AB-INITIO BASED MULTISCALE CALCULATIONS OF EXTENDED DEFECTS IN AND ON GROUP III – NITRIDES

Liverios Lymperakis

January 2005

produced using $\LaTeX\ 2_\varepsilon$

AB-INITIO BASED MULTISCALE CALCULATIONS OF EXTENDED DEFECTS IN AND ON GROUP III – NITRIDES

Dissertation
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
vorgelegt dem
Department Physik der Fakultät für Naturwissenschaften
an der Universität Paderborn

Liverios Lymperakis

January 2005

Tag der Einreichung:	1. Februar 2005
Tag der mündlichen Prüfung:	22. März 2005

Promotionskommission

Vorsitzender	Prof. Dr. rer. nat. Klaus Lischka
Gutachter	Prof. Dr. rer. nat. Thomas Frauenheim
Gutachter	Prof. Dr. rer. nat. Jörg Neugebauer

Καθαρὰ νὰ ξεχωρίσεις κι ἡρωικὰ νὰ δεχτεῖς τὶς πικρὲς γόνιμες τοῦτες, ἀνθρώπινες, σάρκα ἀπὸ τὴ σάρκα μας, ἀλήθειες:
(α) Ὁ νοῦς τοῦ ἀνθρώπου φαινόμενα μονάχα μπορεῖ νὰ συλλάβει, ποτὲ τὴν οὐσία·
(β) κι ὄχι ὅλα τὰ φαινόμενα, παρὰ μονάχα τὰ φαινόμενα τῆς ὕλης·
(γ) κι ἀκόμα στενώτερα: ὄχι κὰν τὰ φαινόμενα τοῦτα τῆς ὕλης, παρὰ μονάχα τοὺς μεταξύ τους συνειρμούς·
(δ) κι οἱ συνειρμοὶ τοῦτοι δὲν εἶναι πραγματικοί, ἀνεξάρτητοι ἀπὸ τὸν ἄνθρωπο· εἶναι κι αὐτοὶ γεννήματα τοῦ ἀνθρώπου·
(ε) καὶ δὲν εἶναι οἱ μόνοι δυνατοὶ ἀνθρώπινοι· παρὰ μονάχα οἱ πιὸ βολικοὶ γιὰ τὶς πραχτικὲς καὶ νοητικές του ἀνάγκες.

Ν. Καζαντζάκης
ΑΣΚΗΤΙΚΗ (SALVATORES DEI) (1922)

Distinguish clearly these bitter yet fertile human truths, flesh of our flesh, and admit them heroically:
(a) the mind of man can perceive appearances only, and never the essence of things;
(b) and not all appearances but only the appearances of matter;
(c) and more narrowly still: not even these appearances of matter, but only relationships between them;
(d) and these relationships are not real and independent of man, for even these are his creations;
(e) and they are not the only ones humanly possible, but simply the most convenient for his practical and perceptive needs.

N. Kazantzakis
ASKITIKI (THE SAVIORS OF GOD) (1922)
translated by Kimon Friar

...dedicated to Anna
for her support and understanding.

Zusammenfassung

Ende 1993 entwickelte S. Nakamura bei Nichia Chemicals die ersten hocheffizienten blauleuchtende Dioden auf GaN-Basis. Dieses Jahr läßt sich als Wendepunkt im Bereich der III-Nitriden Halbleiter bezeichnen, da fortan die Gruppe-III-Nitriden zu einem Schwerpunkt in der Halbleiterforschung wurden. Während enorme Fortschritte in der Qualität der Materialien und bei der Herstellung der Bauelemente erreicht wurden, ist das Verständnis der zugrunde liegenden Eigenschaften und Wachstumsprozesse häufig noch begrenzt.

Auf dem Gebiet der III-Nitride wird der Effekt, den Versetzungen auf die elektronischen Eigenschaften von GaN-Epilayern haben, kontrovers diskutiert. Versetzungen sind durch zwei unterschiedliche Längenskalen charakterisiert: Der Kernbereich der Versetzungen ist stark lokalisiert, während das durch die Versetzung verursachte Spannungsfeld (strain field) auch noch weit entfernt vom Versetzungskern ist. *Ab-initio* basierte Methoden haben sich zu äußerst zuverlässige Verfahren für Berechnungen im Bereich atomarer Größenordnungen entwickelt. Auf der anderen Seite sind die Möglichkeiten, diese Methoden auf große, auf vielen Atomen bestehende Systeme anzuwenden, stark begrenzt (durch die verfügbare Rechenkapazität). Es ist daher notwendig, eine Methode zu entwickeln, die die Beschreibung physikalischer Systeme auf allen relevanten Längenskalen erlaubt. Wir entwickelten daher *ab-initio* basierte Multiskalen-Ansätze, die auf ersten Prinzipien beruhende Berechnungen mit mesoskopischen Konzepten kombinieren. Ein möglicher Ansatz, der hier diskutiert wird, verwendet erste Prinzipien, Elastizitätstheorie und empirische Potentiale, wobei die mikroskopischen mit den makroskopisch relevanten Skalen verbunden werden. Die Genauigkeit dieses Zugangs wird im Prinzip nur durch die Genauigkeit der *ab-initio* Berechnungen begrenzt. Ein weiterer Ansatz wurde angewandt, um Wachstumsprozesse zu untersuchen (*ab-initio* atomare Thermodynamik). Hier werden auf ersten Prinzipien basierte Berechnungen mit dem thermodynamischen Konzept der chemischen Potentiale in Verbindung gebracht. Dies erlaubt es, die thermodynamische Stabilität verschiedener Systeme als Funktion der Wachstumsbedingengen zu berechnen.

Der erste Teil der Dissertation ist der Wachstum von GaN Oberflächen gewidmet. Es wird eine Vielzahl von Stufen-/Nachbarschaftskonfigurationen auf GaN-Oberflächen berücksichtigt. Dies geschiet für die (0001)-Orientierung, die technologisch am interessantesten ist. Dabei wird die thermodynamische Stabilität unter verschiedenen Wachstumsbedingungen untersucht. Ein wichtiges

Ergebnis dieser Untersuchungen ist die Erkenntnis, dass N-reiche Bedingungen eine spontane Ausbildung von Stufen ermöglichen. Weiterhin zeigte sich, dass diese Instabilität unter metallreichen Bedingungen unterdrückt werden kannn: Unter diesen Bedingungen ist eine lateral kontrahierte Bilayer Strukturthermodynamisch am günstigsten. Danach wenden wir uns dem Wachstum unter extrem Ga-reichen Bedingungen zu und untersuchen die Bildung und Stabilität kleiner Ga-Inseln auf den lateral kontrahierten Bilayer. Unsere Berechnungen zeigen, dass diese Inseln Nukleationskeime für die Bildung von Ga-Tropfen bilden. Anhand diese Resultate gelang es verschiedene kontrovers diskutierten Wachstumsuntersuchungen im Rahmen eines einheitlichen Modells zu erklären. Unsere Berechnungen zeigen ferner, dass Strukturen mit mehr als 2 Monolagen sowohl machbar als auch unter extrem Ga-reichen Bedingungen thermodynamisch stabil sind sogar für intrinsisch nicht-polare Oberflächen wie beispielsweise GaN $\left(1\bar{1}00\right)$. Das ist beispielsweise für das Wachstum feldfreier Strukturen deutsam.

Nach dem Studium von Oberflächen richtet sich der Schwerpunkt auf Grenzflächen, wobei eine aus sich abwechselnden Monolagen aufgebaute Struktur (monoatomic-scale ordering) einer AlGaN-Legierung untersucht wird. In Übereinstimmung mit jüngsten Experimenten gelang es dabei zu zeigen, dass eine solche Struktur zu einer Reduktion der Bandlücke führt. Wir fanden, dass für diese Reduktion der Bandlücke eine Lokalisierung der Ladungsträger in einlagigen (monoatomic) GaN-Quantengräben, die zwischen AlN-Barriern eigebettet sind, verantwortlich ist. Im letzten Teil der Arbeit richtet sich der Blick auf ausgedehnte (extended) ein-dimensionale Defekte in GaN (Versetzungen). Mit unserem Multiskalen-Ansatz untersuchen wir eine Vielzahl unterschiedlicher Konfigurationen des Versetzungskerns. Wir zeigen, dass das äußerst starke lokale Spannungsfeld um den Kern der Versetzung in Verbindung mit der kleinen Gitterkonstanten von GaN zu energetisch tiefliegenden Defektzuständen führt und daher elektrisch aktive Stufenversetzungen verursacht, auch wenn alle Atome im Kernbereich vollständig koordiniert sind.

Die im Rahmen dieser Arbeit entwickelten Multiskalen-Ansätze ermöglichen es, eine Vielzahl bislang ungelöster Probleme zu lösen und darüber hinaus ein mikroskopisches Veständnis ausgedehnter Defekte in und auf Gruppe-III-Nitriden zu erhalten. Viele Teile der hier vorgestellten Arbeit wurden in enger Zusammenarbeit mit anderen theoretischen und experimentellen Arbeitsgruppen durchgeführt, wie aus dem Literaturverzeichnis ersichtlich ist. Die Arbeit wurde im Zeitraum Juli 2000 - Dezember 2004 in der Unabhängigen Nachwuchsgruppe des Fritz-Haber-Instituts der Max-Planck-Gesellschaft angefertigt.

Berlin, Januar 2005

preface

In late 1993 S. Nakamura of Nichia Chemicals introduced the first high efficient GaN-based blue light-emitting diode. This year is widely regarded as the turning point in the field of III-Nitride semiconductors. Since then group III-Nitrides are a hot topic in the semiconductors research. While there has been a tremendous progress in materials quality and and device fabrication our understanding concerning fundamental properties and growth phenomena is often still limited.

One of the controversial issues is the effect dislocations have on the electronic properties of the GaN epilayers. Dislocations are characterized by two different length scales: The core structure of the defect is rather localized, while the strain field introduced by the defect is significant even far away from the core. *Ab-initio*-based methods are a reliable tool for atomic scale calculations. However, they are characterized by a rather limited (limited by the available computational power) ability to describe systems consisting of large numbers of atoms. Thus, it is essential to develop methods which allow to describe physical systems on all relevant length scales. We have therefore developed *ab-initio*-based multiscale approaches which combine first-principles calculations with mesoscopic concepts. A first approach combines first-principles, empirical potentials and elasticity theory and links the microscopic with the macroscopically relevant length scales. The accuracy of this approach is in principle limited only by the accuracy of the *ab-initio* calculations. A second approach, which has been applied to study growth is *ab-initio atomistic thermodynamics*: First-principles calculations are combined with the thermodynamic concept of chemical potentials and enable to derive the thermodynamic stability of the various systems as function of the growth conditions.

In the first part of the thesis, the focus will be on the growth of GaN surfaces. A variety of step/vicinal surface configurations on the technologically most relevant GaN (0001) orientation are considered and their thermodynamic stability under different growth conditions is examined. An important conclusion from these studies is that steps may spontaneously form under N-rich conditions, while under more metal-rich conditions they become thermodynamically unfavorable against a Ga rich laterally contracted bilayer structure. Next, we spotlight the Ga-rich growth conditions limit and we study the formation of small Ga islands on the laterally contracted bilayer structure. Our calculations reveal that the islands can be assumed as precursors of Ga droplet formation. Based on these results we explain recent controversially discussed growth studies within a unified model. Our calculations further reveal that structures exceeding 2 MLs of Ga are feasible and thermodynamically stable under extreme Ga rich conditions even on intrinsically non-polar surfaces such as the GaN $\left(1\bar{1}00\right)$ as important for growing field-free structures.

Next, we move from surfaces to interfaces and we study monatomic-scale ordering in AlGaN alloys. In agreement with recent experiments we find that ordering results in bandgap reduction. The origin of the bandgap reduction is shown to be a carrier localization in mono-atomic GaN quantum wells embedded in AlN barriers. Finally, from the two dimensional features (surface/interfaces) we move to one dimensional extended defects in GaN. i.e., dislocations. Based on our multiscale approach we investigate the properties of a variety of different core configurations. We find that the giant local strain-field around the dislocation core, in combination with the small lattice constant of GaN, causes deep defect states and thus electrically active edge dislocations even when all core atoms are fully coordinated.

In summary, in this thesis we developed a multiscale approach and applied it to address a number of hitherto unsolved problems and to get a microscopic understanding of extended defects in and on group III-Nitrides. Most parts of the work presented here have been done in close collaboration with other theoretical and experimental groups, as can be seen in the publication list in page 165. This thesis has been performed during the period July 2000 - December 2004 in the Independent-Junior-Research Group of the Fritz-Haber-Institut der Max-Planck-Gesellschaft.

Berlin, January 2005

Acknowledgements

The writing of a thesis can be assumed as a many-body problem. It is practically impossible to be completed without the crucial interactions with a number of people. Thus, I would like to acknowledge them at the very beginning of this thesis:

First of all, I would like to thank my supervisor Prof. Dr. Jörg Neugebauer for giving me the opportunity to work in the Nachwuchsgruppe of the Fritz-Haber-Institut (FHI) and for his continuous and indispensable interest and guiding during the four years of my doctoral research. I would like also to thank him for reading the final drafts of this work. The discussions with him and his comments were really essential for forming the arguments and discussions included in this thesis. I also would like to acknowledge the director of the Theory Department of the FHI Prof. Dr. Matthias Scheffler for giving me the opportunity to work in the Theory group.

Most of the work present here is in close collaboration with other experimental and theory groups. So, I would like to thank the following people for their collaboration: Dr. Martin Albrecht for the issues of dislocations in GaN and ordering in AlGaN alloys. Also, many thanks for the fruitful discussions on these issues. Prof. Randall Feenstra for the issue of m-plane GaN. Prof. John Northrup for the work on kinetically induced ordering in AlGaN alloys. Prof. Bruno Daudin for the work on Ga adsorption on the GaN (0001) surfaces.

Many thanks also to Prof. Anna Serra and Dr. Antoine Bére for the fruitful discussions on the empirical potential calculations and on the dislocations. I would like to express my gratitude to Dr. Joseph Kioseoglou for the discussions we had on the issue of empirical potentials and for reading partially my thesis and making important comments. Also many thanks to my former supervisor Prof. Hariton Polatoglou and Prof. Philomina Komninou for their support and interest during my doctoral research.

I am thankful to Dr. Andreia Luisa da Rosa, Dr. Evgeni Penev, and Sixten Boeck. Andreia introduced me to the FHI98MD package and helped me with my total energy calculations especially at the beginning of my work in FHI. Evgeni was always providing the necessary help to overcome any barriers regarding Linux, LaTeX, and ps graphics. Sixten was the most appropriate person to introduce me to the newly developed SFHingX package. Also many thanks to Matthias Wahn for trnslating the preface to german.

I would like to express my thanks to all the members of the Nachwuchsgruppe for their useful comments and suggestions during our ”internal” Monday morning talks/discussions.

I would like to thank my parents Athanasia and Kostas for their continuous support and encouragement all these years. At last but not least, many thanks to my partner Anna for her infinite understanding, support, and encouragement during the last four years and for never complaining when I claimed my thesis would be finished in the next month for nearly a year...

This thesis was financial supported by European Union (TMR program IPAM under the contract HPRN-CT-1999-00040) and by Deutsche Forschungsgemeinschaft (DFG-research group ”Nitride based QD-lasers”-University of Bremen).

Contents

Chapter 1

Introduction

Group III-Nitrides (AlN, GaN, and InN) have been considered as promising system for optoelectronic device applications since 1970 [1, 2]. Since then there has been a rapid progress in the field of Nitride semiconductors. The reason for the big interest in these materials is manifold.

