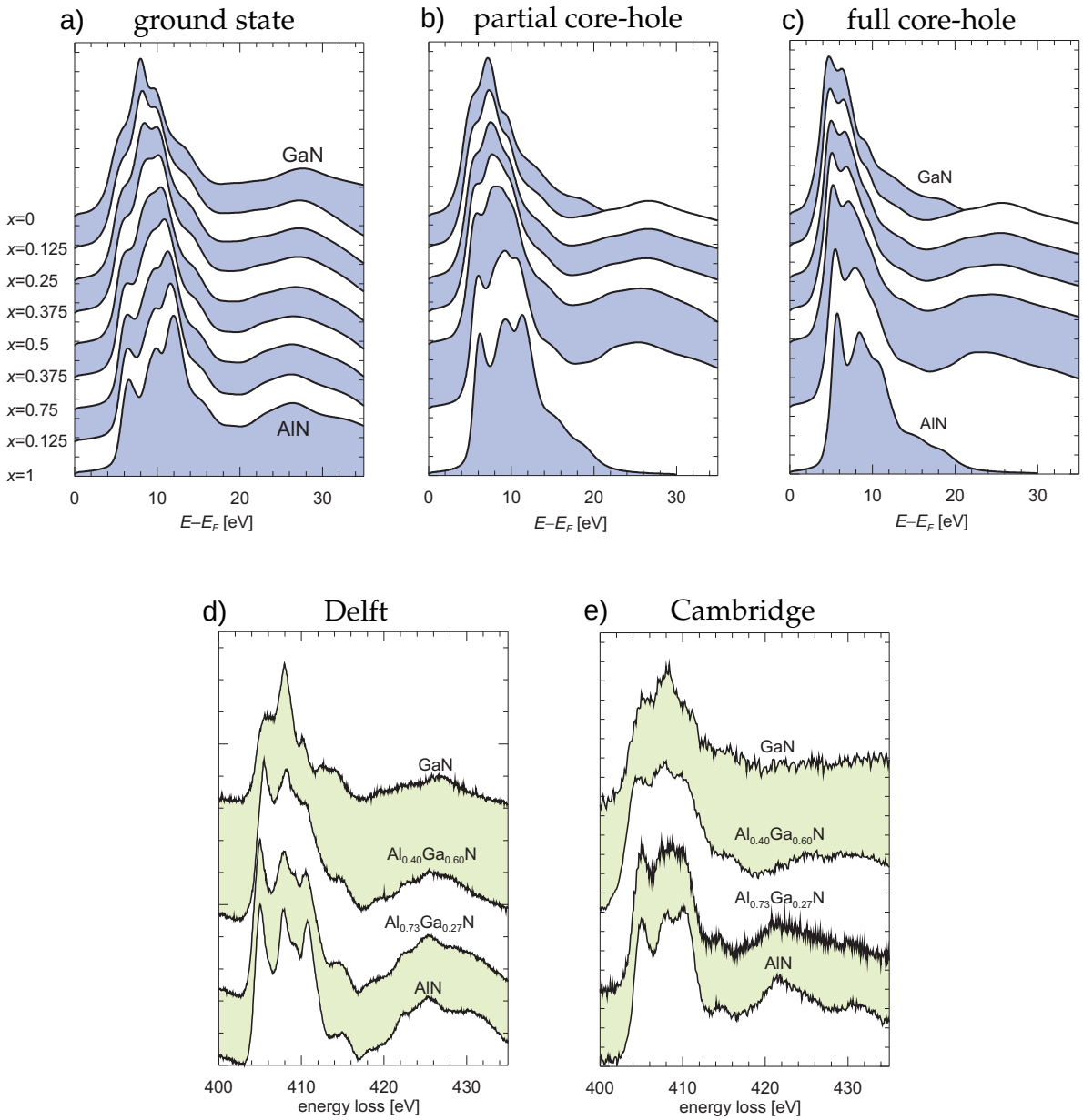


Figure 8.8: Simulated and experimental evolution of N K-edge ELNES of $Al_xGa_{1-x}N$: a) ground state, b) partial core-hole, c) full core-hole calculation and experimental data obtained at d) Delft (Holec et al., 2008a) and e) Cambridge (Cherns, 2007).

obtained on a microscope with a lower energy resolution. This somewhat surprising result is probably due to the fact that the theoretical spectra were calculated using the acquisition parameters of the Cambridge experimental setup. As was demonstrated in section 7.3.5, directionally resolved ELNES present small but important variations in spectra with high energy resolution. Thus, better agreement of theory with spectra of Holec et al. (2008a) with energy resolution as good as 0.1 eV is expected for the directionally resolved ELNES calculations and the actual parameters corresponding from the experiment.

8.2.4 Dependence of ELNES on the local nitrogen atom environment

The supercell representing the $Al_{0.5}Ga_{0.5}N$ structure is shown in Fig. 8.9. By focusing on the four nearest neighbours of each nitrogen atom it can be observed that in this supercell there are several different variants of possible neighbourhoods. These are schematically shown in Fig. 8.10 and can be divided into subsets according to the number of Ga and Al atoms. Further, they are subdivided according to the bond along the *c*-axis; all the other three bonds are equivalent. Five groups containing in total eight subgroups are listed in Table 8.10.

4 Al	3 Al, 1 Ga	2 Al, 2 Ga	1 Al, 3 Ga	4 Ga
N5(21)	N3(19)	N1(17)	N4(20)	N15(31)
	N7(23)	N12(28)	N9(25)	
	N10(26)	N14(30)	N13(29)	
	N16(32)	N2(18)	N6(22)	
		N8(24)		
		N11(27)		

Table 8.10: Different nitrogen atom environments grouped by their similarity.

Fig. 8.11 shows N K-edge ELNES for $Al_{0.5}Ga_{0.5}N$ as resulting from the ground state, partial core hole (0.5 e) and full core hole calculation. The ELNES spectra were calculated for various nitrogen atoms and were subsequently grouped according to the above discussed division of nitrogen atoms (see Table 8.10). Notice that in all these cases, the spectra corresponding to nitrogen atoms from one group are very similar despite the fact that the overall variation of ELNES for different nitrogen atoms is significant. This corresponds well with the fact that ELNES depends strongly on the local environment, and also justifies the use of SQSs with "only" 32 atoms. The small variations in spectra from within one group may be ascribed to the influence of more distant atoms.

8.3 Summary and conclusions

The concept of the special quasi-random structures as described by Wei et al. (1990) was employed to select $2 \times 2 \times 2$ supercells containing 32 atoms to represent wurtzite structures of various compositions. These were subsequently used for the calculation

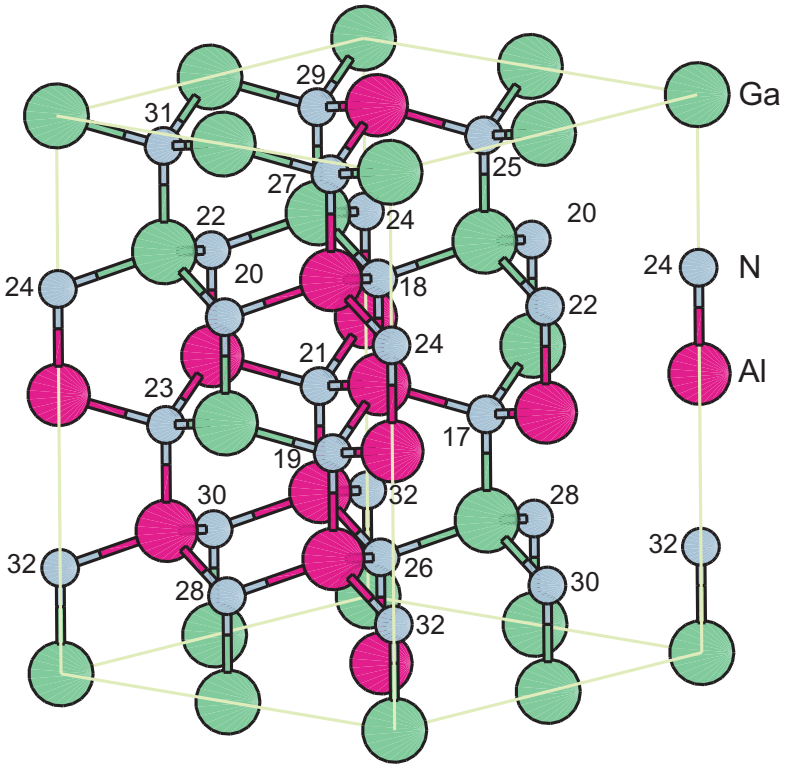


Figure 8.9: A $2 \times 2 \times 2$ supercell representing $Al_{0.5}Ga_{0.5}N$.

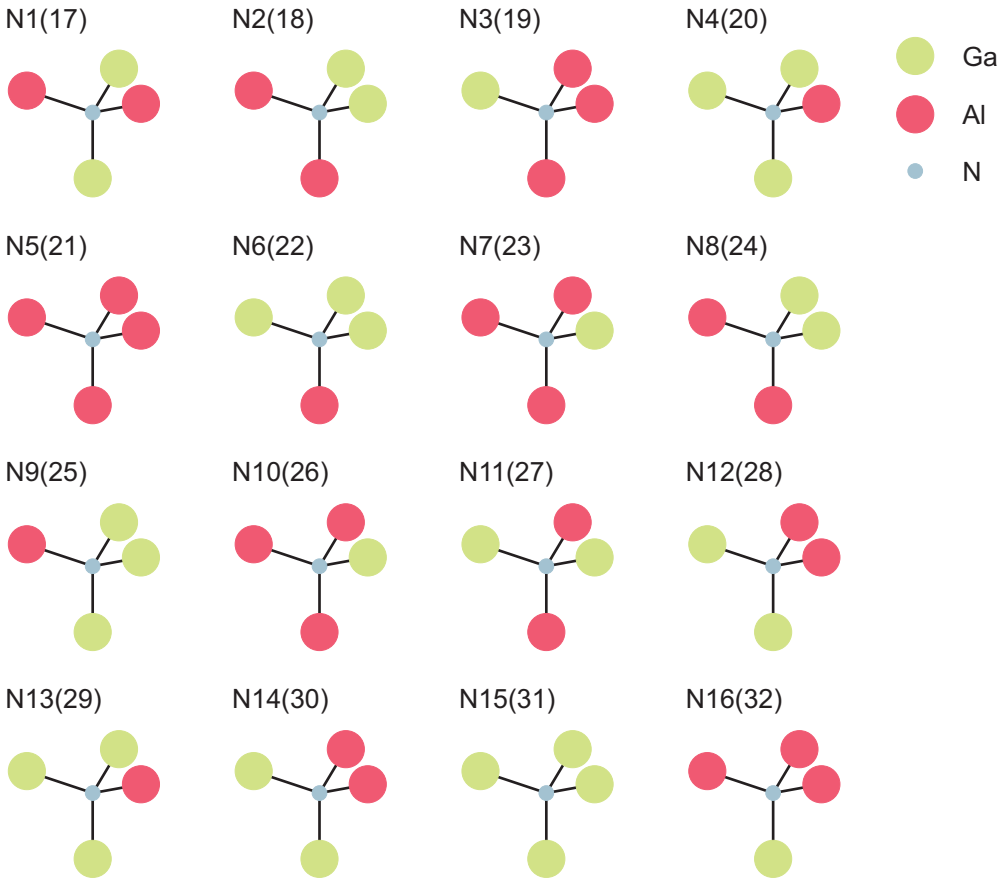


Figure 8.10: Local environments of 16 nitrogen atoms in $2 \times 2 \times 2$ supercell of $Al_{0.5}Ga_{0.5}N$.