Group III-Nitrides crystallize preferentially in the hexagonal wurtzite structure (see Appendix A) and have direct bandgaps in the range between 0.7-6.2 eV (from 0.7-0.8 eV for InN[1] and 3.4 eV for GaN to 6.2 eV for AlN). This range of gap energies covers the entire visible and extends into the ultraviolet (UV) region of the optical spectrum (see Fig. 1.1). Because of their wide bandgap and strong bond strength, this materials system is suitable for violet, blue, and green light emitting devices (e.g., for full color displays, laser printers, high density information storage, or under water communication) and for high-temperature/high-power transistors (as needed e.g., in automobile engines, power distribution systems, or to design all electric vehicles). The fact that they can operate at higher temperatures makes device processing easier. Another important property of III-Nitrides is that InN and AlN can be alloyed with GaN. This property enables bandgap engineering and thus to tune the emission wavelengths. In contrast, conventional III-V semiconductors can not be used for optoelectronic applications in the violet and blue region of the spectrum, since they do not have sufficiently large bandgaps. In contrast to SiC (the material with the longest device history of all the wide bandgap semiconductors [7]), the direct band of III-Nitrides increases the optical transition probability (by at least one order of magnitude compared to the indirect bandgap semiconductor SiC) and thus the adsorption and emission rate of photons.

---

[1] Despite the technological importance of InN its fundamental bandgap is still under debate. Until recently, for the InN bandgap a value of ~ 2 eV was widely accepted and cited in the literature [3, 4]. However, based on very recent experimental data it has been concluded that the gap energy is lower than 1 eV (in the range 0.7-0.8 eV) [5, 6]. This small value gave rise to controversial discussions since it does not obey the *"common-anion bandgap rule"*: For common-anion (or cation) semiconductors the direct bandgap at Γ increases as the cation (or anion) atomic number decreases. The value of 0.7-0.8 eV for the InN bandgap is much smaller than the InP bandgap (1.46 eV). The InN "bandgap-problem" gave rise to animated discussions at the 2004 International Workshop on Nitride Semiconductors (IWN-04): As it is referred in the IWN-04 report (C. Trager-Cowan, MRS BULLETIN October 2004, pp. 762-763) *"The interpretation of available data is hotly contested and, not surprisingly, no definitive conclusions could be reached. It was agreed that InN is a difficult material to grow and its properties vary depending on how and where it is grown."*

Besides the α-phase (wurtzite structure), a metastable β-phase with zincblende structure exists for the group III-Nitrides (see Appendix A). The cubic modification of these materials can be realized by epitaxial growth on cubic substrates such as GaAs or SiC [8, 9, 10, 11]. Group III-Nitrides in the β-phase have physical properties which are favorable for some optoelectronic device applications such as higher electron mobility and easier doping than the hexagonal phase. Furthermore, possible integration of III-Nitrides with the GaAs- (Si-) technology is favored by the use of cubic substrates (see Ref. [12] and Refs. therein). Thus, cubic group III-Nitrides can be assumed as a promising material-system for future optoelectronic applications. However, until now all the commercialized applications of III-Nitrides utilize the hexagonal phase. Therefore, in this thesis we will focus on the hexagonal phase.

1.1 A brief history of GaN

The first commercial application of GaN had been the blue and green Light-Emitting Diodes (LEDs) [13]. However, the commercialization of the first GaN-based LED in 1993 by S. Nakamura at Nichia Laboratories was everything than straightforward and took almost 25 years of research. First attempts to synthesize GaN were made by Juda and Hahn in 1938 [14, 15]. They directed ammonia over hot gallium and produced small needles and platelets. The same technique had been used in 1959 by Grimmeiss and Koelmans to produce small crystals of GaN [16]. The first epitaxial growth of GaN has been reported by Maruska and Tietjen in 1969 [17]. They used a chemical vapor deposition (CVD) technique to make large area layers of GaN on sapphire. However, all samples grown at that time showed large concentrations of electrons even when not deliberately doped. Initially N vacancies have been proposed to be the donors responsible for the n-type conductivity. Later, based on *ab-initio* studies, unintentional doping by Si and O impurities have been shown as origin of the n-type conductivity in GaN samples [18].

The achievement of Maruska and Tietjen was followed by the production of the first blue LED by Pankove and co-workers in 1972 [7, 19]. It was a metal/insulating GaN:Zn/n-GaN ($m-i-n$ type) diode. The wavelength of the emitting light depended on the concentration of Zn in the n-GaN layers. However, Pankove did not succeed in p-type doping of GaN. The lack of p-type doping along with the poor quality of the crystalline layers were the main obstacles in commercializing GaN-based devices. The origin of the poor crystalline quality was due to the lack of suitable lattice matched substrates. Sapphire, which is widely used as substrate for GaN epitaxial growth has a lattice constant which is 15% smaller than GaN and a different thermal expansion coefficient. Both mismatches lead to very high defect densities (e.g., threading dislocations). On the other hand, SiC offers good lattice match with GaN (the difference in lattice constants is only 3.5%) but it is very expensive. In 1983 Yoshida *et al.* introduced a two-step growth method to grow GaN films by Molecular Beam Epitaxy (MBE) [20]. The main idea is to grow in the first step an AlN buffer layer on the sapphire substrate and in a second step the GaN layer on top of the buffer layer. This

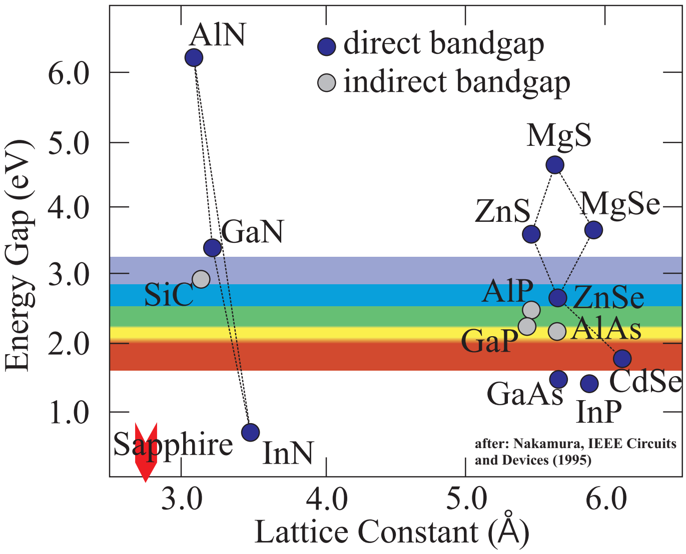


Figure 1.1: Bandgaps as function of lattice constant for various semiconductors [13]. The value of the InN bandgap has been adjusted to 0.8 eV instead of the originally used value of 2 eV (see footnote in page 1).

offers an improved quality of the epitaxial GaN layer. This two-step method had been further explored and improved by Akasasi and co-workers [21]. They demonstrated that high quality GaN layers grown by Metalorganic Vapor Phase Epitaxy (MOVPE) on sapphire can be obtained by first growing a thin AlN buffer layer *at low temperature* on the sapphire substrate. Later Nakamura showed that equally good or better results can be obtained by using a thin low temperature GaN buffer layer [22].

The second key-step in commercializing GaN-based devices was the achievement of p-type doping by Akasaki and co-workers in 1988 [23]. p-(n-)type doping is important for semiconductor devices such as lasers diodes, LEDs, and transistors, since they are based on pn-junctions. n-type doping of Nitrides has never been a problem. However, p-type doping has been very difficult. Akasaki and his team discovered that if a sample of a Mg-doped GaN grown by MOCVD is electron-beam annealed it becomes p-type conducting. Subsequently, S. Nakamura showed that Mg activation can be also achieved by annealing the as-grown Mg-doped GaN samples above 700 ℃ [24]. Van Vechten *et al.* speculated that H plays a crucial role in passivating the Mg acceptors [25]. More detailed theoretical studies based on *ab-initio* density functional theory revealed in detail the mechanism of the Mg acceptor activation: Neugebauer and Van de Walle showed that codoping - incorporation of donors (H) along with acceptors (Mg) - is an effective way of increasing hole concentrations in p-type GaN [26, 27].

1.2 Technological importance of group III-Nitrides

As already mentioned in Sec. 1.1 the first commercial applications of group III-Nitrides were the blue and green LEDs. The economical and technical potential of LEDs is manifold. Until now, the lighting industry is dominated by glass bulbs and glass-fluorescent tubes. This technology

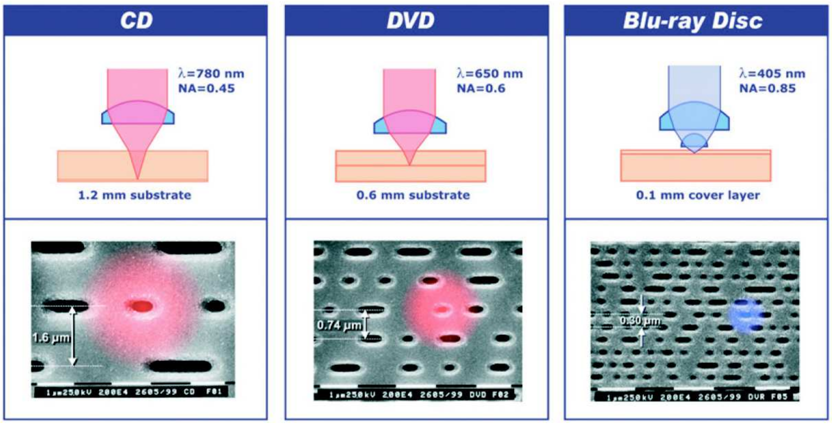


Figure 1.2: Comparison of optical heads and structure size of CD, DVD and Blu-ray disks. The laser wavelength limits the density of the data storage: shorter wavelengths enable higher storage density (Courtesy of Philips research - source: T. Weber, Photonik 4/2003).

is, despite all improvements, essentially 100 years old(!!!) and has many disadvantages such as a short lifetime and insufficient performance (most energy input is converted into heat and blackbody radiation in the invisible infrared spectrum). GaN-based blue, green, and white LEDs allow the replacement of light bulbs and fluorescent tubes by semiconductors (e.g., in traffic lights).

Another important application of GaN-based devices are blue and violet Laser Diodes (LDs) [13, 28]. In optical storage devices such as CDs and DVDs the data storage density is restricted by the laser wavelength (see Fig. 1.2). The size of the diffraction spots in these devices is proportional to the laser wavelength: $D \sim \lambda/\mathrm{NA}$. Here, D is the diameter of the laser focus (FWHM) and NA is the numerical aperture of the optical system. Thus, the amount of data which can be stored on an optical disk scales as $1/\lambda^2$. The use of LDs in the violet region (405 nm) allows to realize a significantly increased data storage than what can be achieved based on the current CD- (780 nm) and DVD- (650 nm) technologies. The next generation of optical data storage devices based on violet LDs (Blu-ray[2] and High-Density-DVD) is expected to dramatically increase the data storage density[3].

LEDs and LDs are not the only applications of III-Nitrides. The AlGaN system has bandgaps covering the range of 3.4-6.2 eV. Taking advantage of the efficient direct optical transitions, it can be used for UV photodetectors. The strong bond strength of III-Nitrides legitimizes them as important materials for high-temperature and high-pressure transistors which are desirable in the automobile industry.

It is interesting to mention the market trends for the GaN-based devices: Since the announce-

---

[2] The spelling of "Blu-ray" is not a mistake. According to the Blu-ray Disk Association: *"Blu" is intentionally spelled without an "e", to allow for a distinctive registration of the trademark name.* For more information on the Blu-ray Disk Association see the official website: `http://www.blu-raydisc.com`.

[3] Both Blu-ray and High Density-DVD (HD-DVD) technologies use LDs at 405 nm. However, in the Blu-ray format the storage density is further increased by using larger values of the numerical aperture of the optical system. On the other hand, in HD-DVD format the numerical aperture is kept close to that CDs and DVDs are using and the disk and device manufacturing process is kept compatible to those. Nevertheless, the storage capacity is 40% lower: single/double layered Blu-ray disks are expected to hold 25/50 GBytes compared to 15/30 GBytes for HD-DVD disks. However, these values must be compared to 0.7 GBytes of a CD and 4.7 GBytes of a single layered DVD.

ment of high brightness GaN-based LEDs in 1993 there has been a rapid increase in Research and Development (R&D) field. Until, 1999 the market had been already grown from zero to more than \$400 million. By the year 2003 the market had been grown to \$1.35 billion and is expected to further grow to \$4.5 billion by 2007 (see Ref. [29]). Detailed discussions and excellent reviews of the structural, electric, optical properties, and applications of group III-Nitrides can be found in Refs. [2, 4, 7, 13, 30, 31, 32].

1.3 Aim and structure of this work

Over the last 10 years, since the first announcement of commercial application of GaN-based semiconductors, there has been rapid progress and extensive and intensive research on the growth and the properties of group III-Nitrides. First-principles calculations have been and still are an indispensable tool in the research of this materials system. They have been extensively used to study various properties and various systems such as bulk properties (e.g., bulk electronic structure and elastic constants [33]), properties of defects [34] and alloys [35, 36, 37], the mechanism of acceptor activation [26, 27], identification of equilibrium surface structures [38, 39, 40, 41, 42, 43], and the formation of interfaces [44]. However, the complexity of some systems of major interest, such as dislocations, imposes the necessity of longer scale length and time approaches.

Ab-initio methods are by now well established for tackling the quantum mechanics of many-body systems (see Ref. [45] and Refs. therein). Although first-principles calculations provide a reliable tool for atomic scale calculations, they suffer from the rather limited ability (limited by the available computational power) to represent systems consisting of a large number of atoms and/or to perform simulations over a large time scale. The challenge of treating different length scales (e.g., in the studies of dislocations) motivated us to develop an *ab-initio* based multiscale approach. This approach combines first-principles calculations with empirical potential calculations and elasticity theory and links microscopic with the macroscopic relevant length scales. The accuracy of the multiscale approach is in principle limited only by the accuracy of the *ab-initio* calculation. Furthermore, the combination of first-principles calculations with the thermodynamic concept of chemical potentials allows to establish a first link between theory and growth experiments. This combination enables to identify the thermodynamic stability of a system as function of the growth conditions.

Based on the aforementioned *ab-initio* multiscale approach, we address the following controversial aspects and unsolved problems in the field of group III-Nitrides:

- Surface phase diagrams for the growth of GaN are by now well established and provide useful information for the surface morphology as function of the growth conditions [46]. Nevertheless, the results regarding the activation barriers for the transitions between the different regimes of a surface phase diagram are contradicting [46, 47, 48, 49].

- Threading dislocations even in device quality GaN are known to exist in high densities between 10^8 and 10^{10} [31]. The electronic structure of these defects is still under debate. Until now, *ab-initio* based studies on dislocations have been focused only in the vicinity of the dislocation line and thus they loose much of the information enclosed in the long-range/elastically-strained region around the core [50, 51, 52, 53, 54, 55]. On the other hand, empirical potential calculations have been successful in describing the long-range behavior but they loose much of the information enclosed in the core region [56, 57]. Studies that will link the different length scales are missing.

- Alloys of group III-Nitrides are of major technological interest. The existence of ordering in AlGaN alloys is well established [58, 59, 60]. However, studies that will link thermodynamics, growth kinetics, and optical properties of these alloys are missing.

In Part II the results of our calculations are presented and discussed along with recent experimental observations. First, we focus on the growth of the technologically most relevant GaN surfaces (i.e., the GaN {0001} surfaces). The morphology of these surfaces is known to be extremely sensitive to the growth conditions: Going from N-rich to more Ga-rich conditions a transformation of the morphology occurs accompanied by smoother surfaces and better film quality [46]. In Chapter 4 we examine the implications surface-steps/vicinal-surfaces may have on the surface morphologies and surface roughening. For growth under Ga-rich conditions it has been shown that the GaN (0001) surfaces are stabilized by the laterally contracted bilayer structure [43]. Technologically this surface turned out to be rather important since the top layer acts like a surfactant and enables 2-dimensional growth. A problem to experimentally realize this surface is that even a small amount of excess Ga can make the surface unstable against droplet formation. In Chapter 5 we therefore address the issues of Ga adsorption and droplet formation (i.e., the formation of small Ga islands on top of the laterally contracted bilayer structure). Although the GaN {0001} surface, is from a technological point view the most important one, non-polar surfaces such as the m-plane GaN surface[4] become more and more important [61]. The growth of non-polar surfaces has been motivated since growth along this axis prevents the buildup of macroscopic fields such as spontaneous and piezoelectric polarization along the polar [0001]-axis. Therefore, in Chapter 6 we present and discuss results of our *ab-initio* calculations on possible reconstructions of the GaN m-plane surface.

Steps and vicinal surfaces on the {0001} surfaces play a crucial role in the formation of ordering in InGaN and AlGaN alloys [43, 62]. Chemically ordered alloys are important since the optical properties of the III-Nitride ternary alloys depend sensitively on fluctuations on the scale of the nearest neighbor distribution. Specifically they principally allow the formation of atomically sharp interfaces or of quantum wells with precise thickness and offer thus the option to control atomic fluctuations on an atomic level. In Chapter 7 we therefor focus on ordering in AlGaN alloys. We

---

[4] The term m-plane corresponds to the $\{10\bar{1}0\}$ GaN surfaces.

examine the thermodynamics, electronic, and optical properties of ordering.

The electronic and optical properties of dislocations in GaN have been and still are one of the most controversial issues in the field of group III-Nitrides. The origin of these extended defects is well known: They are caused by the lack of a suitable lattice and thermal matched substrate. However, their atomic structure, chemical nature and electronic properties are poorly understood. In Chapter 8 we therefor focus on the geometry, energetics and electronic properties of edge type dislocations in GaN. Based on the multiscale approach, different dislocation core geometries are examined and their electronic properties are discussed in detail. We find that the giant local strain field around the dislocation core in combination with the relevant small lattice constant of GaN, introduces deep states in the gap. We also examine the effect strain and vacancies have on the geometry of the dislocation core.

Finally in Chapter 9 we briefly summarize the results presented in this thesis. In Part III the Appendices are given: In Appendix A the calculations on the bulk and elastic properties of GaN are cited while in Appendix B the coordinates of the configurations used to simulate random alloys are given.

Part I

Methods

Chapter 2

Ab-Initio Methods

This thesis addresses phenomena which are described by many-body systems, i.e. they consist of many electrons and nuclei. An N-particle quantum-mechanical system is described by a wavefunction $\Psi\left(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\}\right)$, where $\{\boldsymbol{r}_i\}$ and $\{\boldsymbol{R}_I\}$ correspond to the electronic and ionic coordinates. In real many-body systems, the number of particles N is of the order of Avogadro's constant $N_A = 6.023 \times 10^{23}$. Even with the up-to-date available computational power it is not possible for a problem of this size to be solved. However, approximate methods allow the treatment of such systems. In this and in the following chapter the principles of the approximate methods used here are given.

A Hamiltonian describing a perfect crystal (neglecting relativistic effects) can be written as:

$$H = \sum_{i}^{N_\mathrm{e}} \left[-\frac{1}{2}\nabla_i^2\right] + \sum_{I}^{N_I} \left[-\frac{1}{2M_I}\nabla_I^2\right] + \frac{1}{2}\sum_{I}^{N_I}\sum_{I'}^{N_I}\left[\frac{Z_I Z_{I'}}{|\boldsymbol{R}_I - \boldsymbol{R}_{I'}|}\right] - \\ - \sum_{i}^{N_\mathrm{e}}\sum_{I}^{N_I}\left[\frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|}\right] + \frac{1}{2}\sum_{i}^{N_\mathrm{e}}\sum_{i'}^{N_\mathrm{e}}\left[\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_{i'}|}\right]. \tag{2.1}$$

Here, the first term sums over the number of electrons N_e and the second term over the number of ions N_I. The two terms correspond to the kinetic energies of the electrons and the nuclei respectively, where M_I is the mass of the I_th nucleus. The third sum correspond to the Coulomb repulsion between nuclei, where $\boldsymbol{R}_I$ and Z_I are the position and the atomic number of the nucleus I respectively. The fourth term represents the electron–nucleus Coulomb attraction, where $\mathbf{r}_i$ is the position of the i_th electron. The last term is the repulsive electron–electron Coulomb interaction. Throughout this thesis atomic units have been used, i.e., $\hbar = 1$, $m_\mathrm{e} = 1$, $\frac{e^2}{4\pi\epsilon_0} = 1$.

Eq. (2.1) can be written in a more compact form:

$$H = T_\mathrm{e}\left(\{\boldsymbol{r}_i\}\right) + T_\mathrm{ions}\left(\{\boldsymbol{R}_I\}\right) + V_\mathrm{ions}\left(\{\boldsymbol{R}_I\}\right) + V_\text{e-ions}\left(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\}\right) + V_\mathrm{e}\left(\{\boldsymbol{r}_i\}\right). \tag{2.2}$$

Here, $\{\boldsymbol{r}_i\} = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_{N_\mathrm{e}}\}$ represents the positions of electrons and $\{\boldsymbol{R}_I\} = \{\boldsymbol{R}_1, \dots, \boldsymbol{R}_{N_I}\}$ the positions of ions. The energy of the system is obtained by solving the Schrödinger equation:

$$H\Psi_\xi\left(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\}\right) = E_\xi\Psi_\xi\left(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\}\right). \tag{2.3}$$

Here, Ψ_ξ is the ξ_th eigenfunction (wavefunction) of the system with energy eigenvalue E_ξ. Due to its complexity it is not possible to solve Eq. (2.3) directly and some approximations have to be implemented. The *adiabatic* or *Born–Oppenheimer approximation* is one of the most fundamental approximations in solid–state–physics and will be described in the following section.

2.1 Born-Oppenheimer Approximation (BO)

In 1923, Born and Oppenheimer noted that the nuclei are much more massive than the electrons and move on a time scale which is approximately two orders of magnitude larger than that of the electrons:

$$\frac{\omega_\mathrm{e}}{\omega_\mathrm{ion}} \sim \sqrt{\frac{M}{m}} \sim 100\,.$$

Thus, the electrons can respond to the ionic motion almost instantaneously and any change in the electronic state occurs vary rapidly compared to the nucleus motion. Therefore, it is a good approximation to separate electronic and ionic motion and to perform the calculation of the electronic structure separately from the calculation of the ionic motion. The wavefunction can then be written in the following form:

$$\Psi\left(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_I\}\right) = \psi\left(\{\boldsymbol{r}_i\};\{\boldsymbol{R}_I\}\right)\Xi\left(\{\boldsymbol{R}_I\}\right). \tag{2.4}$$

Here, $\psi\left(\{\boldsymbol{r}_i\};\{\boldsymbol{R}_I\}\right)$ is the electronic wavefunction and $\Xi\left(\{\boldsymbol{R}_I\}\right)$ the nuclear wavefunction. The electronic wavefunction is a function only of $\{\boldsymbol{r}_i\}$ with $\{\boldsymbol{R}_I\}$ as parameters. The parametric form express the assumption of stationery nuclei where the electronic motion takes place in the stationary potential created by the nuclei. The electronic wavefunction $\psi\left(\{\boldsymbol{r}_i\};\{\boldsymbol{R}_I\}\right)$ satisfies the time–independent Schrödinger equation for the electrons in a static array of nuclei:

$$H_\mathrm{e}\psi_\mathrm{e}\left(\{\boldsymbol{r}_i\};\{\boldsymbol{R}_I\}\right) = E_\mathrm{e}\psi_\mathrm{e}\left(\{\boldsymbol{r}_i\};\{\boldsymbol{R}_I\}\right)\,. \tag{2.5}$$

Here, the electronic Hamiltonian is:

$$H_\mathrm{e} = \sum_i^{N_\mathrm{e}}\left[-\frac{1}{2}\nabla_i^2\right] - \sum_i^{N_\mathrm{e}}\sum_I^{N_I}\left[\frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|}\right] + \frac{1}{2}\sum_i^{N_\mathrm{e}}\sum_{i'}^{N_\mathrm{e}}\left[\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_{i'}|}\right]. \tag{2.6}$$

The eigenvalues E_e in Eq. (2.5) depend on the particular values of $\{\boldsymbol{R}_I\}$: $E_\mathrm{e} = E_\mathrm{e}\left(\{\boldsymbol{R}_I\}\right)$. Regarding the Hamiltonian describing the ionic motion, the BO rationale is also used: Since the ionic motion is much slower than the electronic motion, the nuclei experience a smeared out potential from the electrons. Thus, the electronic coordinates can be replaced by their averaged values, averaged over the electronic wavefunction:

$$\langle E_{\mathrm{e}}\left(\{\boldsymbol{R}_I\}\right)\rangle = \left\langle \sum_{i}^{N_{\mathrm{e}}}\left[-\frac{1}{2}\nabla_i^2\right] - \sum_{i}^{N_{\mathrm{e}}}\sum_{I}^{N_I}\left[\frac{Z_I}{|\boldsymbol{r}_i-\boldsymbol{R}_I|}\right] + \frac{1}{2}\sum_{i}^{N_{\mathrm{e}}}\sum_{i'}^{N_{\mathrm{e}}}\left[\frac{1}{|\boldsymbol{r}_i-\boldsymbol{r}_{i'}|}\right]\right\rangle . \tag{2.7}$$

$\langle E_{\mathrm{e}}\left(\{\boldsymbol{R}_I\}\right)\rangle$ can be assumed as an averaged electronic potential experienced by the nuclei. Then, the ionic Hamiltonian can be written in the following form:

$$\begin{aligned} H_{\mathrm{ion}} &= \sum_{I}^{N_I}\left[-\frac{1}{2M_I}\nabla_I^2\right] + \frac{1}{2}\sum_{I}^{N_I}\sum_{I'}^{N_I}\left[\frac{Z_I Z_{I'}}{|\boldsymbol{R}_I-\boldsymbol{R}_{I'}|}\right] + \langle E_{\mathrm{e}}\left(\{\boldsymbol{R}_I\}\right)\rangle \\ &= \sum_{I}^{N_I}\left[-\frac{1}{2M_I}\nabla_I^2\right] + E_{\mathrm{tot}}\left(\boldsymbol{R}_I\right) . \end{aligned} \tag{2.8}$$

From Eq. (2.8) it becomes obvious that the nuclei move on a potential energy surface (PES) derived from the solution of the electronic problem. Although the BO approximation is widely used, it is not valid if the electronic and the nuclei motion cannot be decoupled (e.g. if T_{ions} in Eq. (2.2) becomes important compared to T_{e}), as for example in high-energy atom-surface collisions. In this work we will not deal with fast atomic processes: The BO approximation is thus an excellent approximation.

2.2 The variational principle

The variational principle is extremely powerful and provides the basis for the vast majority of methods aim to find an approximate solution to Scrödinger equation (Eq. (2.3)). According to the variational principle, the expectation value of the Hamiltonian H using a trial wavefunction Ψ_{T} $E\left[\Psi_{\mathrm{T}}\right] = \langle\Psi_{\mathrm{T}}|H|\Psi_{\mathrm{T}}\rangle$[1] is always greater or equal than the true ground state energy E_0:

$$E\left[\Psi_{\mathrm{T}}\right] \geq E_0 \, . \tag{2.9}$$

The true ground state wavefunction Ψ_0 can be approached by minimizing the functional $E\left[\Psi_{\mathrm{T}}\right]$ with respect to the all allowed N_{e}-electron wavefunctions:

$$E_0 = \min_{\Psi} E\left[\Psi\right] . \tag{2.10}$$

Proof of the variational principle can be found in Ref. [63].

---

[1] Here, we assume the wavefunction Ψ_{T} to be normalized. Otherwise the expectation value would be: $E\left[\Psi_{\mathrm{T}}\right] = \langle\Psi_{\mathrm{T}}|H|\Psi_{\mathrm{T}}\rangle \, / \, \langle\Psi_{\mathrm{T}}|\Psi_{\mathrm{T}}\rangle$.

2.3 The Hartree–Fock approximation

Based on the BO electronic and atomic equations of motion are decoupled (see Sec. 2.1). However, the treatment of the many electron problem described by Eq. (2.5) is still not directly possible and one has to rely on approximate theories, such as e.g. the Hartree-Fock (HF) approximation. Using the HF approximation the N_e-electron wavefunction Ψ is replaced by a Slater determinant of single particle wavefunctions:

$$\begin{aligned}\Psi_\mathrm{HF} &= \frac{1}{\sqrt{N_\mathrm{e}!}}\begin{vmatrix} \psi_{\lambda_1}(\boldsymbol{r}_1; s_1) & \psi_{\lambda_2}(\boldsymbol{r}_1; s_1) & \dots & \psi_{\lambda_{N_\mathrm{e}}}(\boldsymbol{r}_1; s_1) \\ \psi_{\lambda_1}(\boldsymbol{r}_2; s_2) & \psi_{\lambda_2}(\boldsymbol{r}_2; s_2) & \dots & \psi_{\lambda_{N_\mathrm{e}}}(\boldsymbol{r}_2; s_2) \\ \vdots & \vdots & & \vdots \\ \psi_{\lambda_1}(\boldsymbol{r}_{N_\mathrm{e}}; s_{N_\mathrm{e}}) & \psi_{\lambda_2}(\boldsymbol{r}_{N_\mathrm{e}}; s_{N_\mathrm{e}}) & \dots & \psi_{\lambda_{N_\mathrm{e}}}(\boldsymbol{r}_{N_\mathrm{e}}; s_{N_\mathrm{e}}) \end{vmatrix} \\ &= \frac{1}{\sqrt{N_\mathrm{e}!}} \det\left[\psi_{\lambda_1}\psi_{\lambda_2}\dots\psi_{\lambda_{N_\mathrm{e}}}\right] \qquad (2.11)\end{aligned}$$

Here, $\psi_{\lambda_i}(\boldsymbol{r}_i; s_i)$ denotes the one-electron wavefunction of the λ_i^th level and depends both on the spatial position $\mathbf{r}_i$ and the spin direction s_i. The one-electron wavefunction $\psi_{\lambda_i}(\boldsymbol{r}_i; s_i)$ form an orthonormal set: $\langle \psi_{\lambda_i} | \psi_{\lambda_j} \rangle = \delta_{ij}$[2]. Ψ_HF gives the antisymmetric wavefunction for N_e indistinguishable particles and serves as a trial function. The ground state energy is determined by a variational principle. However, the HF approximation has the disadvantage that it cannot describe the electronic correlation properly. For example it gives zero density of states at the Fermi level of metals. More detailed discussion on the HF approximation can be found in Refs. [45, 63].