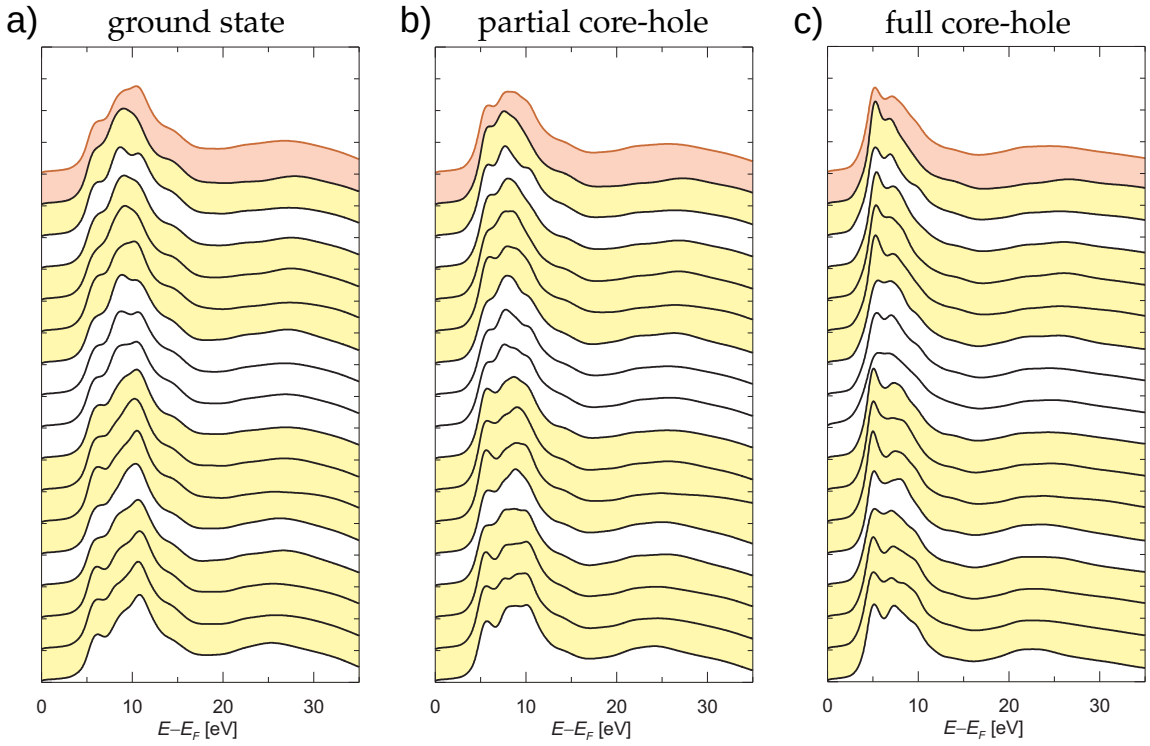


Figure 8.11: Dependence of the shape of N K-edge ELNES of $Al_{0.5}Ga_{0.5}N$ on the local nitrogen atom's environment. The bottom curve corresponds to the nitrogen atom surrounded with 4 Al, whereas the second curve from top corresponds to the nitrogen atom surrounded with 4 Ga atoms. The top curve is the final N K-edge ELNES as obtained by averaging. Graphs correspond to the a) ground state, b) partial core-hole, and c) full core-hole calculations.

of ELNES N K-edge evolution of $Al_xGa_{1-x}N$, $Al_xIn_{1-x}N$, and $In_xGa_{1-x}N$ ternary alloys. The used methodology proved to be able to reasonably reproduce experimental results. Some discrepancies between theory and experiment were found in the case of experimental data with a better energy resolution. These were primarily ascribed to using the directionally averaged ELNES calculation together with a higher broadening parameter which smeared some of the fine features seen in experimental spectra with the high energy resolution.

All the present calculations were done for unrelaxed structures due to limited computing resources. This is an obvious drawback of the present work – the electronic structure, PDOS and thus ELNES is expected to change a bit after the relaxation. However, this effect is not large as the parameters used for the supercell geometries are results of other theoretical calculations published in the literature which included relaxation. More discussion on this topic is given in the next chapter.

The calculations were done employing GGA. It is known that GGA, and DFT in general, gives wrong predictions for the band gap by several eVs. As a consequence, edge onsets are not accurate and are expected to be wrong by about 6 eV for the N K-edge

(Muller, 2006). This is, for example, seen in Fig. 8.7c where the edge onset for $x = 1$ is cut off. Muller (2006) showed that much better agreement (within $\approx 1.5\,\mathrm{eV}$) can be obtained for spin-polarised calculations. Apart from the standard DFT, some recently developed techniques such as perturbation GW method or time dependent DFT are expected to give better results as they implicitly work with excited states. Their utilisation goes beyond the scope of the current work. However, it would certainly be interesting to apply them to the methodology described here and to investigate what improvement could be achieved.

Chapter 9

Strain effects on ELNES and the band gap of AlGaN ternary alloys

9.1 Introduction

Due to different lattice constants of AlN and GaN compounds, the hetero-epitaxy of AlN/Al_xGa_{1-x}N/GaN leads to either strained (unrelaxed) layers, fully relaxed layers or most often partially relaxed layers (the latter two being defective). As a consequence, thin epilayers with varying Al fraction are often under either tensile or compressive strain (depending on a particular sample/device design).

Electron energy loss near edge structure (ELNES), described in chapter 6, is known to reflect the electronic structure of materials. As discussed there, the fine structure of the ELNES edge is given by the projected density of states (PDOS). Therefore, the PDOS can be used as an approximation to the corresponding ELNES fine structure.

Keast et al. (2003) reported on an experimental and theoretical study of strain effects in the N K-edge of ELNES in InGaN alloys. In particular, they concluded that strain enhances the effects of N K-edge evolution due to changing the In fraction. In addition, the edge onset was observed to shift towards lower energies with respect to the central peak due to strain. Several papers reporting in detail on both experimental and theoretical ELNES of the N K-edge in AlN, GaN and/or AlGaN have been published (Radtke et al., 2003, 2004; Mizoguchi et al., 2003, 2004; Gao et al., 2004; Holec et al., 2008a). However, none of these works mentions the effect of strain which is almost certainly present in the experimental epitaxial samples.

In this chapter first principle calculations of AlN, GaN and their ternary alloy AlGaN are reported on. Strain effects on the projected density of states for several deformation modes are investigated aiming to predict the actual influence on the N K-edge. The

conclusions are subsequently confirmed by calculations of the actual ELNES in some special cases. In the end is discussed how the band gap depends on the strain state in the material. The following results were obtained in collaboration with the Max-Planck Institut für Eisenforschung in Düsseldorf, Germany.

9.2 Methodology

9.2.1 Calculation details

The calculations were carried out using the density functional theory as implemented in the plane wave pseudopotential code `S/PHI/nX` (Boeck et al., 2003). Troullier-Martin pseudopotentials (Troullier and Martins, 1991) with the plane-wave cutoff of 70 Ry were used. The parametrisation of Perdew et al. (1996) to the generalised gradient approximation (GGA) was utilised for the exchange-correlation scheme. An equivalent of $6 \times 6 \times 2$ Monkhorst-Pack k-point sampling (Monkhorst and Pack, 1976) for the unit cell was applied. All structures were optimised by a state-by-state conjugate gradient minimisation scheme developed by Payne et al. (1992). The energy convergence factor was set to 10^{-9} Hartree and the mixing method of Pulay (1980) was used for the charge density mixing. The full relaxation of the lattice parameters and internal coordinates was ensured for all the structures.

Binary and ternary alloys were described as $1 \times 1 \times 1$ and $2 \times 2 \times 1$ supercells, respectively. In order to mimic a random structure within a finite-size supercell with periodic boundary conditions, the theory of special quasi-random structures (SQS) was employed (Wei et al., 1990). Specifically, the SQS for $Al_{0.5}Ga_{0.5}N$ was adopted after Lymperakis (2005). The specific configurations of the $Al_{0.25}Ga_{0.75}N$ and $Al_{0.75}Ga_{0.25}N$ SQSs were derived using the same approach and are visualised in Fig. 9.1.

PDOS was calculated for all atoms and for all s, p and d angular components, although only p-PDOS is discussed here. The number of empty states was 30. The energy range

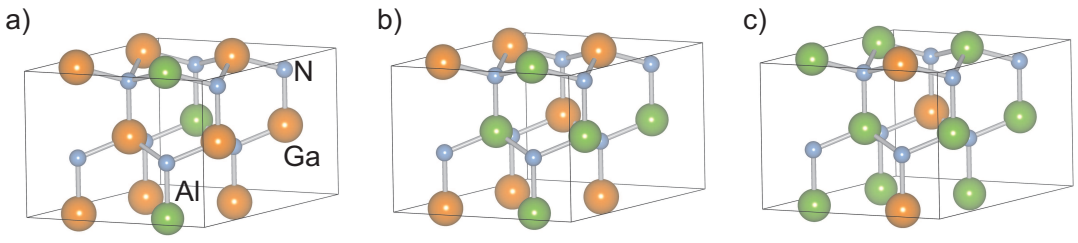


Figure 9.1: Configurations of atoms in SQS of ternary alloys: a) $Al_{0.25}Ga_{0.75}N$ b) $Al_{0.5}Ga_{0.5}N$ c) $Al_{0.75}Ga_{0.25}N$.

used in our calculations was from 0 eV to 30 eV sampled with 3 000 points. The top of the valence band after the self-consistent calculations of charge density was ≈1–5 eV. Therefore, the chosen range guarantees coverage of the unoccupied states up to 20 eV, which is the most interesting range for ELNES. The 2.5 Bohr cut-off radius as radial localisation and 0.01 Bohr smearing factor was used. The Gaussian broadening was 0.05 eV.

All band gaps in the section 9.3.3 were evaluated at the Γ-point.

The all electron full potential `Wien2k` code (Blaha et al., 2000) was used for the calculations of actual ELNES spectra of binary alloys. The GGA-PBE parametrisation with up to 4 000 k-points in the unit cell was used. A $2 \times 2 \times 2$ supercell was utilised for the core hole calculations.

9.2.2 Deformation modes

The ground state wurtzite structures were optimised by the multiple step process. First, the lattice constant a, the ratio c/a and the internal parameter u were optimised. The ideal c/a ratio ($\sqrt{8/3}$) and the parameter $u = 3/8$ were assumed for the initial geometry. The equilibrium volume was determined by variation of the a and c parameters. The optimal value of parameter u was guaranteed by the full relaxation of the atomic coordinates during each calculation.

Several deformation modes were simulated in order to determine a response of the PDOS spectra to the strain in structures. These were the uni-axial strain along the c-direction, bi-axial strain in the c-plane, deformation with keeping the unit cell volume V_0 constant, and the bi-axial stress case with $\sigma_{zz} = 0$. They are shown schematically in Fig. 9.2 and the ranges for deformations are given in Table 9.1. The first two uni-axial deformation modes are difficult to realise experimentally, however they are useful for

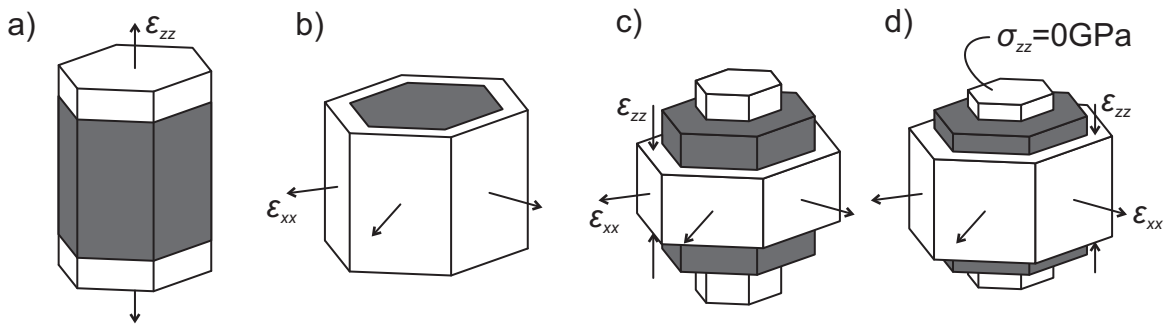


Figure 9.2: Deformation modes investigated in this paper: a) uni-axial strain along the c-direction, b) bi-axial strain in the c-plane, c) deformation with constant unit cell volume V_0, and d) bi-axial stress with $\sigma_{zz} = 0$.

understanding the more complex and more realistic deformation modes. The different deformations were controlled by variation of corresponding lattice parameters. The absolute strain values about 0.01 correspond approximately to unrelaxed $Al_{0.5}Ga_{0.5}N$ on either GaN or AlN and thus are perfectly realistic.

	a) $a = \text{const.}$ $\updownarrow\ c$	b) $\leftrightarrow\ a$ $c = \text{const.}$	c) $\leftrightarrow\ a$ $\updownarrow\ c$	d) $\leftrightarrow\ a$ $\updownarrow\ c$ (relaxed)
$\epsilon_{xx}^{\min}$	0.000	−0.012	−0.012	−0.012
$\epsilon_{xx}^{\max}$	0.000	0.012	0.012	0.012
$\epsilon_{zz}^{\min}$	−0.023	0.000	−0.023	−0.006
$\epsilon_{zz}^{\max}$	0.024	0.000	0.024	0.007
$\sigma_{xx}^{\min}$ [GPa]	−8.75	−2.18	−6.45	−8.14
$\sigma_{xx}^{\max}$ [GPa]	8.57	2.30	6.40	7.97
$\sigma_{zz}^{\min}$ [GPa]	−2.26	−8.00	−5.79	0.00
$\sigma_{zz}^{\max}$ [GPa]	2.21	8.47	6.21	0.00

Table 9.1: Ranges of strains and stresses used in the present calculations. a) Uni-axial strain along the c-direction, b) bi-axial strain in the c-plane, c) deformation with constant unit cell volume V_0, and d) bi-axial stress with $\sigma_{zz} = 0$.

9.3 Discussion of results

9.3.1 Projected density of states

The site and symmetry projected density of unoccupied states (PDOS) up to 15–20 eV above the top of the valence band was calculated for all four deformation modes and five Al fractions. Since the PDOS is highly sensitive to the local atom environment, it varies for different N sites in the alloy. The electron loss process can take place on all of these atoms with the same probability. As a consequence, the PDOSs were averaged over all N atoms in order to obtain one representative spectrum for the bulk material.

Figure 9.3 shows N p-PDOS of unstrained GaN and AlN, and of AlN compressively strained to $\epsilon_{xx} = -0.012$ in the mode with constant unit cell volume. Dotted lines represent spectra with a small broadening parameter of 0.05 eV. Although many features can be observed in the spectrum as well as the overall triple peak (AlN) and single central peak (GaN) shape, it is very hard to extract any trends between them. A much clearer picture of the main features is obtained for a bigger broadening of 1.0 eV (solid

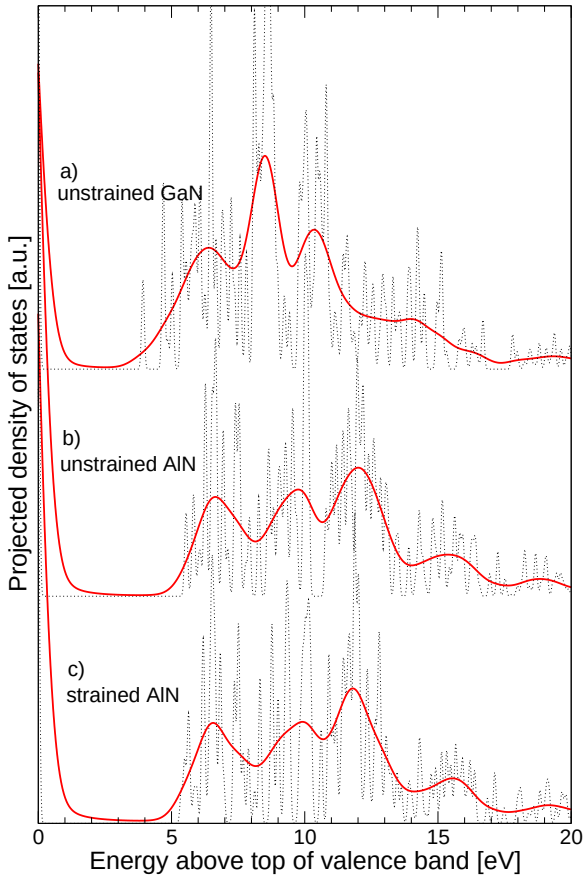


Figure 9.3: Calculated N p-PDOS of a) relaxed GaN, b) relaxed AlN, and c) strained AlN (see text for more details) using broadening of 0.05 eV (dotted lines) and 1.0 eV (solid lines).

lines). As such shapes are much closer to what can be obtained experimentally, the rest of this study focuses on the broadened spectra.

The evolution of unoccupied N p-states with aluminium fraction is shown in Fig. 9.4. The structures were deformed using the constant unit cell volume mode in which the tensile strain in the c-plane is coupled with the compressive strain along the c-direction and vice versa. On the Al rich side (Fig. 9.4a and 9.4b) it is clearly observed that when going from tension ($\epsilon_{xx} > 0$) to compression ($\epsilon_{xx} < 0$) in the c-plane (i) the first (at about 6 eV) and third (at about 11–12 eV) peak sharpen and, (ii) at the same time the second peak (at about 9–10 eV) diminishes. On the other hand, inspection of the Ga rich side (Fig. 9.4e) reveals only very small strain induced variations in the spectrum shape.

Figure 9.5 shows responses of the unoccupied N p-states of AlN to various deformation modes. The uni-axial stain along the c-axis and the bi-axial strain in the c-plane are depicted in Fig. 9.5a and 9.5b. It follows that the response is much more pronounced in the case of the deformation along the c-direction than in the bi-axial strain case.

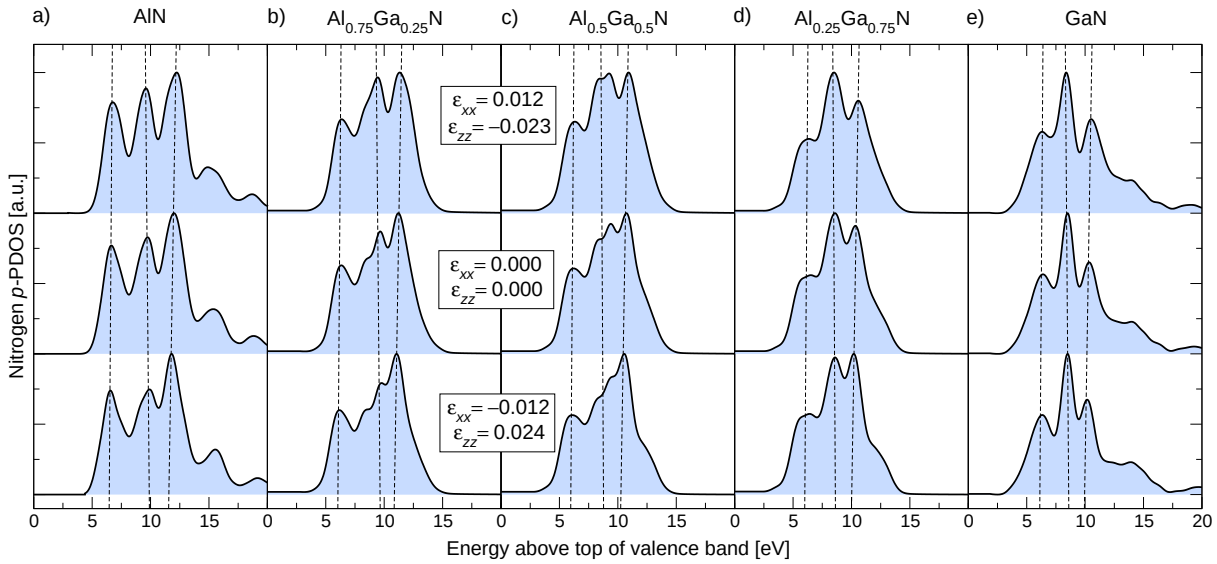


Figure 9.4: Unoccupied broadened N p-PDOS of $Al_xGa_{1-x}N$ for a) $x = 1$, b) $x = 0.75$, c) $x = 0.5$, d) $x = 0.25$, and e) $x = 0$ corresponding to the deformation mode with constant unit cell volume.

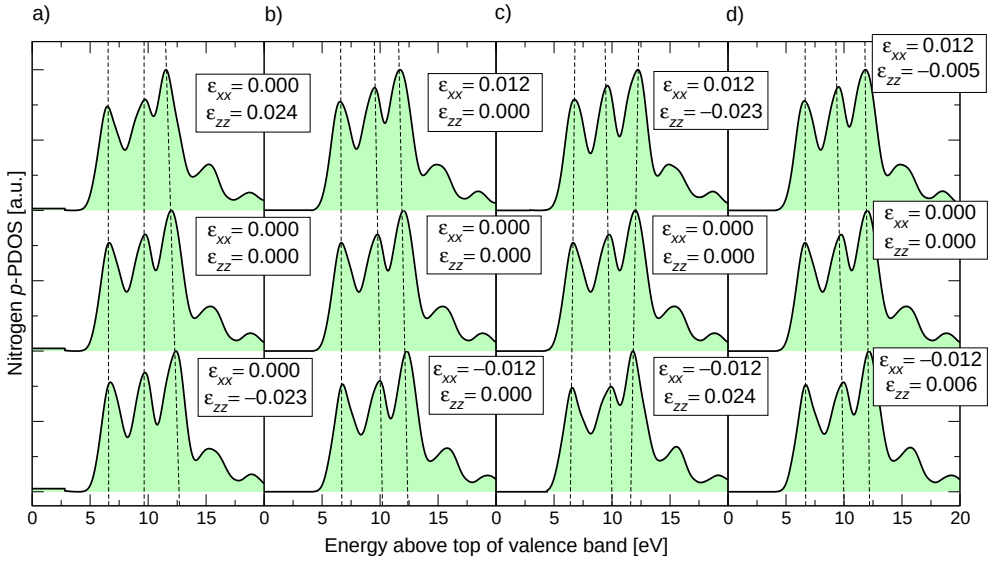


Figure 9.5: Effect of various strain/stress modes on the unoccupied broadened N p-PDOS of AlN: a) uni-axial strain along the c-direction (a was kept constant), b) bi-axial strain in the c-plane (c was kept constant), c) deformation with unit cell volume kept constant, and d) bi-axial stress in the c-plane with $\sigma_{zz} = 0\,\text{GPa}$.

Fig. 9.5c corresponding to the deformation mode with constant unit cell volume reveals the same behaviour as the uni-axial strain case. It can be concluded therefore, that in this case, it is the deformation along the c-axis which is responsible for the spectral changes. The last graph in Fig. 9.5d shows the PDOS redistribution in the bi-axial stress state. In this case, the spectrum varies only slightly which is consistent with only very small deformations along the c-axis (see Table 9.1).

Apart from the changing peak shapes, systematic shifts of the peak positions are also predicted. Table 9.2 lists energies and their differences for the main three peaks of AlN in all the four deformation modes. In the uni-axial and bi-axial strain deformation modes, the energy difference between the first and the second peak changes only very little. The same holds also for the energy difference between the second and third peak in the case of the bi-axial strain. In contrast, a shift as big as 0.7 eV over the chosen strain range is obtained in the case of uni-axial deformation. The trends of these two modes add up for the deformation with constant unit cell volume. Since the uni-axial and bi-axial strain deformations act against each other (e.g. tensile strain along the c-axis requires compressive strain in the c-plane), the distance of the second and third peak varies only about 0.6 eV. At the same time, a pronounced effect on the distance between the first and second peak is obtained. Finally, in the case of the bi-axial stress deformation, the peak positions remain practically invariant.

	ϵ_{xx}	ϵ_{zz}	E_1 [eV]	E_2 [eV]	E_3 [eV]	$E_2 - E_1$ [eV]	$E_3 - E_2$ [eV]	$E_3 - E_1$ [eV]
a = const.	0.000	0.021	6.5	9.7	11.6	3.2	1.9	5.1
	0.000	0.000	6.6	9.8	12.0	3.2	2.2	5.4
	0.000	−0.020	6.8	9.8	12.4	3.0	2.6	5.6
c = const.	−0.010	0.000	6.8	10.0	12.3	3.2	2.3	5.5
	0.000	0.000	6.7	9.8	12.0	3.1	2.2	5.3
	0.010	0.000	6.6	9.6	11.8	3.0	2.2	5.2
V = const.	−0.010	0.021	6.6	9.9	11.8	3.3	1.9	5.2
	0.000	0.000	6.6	9.8	12.0	3.2	2.2	5.4
	0.010	−0.020	6.7	9.6	12.1	2.9	2.5	5.4
bi-axial stress	−0.010	0.006	6.7	9.9	12.1	3.2	2.2	5.4
	0.000	0.000	6.6	9.8	12.0	3.2	2.2	5.4
	0.010	−0.005	6.6	9.6	11.9	3.0	2.3	5.3

Table 9.2: Energies and relative distances between the first three peaks in N p-PDOS in different deformation modes.

Lawniczak-Jablonska et al. (2000) and Radtke et al. (2004) showed that the π^* anti-bonding orbital corresponding to the N p_z-PDOS and orientated along the c-axis contributes mainly to the first and third peak of corresponding N K-edge ELNES, whereas the in-plane σ^* orbitals contribute to all three peaks. This is consistent with the above described behaviour of the p-PDOS. Deformations along the c-axis are expected to influence mainly the p_z states (the π^* orbital), i.e. to change the intensity of the first and

third peak. As a consequence, the relative intensities of the three peaks are expected to change. On the other hand, deformations within the c-plane would affect mainly the $p_x + p_y$ states (σ^* orbitals). Therefore, all three peaks would change in more or less the same way, i.e. the relative intensities are not expected to vary significantly.

The above described observations have several implications. (i) The second peak in the AlN triplet seems to be highly sensitive to strain along the c direction. (ii) Since the physically most relevant strain/stress state for polar (Al,Ga)N material, the bi-axial stress $\sigma_{xx} = \sigma_{yy} \neq 0\,\text{GPa}$ with $\sigma_{zz} = 0\,\text{GPa}$, shows only a very weak dependence on the actual strain, the experimental N K-edge of AlN is expected to be practically strain-independent. The same behaviour was predicted for all studied alloys with different Al fraction (not shown here). (iii) When switching from polar to non- and semi-polar material, which are currently of a great experimental interest, the actual strain-induced changes are expected to emerge due to different (and much more pronounced) deformations along the c-axis.

9.3.2 Nitrogen K-edge ELNES

In order to confirm the conclusions drawn for N p-PDOS, the N K-edge ELNES spectra for several special cases were calculated using the `Wien2k` code. The correlation between the PDOS and ELNES is shown in Fig. 9.6 for AlN and GaN. In these cases, the

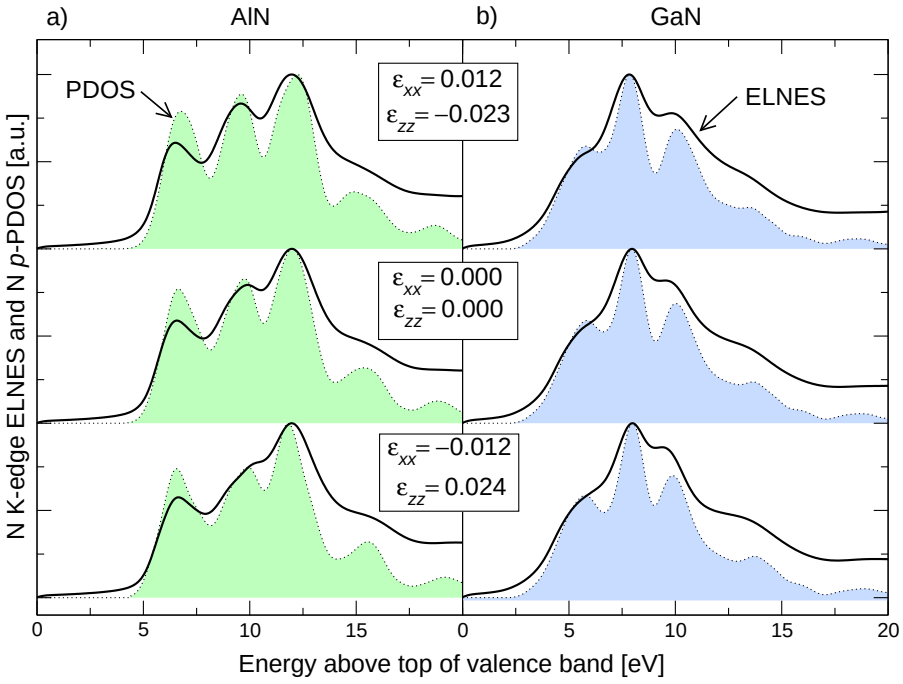


Figure 9.6: Comparison of broadened unoccupied p-PDOS of nitrogen with N K-edge ELNES: a) AlN and b) GaN.