2.4 Density Functional Theory

Alternatively to the HF approximation, one may consider the *Density Functional Theory (DFT)*. In DFT the many–body problem is reformulated in terms of the electron density [45, 63]. Electron density compared to the many-body wavefunction is more attractive, since it drastically reduces the degrees of freedom: it depends only on x, y, z spatial coordinates and there may be two densities for spin polarized systems – $\rho_\uparrow$ (spin up) and $\rho_\downarrow$ (spin down). In contrast, the many–body wavefunction depends on all coordinates of all particles (i.e., for N_e electrons depends on $3N_\mathrm{e}$ coordinates or $4N_\mathrm{e}$ if we include the spin).

Although, DFT originally applied in atoms and simple molecules, nowadays can be assumed as a work horse for more complex systems in the chemical and materials sciences [64]. The origins of DFT can be tracked buck to the works of Thomas and Fermi [65, 66, 67, 68]. In Thomas–Fermi

---

[2] δ_{ij} is the Kronecker delta and it is defined as:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad (2.12)$$

model the electron density $n(\boldsymbol{r})$ is the "basic variable" for treating the many–body problem. The basic framework of DFT was established by Hohenberg and Kohn [69].

2.4.1 The Hohenberg–Kohn theorems

The 1st theorem Let us assume a system of an arbitrary number of electrons N_e moving under the influence of an external potential $\upsilon(\boldsymbol{r})$. The Hamiltonian of the system can be written in the following form:

$$H_\mathrm{e} = T_\mathrm{e} + V_\mathrm{ee} + \upsilon \tag{2.13}$$

Here, T_e is the kinetic energy term and V_ee corresponds to the electron–electron Coulomb interaction. The first Hohenberg–Kohn theorem states that the external potential $\upsilon(\boldsymbol{r})$ is – within a trivial additive constant– unique functional of the electron density $n(\boldsymbol{r})$. The electron charge density distribution is related to the N_e-body wavefunction $\Psi(\boldsymbol{r})$ via

$$n(\boldsymbol{r}) = \left\langle \Psi \left| \sum_i^{N_\mathrm{e}} \delta(\boldsymbol{r} - \boldsymbol{r}_i) \right| \Psi \right\rangle . \tag{2.14}$$

The electron density $n(\boldsymbol{r})$ determines the number of electrons. Thus, it determines the ground state wavefunction Ψ and all other properties of the ground state (e.g., eigenvalues, Hamiltonian). Thus, we can conclude that: Every observable of a stationary quantum system, can be calculated, in principle exactly, by the ground state electron density $n(\boldsymbol{r})$ alone. Thus, it is possible to write an *energy functional* $E_\upsilon[n(\boldsymbol{r})]$ in terms of the electron density $n(\boldsymbol{r})$:

$$\begin{aligned} E_\upsilon[n] &= T_\mathrm{e}[n] + V_{ne}[n] + V_{ee}[n] \\ &= \int \upsilon(\boldsymbol{r})\, n(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + F[n] . \end{aligned} \tag{2.15}$$

Here,

$$F[n] = T_\mathrm{e}[n] + V_{ee}[n] . \tag{2.16}$$

$F[n]$ is a *universal functional* independent of the number of particles or the external potential.

The 2nd theorem The 2nd Hohenberg–Kohn theorem provides the energy variational principle and reads:
The total energy $E_\upsilon[\widetilde{n}]$, calculated by Eq. (2.15) and a trial density $\widetilde{n}(\boldsymbol{r})$, which has to obey charge conservation $\int \widetilde{n}(\boldsymbol{r})\ u d\boldsymbol{r} = N_\mathrm{e}$, is always greater or equal to the true ground state total energy E_0 of the system:

$$\begin{aligned} E_\upsilon[\widetilde{n}] &\geq E_0 \\ \int \widetilde{n}[\boldsymbol{r}]\, \upsilon(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + F[\widetilde{n}] &\geq E_\upsilon[n_0] \end{aligned} \tag{2.17}$$

If the functional $F\left[n\left(\boldsymbol{r}\right)\right]$ was known, the problem of calculating the ground state energy and density could be easily solved: All one would have to do would be to minimize the functional $E_{\upsilon}\left[n\left(\boldsymbol{r}\right)\right]$ as given by Eq. (2.15) with respect to the three dimensional density $n\left(\boldsymbol{r}\right)$.

2.4.2 The Kohn–Sham equations

The Hohenberg-Kohn theorems states that all quantities of a physical system can be calculated from the electron density alone via the universal functional $F\left[n\right]$. However, the problem of obtaining an expression for the kinetic energy of the electrons in terms of the charge density $T_{\mathrm{e}}n\left(\boldsymbol{r}\right)$ remains. Kohn-Sham equations tackle this problem by assuming an auxiliary problem of *"non-interacting particles"* [70].

Following the Kohn-Sham approach, the functional $E_{\upsilon}\left[n\left(\boldsymbol{r}\right)\right]$ in Eq. (2.15) can be reformulated as follows:

$$E\left[n\left(\boldsymbol{r}\right)\right] = \int \upsilon\left(\boldsymbol{r}\right) n\left(\boldsymbol{r}\right) \mathrm{d}\boldsymbol{r} + J\left[n\left(\boldsymbol{r}\right)\right] + G\left[n\left(\boldsymbol{r}\right)\right] . \tag{2.18}$$

Here,

$$J\left[n\left(\boldsymbol{r}\right)\right] = \frac{1}{2}\int\int \frac{n\left(\boldsymbol{r}\right) n\left(\boldsymbol{r}'\right)}{\left|\boldsymbol{r}-\boldsymbol{r}'\right|} \mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{r}' \tag{2.19}$$

corresponds to the classical electrostatic repulsion (Hartree term). $G\left[n\right]$ is a universal functional of the density which has the following form:

$$G\left[n\left(\boldsymbol{r}\right)\right] \equiv T_s\left[n\left(\boldsymbol{r}\right)\right] + E_{\mathrm{XC}}\left[n\left(\boldsymbol{r}\right)\right] . \tag{2.20}$$

Here, $T_s\left[n\right]$ is the kinetic energy functional of a system of non-interacting particles and $E_{\mathrm{XC}}\left[n\right]$ is the so-called *exchange and correlation* (XC) energy of a system of interacting particles. By introducing single-particle wavefunctions $\psi_i\left(\boldsymbol{r};s\right)$ (which take into account that we have now to deal with non-interacting electrons) the density and the kinetic energy can be written:

$$n\left(\boldsymbol{r}\right) = \sum_{i}^{N_{\mathrm{occ}}} \left\langle \psi_i\left(\boldsymbol{r};s\right) \middle| \psi_i\left(\boldsymbol{r};s\right) \right\rangle , \quad T_s\left[n\left(\boldsymbol{r}\right)\right] = \sum_{i}^{N_{\mathrm{occ}}} \left\langle \psi_i \middle| -\frac{1}{2}\nabla^2 \middle| \psi i \right\rangle . \tag{2.21}$$

Here, the sums run over all the occupied states N_{occ}. It has been further assumed that the orbitals are orthonormal:

$$\left\langle \psi_i \middle| \psi_j \right\rangle = \delta_{ij} . \tag{2.22}$$

Minimization of Eq. (2.18) is subject to the constrains of Eq. (2.22) and leads to the equations:

$$\widehat{h}_{\mathrm{eff}}\,\psi_i = \left[-\frac{1}{2}\nabla^2 + \upsilon_{\mathrm{eff}}\right]\psi_i = \sum_{j}^{N_{\mathrm{occ}}} \epsilon_{ij}\psi_j . \tag{2.23}$$

Here, $\widehat{h}_{\mathrm{eff}}$ is a one-electron Hamiltonian, ϵ_{ij} are the Lagrange multipliers due to the orthonormalization constrain (2.22), and υ_{eff} is the single electron effective potential:

$$\begin{aligned}\upsilon_{\text{eff}}(\boldsymbol{r}) &= \upsilon(\boldsymbol{r}) + \frac{\delta J[n(\boldsymbol{r})]}{n(\boldsymbol{r})} + \frac{\delta E_{\text{XC}}[n(\boldsymbol{r})]}{n(\boldsymbol{r})} \\ &= \upsilon(\boldsymbol{r}) + \int \frac{n(\mathbf{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\mathrm{d}\boldsymbol{r}' + \upsilon_{\text{XC}}(\boldsymbol{r}) \ . \qquad (2.24)\end{aligned}$$

Here, $\upsilon(\boldsymbol{r})$ is the external potential and $\upsilon_{\text{XC}}(\boldsymbol{r})$ is the exchange–correlation potential given by the functional derivative of the exchange-correlation energy $E_{\text{XC}}[n(\boldsymbol{r})]$:

$$\upsilon_{\text{XC}}(\boldsymbol{r}) = \frac{\delta E_{\text{XC}}[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} \ . \qquad (2.25)$$

Since the Hamiltonian $\widehat{h}_{\text{eff}}$ is Hermitian, the matrix (ϵ_{ij}) is Hermitian, and can be diagonalized by a unitary transformation of the orbitals. Thus Eq. (2.23) can be reformulated:

$$\left[-\frac{1}{2}\nabla^2 + \upsilon_{\text{eff}}\right]\psi_i = \epsilon_i\psi_i \qquad (2.26)$$

Eqs. (2.21) and (2.24)–(2.26) are the so-called *Kohn–Sham equations* (KS) and need to be solved self-consistently, due to the density dependence of the effective one-electron potential $\upsilon_{\text{eff}}[n(\boldsymbol{r})]$. The procedure is as follows: One starts with a guess for the electron density $n^0(\boldsymbol{r})$, then the exchange–correlation and the effective potentials are calculated using Eqs. (2.25) and (2.24) respectively. The single-particle orbitals are obtained from Eq. (2.26) and a new density is generated by Eq. (2.21). The above procedure is repeated until convergence is achieved. The ground-state total energy is given by Eqs. (2.18) and (2.20) or equivalently by:

$$E = \sum_i^{N_{\text{occ}}} \epsilon_i - \frac{1}{2}\int\int \frac{n(\boldsymbol{r})\,n(\mathbf{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{r}' + E_{\text{XC}}[n(\boldsymbol{r})] - \int \upsilon_{\text{XC}}(\boldsymbol{r})\,n(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \qquad (2.27)$$

The KS theory is exact in principle and (in principle) fully incorporates the exchange-correlation effect of the electrons. The KS equations will give the exact ground state density and energy, if the exchange–correlation functional $E_{\text{XC}}[n(\boldsymbol{r})]$ is known precisely. A large number of approximations to $E_{\text{XC}}[n(\boldsymbol{r})]$ have been developed over the years and research in this field is ongoing.

2.4.3 Approximations to E_{XC} : LDA and GGA

The simplest method of describing the exchange and correlation energy is the *local-density approximation* (LDA) [70]. LDA assumes the XC energy per electron at a point $\boldsymbol{r}$ in the electron gas, $\epsilon_{\text{xc}}[n(\boldsymbol{r})]$, to be the same as the XC energy per electron in a homogeneous electron gas that has the same density as the electron gas at the same point $\boldsymbol{r}$: $\epsilon_{\text{xc}}^{\text{hom}}[n(\boldsymbol{r})]$. Despite the simplicity of the LDA, it has been and is widely used to study a large variety of systems. More recently a more complex method that takes into account additionally the gradient terms in charge density, ∇n, has been developed (*generalized–gradient–approximation* – GGA).

Local Density Approximation

The XC energy in LDA is given by:

$$E_{\mathrm{XC}}^{\mathrm{LDA}} = \int \epsilon_{\mathrm{xc}}^{\mathrm{hom}} \left[n\left(\boldsymbol{r}\right) \right] n\left(\boldsymbol{r}\right) \mathrm{d}\mathbf{r} \,. \tag{2.28}$$

Here, the XC energy per electron, $\epsilon_{\mathrm{xc}}^{\mathrm{hom}} \left[n\left(\boldsymbol{r}\right) \right]$, can be written as the sum of the exchange and correlation energies per electron: $\epsilon_{\mathrm{xc}}^{\mathrm{hom}} = \epsilon_{\mathrm{x}}^{\mathrm{hom}} + \epsilon_{\mathrm{c}}^{\mathrm{hom}}$. The exchange part is given by the Dirac exchange-energy functional [63, 71]:

$$\epsilon_{\mathrm{x}}^{\mathrm{hom}} = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} n\left(\boldsymbol{r}\right)^{1/3} \tag{2.29}$$

The exact form of the correlation part is unknown. Analytical expressions are available as asymptotic expansions in the low ($r_{\mathrm{s}} \rightarrow 0$) [72] and high ($r_{\mathrm{s}} \rightarrow \infty$) [73] density limits. Here, $r_{\mathrm{s}} = \left(\frac{4\pi}{3} n \right)^{-1/3}$ is the Wigner-Seitz radius. However, more recent representations based on Quantum-Monte-Carlo calculations on the uniform gas reported by Ceperley and Alder [74] may be used. The most widely used parametrization has been proposed by Perdew and Zunger and is used in this thesis [75]:[3]

$$\epsilon_{\mathrm{c}}^{\mathrm{hom}} = \begin{cases} -0.1423/\left(1 + 1.0529\sqrt{r_{\mathrm{s}}} + 0.3334 r_{\mathrm{s}}\right) & \text{if} \quad r_{\mathrm{s}} \geq 1 \\ -0.048 + 0.0311 \ln r_{\mathrm{s}} - 0.0116 r_{\mathrm{s}} + 0.002 r_{\mathrm{s}} \ln r_{\mathrm{s}} & \text{if} \quad 0 \leq r_{\mathrm{s}} \leq 1 \end{cases} \tag{2.30}$$

Despite the simplicity of the LDA functional it has been successful in describing the atomic structure, the formation enthalpy and the elastic and vibrational properties of a wide range of systems. A common problem of LDA is that it ”overbinds” and underestimates the lattice constants within a 1% ∼ 3% accuracy. Although in most cases it gives correct band structures, the calculated bandgaps are smaller, but this a more general problem not explicitly connected to LDA. A next step beyond the LDA is to take into account, besides the local density n, the gradient terms in the charge density ($\nabla n\left(\mathbf{r}\right)$).

Generalized Gradient Approximation

If the charge density becomes non-uniform and undergoes rapid changes, it is convenient to include in the XC-energy the gradient ($\nabla n\left(\boldsymbol{r}\right)$) and higher spatial derivatives of the total charge

---

[3] The formula presented here is valid for spin-unpolarized systems ($\zeta \equiv \frac{n\uparrow - n\downarrow}{n} = 0$). For spin-polarized systems it is necessary to treat the spin-up ($n\uparrow$) and spin-down ($n\downarrow$) densities separately and the *local-spin-density approximation* (LSDA) should be used. The LSDA has been first proposed by Barth and Hedin [76].

density. The *Generalized-Gradient-Approximation* (GGA) is based on this idea and can be symbolically written as:

$$E_{\mathrm{XC}} = E_{\mathrm{XC}}^{\mathrm{GGA}}\left[n\left(\boldsymbol{r}\right);\nabla n\left(\boldsymbol{r}\right)\right] = \int F_{\mathrm{XC}}\left[n\left(\boldsymbol{r}\right);\left|\nabla n\left(\boldsymbol{r}\right)\right|\right]\mathrm{d}\boldsymbol{r} \tag{2.31}$$

Several approximations have been proposed for the F_{XC} functional, as for example those by Becke [77] (exchange part) and Perdew[78] (correlation part) (BP), Becke [77] (exchange part) and Lee-Yang-Parr [79] (exchange part) (BLYP), Perdew and Wang [80] (PW91) or by Perdew, Burke, and Ernzehof [81] (PBE).