ELNES spectra correspond to ground state calculations (zero core hole), and thus are compatible with the pseudopotential based PDOSs. The same features and the same trends including sharpening the peaks, peak position shifts and relative peak intensity changes are observed. Therefore, conclusions from the previous section are expected to hold for N K-edge ELNES.

A series of N K-edge ELNES spectra for AlN and GaN under bi-axial stress with $\sigma_{zz} = 0\,\text{GPa}$ was calculated. A core hole charge of 0.5 e was used for the calculations employing $2 \times 2 \times 2$ supercells. The predictions for ELNES behaviour based on p-PDOS are qualitative. Since the variations in p-PDOS in the case of the bi-axial stress were only very small, it was concluded in the previous section that ELNES would not be sensitive to strain in this deformation mode. However, the precise calculations of ELNES do reveal some slight variations which, in principle, should be possible to observe experimentally.

The results of the calculations are shown in Fig. 9.7. The most pronounced variation over the range of calculated deformations is the shift of the peak at about 25–30 eV, highlighted by arrows in the figure. Unfortunately, the pseudopotential calculations of p-PDOS did not cover this peak due to computational limitations. In both cases, AlN and GaN, it can be observed the this peak shifts more than 2 eV when changing from tensile ≈ 0.02 to compressive ≈ -0.02 strain. There is also a slight shift of the edge onset energy for both materials. This is consistent with experimental observations and theoretical results of Keast et al. (2003). The main three-peak structure of AlN widens with increasing tensile strain. On the contrary, the second shoulder at about 13–14 eV appears in the GaN spectrum when changing from compression to tension in the c-plane.

9.3.3 Band gap of strained AlGaN

AlN, GaN and their ternary alloy are direct band gap semiconductors (Shan et al., 1999). Therefore, energy gaps between the top of the valence band and the bottom of the conduction band at the Γ-point were calculated. The resulting values as functions of the in-plane strain ϵ_{xx} in the bi-axial stress deformation mode are plotted in Fig. 9.8a. The absolute values of band gap are underestimated by 1–2 eV when compared with the experimental values. It can be observed that the band gaps are largest for compressive strains in the c-plane and gradually decrease with ϵ_{xx}, changing through the unstrained state to tension. The variations are as big as 0.2 eV and are largest for AlN, decreasing slightly with decreasing Al fraction. The strains used in the present calculations correspond roughly to $-8\,\text{GPa} < \sigma_{xx} < 7\,\text{GPa}$.

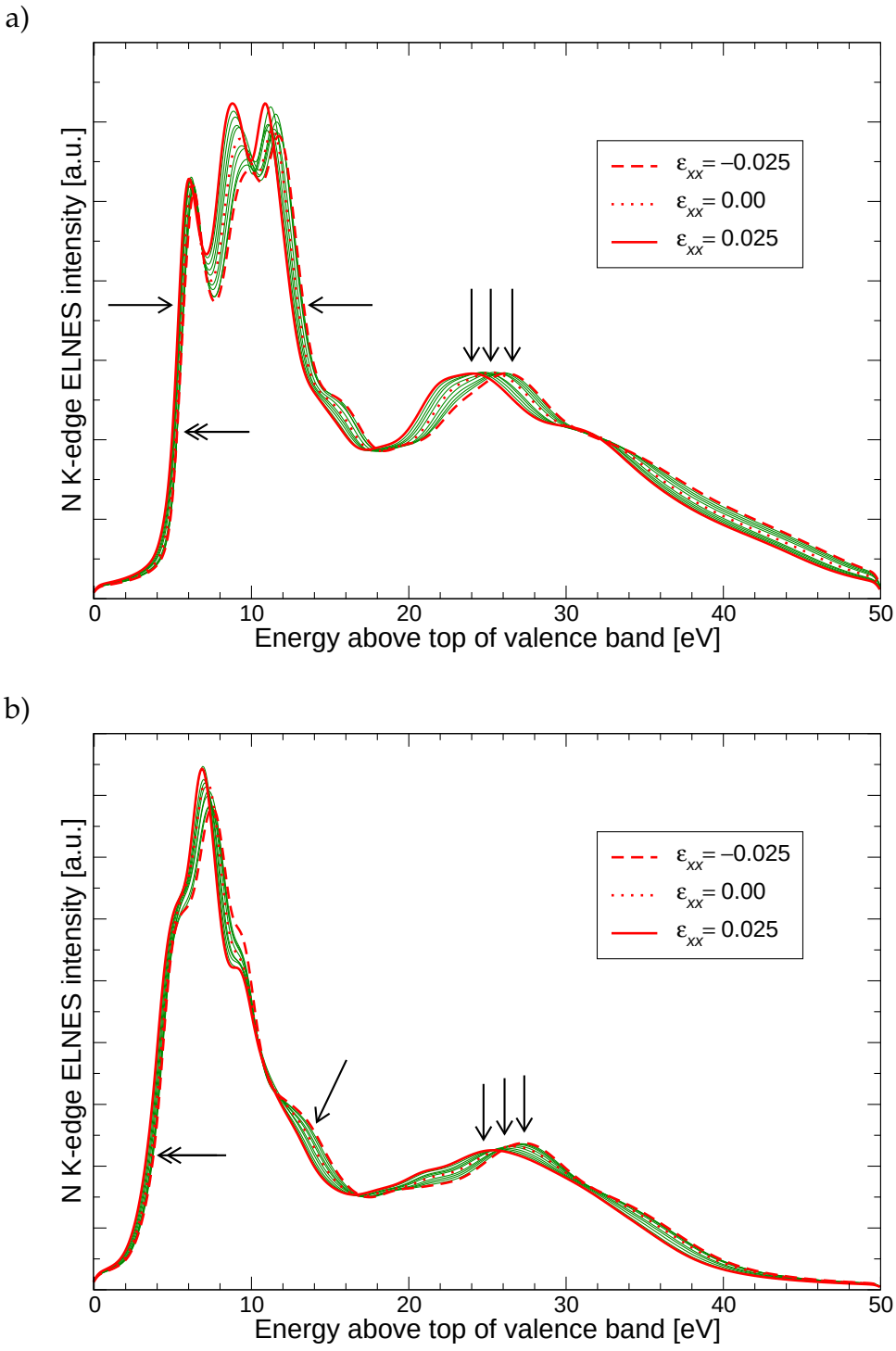


Figure 9.7: Calculated strain-induced variations of the N K-edge ELNES: a) AlN and b) GaN.

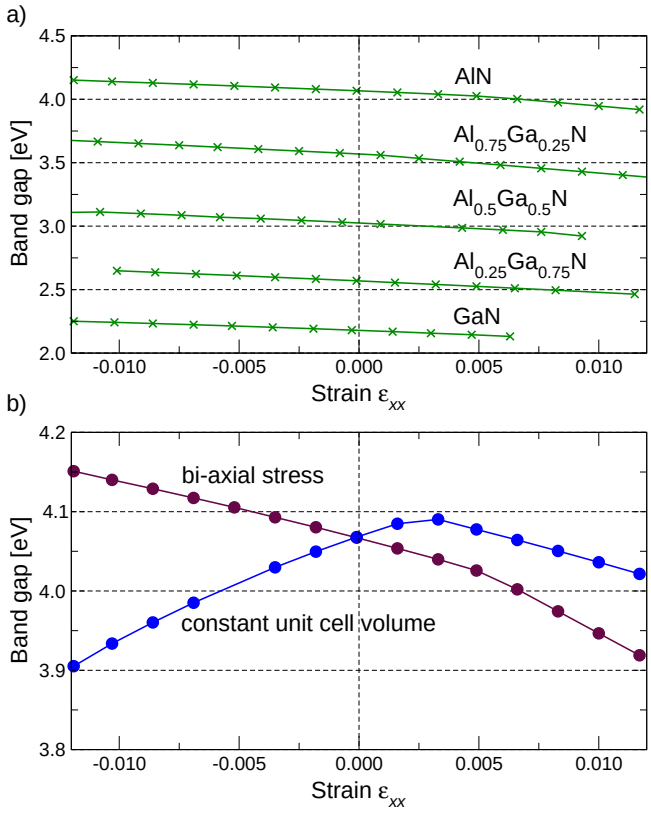


Figure 9.8: a) Calculated band gap of $Al_xGa_{1-x}N$ ternary alloy under bi-axial stress with $\sigma_{zz} = 0\,\text{GPa}$. b) Band gap of AlN in two different deformation modes (see text for more details).

Shan et al. (1999) reported on the experimental determination of the band gap of the ternary $Al_xGa_{1-x}N$ alloy as a function of applied hydrostatic pressure. They observed variations of about 0.2–0.3 eV for the range of applied hydrostatic pressures 0–9 GPa. At the same time, the largest variations were observed in the sample with the largest Al fraction. Although their deformation mode is different from the bi-axial stress used in the present study, the results are in very good agreement.

Lastly, the band gap as a function of the in-plane strain ϵ_{xx} is plotted in Fig. 9.8b for two different deformation modes, with the constant unit cell volume and for the bi-axial stress with $\sigma_{zz} = 0\,\text{GPa}$. In the bi-axial stress case, the dependence is linear although the slope changes at about $\epsilon_{xx} = 0.005$. The range of band gap values for the constant volume deformation mode is practically the same, however the quality of the dependence changes. The band gap increases when going from the compressively strained structure to the unstrained structure, reaches a maximum value for slightly tensile strain of about 0.003, and then decreases again. This result indicates that the band gap is very sensitive to the actual type of the strain/stress state in the material.

9.4 Conclusions

This chapter reports on first principle calculations of strained ternary $Al_xGa_{1-x}N$ alloys employing pseudopotential and all-electron full-potential methods. In particular, the relationship between the broadened unoccupied N *p*-PDOS and the N K-edge ELNES was studied. The following main conclusions were drawn:

- The second peak at just below 10 eV above the top of the valence band is highly sensitive to deformations along the *c*-axis.
- The energy difference between the second and third peak varies with strain: it increases when going from compressive to tensile strain. The variations are more sensitive to deformation along the *c*-axis than to the strain in the *c*-plane.
- Only very small variations in the broadened N *p*-PDOS in the bi-axial stress mode were observed. As a consequence, the N K-edge is expected to be mostly strain-invariant in *c*-plane AlGaN structures in a bi-axial stress state. However, precise calculations of theoretical N K-edge ELNES revealed some fine trends depending on the applied strain.
- Strain-induced shape variations of the N K-edge ELNES are expected for the non- and semi-polar material.
- The calculations also showed significant variations of the band gap, which strongly depend on the actual strain state in the material.

Chapter 10

Conclusions and future work

10.1 Summary of results

This thesis presents an overview of theoretical materials modelling on two different scales. The first part contains results of the general theory of dislocations in elastic continuum applied to wurtzite III-nitride material with a subsequent utilisation in studies of dislocation behaviour in heterostructures. The second part deals with *ab initio* calculation of electronic properties of binary and ternary III-N compounds with respect to the modelling of the electron energy loss near edge structures.

A summary of anisotropic theory of dislocations, as derived within the linear theory of elasticity, was given in chapter 2. The derived formulae were used for investigation of how geometrical properties of straight dislocations in an infinite medium differ in hexagonal materials with respect to isotropic media. It was shown that the often used argument "the screw-type dislocation has lower energy than the mixed- and edge-type dislocations" does not hold in the case of wurtzite GaN. The second study focused on an estimate of the energetically optimal dislocation line directions near a free surface of a finite crystal. Here, the hexagonal theory predicts that the dislocation line along the $[0001]$ direction possesses the lowest energy configuration regardless of the dislocation type which is in contrast to the isotropic theory where the $(a+c)$-type dislocations are predicted to be inclined from the the $[0001]$ direction.

In chapter 3, the equilibrium critical thickness for generation of misfit dislocations of AlGaN/GaN and InGaN/GaN systems was calculated using the energy balance model. This model was corrected for the hexagonal symmetry of wurtzite III-nitrides. The best agreement with experimental data was obtained when the dislocation core energy was included. The dislocation core energy had been taken from published atomistic studies. This connection of continuum and atomistic level modelling was

done here for the first time for III-nitrides and proved to be an important improvement towards agreement of the predicted with the experimental results.

A 2D model for predicting the threading dislocation density reduction during the growth of GaN was presented in chapter 4. The main purpose of the study was to estimate qualitatively the influence of islands on the TDD reduction, and to investigate trends linked to various parameters (such a geometry) of the proposed model. A good qualitative agreement with experiments was obtained. Therefore it is believed that the model as it stands, whilst highly simplified, is capable of *qualitative* predictions of various trends.

Chapter 7 presented a systematic study of nitrogen K-edge electron energy loss near edge structures of AlN, GaN and InN. The concept of the core hole was employed in these studies. It was noted that the core hole is an appropriate approach to the approximation of the excited state of a material using ground state density functional theory. A comparison with several available experimental spectra from the literature revealed that the best agreement is obtained for a core hole charge of about 0.6–0.7 e for AlN, about 0.4–0.5 e for GaN and about 0.4 e for InN. The variations in N K-edge shape with core hole charge were traced to rearrangement of the final p states of nitrogen. Generally speaking, they move towards lower energies with increasing core hole charge. This effect was ascribed to a weaker shielding effect in the presence of a core hole (i.e. in the absence of a (fraction of a) core electron). Differences in bonding in terms of the valence change density and the overlap of valence states of AlN, GaN, and InN were discussed as well.

The concept of special quasi-random structures (SQS) was employed in chapter 8 to select $2 \times 2 \times 2$ supercells containing 32 atoms to represent wurtzite structures of various compositions. Although the SQS method is not new, a description of the supercells for utilisation with the wurtzite III-nitrides is missing in the literature. These were subsequently used for the calculation of ELNES N K-edge evolution of $Al_xGa_{1-x}N$, $Al_xIn_{1-x}N$, and $In_xGa_{1-x}N$ ternary alloys. The methodology employed proved to be able to reasonably reproduce experimental results. A strong dependence on the nearest neighbourhood of nitrogen atoms was found which we suggest justifies the use of $2 \times 2 \times 2$ supercells which are often regarded as too small for the core hole/alloy calculations.

In the last chapter with results, chapter 9, first principle calculations of strained ternary $Al_xGa_{1-x}N$ alloys were reported. A study combining the advantages of the fast pseudopotential methods (which are, however, unable to address the core electron properties) and the expensive all electron calculation was used. The relationship between

the broadened unoccupied N p-PDOS and the N K-edge ELNES was studied for four deformation modes (two uni-axial modes, a bi-axial mode and a mode with constant volume). It was concluded that for the realistic values of bi-axial strain/stress the N K-edge of AlGaN structures is influenced only very little. However, other deformation modes predict much stronger dependence of ELNES on the actual strain state and thus strain-induced shape variations of the N K-edge ELNES are predicted for the non- and semi-polar material. The calculations also showed significant variations of the band gap, which strongly depend on the actual strain state in the material.

10.2 Proposed future work

This work covers a wide range of problems. As a consequence, it leaves a long "to do" list. In this section are suggested some directions of possible future studies based on the conclusions of this work.

The critical thickness calculations are of a great interest for experimentalists. In order to improve the current models (not only for III-nitrides), it would be interesting to account for the dislocation generation mechanism(s) and thus to provide a theory which goes beyond the equilibrium critical thickness as presented here. Other interesting improvements would include the free surface (modification of the dislocation stress field) and would address the differences in the material properties of substrates and epilayers. However, the wide scatter of experimental data suggests that an important role is played by the actual growth technique which goes beyond the theory of dislocations. Another possible direction would be to try to incorporate into the model other defects such as the V-pits.

The modelling of threading dislocation density reduction as described here was only a small step towards a model capable of quantitative predictions. Future work may include further studies on the effects of different parameters of the model as well as new geometries (including different facets, different growth rates of the facets). In addition, using the knowledge gained with the 2D model, an importance advance can be achieved by extending this model to 3D. Another significant improvement toward realism would be to consider the actual dislocation–dislocation interactions. This would involve calculation of dislocation stress fields and, in fact, employing some methods of dislocation dynamics. Such work would involve a major effort in the development and employment of new analytical and numerical methods and significant optimisation of the code used for simulations. However, a model capable of *quantitative* predictions is expected as an outcome.

In the present calculations of ELNES, all the structures were unrelaxed due to limited computing resources. This is an obvious drawback of the present work – the electronic structure, PDOS and thus ELNES is expected to change after the relaxation. Therefore, it would be interesting to see how much the III-nitrides properties actually change after the relaxation is performed. It worth noting that the structures (the lattice and u parameters) used here were taken from published studies in which the structures had been optimised. As a consequence, no significant deviations are expected, certainly not for the greatly broadened ELNES spectra. More interesting work could be done with the strained effects. Band structures of the deformed materials should be studied to reveal differences in electronic properties (e.g. if the band gap stays direct with the deformation). Additionally, similar studies may be applied to other materials of interest, namely InGaN and AlInN. Another possible improvement is to employ methods such as GGA+U for more precise studies of the electronic structure.

With the steadily increasing computational power available it may become possible in the near future to combine knowledge of the dislocation theory with *ab initio* calculations to study (electronic) properties of dislocations.

Appendix A

Convergence tests of the Wien2k calculations

Convergence results of AlN and GaN calculations using the `Wien2k` code with respect to the number of k-points and $R_{MT}K_{\max}$ (as described in section 7.2).

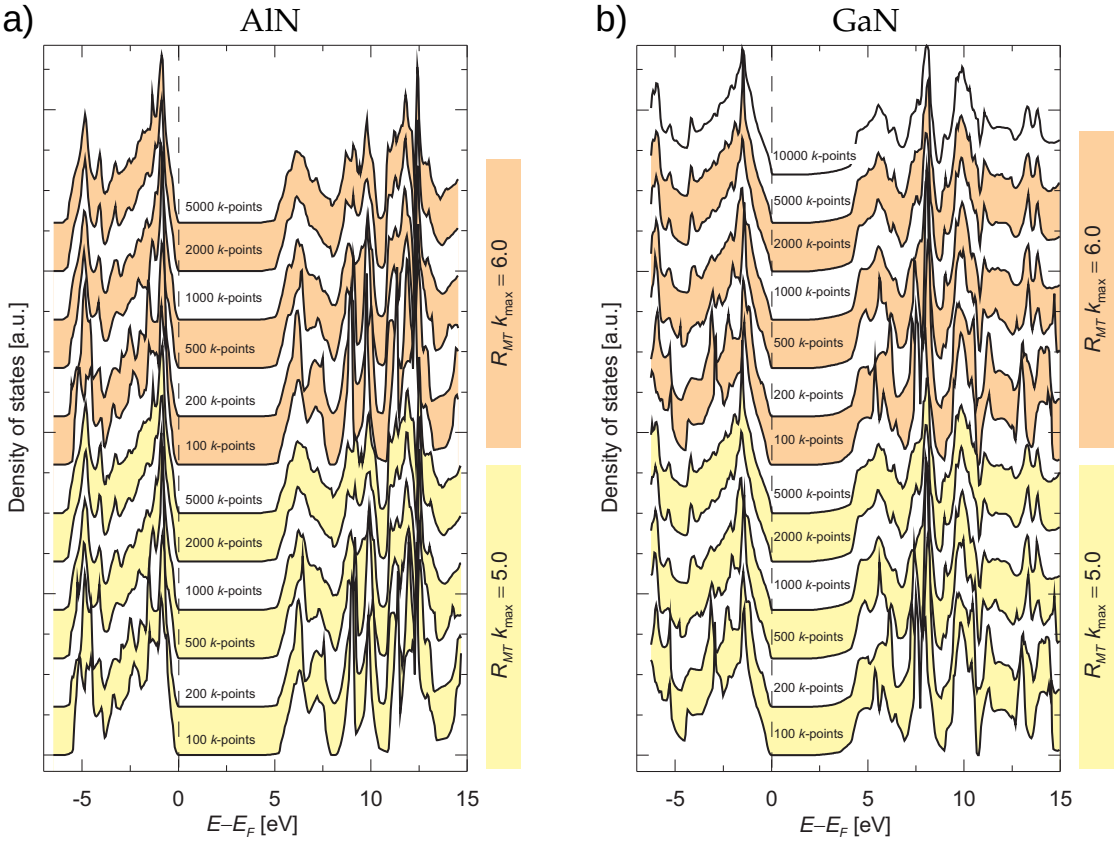


Figure A.1: Dependence of the total density of states on the number of k-points and $R_{\mathrm{MT}}K_{\max}$ for the ground state of a) AlN and b) GaN.

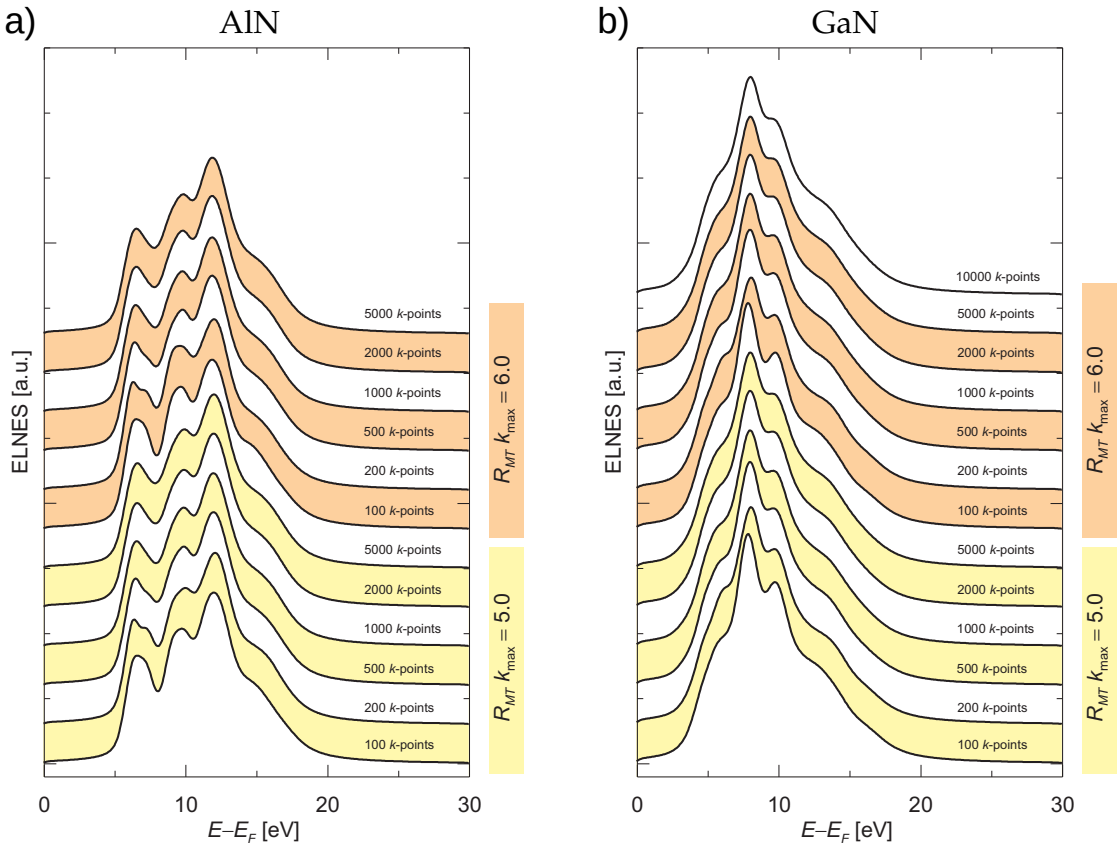


Figure A.2: Dependence of the N K-edge ELNES on the number of k-points and $R_{\mathrm{MT}}K_{\mathrm{max}}$ for the ground state of a) AlN and b) GaN.

Appendix B

Features of the experimental N K-edge ELNES of AlN and GaN

The following two tables provide data from the analysis of the experimental N K-edge ELNES spectra of AlN and GaN as described in section 7.3.1. The data are graphically visualised in Figs. 7.2 and 7.3.

	I_1 [a.u.]	I_2 [a.u.]	I_3 [a.u.]	I_1/I_2	I_3/I_2	I_1 [eV]	I_2 [eV]	I_3 [eV]	I_2-I_1 [eV]	I_3-I_1 [eV]	$\frac{I_3-I_1}{I_2-I_1}$
0 e	2.55	4.70	3.80	0.543	0.809	5.54	7.95	9.49	2.41	3.95	1.640
0.1 e	2.71	4.71	3.71	0.575	0.787	5.45	7.76	9.30	2.31	3.86	1.667
0.2 e	2.91	4.71	3.51	0.618	0.745	5.45	7.66	9.30	2.22	3.86	1.739
0.3 e	3.12	4.72	3.42	0.661	0.724	5.40	7.52	9.30	2.12	3.90	1.841
0.4 e	3.37	4.72	3.27	0.714	0.693	5.35	7.33	9.30	1.98	3.95	2.000
0.5 e	3.58	4.68	3.08	0.765	0.658	5.30	7.13	9.25	1.83	3.95	2.158
0.6 e	3.78	4.73	2.98	0.799	0.630	5.30	6.99	9.20	1.69	3.90	2.314
0.7 e	4.19	4.74	2.79	0.884	0.588	5.06	6.80	9.11	1.73	4.05	2.333
0.8 e	4.64	4.74	2.64	0.979	0.557	4.77	6.55	9.11	1.78	4.34	2.432
0.9 e	4.70	4.68	2.45	1.004	0.523	4.92	6.41	9.06	1.49	4.14	2.774
1 e	4.70	4.35	2.25	1.080	0.517	4.67	6.31	8.96	1.64	4.29	2.618
Radtke et al. (2004)	1.75	2.15	1.55	0.814	0.721	401.50	403.25	405.00	1.75	3.50	2.000
Lazar et al. (2004)	2.25	3.55	1.95	0.634	0.549	5.00	7.37	10.00	2.37	5.00	2.110
Delft	1.65	2.50	1.35	0.660	0.540	405.65	408.00	410.17	2.35	4.52	1.926
Cambridge	2.75	3.40	2.55	0.809	0.750	405.00	407.84	410.54	2.84	5.54	1.952

Table B.1: Comparison of peak positions, relative positions, intensities and relative intensities of the GaN N K-edge calculated for various core hole charges with the experimental observation.