In general the GGA functional can be written in terms of the local exchange that is modified by an enhancement factor:

$$E_{\mathrm{XC}}^{\mathrm{GGA}}\left[n\left(\boldsymbol{r}\right)\right] = \int n\left(\boldsymbol{r}\right)\epsilon_{\mathrm{x}}^{\mathrm{hom}}\left[n\left(\boldsymbol{r}\right)\right]F_{\mathrm{XC}}\left(r_{\mathrm{s}},s\right)\mathrm{d}\boldsymbol{r} \tag{2.32}$$

where $s\left(\boldsymbol{r}\right) = \left|\nabla n\right|/\left(2k_{\mathrm{F}}n\right)$ with $k_{\mathrm{F}} = \left(3\pi^2 n\right)^{1/3}$ the reduced density gradient.

In the PBE approximation the correlation part of the energy in GGA is:[4]

$$E_{\mathrm{C}}^{\mathrm{GGA}}\left[n\left(\boldsymbol{r}\right)\right] = \int n\left(\boldsymbol{r}\right)\left(\epsilon_{\mathrm{c}}^{\mathrm{hom}}\left(r_{\mathrm{s}}\right) + H_{\mathrm{C}}\left(r_{\mathrm{s}},t\right)\right)\mathrm{d}\boldsymbol{r}\,. \tag{2.33}$$

Here, $t = \left|\nabla n\right|/\left(2k_{\mathrm{s}}n\right)$ is a dimensionless density gradient, with $k_{\mathrm{s}} = \sqrt{4k_{\mathrm{F}}/\pi\alpha_0}$ and $\alpha_0 = \hbar^2/me^2$. For the term $H_{\mathrm{C}}\left(r_{\mathrm{s}},t\right)$ we have:

$$H_{\mathrm{C}}\left(r_{\mathrm{s}},t\right) = \gamma\ln\left[1 + \frac{\beta}{\gamma}t^2\left(\frac{1+At^2}{1+At^2+A^2t^4}\right)\right], \tag{2.34}$$

$$A = \frac{\beta}{\gamma}\frac{1}{\exp\left(-\epsilon_{\mathrm{c}}^{\mathrm{hom}}/\gamma\right)-1},$$

$$\gamma = \frac{1-\ln 2}{\pi^2}$$

where $\beta \simeq 0.066725$.

The exchange part is given by:

$$E_{\mathrm{X}}^{\mathrm{GGA}}\left[n\left(\boldsymbol{r}\right)\right] = \int n\left(\boldsymbol{r}\right)\epsilon_{\mathrm{x}}^{\mathrm{hom}}\left[n\left(\boldsymbol{r}\right)\right]\left(1+\kappa-\frac{\kappa}{1+\mu s^2/\kappa}\right)\mathrm{d}\boldsymbol{r}\,. \tag{2.35}$$

Here, $\mu = \beta\left(\pi^2/3\right)$ and $\kappa = 0.804$.

For many solids GGA was found to correct the LDA-overbinding problem and significantly improve cohesive energies for solids (see Ref. [82] and Refs. therein). Moreover, it gives a better description of the activation energy barriers in chemical reactions [83, 84]. On the other hand,

---

[4] As in the case of LDA functional, we refer again for spin-unpolarized systems.

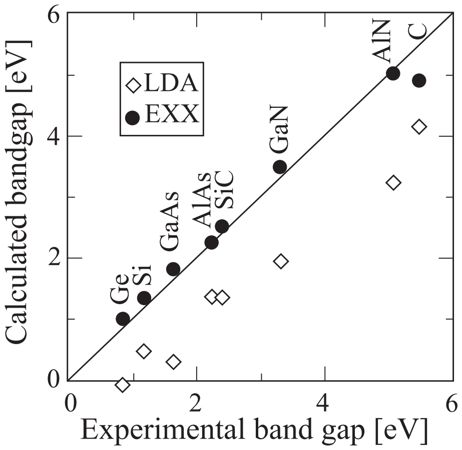


Figure 2.1: Comparison of self-consistently calculated LDA and EXX bandgaps with experimental data for various semiconductors. After Städele *et al.* [87].

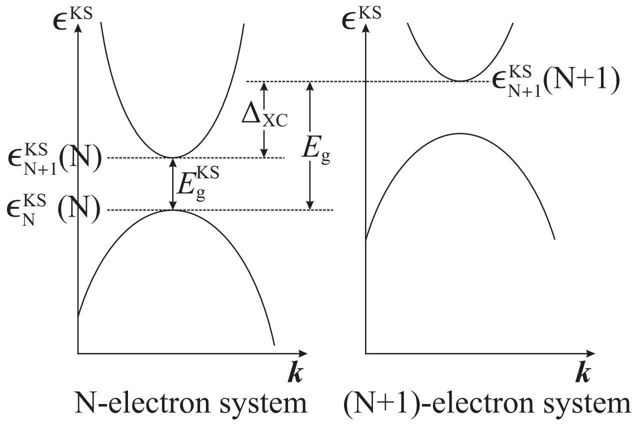


Figure 2.2: Illustration of the exact DFT Kohn-Sham one-electron energies shown in the form of a band structure for the N- and $(N+1)$-particle systems. After Godby *et al.* [88].

GGA in solids softens the bonds yielding to an increase of the bond lengths and subsequently a decrease in the bulk moduli [85, 86]. In GGA the problem of the bandgap underestimation remains (see Refs [45, 64]).

A few remarks on the bandgap problem in DFT.

It has been already mentioned that LDA and GGA DFT calculations systematically underestimate the bandgap of semiconductors and insulators (see Fig. 2.1). The bandgap E_{g} of an N electron system is defined in terms of the ionization potential $I = -\left(E\left[N\right] - E\left[N-1\right]\right)$ and the electron affinity $A = -\left(E\left[N+1\right] - E\left[N\right]\right)$ with $E\left[N+1\right]$, $E\left[N\right]$, and $E\left[N-1\right]$ being the ground-state total energies of the $(N+1)$-, (N)-, and $(N-1)$-electron systems respectively:

$$E_{\mathrm{g}} = I - A = E\left[N+1\right] + E\left[N-\right] - 2E\left[N\right]. \tag{2.36}$$

It has been shown that the bandgap can be expressed in the following form [89, 90]:

$$E_{\mathrm{g}} = E_{\mathrm{g}}^{\mathrm{KS}} + \Delta_{\mathrm{XC}}. \tag{2.37}$$

Here,

$$E_{\mathrm{g}}^{\mathrm{KS}} = \epsilon_{N+1}^{\mathrm{KS}}\left(N\right) - \epsilon_{N}^{\mathrm{KS}}\left(N\right) \tag{2.38}$$

is the Kohn-Sham bandgap (the gap of a non-interacting system). $\epsilon_{i}^{\mathrm{KS}}\left(M\right)$ corresponds to the i-th single-particle eigenvalue of the M-electron system. Δ_{XC} is just the difference between the $(N+1)$-th eigenvalues in the $(N+1)$- and N-particle systems: $\Delta_{\mathrm{XC}} = \epsilon_{N+1}^{\mathrm{KS}}\left(N+1\right) - \epsilon_{N+1}\left(N\right)$.

Sham and Schlüter showed that the difference Δ_{XC} arises from the discontinuity in the exchange-correlation potential under adding and subtracting an infinitesimal fraction η of the integer particle number N:

$$\Delta_{\mathrm{XC}} = \lim_{\eta \to 0} \left[\left. \frac{\delta E_{\mathrm{XC}}[n]}{\delta n} \right|_{N+\eta} - \left. \frac{\delta E_{\mathrm{XC}}[n]}{\delta n} \right|_{N-\eta} \right] \tag{2.39}$$

The situation described by Eq. (2.37) is illustrated in Fig. 2.2. The exact DFT Kohn-Sham one-electron eigenvalues are shown in form of a band-structure for the N- and $(N+1)$-electron systems. The two band-structures differ in a shift of the eigenvalues by Δ_{XC} which is only weakly dependent on $\boldsymbol{k}$. The XC-potential discontinuity cannot be accounted for by continuous functionals like the LDA or GGA. However, it is still an open and controversial question if DFT could be solved exactly, would provide the correct bandgap (see Ref. [64] and Refs. therein). Various approaches have been developed to remedy the bandgap problem including the application of exact-exchange potential (EXX) (see Fig. 2.1 and Ref. [87]), Green-function methods (GW approximation) [91, 92], and Time Dependent DFT (TDDFT) [93]. Nevertheless, these methods are computationally expensive when treating solids and in this thesis we will restrict on LDA/GGA calculations only. The poorly estimated bandgap does not effect ground state properties like, charge density, total energy, and forces. Additionally, the weak dependence on $\boldsymbol{k}$ of the value of Δ_{XC} leaves the band dispersion unaffected [88]. Here (Chapters 7 and 8), we will assume that the eigenvalues predicted by DFT are representative of the electronic band structure.

2.5 The plane-wave representation

Density Functional Theory maps the many body problem onto an effective single particle problem. Nevertheless, it is impossible to handle a problem consisting of an infinite number of non-interacting electrons moving in a static potential caused by an infinite number of nuclei. The problem can be solved by chosing a minimum representation of the system and model the "infinitely" large system using *periodic boundary conditions* (PBC).

The periodicity of the system can be expressed in terms of a set of three linear independent vectors $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\alpha}_3$ (*primitive vectors*). The prism spanned by the three primitive vectors is called *supercell* and corresponds to the minimum possible representation of the system under consideration[5]. Hence any lattice point $\boldsymbol{R}'$ can be obtained by another point $\boldsymbol{R}$ that lies inside the supercell by $\boldsymbol{R}' = \boldsymbol{R} + \mathrm{n}_1\boldsymbol{\alpha}_1 + \mathrm{n}_2\boldsymbol{\alpha}_2 + \mathrm{n}_3\boldsymbol{\alpha}_3$, with n_1, n_2, n_3 being integers. In general, with PBC we can treat bulk systems as well as surfaces and isolated atoms, molecules or clusters of atoms. Fig. (4.7) shows how a surface can be modelled using PBC with a repeated slab geometry.

---

[5] The supercell does not necessary coincides with the primitive unit cell. The primitive unit cell is the smallest possible volume including the basis set (minimal set of atoms) that reproduces the corresponding Bravais lattice.

Bloch's Theorem

In a system with translational invariance, the effective one electron Hamiltonian obeys also translational invariance: $H = -\frac{1}{2}\nabla^2 + V(\boldsymbol{r})$, where $V(\boldsymbol{r}) = V(\boldsymbol{r} + \boldsymbol{R})$. Bloch's theorem states that in a periodic solid, due to the translational invariance of the Hamiltonian, any effective one-electron wavefunction can be written as a product of a wave-like part and a periodic function:

$$\psi_{\boldsymbol{k}\lambda}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\,\boldsymbol{k}\cdot\mathbf{r}}\, u_{\boldsymbol{k}\lambda}(\mathbf{r}) \; . \tag{2.40}$$

Here, λ is the band index, $\mathbf{k}$ is the wave vector that lies within the 1^{st} Brillouin zone[6] (BZ) and $u_{\boldsymbol{k}\lambda}(\boldsymbol{r})$ is a function with the same periodicity as the real space lattice: $u_{\boldsymbol{k}\lambda}(\boldsymbol{r}) = u_{\boldsymbol{k}\lambda}(\boldsymbol{r} + \boldsymbol{R})$. The periodic function can be expanded as a Fourier series:

$$u_{\boldsymbol{k}\lambda}(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\boldsymbol{G}} \widetilde{u}_{\boldsymbol{k}\lambda}(\boldsymbol{G})\, \mathrm{e}^{\mathrm{i}\,\boldsymbol{G}\cdot\boldsymbol{r}} \tag{2.41}$$

$$\widetilde{u}_{\boldsymbol{k}\lambda}(\boldsymbol{G}) = \frac{1}{\sqrt{\Omega}} \int_{\text{cell}} \mathrm{e}^{-\,\mathrm{i}\,\boldsymbol{G}\cdot\boldsymbol{r}}\, u_{\boldsymbol{k}\lambda}(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}\, . \tag{2.42}$$

Here, Ω is the volume of the unit cell and $\boldsymbol{G}$ is a reciprocal lattice vector defined by $\boldsymbol{G} = l_1\boldsymbol{b}_1 + l_2\boldsymbol{b}_2 + l_3\boldsymbol{b}_3$ with $\boldsymbol{b}_i$ being the primitive reciprocal lattice vectors that satisfy $\boldsymbol{\alpha}_i \cdot \boldsymbol{b}_j = 2\pi\delta_{ij}$.

Plane Waves Representation

Substituting Eq. (2.41) into the Bloch's theorem (2.40), the effective one-electron wavefunction can be expanded by plane waves (PW):

$$\psi_{\boldsymbol{k}\lambda}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} C_{\boldsymbol{k}\lambda}(\boldsymbol{G}) \cdot |\, \boldsymbol{k} + \boldsymbol{G}\, \rangle\, . \tag{2.43}$$

Here, $|\, \boldsymbol{k} + \boldsymbol{G}\, \rangle = \mathrm{e}^{\mathrm{i}(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}}$ denotes a single PW. The Kohn-Sham equation in reciprocal space can be written as:

$$\sum_{\boldsymbol{G}'} \left\{ -\frac{1}{2} |\boldsymbol{k} + \boldsymbol{G}|^2\, \delta_{\boldsymbol{G},\boldsymbol{G}'} + v_{\text{eff}}\left(\boldsymbol{G}, \boldsymbol{G}'\right) \right\} C_{\boldsymbol{k}\lambda}\left(\boldsymbol{G}'\right) = \epsilon_{\boldsymbol{k}\lambda} C_{\boldsymbol{k}\lambda}(\boldsymbol{G})\, . \tag{2.44}$$

In practice, in the plane wave expansion, the maximum value of the kinetic energy $\frac{1}{2}|\boldsymbol{k} + \boldsymbol{G}|^2$ is limited to some value which is referred to the *cutoff energy*: $\frac{1}{2}|\boldsymbol{k} + \boldsymbol{G}|^2 < E_{\text{cutoff}}$. E_{cutoff} is a parameter that controls the quality of the basis set and should be carefully optimized prior to the calculations.

---

[6] A single unit cell in the reciprocal space.

Brillouin Zone Integration

One important consequence emerging from the spatial periodicity within a supercell approach and Bloch's theorem is that charge density and other quantities can be calculated by integration over the 1st BZ:

$$n\left(\boldsymbol{r}\right)=\frac{1}{\Omega_{\mathrm{BZ}}}\sum_{\lambda}^{\mathrm{occ}}\int_{\mathrm{BZ}}f\left(\epsilon_{\boldsymbol{k}\lambda}-\epsilon_{\mathrm{F}}\right)\left|\psi_{\boldsymbol{k}\lambda}\left(\boldsymbol{r};\boldsymbol{k}\right)\right|^{2}\mathrm{d}^{3}\boldsymbol{k}. \tag{2.45}$$

Here, Ω_{BZ} is the volume of the BZ, ϵ_{F} is the Fermi energy and $f\left(\epsilon_{\boldsymbol{k}\lambda}-\epsilon_{\mathrm{F}}\right)$ is the Fermi distribution function:

$$f\left(\epsilon_{\boldsymbol{k}\lambda}-\epsilon_{\mathrm{F}}\right)=\frac{1}{\mathrm{e}^{\left(\epsilon_{\boldsymbol{k}\lambda}-\epsilon_{\mathrm{F}}\right)/k_{\mathrm{B}}T}+1}.$$

The sum in Eq. (2.45) is over all the occupied states λ and it is valid for insulators and semiconductors.