	I_1 [a.u.]	I_2 [a.u.]	I_3 [a.u.]	I_1 / I_2	I_3 / I_2	I_1 [eV]	I_2 [eV]	I_3 [eV]	$I_2 - I_1$ [eV]	$I_3 - I_1$ [eV]	$\frac{I_3 - I_1}{I_2 - I_1}$
0 e	0.81	1.07	1.34	0.756	1.252	6.55	9.90	11.95	3.35	5.40	1.612
0.1 e	0.42	0.54	0.66	0.780	1.220	6.50	9.80	11.85	3.30	5.35	1.621
0.2 e	0.44	0.55	0.64	0.802	1.161	6.40	9.70	11.70	3.30	5.30	1.606
0.3 e	0.47	0.56	0.62	0.832	1.113	6.35	9.65	11.60	3.30	5.25	1.591
0.4 e	0.49	0.56	0.60	0.871	1.072	6.30	9.55	11.45	3.25	5.15	1.585
0.5 e	0.51	0.57	0.58	0.909	1.029	6.20	9.30	11.35	3.10	5.15	1.661
0.6 e	0.54	0.58	0.56	0.929	0.975	6.15	9.20	11.15	3.05	5.00	1.639
0.7 e	0.56	0.58	0.53	0.975	0.923	6.10	9.00	11.00	2.90	4.90	1.690
0.8 e	0.58	0.56	0.49	1.039	0.873	6.00	8.80	10.85	2.80	4.85	1.732
0.9 e	0.62	0.56	0.47	1.096	0.828	5.85	8.60	10.65	2.75	4.80	1.745
1 e	0.64	0.56	0.45	1.134	0.802	5.75	8.45	10.45	2.70	4.70	1.741
Radtke et al. (2004)	2.15	2.15	2.35	1.000	1.093	405.00	408.10	410.80	3.10	5.80	1.871
Delft	2.55	2.45	2.30	1.041	0.939	405.00	407.65	409.90	2.65	4.90	1.849
Cambridge	2.85	3.00	3.10	0.950	1.033	405.00	407.94	410.00	2.94	5.00	1.700

Table B.2: Comparison of peak positions, relative positions, intensities and relative intensities of the AlN N K-edge calculated for various core hole charges with the experimental observation.

Appendix C

Structural correlation coefficients of the special quasi-random structures

Summary of structural correlation coefficients for various compositions x as obtained for the random alloy and for the best matching special quasi-random structures. The second table presents average numbers of opposite type atoms in several nearest neighbour distance as calculated for an ideal random alloy and for the selected special quasi-random structure. The methodology was described in section 8.1.3.

		$\overline{\Pi}_{(1,0)}$	$\overline{\Pi}_{(2,1)}$	$\overline{\Pi}_{(2,2)}$	$\overline{\Pi}_{(2,3)}$	$\overline{\Pi}_{(2,4)}$	$\overline{\Pi}_{(3,2)}$	$\overline{\Pi}_{(3,3)}$	$\overline{\Pi}_{(4,2)}$
$x = 0.0$	RA	−1.000	1.000	1.000	1.000	1.000	−1.000	−1.000	1.000
	SQS	−1.000	1.000	1.000	1.000	1.000	−1.000	−1.000	1.000
$x = 0.0625$	RA	−0.875	0.766	0.766	0.766	0.766	−0.670	−0.670	0.586
	SQS	−0.875	0.750	0.750	0.750	0.750	−0.625	−0.625	0.500
$x = 0.125$	RA	−0.750	0.563	0.563	0.563	0.563	−0.422	−0.422	0.316
	SQS	−0.750	0.583	0.500	0.500	0.500	−0.375	−0.333	0.250
$x = 0.1875$	RA	−0.625	0.391	0.391	0.391	0.391	−0.244	−0.244	0.153
	SQS	−0.625	0.333	0.417	0.500	0.250	−0.250	−0.292	0.250
$x = 0.25$	RA	−0.500	0.250	0.250	0.250	0.250	−0.125	−0.125	0.063
	SQS	−0.500	0.250	0.167	0.250	0.000	−0.063	−0.083	0.000
$x = 0.3125$	RA	−0.375	0.141	0.141	0.141	0.141	−0.053	−0.053	0.020
	SQS	−0.375	0.083	0.083	0.250	0.250	0.000	−0.042	0.000
$x = 0.375$	RA	−0.250	0.063	0.063	0.063	0.063	−0.016	−0.016	0.004
	SQS	−0.250	0.000	0.000	0.000	0.000	0.000	−0.083	0.000
$x = 0.4375$	RA	−0.125	0.016	0.016	0.063	0.016	−0.002	−0.002	0.000
	SQS	−0.125	0.000	−0.083	0.000	−0.250	0.063	−0.042	0.000
$x = 0.5$	RA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SQS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$x = 0.5625$	RA	0.125	0.016	0.016	0.016	0.016	0.002	0.002	0.000
	SQS	0.125	0.000	−0.083	0.000	−0.025	−0.063	0.042	0.000
$x = 0.625$	RA	0.250	0.063	0.063	0.063	0.063	0.016	0.016	0.004
	SQS	0.250	0.000	0.000	0.000	0.000	0.000	−0.083	0.000
$x = 0.6875$	RA	0.375	0.141	0.141	0.141	0.141	0.053	0.053	0.020
	SQS	0.375	0.083	0.083	0.250	0.250	0.000	0.042	0.000
$x = 0.75$	RA	0.500	0.250	0.250	0.250	0.250	0.125	0.125	0.063
	SQS	0.500	0.250	0.167	0.250	0.000	0.063	0.083	0.000
$x = 0.8125$	RA	0.625	0.391	0.391	0.391	0.391	0.244	0.244	0.153
	SQS	0.625	0.333	0.417	0.500	0.250	0.250	0.292	0.250
$x = 0.875$	RA	0.750	0.563	0.563	0.563	0.563	0.423	0.423	0.316
	SQS	0.750	0.583	0.500	0.500	0.500	0.375	0.333	0.250
$x = 0.9375$	RA	0.875	0.766	0.766	0.766	0.766	0.670	0.670	0.586
	SQS	0.875	0.750	0.750	0.750	0.750	0.625	0.625	0.500
$x = 1.0$	RA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	SQS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table C.1: Structural correlation coefficients of the first eight non-zero figures for various compositions x as obtained for the random alloy (RA) and for the best matching special quasi-random structures (SQS).

		nn	nn2	nn3	nn4
$x = 0.0$	RA	0.000	0.000	0.000	0.000
	SQS	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
$x = 0.0625$	RA	0.703	0.703	0.703	0.234
	SQS	0.75 ± 1.48	0.75 ± 1.61	0.75 ± 1.73	0.25 ± 0.68
$x = 0.125$	RA	1.313	1.313	1.313	0.438
	SQS	1.25 ± 1.53	1.50 ± 2.00	1.50 ± 2.19	0.50 ± 0.89
$x = 0.1875$	RA	1.828	1.828	1.828	0.609
	SQS	2.00 ± 1.75	1.75 ± 2.05	1.50 ± 1.90	0.75 ± 1.00
$x = 0.25$	RA	2.250	2.250	2.250	0.750
	SQS	2.25 ± 1.44	2.50 ± 2.00	2.25 ± 2.05	1.00 ± 1.03
$x = 0.3125$	RA	2.578	2.578	2.578	0.859
	SQS	2.75 ± 1.39	2.75 ± 1.91	2.25 ± 2.32	0.75 ± 1.00
$x = 0.375$	RA	2.813	2.813	2.813	0.938
	SQS	3.00 ± 1.26	3.00 ± 1.79	3.00 ± 2.19	1.00 ± 1.03
$x = 0.4375$	RA	2.953	2.953	2.953	0.984
	SQS	3.00 ± 0.82	3.25 ± 1.61	3.00 ± 1.90	1.25 ± 1.00
$x = 0.5$	RA	3.000	3.000	3.000	1.000
	SQS	3.00 ± 0.73	3.00 ± 1.79	3.00 ± 2.19	1.00 ± 1.03
$x = 0.5625$	RA	2.953	2.953	2.953	0.984
	SQS	3.00 ± 0.82	3.25 ± 1.61	3.00 ± 1.90	1.25 ± 1.00
$x = 0.625$	RA	2.813	2.813	2.813	0.938
	SQS	3.00 ± 1.26	3.00 ± 1.79	3.00 ± 2.19	1.00 ± 1.03
$x = 0.6875$	RA	2.578	2.578	2.578	0.859
	SQS	2.75 ± 1.39	2.75 ± 1.91	2.25 ± 2.32	0.75 ± 1.00
$x = 0.75$	RA	2.250	2.250	2.250	0.750
	SQS	2.25 ± 1.44	2.50 ± 2.00	2.25 ± 2.05	1.00 ± 1.03
$x = 0.8125$	RA	1.828	1.828	1.828	0.609
	SQS	2.00 ± 1.75	1.75 ± 2.05	1.50 ± 1.90	0.75 ± 1.00
$x = 0.875$	RA	1.313	1.313	1.313	0.438
	SQS	1.25 ± 1.53	1.50 ± 2.00	1.50 ± 2.19	0.50 ± 0.89
$x = 0.9375$	RA	0.703	0.703	0.703	0.234
	SQS	0.75 ± 1.48	0.75 ± 1.61	0.75 ± 1.73	0.25 ± 0.68
$x = 1.0$	RA	0.000	0.000	0.000	0.000
	SQS	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Table C.2: Average numbers of opposite type atoms in the nearest neighbour distance (nn), second nearest neighbour distance (nn2), third nearest neighbour distance (nn3) and fourth nearest neighbour distance (nn4) in random alloy (RA) and special quasi-random structure (SQS) that best matches random alloy for a given composition.

Bibliography

Akasaki I. and Amano H.: Breakthroughs in improving crystal quality of GaN and invention of the *p–n* junction blue-light-emitting diode. *Jpn. J. Appl. Phys. Part 1*, 45(12):9001–9010, 2006.

Albrecht M., Nikitina I.P., Nikolaev A.E., Melnik Yu.V., Dmitriev V.A., and Strunk H.P.: Dislocation reduction in AlN and GaN bulk crystals grown by HVPE. *Phys. Stat. Sol. (a)*, 176(1):453–458, 1999.

Amano H., Kito M., Hiramatsu K., and Akasaki I.: *p*-type conduction in Mg-doped GaN treated with low-energy electron-beam irradiation (LEEBI). *Jpn. J. Appl. Phys. Part 2*, 28(12):L2112–L2114, 1989.

Andersen O.K.: Linear methods in band theory. *Phys. Rev. B*, 12(8):3060–3083, 1975.

Ankudinov A.L., Ravel B., Rehr J.J., and Conradson S.D.: Real-space multiple-scattering calculation and interpretation of x-ray-absorption near-edge structure. *Phys. Rev. B*, 58(12):7565–7576, 1998.

Bean J.C., Feldman L.C., Fiory A.T., Nakahara S., and Robinson I.K.: Ge_xSi_{1-x}/Si strained-layer superlattice grown by molecular-beam epitaxy. *J. Vac. Sci. Technol. A*, 2(2):436–440, 1984.

Beaumont B., Vennéguès P., and Gibart P.: Epitaxial lateral overgrowth of GaN. *Phys. Stat. Sol. (b)*, 227(1):1–43, 2001.

Belabbas I., Bere A., Chen J., Petit S., Belkhir M.A., Ruterana P., and Nouet G.: Atomistic modeling of the $(a + c)$-mixed dislocation core in wurtzite GaN. *Phys. Rev. B*, 75(11):115201, 2007.

Bethe H.: Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie. *Annalen d. Phys.*, 397(3):325–400, 1930.

Bethoux J.-M. and Vennéguès P.: Ductile relaxation in cracked metal-organic chemical-vapor-deposition-grown AlGaN films on GaN. *J. Appl. Phys.*, 97(12): 123504, 2005.

Blaha P., Schwarz K., Madsen G., Kvasnicka D., and Luitz J.: *Wien2k, an improved and updated version of ab initio calculation package originally published as: P. Blaha, K. Schwarz, P.I. Sorantin, S.B. Trickey, Full-potential linearized augmented plane wave programs for crystalline systems, Comput. Phys. Commun. 59 (1990) 399–415.*, 2000.

Boeck S., Dick A., Freysoldt Ch., Grzegorzewski F., Hammerschmidt T., Ismer L., Lymperakis L., and Wahn M.: *SFHIngX - Users' Guide*. http://www.sfhingx.de, 2003.

Bougrov V.E., Odnoblyudov M.A., Romanov A.E., Lang T., and Konstantinov O.V.: Threading dislocation density reduction in two-stage growth of GaN layers. *Phys. Stat. Sol. (a)*, 203(4):R25–R27, 2006.

Brdička M., Samek L., and Sopko B.: *Mechanika kontinua*. Academia, Praha, 2000.

Cherns P.D.: *A Transmission Electron Microscopy Study of AlGaN/GaN Heterostructures*. PhD thesis, University of Cambridge, 2007.

Cherns P.D., McAlesee C., M.J.Kappers, and C.J.Humphreys: Strain relaxation in an AlGaN/GaN quantum well system. In *Microscopy of Seminconducting Materials 2007 conference*, 2007.