The integral in Eq. (2.45)can be approximated by a weighted sum over special $\boldsymbol{k}$-points:

$$n\left(\boldsymbol{r}\right)\approx\sum_{\lambda}^{\mathrm{occ}}\sum_{i}^{N_{\mathrm{kpt}}}w_{i}\left|\psi_{\boldsymbol{k}_{i}\lambda}\left(\boldsymbol{r};\boldsymbol{k}_{i}\right)\right|^{2} \tag{2.46}$$

The trick applied here is based on the fact that the electronic wavefunction at $\boldsymbol{k}$-points that are very close together will be almost identical (most general: all quantities are smooth in the $\boldsymbol{k}$-space). Thus choosing a dense enough mesh the integral in Eq. (2.45) can be accurately approximated by Eq. (2.46).

Several methods have been developed in calculating the charge density and thus the total energy in a finite discrete set of $\boldsymbol{k}$-points [94, 95]. In this work the Monkhorst-Pack scheme is used for contracting the $\boldsymbol{k}$-points mesh inside the BZ zone [95]. In general the $\boldsymbol{k}$-points are specified as:

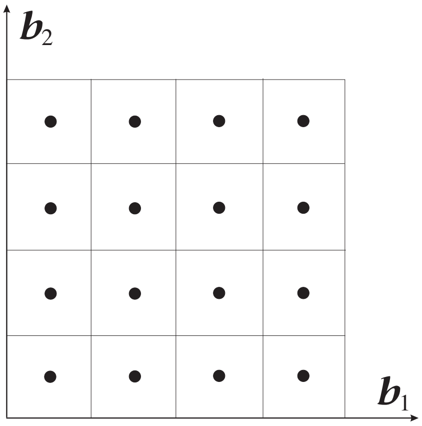


Figure 2.3: Schematic view of a 4 × 4 Monkhorst-Pack mesh for sampling the 2D BZ of a surface with cubic symmetry. The special $\boldsymbol{k}$-points are denoted by the filled circles and located at the center $\left(\frac{1}{2},\frac{1}{2}\right)$ of each of the 16 polyhedra that tile the BZ. In this work the surfaces are modelled using a slab geometry. Since there is no dispersion of the electronic band along the surface normal direction, we require the mesh to lie in the xy-plane. Thus we refer to this mesh as 4 × 4 × 1 sampling with special $\boldsymbol{k}$-points at $\left(\frac{1}{2},\frac{1}{2},0\right)$.

$$\boldsymbol{k}_{i}=k_{1}^{(i)}\boldsymbol{b}_{1}+k_{2}^{(i)}\boldsymbol{b}_{2}+k_{3}^{(i)}\boldsymbol{b}_{3},\qquad i=1,\ldots,N_{\mathrm{kpt}}. \tag{2.47}$$

Here:

$$\boldsymbol{k}_{j}^{(i)}=\frac{2p_{j}^{(i)}+2q_{j}^{i}-l_{j}-2}{2l_{j}}\qquad\left\{\begin{array}{lcl}q_{j}&=&1,2,\ldots,l_{j}\\ i&=&1,2,\ldots,N_{\mathrm{kpt}}\end{array}\right.. \tag{2.48}$$

Within the Monkhorst-Pack scheme the BZ is splited into $l_1\times l_2\times l_3$ identical polyhedra (l_i times along each direction $\boldsymbol{k}_i$). In each of the polyhedra a set of $N_{\mathrm{kpt}}^{\mathrm{s}}$ $\boldsymbol{k}$-points with local coordinates

$\left(p_1^{(1)}, p_2^{(1)}, p_3^{(1)}\right), \ldots, \left(p_1^{(N_{\mathrm{kpt}}^{\mathrm{s}})}, p_2^{(N_{\mathrm{kpt}}^{\mathrm{s}})}, p_3^{(N_{\mathrm{kpt}}^{\mathrm{s}})}\right)$ is defined. The $N_{\mathrm{kpt}}^{\mathrm{s}}$ $\boldsymbol{k}$-points correspond to the so called *special* $\boldsymbol{k}$*-points*. In Fig. (2.3) a 2D BZ sampling of a surface with a cubic symmetry by a 4 × 4 Monkhorst-Pack grid is shown. The special $\boldsymbol{k}$-points, indicated by the filled circles, are located at the center $((\frac{1}{2}, \frac{1}{2}))$ of each polyhedron.

The Monkhorst-Pack mesh reproduces a mesh of $N_{\mathrm{kpt}} = (l_1 \times l_2 \times l_3) \times N_{\mathrm{kpt}}^{\mathrm{s}}$ $\boldsymbol{k}$-points. The number of $\boldsymbol{k}$-points can be further reduced if the point group symmetry of the crystal under investigation is taken into account. Then the sum in Eq. (2.46) goes only over the reduced number of $\boldsymbol{k}$-points.

In general we are not usually interested in the total energies themselves, but in total energy differences between different systems. Thus *equivalent* $\boldsymbol{k}$-point samplings must be used for sampling the different BZs: The different BZs for the different systems are sampled using equally uniform meshes. Moreover the $\boldsymbol{k}$-point's sampling must be tested prior to the calculations, i.e., it should be one of the objectives of the convergence tests. It should be noted here that the $\boldsymbol{k}$-point's mesh does not follow the variational principle: The total energy does not necessarily decrease by increasing the number of $\boldsymbol{k}$-points. Additional information concerning the $\boldsymbol{k}$-point sampling used in this work can be found in the two manuals of the codes used here [96, 97].

2.6 The pseudopotential concept

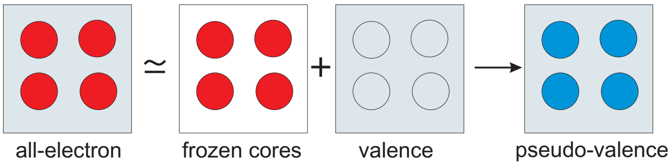


Figure 2.4: The frozen-core approximation used to distinguish the core and the valence electrons.

We have seen in the previous section that the PW basis set is attractive thanks to the simplicity of the representation and the ease by which a truncated basis set can be improved systematically. A disadvantage is that the use of PWs in full-potential all-electron approaches is computationally not feasible for most systems of interest. At small distances $\boldsymbol{r}$ in the atomic core the potential becomes very strong and a huge number of PWs will be required to represent satisfactory the strong wavelength oscillations inside the core [98].

To overcome this problem it is used that the core electrons are not important in describing the nature of the bonding between atoms in a crystal. The chemical bonding is mainly determined by the valence electrons while the core electrons matter only indirectly. Thus the inner (core) electrons can be considered as inert. The core electrons can be treated within the frozen core approximation (see Fig. (2.4)). The inner core electrons are eliminated within the frozen-core approximation by assuming them to be rigid bodies which are kept fixed with the nuclei. The electrostatic and quantum mechanical interactions of the valence electrons with the cores are accounted for by pseudopotentials.

The distinction between the core and the valence electrons is based upon the following criteria:

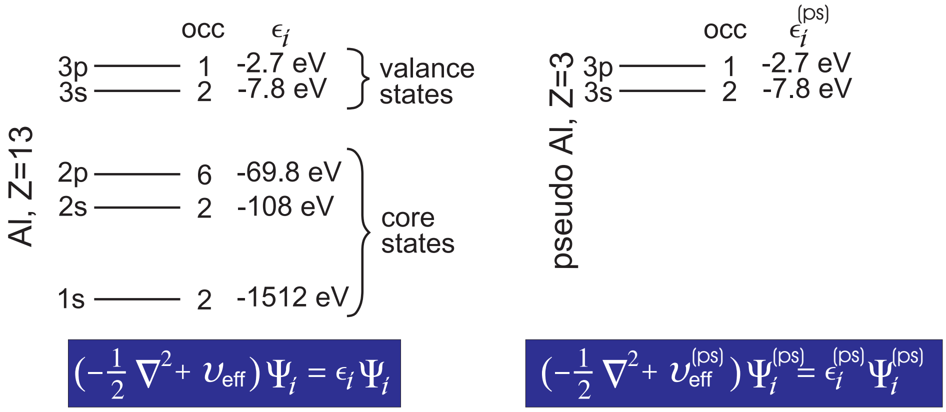


Figure 2.5: Removal of the core states from the spectrum and construction of the pseudo–Hamiltonian.

- In the chemical bonding only the valence electrons contribute directly.
- The core and the valence electron energies differ by at least one order of magnitude (see Fig (2.5)).
- The density distribution of the core electrons is mostly localized, compared to the valence electrons that might be delocalized and spread out over many atomic sites.

The pseudopotentials reproduce the true eigenenergies and eigenfunctions for the valence electrons outside the core region. Inside the core the valence pseudo-wavefunctions are smooth and nodeless (reducing the number of PWs required to describe them), see Fig. (2.6), but reproduce the true electron density inside the core.

2.6.1 Phillips-Kleinman Construction

The pseudopotential approach has its origin in the orthogonalized plane wave method (OPW). The OPW denotes a PW orthogonalized to all the core orbitals $|\psi_c\rangle$. It is convenient to describe the valance orbitals with OPWs since the former are all orthogonal to the core orbitals. An OPW is constructed by a superposition of a PW and the core orbitals. Writing the projection operator to the core orbitals as

$$\mathrm{P} = \sum_{\mathrm{c}}^{\mathrm{core}} |\psi_{\mathrm{c}}\rangle\langle\psi_{\mathrm{c}}|. \tag{2.49}$$

An OPW is introduced as:

$$|\boldsymbol{k}+\boldsymbol{G}\rangle_{\mathrm{OPW}} = (1-\mathrm{P})\,|\boldsymbol{k}+\boldsymbol{G}\rangle = |\boldsymbol{k}+\boldsymbol{G}\rangle - \sum_{\mathrm{c}}^{\mathrm{core}} |\psi_{\mathrm{c}}\rangle\langle\psi_{\mathrm{c}}|\boldsymbol{k}+\boldsymbol{G}\rangle \tag{2.50}$$

The true wavefunctions $\big|\boldsymbol{k}\lambda\big\rangle$ are given by linear combinations of OPWs:

$$\big|\boldsymbol{k}\lambda\big\rangle = (1 - \mathrm{P})\,\big|\boldsymbol{k}\lambda\big), \quad \text{where}: \big|\boldsymbol{k}\lambda\big) = \sum_{\boldsymbol{G}} \tilde{u}_{\boldsymbol{k}\lambda}(\boldsymbol{G})\,\big|\boldsymbol{k}+\boldsymbol{G}\big\rangle. \tag{2.51}$$

Inserting the wavefunction as defined above into the Kohn-Sham equations, we get:

$$\begin{gathered}\left[-\frac{1}{2}\nabla^2 + V_{\mathrm{PP}}^{\mathrm{PK}}\right]\big|\boldsymbol{k}\lambda\big) = \epsilon_{\boldsymbol{k}\lambda}\big|\boldsymbol{k}\lambda\big) \\ V_{\mathrm{PP}}^{\mathrm{PK}}(\boldsymbol{r}) = V(\boldsymbol{r}) + \sum_{\mathrm{c}}^{\mathrm{core}}(\epsilon_{\boldsymbol{k}\lambda} - \epsilon_{\mathrm{c}})\,\big|\psi_{\mathrm{c}}\big\rangle\big\langle\psi_{\mathrm{c}}\big|\end{gathered} \tag{2.52}$$

where $V(\boldsymbol{r})$ is the self-consistent potential and $\epsilon_{\boldsymbol{k}\lambda}$ and ϵ_{c} are the valance and core eigenvalues respectively. The potential described by Eq. (2.52) is the Phillips-Kleinman (PK) pseudopotential [99]. It can be regarded as the potential seen by the PWs and it is an example of pseudopotential. Outside the core region the PK pseudopotential becomes equal to $V(\boldsymbol{r})$ as the core wavefunctions vanish. Thus there is some radius r_{c} around an atom beyond which the contribution of the core states becomes negligible. The main advantage of the above formulation is clear: the all-electron problem can be replaced by an effective Hamiltonian acting on the smooth valence wavefunctions.

The pseudopotential defined by Eq. (2.52), contains, besides the term $V(\boldsymbol{r})$, the additional projector to the core states: $\widehat{\wp} = \sum_{\mathrm{c}}^{\mathrm{core}}(\epsilon_{\boldsymbol{k}\lambda} - \epsilon_{\mathrm{c}})\,\big|\psi_{\mathrm{c}}\big\rangle\big\langle\psi_{\mathrm{c}}\big|$, which is nonlocal[7]. The projector operator acts differently on the different angular momenta that are present in the core, and the nonlocality of these potential is not convenient (because it causes the predominant computational effort in calculating large systems. The next sections treat the generation of fully separable, norm-conserving pseudopotentials.

2.6.2 Norm-Conserving Pseudopotentials

Pseudopotential generation itself has been and is the subject of intensive studies. The principal objectives in generating a pseudopotential are the following:

Transferability The transferability refers to the ability of the pseudopotential to accurately describe the valence electrons in different chemical environments.

Efficiency A pseudopotential should keep the computational demands as low as possible. That is to compute wavefunctions and electron density with as few basis functions as possible.

For a given set of quantum numbers n, l, m the wavefunction is $\psi_{nlm} = \mathrm{Y}_{lm}(\theta,\phi)\cdot\mathrm{R}_{nl}(r)$. Here $\mathrm{Y}_{lm}(\theta,\phi)$ is the corresponding spherical harmonic and $\mathrm{u}_{nl}(r) = r\cdot\mathrm{R}_{nl}(r)$ is the radial part that solves the radial Shrödinger equation (or equivalently solves self-consistently the radial Kohn-Sham

---

[7] It is nonlocal in terms of the angular momentum l. Moreover it is short-range, like the core orbitals.

equation):

$$\left[-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}r^2}+\frac{l\left(l+1\right)}{2r^2}+\upsilon_{\text{eff}}\right]\mathrm{u}_{nl}\left[r;\epsilon_{nl}\right]\left(r\right)=\epsilon_{nl}\mathrm{u}_{nl}\left(r;\epsilon_{nl}\right). \tag{2.53}$$

Here $\upsilon_{\text{eff}}\left[n;r\right]$ is the self-consistent one-electron potential:

$$\upsilon_{\text{eff}}\left[n;r\right]=\upsilon_{\text{ext}}+\upsilon_{\text{H}}\left[n;r\right]+\upsilon_{\text{XC}}\left[n;r\right]. \tag{2.54}$$

$\upsilon_{\text{ext}}\left(r\right)$ is the external potential, $\upsilon_{\text{H}}\left[n;r\right]$ is the electrostatic Hartree potential, and $\upsilon_{\text{XC}}\left[n;r\right]$ is the exchange and correlation potential.

Constructing the logarithmic derivative

$$\mathrm{L}_l\left(r;\epsilon\right)=-\frac{\mathrm{d}\log\mathrm{R}_l\left(r;\epsilon\right)}{\mathrm{d}r} \tag{2.55}$$

the following statement relating the logarithmic derivative $\mathrm{L}_l\left(r;\epsilon\right)$ with the norm of the corresponding eigensolution $\mathrm{R}_l\left(r;\epsilon\right)$, holds:

$$\frac{\partial\mathrm{L}_l\left(r;\epsilon\right)}{\partial\epsilon}=-\frac{1}{2\mathrm{R}_l^2\left(r;\epsilon\right)}\int_0^r\left|\mathrm{R}_l\left(r';\epsilon\right)\right|^2\mathrm{d}r' \tag{2.56}$$

Moreover the scattering of a PW from a spherically-symmetric potential localized within a radius r_{c} and centered at the origin, is fully characterized in terms of the logarithmic derivatives. Thus, the logarithmic derivatives can be used as a criterion of the pseudopotential's *transferability* and its ability to correctly describe the scattering due to the ion core in a variety of chemical environments.

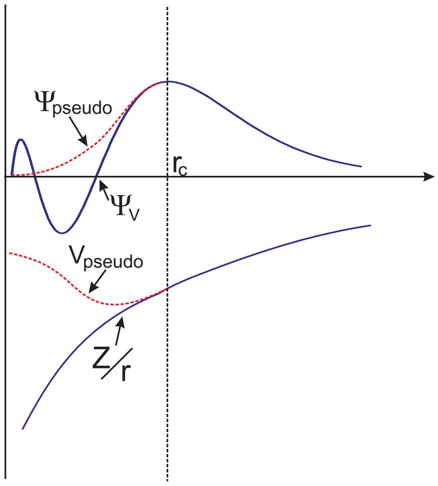


Figure 2.6: Schematic representation of the all-electron (solid lines) and the pseudo (dashed lines) wavefunctions and the corresponding potentials.

Transferability is related to the so called *norm-conservation* of the pseudopotential: Assume, that $\mathrm{u}_{nl}\left[r;\epsilon_{nl}\right]$ is the solution of the all-electron radial Schrödinger equation (Eq. (2.53)) and $\mathrm{u}_l^{\text{ps}}\left[r;\epsilon_l^{\text{ps}}\right]$ is the solution of the following equation:

$$\left[-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}r^2}+\frac{l\left(l+1\right)}{2r^2}+V_l^{(\text{ps, scr})}\right]\mathrm{u}_l^{\text{ps}}\left[r;\epsilon_l^{\text{ps}}\right]=\epsilon_l^{\text{ps}}\mathrm{u}_l^{\text{ps}}\left[r;\epsilon_l^{\text{ps}}\right]. \tag{2.57}$$

Further, it should give the same eigenvalue

$$\epsilon_l^{\text{ps}}\equiv\epsilon_{nl}\ \left(\equiv\epsilon\right). \tag{2.58}$$

Finally, the following two relations are assumed:

$$\mathrm{u}_l^{\text{ps}}\left[r;\epsilon_l^{\text{ps}}\right]\to\mathrm{u}_{nl}\left[r;\epsilon_{nl}\right]\quad\text{for } r>r_{\text{c}},\quad\text{and} \tag{2.59}$$

$$\int_0^{r_\mathrm{c}} \left|\mathrm{u}_l^\mathrm{ps}\left[r;\epsilon_l^\mathrm{ps}\right]\right|^2 \mathrm{d}r = \int_0^{r_\mathrm{c}} \left|\mathrm{u}_{nl}\left[r;\epsilon_{nl}\right]\right|^2 \mathrm{d}r. \tag{2.60}$$

Then Eq. (2.55) guarantees that their logarithmic derivatives are identical in an energy range around ϵ. Fulfilling Eq. (2.60) is identical to the statement that the pseudopotential is norm-conserving. The norm-conservation statement has a clear physical meaning: The total amount of charge density enclosed inside a sphere of radius r_c by the all-electron wavefunction and the pseudo-wavefunction is exactly the same. The last statement ensures that the electrostatic energy of the valence electrons is well approximated and that the long-range tail of the electrostatic potential arising from the nucleus plus the core electrons is correct [100].

The generation of norm-conserving pseudopotentials is given schematically in Fig. 2.7 and consists in general of the following steps:

1. An all-electron calculation of the free atom in a reference configuration (usually the ground state) is carried out. The eigenenergies ϵ_{nl} and the corresponding eigenfunctions $\mathrm{u}_{nl}\left[r;\epsilon_{nl}\right]$ are obtained.

2. For each angular momentum l a cutoff radius r_c^l is chosen. The all-electron and the pseudo-wavefunctions must be equal (or exponentially converge to each other) for $r > r_\mathrm{c}$. Moreover the pseudo-wavefunctions must be nodeless and thus r_c must be greater than the distance at which the last node of the all-electron wavefunction is obtained.

3. For $r < r_\mathrm{c}$ a parametrization form is chosen for $\mathrm{u}_l^\mathrm{ps}\left[r;\epsilon_l^\mathrm{ps}\right]$. Different parametrization schemes exist. In the present work the Troullier-Martins scheme is used [101]. Details will be given in the paragraph below.

4. Since the pseudo-wavefunctions contain no radial nodes Eq. (2.57) can be inverted and the corresponding pseudopotential is obtained:

$$V_l^\mathrm{ps,scr} = \epsilon_l^\mathrm{ps} - \frac{l\,(l+1)}{2r^2} + \frac{1}{2\upsilon_l^\mathrm{ps}\,(r)} \cdot \frac{\mathrm{d}^2}{\mathrm{d}r^2}\upsilon_l^\mathrm{ps}\,(r) \tag{2.61}$$

   where $\epsilon_l^\mathrm{ps} = \epsilon_{nl}$.

5. The final ionic potential is determined by subtracting from the screened pseudopotentials (Eq. (2.61)) the electrostatic (Hartree) and the exchange-correlation contributions due to the valence electrons:

$$\begin{aligned} V_l^\mathrm{ps}\,(r) &= V_l^\mathrm{ps,scr}\,(r) - \upsilon_\mathrm{H}\left[n_0^\mathrm{ps};r\right] - \upsilon_\mathrm{xc}\left[n_0^\mathrm{ps};r\right], \\ \text{with } n_0^\mathrm{ps}\,(r) &= \frac{1}{4\pi}\sum_{l=0}^{l_\mathrm{max}} f_l \left|\frac{\mathrm{u}_l^\mathrm{ps}\left[r;\epsilon_l^\mathrm{ps}\right]}{r}\right|^2. \end{aligned} \tag{2.62}$$

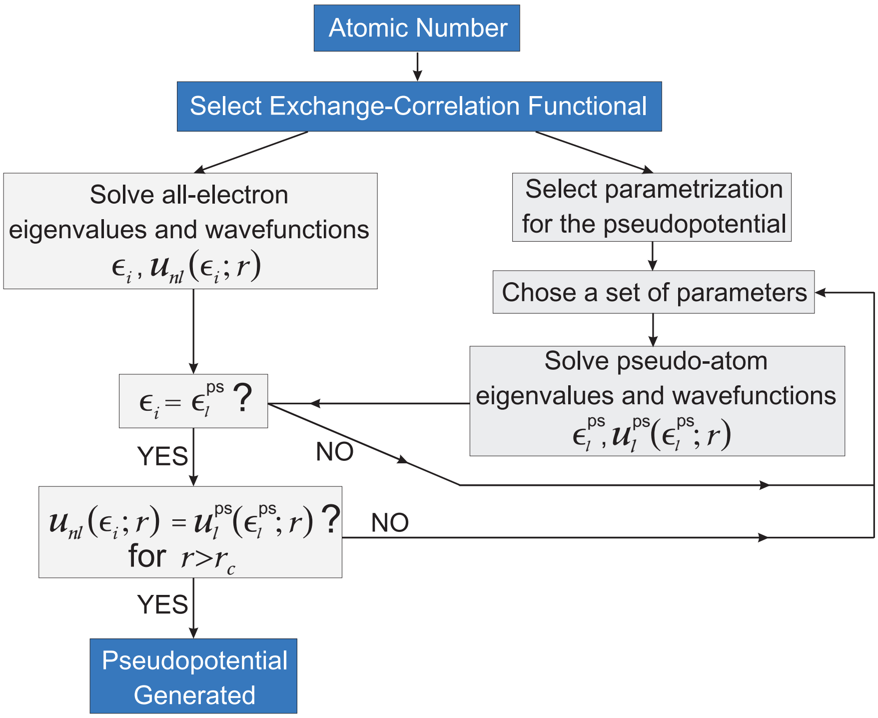


Figure 2.7: Flow chart of a pseudopotential generation schema.

Here f_l are the occupation numbers. Due to the spherical symmetry, states with the same quantum number $n = n_i$ and $l = l_i$, but different m_i are energetically degenerate: $\epsilon_i = \epsilon_{n_i l_i m_i} = \epsilon_{nl}$. Thus, the occupation numbers as function of the energy are:

$$\begin{cases} f_{nl} = 2\,(2l+1) & \text{for } \epsilon_{nl} < \epsilon_N, \\ 0 \leq f_{nl} \leq 2\,(2l+1) & \text{for } \epsilon_{nl} = \epsilon_N, \\ f_{nl} = 0 & \text{for } \epsilon_{nl} > \epsilon_N. \end{cases} \tag{2.63}$$

Here N corresponds to the total number of electrons: $N = \sum_{nl} f_{nl}$

Trouiller-Martins pseudopotentials

As mentioned above prior to the generation of the pseudopotential a parametrization for $r < r_l^c$ has to be chosen. In the literature various parametrization schemes have been reported [101, 102, 103, 104, 105, 106, 107, 108]. The pseudopotentials used in this work were generated using the Troullier and Martins [101] parametrization which can be considered as generalization of the method proposed by Kerker [105]. The form of the pseudo-wavefunction proposed by Troullier and Martins for $r < r_l^c$ is:

$$\mathrm{u}_l^{\mathrm{ps}}(r) = r^{l+1}\,\mathrm{e}^{p(r)}\ , \tag{2.64}$$

where:

$$p\left(r\right)=c_0+c_2r^2+c_4r^4+c_6r^6+c_8r^8+c_{10}r^{10}+c_{12}r^{12}\,. \tag{2.65}$$

The five coefficients of the above polynomial are defined in order to meet the following conditions:

- Norm conservation of the charge density within the core radius r_l^{c},
- Continuity of the pseudo-wavefunctions and its first four derivatives at r_l^{c}, and
- Zero curvature of the screened pseudopotential at the origin: $\frac{\mathrm{d}^2}{\mathrm{d}r^2}V_l^{\mathrm{ps,scr}}\left(r\right)|_{r=0}$.

The non-linear core correction

With the unscreened-ionic pseudopotential, defined by Eq. (2.62), the total energy of a system is:

$$E^{\mathrm{tot}}=\sum_i f_i\left\langle\psi_i\left|\widehat{\mathrm{T}}+\widehat{\mathrm{V}}^{\mathrm{ps}}\right|\psi_i\right\rangle+E^{\mathrm{H}}\left[n^{\mathrm{ps}}\right]+E^{\mathrm{XC}}\left[n^{\mathrm{ps}}\right],$$

$$\text{with } n^{\mathrm{ps}}\left(\boldsymbol{r}\right)=\sum_i f_i\left|\psi_i\left(\boldsymbol{r}\right)\right|^2. \tag{2.66}$$

Here E^{XC} and E^{H} refer to the exchange-correlation and the Hartree interactions respectively. The interactions between the valence electrons and the core electrons are included in the pseudopotential and are given in terms of a linear dependence on the valence density n^{ps}. The electrostatic part is linear in the valence electron density and obeys the superposition principle ($\upsilon^{\mathrm{H}}\left(n^{\mathrm{core}}+n^{\mathrm{ps}}\right)=\upsilon^{\mathrm{H}}\left(n^{\mathrm{core}}\right)+\upsilon^{\mathrm{H}}\left(n^{\mathrm{ps}}\right)$). On the other hand the exchange-correlation energy is a non-linear function of the total electron density. However, it is common to use the above (linearized) approximation, both in LDA and GGA [109]. If the core and the valence electron densities are well separated in space the "linearization" will not introduce serious errors. But if there is a significant overlap between core and valance electron densities, the linearization will reduce the transferability of the pseudopotential and will cause systematic errors in the calculated energies. In order to overcome the above inconsistency, the dependence of E^{XC} on the total charge density should be restored: $E^{\mathrm{XC}}\left[n_0^{\mathrm{ps}}\right]\rightarrow E^{\mathrm{XC}}\left[n_0^{\mathrm{ps}}+n_0^{\mathrm{core}}\right]$. That means that the core charge density should be added to the valance charge density whenever E^{XC} or υ^{xc} are computed. Louie *et al.* suggested that it is not necessary to add the full core density n_0^{core}, but it suffices to add a partial core density $\widetilde{n}_0^{\mathrm{core}}$ that is equal to the true charge density outside some radius r^{nlc} [110]. Inside r^{nlc} it was taken to be equal to some function that matches onto the true charge density at r^{nlc}. The function is chosen such that it is easy to perform a Fourier transform action, and to give the smallest possible integrated charge density. Louie *et al.* suggested the following formula [110]:

$$\widetilde{n}_0^{\mathrm{core}}\left(r\right)=\begin{cases} n_0^{\mathrm{core}}\left(r\right) & \text{for } r>r^{\mathrm{nlc}} \\ A\sin\left(r\right)/r & \text{for } r<r^{\mathrm{nlc}} \end{cases}. \tag{2.67}$$

Here r^{nlc} is chosen to be the radius at which the core charge density is between 1 to 2 times larger than the valence charge density. The parameters A and B are determined by the value and the gradient of the core charge density at r^{nlc}.

In this work pseudopotentials with a different parametrization in the *non-linear core correction* are used (for more details see Ref. [100]):

$$\widetilde{n}_0^{\text{core}}(r) = \begin{cases} n_0^{\text{core}} & \text{for } r \geq r^{\text{nlc}} \\ c_0 + \sum_{i=3}^{6} c_i r^i & \text{for } r < r^{\text{nlc}}, \text{ with } \widetilde{n}_0^{\text{core}} < n_0^{\text{core}} \end{cases}. \tag{2.68}$$

Here the coefficients c_i are taken such that $\widetilde{n}_0^{\text{core}}$ has zero slope and curvature at the origin, decays monotonically, and joins the full core density continuously up to the third derivative. The exchange-correlation term entering Eq. (2.62) should be redefined as follows:

$$\upsilon_{\text{xc}}\left[n_0^{\text{ps}}; r\right] \rightarrow \upsilon_{\text{xc}}\left[n_0^{\text{ps}} + \widetilde{n}_0^{\text{core}}; r\right]. \tag{2.69}$$

Separable pseudopotentials: The Kleinman-Bylander representation

The pseudopotentials derived by Eq. (2.62) are angular momentum dependent. Thus, a pseudopotential operator should have in general the form:

$$\widehat{\text{V}}^{\text{ps}}(r) = \sum_{lm} V_l^{\text{ps}} \widehat{\wp}_{l,m}. \tag{2.70}$$

Here $\widehat{\wp}_{l,m}$ is the projector on the angular momentum state $\{l, m\}$. In polyatomic systems the projection onto an angular momentum expansion is not convenient, since it would dominate the computational effort. However, what can be used is that at large radius r all components V_l^{ps} reduce to the ionic Coulomb potential $-\text{Z}_{\text{ion}}/r$, and get independent of l. Hence, a good approximation would be to express the pseudopotential as a multiplicative local potential plus only a few l-dependent terms for $l < l_{\text{max}}$:

$$\widehat{\text{V}}^{\text{ps}}(r) = \widehat{\text{V}}_{\text{loc}} + \sum_{l=0}^{l_{\text{max}}} \sum_{m=-l}^{l} |lm\rangle\, \delta\widehat{\text{V}}_l \,\langle lm|\,. \tag{2.71}$$

Here $|lm\rangle$ denotes the spherical harmonic Y_{lm}. The local term is taken as one of the semilocal pseudopotential components, $\widehat{\text{V}}_{\text{loc}} = \widehat{\text{V}}_{l=l_{\text{loc}}}^{\text{ps}}$, and the nonlocal term $\delta\widehat{\text{V}}_l = \widehat{\text{V}}_l^{\text{ps}} - \widehat{\text{V}}_{\text{loc}}$ vanishes beyond r_l^{c}. The expansion in Eq. (2.71) is truncated at some l_{max}. The local component of the pseudopotential should reproduce the scattering properties of the higher angular momentum channels and the computational effort can be reduced by choosing $l_{\text{loc}} = l_{\text{max}}$.