Chichibu S.F., Uedono A., Onuma T., Haskell B.A., Chakraborty A., Koyama T., Fini P.T., Keller S., Denbaars S.P., Speck J.S., Mishra U.K., Nakamura S., Yamaguchi S., Kamiyama S., Amano H., Akasaki I., Han J., and Sota T.: Origin of defect-insensitive emission probability in In-containing (Al, In, Ga) N alloy semiconductors. *Nat. Mater.*, 5(10):810–816, 2006.

Cho H.K.N. and Yang G.M.: Generation of misfit dislocations in high indium content InGaN layer grown on GaN. *J. Cryst. Growth*, 243(1):124–128, 2002.

Chou Y. T. and Eshelby J. D.: The energy and line tension of a dislocation in a hexagonal crystal. *J. Mech. Phys. Solids*, 10(1):27–34, 1962.

Costa P.M.F.J., Datta R., Kappers M.J., Vickers M.E., Humphreys C.J., Graham D.M., Dawson P., Godfrey M.J., Thrush E.J., and Mullins J.T.: Misfit dislocations in In-rich InGaN/GaN quantum well structures. *Phys. Stat. Sol. (a)*, 203(7):1729–1732, 2006.

Cottenier S.: *Density Functional Theory and the family of (L)APW-methods: a step-by-step introduction*. Instituut voor Kern- en Stralingfysica, K.U. Leuven, Belgium, 2002.

Cottrell A.H.: *Theory of Crystal Dislocations*. Blackie and Son, London, 1964.

Dadgar A., Schulze F., Wienecke M., Gadanecz A., Blasing J., Veit P., Hempel T., Diez A., Christen J., and Krost A.: Epitaxy of GaN on silicon–impact of symmetry and surface reconstruction. *New J. Phys.*, 9(10):389, 2007.

Daniels H., Brown A., Scott A., Nichells T., Rand B., and Brydson R.: Experimental and theoretical evidence for the magic angle in transmission electron energy loss spectroscopy. *Ultramicroscopy*, 96(3-4):523–534, 2003.

Daruka I. and Barabási A.-L.: Dislocation-free island formation in heteroepitaxial growth: A study at equilibrium. *Phys. Rev. Lett.*, 79(19):3708–3711, 1997.

Dodson B.W. and Tsao J.Y.: Relaxation of strained-layer semiconductor structures via plastic-flow. *Appl. Phys. Lett.*, 51(17):1325–1327, 1987.

Dridi Z., Bouhafs B., and Ruterana P.: First-principles investigation of lattice constants and bowing parameters in wurtzite $Al_xGa_{1-x}N$, $In_xGa_{1-x}N$ and $In_xAl_{1-x}N$ alloys. *Semicond. Sci. Technol.*, 18(9):850–856, 2003.

Dudešek P., Benco Ľ., Daul C., and Schwarz K.: *d*-to-*s* bonding in GaN. *J. Phys.: Condens. Matter*, 10(32):7155–7162, 1998.

Eaglesham D.J., Kvam E.P., Maher D.M., Humphreys C.J., Green G.S., Tanner B.K., and Bean J.C.: X-ray topography of the coherency breakdown in $Ge_xSi_{1-x}/Si(100)$. *Appl. Phys. Lett.*, 53(21):2083–2085, 1988.

Egerton R.F.: *Electron Energy-Loss Spectroscopy in the Electron Microscope*. Plenum Press, New York and London, 2 edition, 1996.

Fischer A.: Stability constraints in SiGe epitaxy. In Cressler J.D., editor, *Silicon Heterostructure Handbook – Materials, Fabrication, Devices, Circuits, and Applications of SiGe and Si Strained-Layer Epitaxy*, chapter 2.7, pages 127–143. CRC Press, Taylor&Francis Group, Boca Raton, 2006.

Fischer A. and Richter H.: On plastic flow and work hardening in strained layer heterostructures. *Appl. Phys. Lett.*, 64(10):1218–1220, 1994.

Fischer A., Kühne H., and Richter H.: New approach in equilibrium-theory for strained-layer relaxation. *Phys. Rev. Lett.*, 73(20):2712–2715, 1994.

Floro J.A., Follstaedt D.M., Provencio P., Hearne S.J., and Lee S.R.: Misfit dislocation formation in the AlGaN/GaN heterointerface. *J. Appl. Phys.*, 96(12):7087–7094, 2004.

Foreman A.J.E: Dislocation energies in anisotropic crystals. *Acta Metall.*, 3(4):322–330, 1955.

Frank F.C. and van der Merwe J.H.: One-dimensional dislocations. I. Static theory. *Proc. R. Soc. London, Ser. A*, 198(1053):205–216, 1949a. ISSN 00804630.

Frank F.C. and van der Merwe J.H.: One-dimensional dislocations. II. Misfitting monolayers and oriented overgrowth. *Proc. R. Soc. London, Ser. A*, 198(1053):216–225, 1949b. ISSN 00804630.

Freund L.B. and Hull R.: On the Dodson-Tsao excess stress for glide of a threading dislocation in a strained epitaxial layer. *J. Appl. Phys.*, 71(4):2054–2056, 1992.

Freund L.B. and Suresh S.: *Thin film materials: stress, defect formation, and surface evolution*. Cambridge University Press, Cambridge, 2003.

Fuggle J.C. and Inglesfield J.E., editors, *Unoccupied Electronic States*. Springer, Berlin, 1992.

Gao S.-P., Zhang A., Zhu J., and Yuan J.: Anisotropic spectroscopy of nitrogen K-edge in group-III nitrides. *Appl. Phys. Lett.*, 84(15):2784–2786, 2004.

Gherasimova M., Cui G., Ren Z., Su J., Wang X.-L., Han J., Higashimine K., and Otsuka N.: Heteroepitaxial evolution of AlN on GaN grown by metal-organic chemical vapor deposition. *J. Appl. Phys.*, 95(5):2921–2923, 2004.

Gibart P.: Metal organic vapour phase epitaxy of GaN and lateral overgrowth. *Rep. Prog. Phys.*, 67(5):667–715, 2004.

Gil B., editor, *Group III Nitride Semiconductor Compounds – Physics And Applications*. Clarendon Press (Oxford Science Publications), 1998.

Gourley P.L., Fritz I.J., and Dawson L.R.: Controversy of critical layer thickness for InGaAs/GaAs strained-layer epitaxy. *Appl. Phys. Lett.*, 52(5):377–379, 1988.

Graham D.M., Vala A.S., Dawson P., Godfrey M.J., Kappers M.J., Smeeton T.M., Barnard J.S., Humphreys C.J., and Thrush E.J.: Exciton localization in InGaN/GaN single quantum well structures. *Phys. Stat. Sol. (b)*, 240(2):344–347, 2003.

Grimmeiss H. and Koelmans Z.H.: *Nature*, 14a:264, 1959.

Hébert C.: Practical aspects of running the `Wien2k` code for electron spectroscopy. *Micron*, 38(1):12–28, 2007.

Hébert C., Luitz J., and Schattschneider P.: Improvement of energy loss near edge structure calculation using `Wien2k`. *Micron*, 34(3-5):219–225, 2003.

Hiramatsu K., Nishiyama K., Motogaito A., Miyake H., Iyechika Y., and Maeda T.: Recent progress in selective area growth and epitaxial lateral overgrowth of III-nitrides: Effects of reactor pressure in MOVPE growth. *Phys. Stat. Sol. (a)*, 176(1): 535–543, 1999.

Hirth J.P. and Lothe J.: *Theory of dislocations*. Krieger Publishing Company, Malabar, Florida, second edition, 1982.

Hitzel F., Klewer G., Lahmann S., Rossow U., and Hangleiter A.: Localized high-energy emissions from the vicinity of defects in high-efficiency $Ga_xIn_{1-x}N$/GaN quantum wells. *Phys. Rev. B*, 72(8):081309, 2005.

Hohenberg P. and Kohn W.: Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864–B871, 1964.

Holec D.: Analytical formula for dislocation energy in the hexagonal system lying along the *c*-direction. Internal report, 2006a.

Holec D.: *Critical Thickness Calculations for $In_xGa_{1-x}N$/GaN Systems*. CPGS thesis, University of Cambridge/Selwyn College, 2006b.

Holec D. and Humphreys C.J.: Calculations of equilibrium critical thickness for non-polar wurtzite InGaN/GaN systems. *Mater. Sci. Forum*, 567–568:209–212, 2007.

Holec D., Costa P.M.F.J., Kappers M.J., and Humphreys C.J.: Critical thickness calculations for InGaN/GaN. *J. Cryst. Growth*, 303:314–317, 2007.

Holec D., Costa P.M.F.J., Cherns P.D., and Humphreys C.J.: Electron energy loss near edge structure (ELNES) spectra of AlN and AlGaN: A theoretical study using the `Wien2k` and `Telnes` programs. *Micron*, In Press, Corrected Proof:–, 2008a.

Holec D., Costa P.M.F.J., Cherns P.D., and Humphreys C.J.: A theoretical study of ELNES spectra of $Al_xGa_{1-x}N$ using `Wien2k` and `Telnes` programs. *Comput. Mater. Sci.*, In Press, Corrected Proof:–, 2008b.

Hu S.M.: Misfit dislocations and critical thickness of heteroepitaxy. *J. Appl. Phys.*, 69 (11):7901–7903, 1991.

Hull D. and Bacon D.J.: *Introduction to dislocations*. Butterworth-Heinemann, Oxford, reprint of 4th edition, 2002.

Jahnen B., Albrecht M., Dorsch W., Christiansen S., Strunk H.P., Hanser D., and Davis R.F.: Pinholes, dislocations and strain relaxation in InGaN. *MRS Internet J. Nitride Semicond. Res.*, 3(39), 1998.

Jain S.C., Harker A.H., and Cowley R.A.: Misfit strain and misfit dislocations in lattice mismatched epitaxial layers and other systems. *Phil. Mag. A*, 75:1461–1515, 1997.

Jain S.C., Willander M., Narayan J., and van Overstraeten R.: III-nitrides: Growth, characterization, and properties. *J. Appl. Phys.*, 87(3):965–1006, 2000.

Jasinski J. and Liliental-Weber Z.: Extended defects and polarity of hydride vapor phase epitaxy GaN. *J. Electron. Mater.*, 31(5):429–436, 2002.

Johnston C.F., Kappers M.J., Barnard J.S., and Humphreys C.J.: Morphological study of non-polar $(11\bar{2}0)$ GaN grown on *r*-plane $(1\bar{1}02)$ sapphire. *Phys. Stat. Sol. (c)*, 2008. Proceedings of ICNS-7 conference. In press.

Jorissen K.: *Ab initio* core loss spectra (EELS and XAS). Lecture at the `Wien2k` workshop, Penn State University, 2007.

Jouffrey B., Schattschneider P., and Hébert C.: The magic angle: a solved mystery. *Ultramicroscopy*, 102(1):61–66, 2004.

Juza R. and Hahn H.: Über die Kristallstrukturen von Cu_3N, GaN und InN Metallamide und Metallnitride. *Z. Anorg. Allg. Chem.*, 239(3):282–287, 1938.

Kappers M.J., Datta R., Oliver R.A., Rayment F.D.G., Vickers M.E., and Humphreys C.J.: Threading dislocation reduction in (0001) GaN thin films using SiN_x interlayers. *J. Cryst. Growth*, 300(1):70–74, 2007.

Keast V.J., Scott A.J., Brydson R., Williams D.B., and Bruley J.: Electron energy-loss near-edge structure – a tool for the investigation of electronic structure on the nanometre scale. *J. Microsc. (Oxf.)*, 203(2):135–175, 2001.

Keast V.J., Scott A.J., Kappers M.J., Foxon C.T., and Humphreys C.J.: Electronic structure of GaN and $In_xGa_{1-x}N$ measured with electron energy-loss spectroscopy. *Phys. Rev. B*, 66(12):125319, 2002.

Keast V.J., Kappers M.J., and Humphreys C.J.: Electron energy-loss near edge structure (ELNES) of InGaN quantum wells. *J. Microsc. (Oxf.)*, 210(Part 1):89–93, 2003. ISSN 0022-2720.

Kioseoglou J., Komninou Ph., and Karakostas Th.: Study of structural properties of planar defects and *a*-edge threading dislocations in III(Al,Ga,In)-N by interatomic potential calculations. Unpublished.

Kohn W. and Sham L.J.: Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, 1965.

Kroupa F. and Brown L.M.: Elastic self- and interaction energies of dislocation jogs. *Phil. Mag.*, 6(70):1297–1270, 1961.

Lang T., Odnoblyudov M.A., Bougrov V.E., Romanov A.E., Suihkonen S., Sopanen M., and Lipsanen H.: Multistep method for threading dislocation density reduction in MOCVD grown GaN epilayers. *Phys. Stat. Sol. (a)*, 203(10):R76–R78, 2006.

Lang T., Odnoblyudov M.A., Bougrov V.E., Suihkonen S., Svensk O., Torma P.T., Sopanen M., and Lipsanen H.: Reduction of threading dislocation density in $Al_{0.12}Ga_{0.88}N$ epilayers by a multistep technique. *J. Cryst. Growth*, 298:276–280, 2007.

Lawniczak-Jablonska K., Suski T., Gorczyca I., Christensen N.E., Attenkofer K.E., Perera R.C.C., Gullikson E.M., Underwood J.H., Ederer D.L., and Liliental-Weber Z.: Electronic states in valence and conduction bands of group-III nitrides: Experiment and theory. *Phys. Rev. B*, 61(24):16623–16632, 2000.

Lazar S., Hébert C., and Zandbergen H.W.: Investigation of hexagonal and cubic GaN by high-resolution electron energy-loss spectroscopy and density functional theory. *Ultramicroscopy*, 98(2-4):249–257, 2004.

Lee S.R., Koleske D.D., Cross K.C., Floro J.A., Waldrip K.E., Wise A.T., and Mahajan S.: In situ measurements of the critical thickness for strain relaxation in AlGaN/GaN heterostructures. *Appl. Phys. Lett.*, 85(25):6164–6166, 2004.

Liu R., Mei J., Srinivasan S., Omiya H., Ponce F.A., Cherns D., Narukawa Y., and Mukai T.: Misfit dislocation generation in InGaN epilayers on free-standing GaN. *Jpn. J. Appl. Phys.*, 45(22):L549–L551, 2006a.

Liu R., Mei J., Srinivasan S., Ponce F.A., Omiya H., Narukawa Y., and Mukai T.: Generation of misfit dislocations by basal-plane slip in InGaN/GaN heterostructures. *Appl. Phys. Lett.*, 89(20):201911, 2006b.

Lü W., Li D.B., Li C.R., and Zhang Z.: Generation and behavior of pure-edge threading misfit dislocations in $In_xGa_{1-x}N$/GaN multiple quantum wells. *J. Appl. Phys.*, 96(9):5267–5270, 2004.

Luitz J.: *Simulation of core level spectra using Density Functional Theory*. PhD thesis, Technischen Universität Wien, 2006.

Lymperakis L.: *Ab-Initio Based Multiscale Calculations Of Extended Defects In And On Group III-Nitrides*. PhD thesis, Department Physik der Fakultät für Naturwissenschaften an der Universität Paderborn, 2005.

Madelung O., editor, *Semiconductors: Data Handbook*. Springer-Verlag, Berlin–Heidelberg–New York, 3rd edition, 2004.

Maruska H.P. and Tietjen J.J.: The preparation and properties of vapor-deposited single-crystal-line GaN. *Appl. Phys. Lett.*, 15(10):327–329, 1969.

Mathis S.K., Romanov A.E., Chen L.F., Beltz G.E., Pompe W., and Speck J.S.: Modeling of threading dislocation reduction in growing GaN layers. *J. Cryst. Growth*, 231(3): 371–390, 2001.

Matthews J.W. and Blakeslee A.E.: Defects in epitaxial multilayers: 1. misfit dislocations. *J. Cryst. Growth*, 27:118–125, 1974.

Miller D.A.B., Chemla D.S., Damen T.C., Gossard A.C., Wiegmann W., Wood T.H., and Burrus C.A.: Band-edge electroabsorption in quantum well structures – the quantum-confined Stark-effect. *Phys. Rev. Lett.*, 53(22):2173–2176, 1984. ISSN 0031-9007.

Mizoguchi T., Tanaka I., Kunisu M., Yoshiya M., Adachi H., and Ching W.Y.: Theoretical prediction of ELNES/XANES and chemical bondings of AlN polytypes. *Micron*, 34(3-5):249–254, 2003.

Mizoguchi T., Tanaka I., Yoshioka S., Kunisu M., Yamamoto T., and Ching W.Y.: First-principles calculations of ELNES and XANES of selected wide-gap materials: Dependence on crystal structure and orientation. *Phys. Rev. B*, 70(4):045103, 2004.

Monkhorst H.J. and Pack J.D.: Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13(12):5188–5192, 1976.

Moram M.A., Kappers M.J., Barber Z.H., and Humphreys C.J.: Growth of low dislocation density GaN using transition metal nitride masking layers. *J. Cryst. Growth*, 298:268–271, 2007.

Moreau P., Boucher F., Goglio G., Foy D., Mauchamp V., and Ouvrard G.: Electron energy-loss spectra calculations and experiments as a tool for the identification of a lamellar C_3N_4 compound. *Phys. Rev. B*, 73(19):195111, 2006.

Muller D.A.: Interpreting eels fine structure. Lecture at the Summer School and Workshop "Electron Microscopy: Fundamental Limits and New Science", 2006.

Muller D.A., Singh D.J., and Silcox J.: Connections between the electron-energy-loss spectra, the local electronic structure, and the physical properties of a material: A study of nickel aluminum alloys. *Phys. Rev. B*, 57(14):8181–8202, 1998.

Nakamura S., Senoh M., and Mukai T.: p-GaN/n-InGaN/n-GaN double-heterostructure blue-light-emitting diodes. *Jpn. J. Appl. Phys., Part 2*, 32(1A-B): L8–L11, 1993.

Nelhiebel M., Louf P.-H., Schattschneider P., Blaha P., Schwarz K., and Jouffrey B.: Theory of orientation-sensitive near-edge fine-structure core-level spectroscopy. *Phys. Rev. B*, 59(20):12807–12814, 1999.

Nordheim L.: The electron theory of metals. *Ann. Phys.*, 9:607–40,641–78, 1931.

Oliver R.A., Kappers M.J., Sumner J., Datta R., and Humphreys C.J.: Highlighting threading dislocations in MOVPE-grown GaN using an in situ treatment with SiH_4 and NH_3. *J. Cryst. Growth*, 289(2):506–514, 2006.

Park S.-E., O B., and Lee C.-R.: Strain relaxation in $In_xGa_{1-x}N$ epitaxial films grown coherently on GaN. *J. Cryst. Growth*, 249(3–4):455–460, 2003.

Parker C.A., Roberts J.C., Bedair S.M., Reed M.J., Liu S.X., and El-Masry N.A.: Determination of the critical layer thickness in the InGaN/GaN heterostructures. *Appl. Phys. Lett.*, 75(18):2776–2778, 1999.

Payne M.C., Teter M.P., Allan D.C., Arias T.A., and Joannopoulos J.D.: Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, 64(4):1045–1097, 1992.

People R. and Bean J.C.: Calculation of critical layer thickness versus lattice mismatch for Ge_xSi_{1-x}/Si strained-layer heterostructures. *Appl. Phys. Lett.*, 47(3):322–324, 1985.

Perdew J.P., Burke K., and Ernzerhof M.: Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865–3868, 1996.

Ponce F.A.: Structural defects and materials performance of the III–V nitrides. In Gil Bernard, editor, *Group III Nitrides Semiconductor Compounds*, chapter 4, pages 123–157. Clarendon Press, Oxford, 1998.

Pulay P.: *Chem. Phys. Lett.*, 73:393, 1980.

Radtke G., Epicier T., Bayle-Guillemaud P., and Le Bosse J. C.: N-K ELNES study of anisotropy effects in hexagonal AlN. *J. Microsc. (Oxf.)*, 210(1):60–65, 2003.

Radtke G., Bayle-Guillemaud P., and Thibault J.: EELS study of near edge fine structure in $Al_xGa_{1-x}N$ alloys. In *Electron Microscopy And Analysis 2003 conference*, number 179 in Institute of Physics Conference Series, pages 303–306, 2004.

Reed M.J., El-Masry N.A., Parker C.A., Roberts J.C., and Bedair S.M.: Critical layer thickness determination of GaN/InGaN/GaN double heterostructures. *Appl. Phys. Lett.*, 77(25):4121–4123, 2000.

Romanov A.E., Pompe W., Beltz G., and Speck J.S.: Modeling of threading dislocation density reduction in heteroepitaxial layers: I. Geometry and crystallography. *Phys. Stat. Sol. (b)*, 198(2):599–613, 1996.

Romanov A.E., Pompe W., Beltz G., and Speck J.S.: Modeling of threading dislocation density reduction in heteroepitaxial layers: II. Effective dislocation kinetics. *Phys. Stat. Sol. (b)*, 199(1):33–49, 1997.

Sanchez J.M., Ducastelle F., and Gratias D.: Generalized cluster description of multicomponent systems. *Physica A*, 128(1-2):334–350, 1984.

Sasaoka C., Sunakawa H., Kimura A., Nido M., Usui A., and Sakai A.: High-quality InGaN MQW on low-dislocation-density GaN substrate grown by hydride vapor-phase epitaxy. *J. Cryst. Growth*, 189-190:61–66, 1998.

Schattschneider P., Hébert C., and Jouffrey B.: Orientation dependence of ionization edges in EELS. *Ultramicroscopy*, 86(3-4):343–353, 2001.

Schattschneider P., Hébert C., Franco H., and Jouffrey B.: Anisotropic relativistic cross sections for inelastic electron scattering, and the magic angle. *Phys. Rev. B*, 72(4): 045142, 2005.

Shan W., Ager J.W., Yu K.M., Walukiewicz W., Haller E.E., Martin M.C., McKinney W.R., and Yang W.: Dependence of the fundamental band gap of $Al_xGa_{1-x}N$ on alloy composition and pressure. *J. Appl. Phys.*, 85(12):8505–8507, 1999.

Sjöstedt E., Nordström L., and Singh D.J.: An alternative way of linearizing the augmented plane-wave method. *Solid State Commun.*, 114(1):15–20, 2000.

Soven P.: Coherent-potential model of substitutional disordered alloys. *Phys. Rev.*, 156 (3):809–813, 1967.

Srinivasan S., Geng L., Liu R., Ponce F.A., Narukawa Y., and Tanaka S.: Slip systems and misfit dislocations in InGaN epilayers. *Appl. Phys. Lett.*, 83(25):5187–5189, 2003.

Steeds J.W.: *Introduction to anisotropic elasticity theory of dislocations*. Clarendon Press, Oxford, 1973.

Stringfellow G.B.: *Organometallic Vapor-Phase Epitaxy: Theory and Practice*. Academic Press, 2nd edition, 1999.

Sumner J., Das Bakshi S., Oliver R.A., Kappers M.J., and Humphreys C.J.: Unintentional doping in GaN assessed by scanning capacitance microscopy. *Phys. Stat. Sol. (b)*, 245(5):896–898, 2008.

Tersoff J. and Legoues F.K.: Competing relaxation mechanisms in strained layers. *Phys. Rev. Lett.*, 72(22):3570–3573, 1994.

Troullier N. and Martins J.L.: Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, 43(3):1993–2006, 1991.

van der Merwe J.H.: Crystal interfaces. Part I. Semi-infinite crystals. *J. Appl. Phys.*, 34 (1):117–122, 1963a.

van der Merwe J.H.: Crystal interfaces. Part II. Finite overgrowths. *J. Appl. Phys.*, 34 (1):123–127, 1963b.

Velický B., Kirkpatrick S., and Ehrenreich H.: Single-site approximations in the electronic theory of simple binary alloys. *Phys. Rev.*, 175(3):747–766, 1968.

Vennéguès P., Bougrioua Z., Bethoux J. M., Azize M., and Tottereau O.: Relaxation mechanisms in metal-organic vapor phase epitaxy grown Al-rich (Al,Ga)N/GaN heterostructures. *J. Appl. Phys.*, 97(2):024912, 2005.

Vickers M. E., Kappers M. J., Smeeton T. M., Thrush E. J., Barnard J. S., and Humphreys C. J.: Determination of the indium content and layer thicknesses in InGaN/GaN quantum wells by x-ray scattering. *J. Appl. Phys.*, 94(3):1565–1574, 2003.

Waltereit P., Brandt O., Trampert A., Grahn H.T., Menniger J., Ramsteiner M., Reiche M., and Ploog K.H.: Nitride semiconductors free of electrostatic fields for efficient white light-emitting diodes. *Nature*, 406(6798):865–868, 2000. ISSN 0028-0836.

Wei S.-H., Ferreira L.G., Bernard J.E., and Zunger A.: Electronic properties of random alloys: Special quasirandom structures. *Phys. Rev. B*, 42(15):9622–9649, 1990.

Weijs P.J.W., Czyżyk M.T., van Acker J.F., Speier W., Goedkoop J.B., van Leuken H., Hendrix H.J.M., de Groot R.A., van der Laan G., Buschow K.H.J., Wiech G., and Fuggle J.C.: Core-hole effects in the x-ray-absorption spectra of transition-metal silicides. *Phys. Rev. B*, 41(17):11899–11910, 1990.

Williams D.B. and Carter C.B.: *Transmission Electron Microscopy*. Plenum Press, 1996.

Willis J.R., Jain S.C., and Bullough R.: The energy of an array of dislocations – implications for stain relaxation in semiconductor heterostructures. *Phil. Mag. A*, 62(1): 115–129, 1990.

Wright A.F. and Nelson J.S.: Consistent structural properties for AlN, GaN, and InN. *Phys. Rev. B*, 51(12):7866–7869, 1995.

Xin Y., Pennycook S.J., Browning N.D., Nellist P.D., Sivananthan S., Ommès F., Beaumont B., Faurie J.P, and Gibart P.: Direct observation of the core structure of threading dislocations in GaN. *Appl. Phys. Lett.*, 72:2680–2682, 1998.

Yoshida S., Misawa S., and Gonda S.: Improvements on the electrical and luminescent properties of reactive molecular beam epitaxially grown GaN films by using AlN-coated sapphire substrates. *Appl. Phys. Lett.*, 42(5):427–429, 1983.

Zakharov D.N., Liliental-Weber Z., Wagner B., Reitmeier Z.J., Preble E.A., and Davis R.F.: Structural TEM study of nonpolar a-plane gallium nitride grown on

$(11\bar{2}0)$ 4H-SiC by organometallic vapor phase epitaxy. *Phys. Rev. B*, 71(23):235334, 2005.