The computational effort can be further reduced by using the fully separable form of Kleinman and Bylander (KB) [111]:

$$\begin{aligned}\langle \boldsymbol{r}|\, \widehat{\mathrm{V}}^{\mathrm{ps}} \left|\boldsymbol{r}'\right\rangle &= \langle \boldsymbol{r}|\, \widehat{\mathrm{V}}_{\mathrm{loc}} + \delta\widehat{\mathrm{V}}_{\mathrm{KB}} \left|\boldsymbol{r}'\right\rangle \\ &= V_{\mathrm{loc}}\left(r\right)\delta\left(\boldsymbol{r}-\boldsymbol{r}'\right) + \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} \langle \boldsymbol{r}|\, \chi_{lm}^{\mathrm{KB}} \,\rangle\, E_l^{\mathrm{KB}} \left\langle \chi_{lm}^{\mathrm{KB}} \right|\boldsymbol{r}'\rangle\,. \end{aligned} \tag{2.72}$$

Here $\langle \boldsymbol{r}|\, \chi_{lm}^{\mathrm{KB}}\, \rangle$ are the projector functions given by:

$$\left\langle \boldsymbol{r}\middle|\chi_{lm}^{\mathrm{KB}}\right\rangle = \frac{1}{r}\chi_l\left(r\right)\mathrm{Y}_{lm}\left(\theta,\phi\right) = \frac{1}{r}\frac{\mathrm{u}_l^{\mathrm{ps}}\left(r\right)\delta V_l\left(r\right)}{\|\mathrm{u}_l^{\mathrm{ps}}\delta V_l\|}\mathrm{Y}_{lm}\left(\theta,\phi\right). \tag{2.73}$$

The KB-energies E_l^{KB} are given by:

$$E_l^{\mathrm{KB}} = \frac{\|\mathrm{u}_l^{\mathrm{ps}}\delta V_l\|}{\left\langle \psi_{lm}^{\mathrm{ps}}\middle|\chi_{lm}^{\mathrm{KB}}\right\rangle}. \tag{2.74}$$

The KB-energy measures the strength of a non-local component relative to the local part of the pseudopotential.

The KB form has the disadvantage that it may give rise to unphysical ”ghost” states, which are additional states at energies below or close to the physical valance states [112]. Fortunately, the problem with a ghost state occurring for some l can be tackled by choosing a different l_{loc} and changing the cutoff radius r_l^{c}. By doing this the transferability of the potential must be reserved. It is desirable to take $l_{\mathrm{loc}} = l_{\max}$ in order to keep the computational effort as low as possible.

Chapter 3

Valance Force Fields Methods

Ab-initio based methods, as the DFT approach presented in the previous chapter, provide a reliable tool for accurate atomic scale calculations [45, 63]. A major drawback of DFT calculations is their rather limited ability to represent systems consisting of large numbers of atoms and/or to perform simulations over a large timescale, in the case of Molecular Dynamics (MD). Today calculations based on DFT for systems with more than a few hundred atoms are not feasible. Thus it is crucial to develop and introduce numerically simpler methods that will allow to treat *large scale* systems and/or *long time scale* simulations on the atomic scale. An established method to do this is the application of *empirical potentials* (EP).

3.1 What is an Empirical Potential?

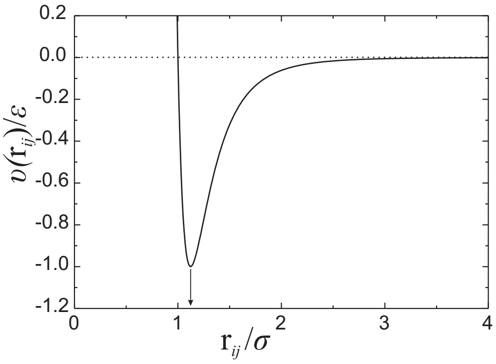


Figure 3.1: The Lennard–Jones interatomic potential.

In order to define EP we are obligated to introduce a crude assumption: The solid consists of well defined isolated atoms sticking to each other at preferred distance and angles. However, this assumption is rather arbitrary, since it does not take into account the crucial role electrons play in cohesion. For example, in covalent solids the valence electrons are partially localized along the bonds and more or less are equally shared between the nuclei in the bonds or may also resonate among a number of nearby bonding states. Nevertheless, our assumption can be justified if we take into account the Born-Oppenaheimer approximation (see Sec. 2.1): The nuclei and electronic motions can be decoupled. The *"ball and spring"* model can be used in order to treat the atoms and bonds between the atoms. The atoms are assumed as balls, centered at the nuclear positions, and the springs represent the interacting force between the atoms. The interacting force is given by an analytical formula, the energy functional, that contains a number of adjustable parameters. The parameters may be fitted to experimental data or to the results of *ab-initio* calculations, usually in order to reproduce as accurately as possible energy surface, elastic constants and/or cohesive energy curves for different phases of the solid under consideration.

The form of the interatomic potential depends on the type of the system under considerations. For example, one of the most famous EP is the Lennard–Jones potential (LJP), which is given by the following energy functional (see also Fig. 3.1):

$$\upsilon_{\mathrm{LJ}}\left(r_{ij}\right) = -4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right]. \tag{3.1}$$

Herr r_{ij} is the distance between the i_{th} and the j_{th} atoms and ε is the binding energy at the equilibrium distance σ. For short distances the term $(1/r)^{12}$ dominates and gives a repulsion between atoms that are brought close to each other. The physical origin of this term is related to the Pauli principle: When the atoms are brought close enough the electronic clouds overlap and the energy of the system increases. On the other hand, for large distances the $-(1/r)^{6}$ term related to the van-der-Waals attractive interaction between closed-shell atoms dominates.

The LJP has been applied successfully for describing the liquified nobel gases (Ar, Kr, Xe) that are stabilized in close-packed crystals. On the other hand, it is not suitable to describe solids with open shell atoms where localized bonds may form (covalent systems) or a delocalized "electron sea" exists (metals). For example using LJP it is not possible to stabilize the diamond structure as characteristic of tetrahedral semiconductors.

In general any potential energy function describing interatomic interactions can be written as a sum of one-body, two-body, three-body, etc. contributions:

$$\begin{aligned} E\left(1,2,\ldots,N\right) &= \sum_{i}\upsilon_1\left(i\right) + \sum_{i,j}\upsilon_2\left(i,j\right) + \sum_{i,j,k}\upsilon_3\left(i,j,k\right) + \ldots \\ &\quad + \sum_{i,j,k,\ldots,N}\upsilon_N\left(1,2,3,\ldots,N\right). \end{aligned} \tag{3.2}$$

In the above expansion the contributions υ_n converge quickly to zero as n increases. The one-body term in Eq. (3.2) corresponds to external forces acting on the system, and can be in general excluded. As becomes obvious from the LJP, the use of three-body and maybe also of higher order terms in the energy functional, is necessary to study tetrahedral semiconductors. In the next section some of the most widely used EP for covalent materials are briefly presented. For a more complete and detailed presentation, we refer to S. Erkoc [113] and R. M. Nieminen *et al.* [114].

3.2 Major types of empirical potentials for GaN

One of the oldest EP for tetrahedral semiconductors is the Keating potential and consists of two- and three-body terms [115]. The advantage of the Keating potential is that for a bulk system consisting of a single species only three parameters need to be adjusted. The one corresponds to the ideal bond length and the other two can be easily calculated by the elastic constants. The interaction range of the EP is limited to the first shell of neighbors and consequently for each atom only the first nearest neighbors (four for tetrahedral semiconductors) are taken into to account. For small distortions of the equilibrium geometry the potential is expected to work well. However, for more complicated reconstructions a better model is required. For describing group III-Nitrides [116, 117, 118], their alloys [119] and superlattices [120] a popular choice has been the Keating model. Grosse and Neugebauer [119] have examined the accuracy of the Keating potential for $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}$ alloys. They find that the use only of the nearest neighbors interactions is a drawback for achieving a correct description of the nonideal wurtzite crystals. But, despite this, the model is capable to give a correct description of the formation energies and atomic relaxation of the alloys and to correctly calculate the differences between the zincblende and the wurtzite structure.

Another widely used family of empirical models that utilizes a different approach are the potentials developed by Tersoff [121, 122, 123, 124]. The Tersoff potentials are based on a work by Abell[125]. Abell showed that the universality behavior of bonding energy curves observed by Rose, Smith, and Ferrante [126] can be modelled by pairwise interactions moderated by the local environment[1]. Tersoff, following Abell, introduced Morse-type pair potentials in the energy function:

$$\begin{aligned} E &= \tfrac{1}{2}\textstyle\sum_{i,j\neq i} \upsilon\left(i,j\right), \\ \upsilon\left(i,j\right) &= f_c\left(r_{ij}\right)\left[A\exp\left(-\lambda_1 r_{ij}\right) - B_{ij}\exp\left(-\lambda_2 r_{ij}\right)\right]. \end{aligned} \tag{3.3}$$

Here $\upsilon\left(i,j\right)$ is the interaction energy between the atoms i and j, separated by distance r_{ij}, and A, λ_1, and λ_2 are positive parameters, with $\lambda_1 > \lambda_2$. $f_c\left(r_{ij}\right)$ is a cutoff function that restricts the range of the potential. The second attractive term in Eq. (3.3) is moderated by the factor B_{ij}. This factor include all the information concerning the local environment, the coordination number and the bond bending contribution. For example the bonding strength factor B_{ij} between the i and j atoms should monotonically decrease with the number of competing bonds.

Based on the Abell's idea of modelling the binding energies by pairwise interactions moderated by the local environment, different EPs have been used in the study of III–Nitride semiconductors [127, 128, 129, 130, 131, 132, 133]. The differences are met at the different forms of the

---

[1] Rose, Smith, and Ferrante showed that a universal relationship between total energies and distance between atoms in solids with metallic and covalent bonds in adhesion, chemisorption, and cohesion exists. The energy relation has the simple form: $E\left(\alpha\right) = \Delta E E^*\left(\left(\alpha-\alpha_m\right)/l\right)$, where $E\left(\alpha\right)$ is the total energy as a function of the interatomic separation distance α, and ΔE the equilibrium binding energy. E^* is an approximately universal function of the scaled length $\alpha^* = \left(\alpha-\alpha_m\right)/l$, with α_m being the equilibrium separation distance and l a parameter that is specified in order the second derivative of the total energy with respect to atomic separation to be equal to 1.

cutoff function $f_c(r_{ij})$ and the bonding strength factor B_{ij} (see Eq. (3.3)) and/or the different parameterizations.

Ito [127] parameterized the Khor-Das Sarma [134] EP for zincblende AlN, GaN, and InN. This parameterizations has been successful in reproducing the elastic constants and their pressure derivatives. The major drawback of this potential is that it is not able to reproduce the relative stability between the wurtzite and zincblende structures. Ito and Kangawa [129] attempted to improve the EP proposed by Ito, by adding a term that corresponds to contributions beyond the second nearest neighbors and is described as electrostatic energy. The new formalism successfully describes the relative stability of the wurtzite structure over the zincblende counterpart.

The EP proposed by Ito and all thereafter potentials based on it, treated GaN as a one-component system. The same approach followed by Benkabou *et al.* [133], who used the formalism of the bond-order Tersoff interatomic potential [124] to describe structural and dynamical and thermal properties of the β-GaN.

The intrinsic problem of treating a binary system as a single-component system, where the same set of parameters is applied to both species, is the transferability of the potential. Nord *et al.* [131] proposed an EP for GaN based on the Tersoff-Brenner formulation [135], where different sets of parameters are used for the three different interactions between the two species. This work follows the work by Albe *et al.* in GaAs [136]. The potential by Nord *et al.* is successful in describing the structural properties and the elastic constants of $\alpha-$ and $\beta-$ GaN, but it gives the two phases to be energetically degenerate. Moreover, it is capable to give with good accuracy the energetics of N-vacancy defects, but it fails to describe other intrinsic defects like $\mathrm{Ga_N}$ interstitials.

Another widely used EP for tetrahedral semiconductors are the Stillinger-Weber (SWP) potentials [137]. The SWP consist of two- and three- body terms (the terms denoted as υ_2 and υ_3 in Eq. (3.2)) and initially were used for studies of structural and dynamical properties of silicon. SWP are short ranged and their interaction is restricted to a distance between first and second nearest neighbors. In the initial version of SWP in GaN, a Gaussian term was added to the potential to include interactions beyond the second nearest neighbors [138]. More recently a modification without the Gaussian term was proposed by Aïchoune *et al.* [139]. Based on this modification, SWP have been extensively used in GaN [139, 140, 141, 142]. A more detailed description of the SWP potential and the parameterization used in this work will be given in the next section.

3.3 Stillinger-Weber Empirical Potential

The SWP is one of the most widely used models for tetrahedrally bonded semiconductors. Initially it was fitted to experimental properties of crystalline and liquid silicon phases [137]. The

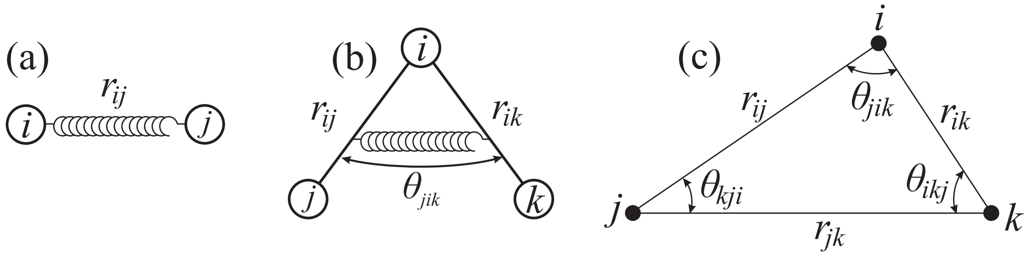


Figure 3.2: Illustration of (a) bond stretching, (b) bond bending and (c) triplet of atoms used in the three body term.

energy functional of the SWP consist of two- and three- body terms:

$$E = \sum_{i,j(i<j)} \upsilon_2(i,j) + \sum_{i,j,k(i<j<k)} \upsilon_3(i,j,k) \tag{3.4}$$

where:

$$\begin{aligned} \upsilon_2(i,j) &= \varepsilon f_2(r_{ij}/\sigma) \\ \upsilon_3(i,j,k) &= \varepsilon f_3(\boldsymbol{r}_{ij}/\sigma, \boldsymbol{r}_{ik}/\sigma) . \end{aligned} \tag{3.5}$$

Here $r_{ij} = |\boldsymbol{r}_j - \boldsymbol{r}_i|$, and $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$. ε and σ are the energy and length units. The minimum of f_2 is -1. The pair interaction f_2 is a function only of the scalar distance r_{ij} and describes the forces against bond-stretching. It should have a deep minimum at the equilibrium bond length. The proposed form of the two-body term is the following:

$$f_2(r_{ij}) = \begin{cases} A\left(B(r_{ij}/\sigma)^{-p} - (r_{ij}/\sigma)^{-q}\right)\exp\left[\left((r_{ij}/\sigma) - \alpha\right)^{-1}\right] & , \quad r_{ij}/\sigma < \alpha \\ 0 & , \quad r_{ij}/\sigma > \alpha \end{cases} \tag{3.6}$$

Here A, B, p, and α are positive parameters. The above form converges smoothly to zero, without discontinuities, at $r_{ij}/\sigma = \alpha$ (see Fig. 3.4).

The range of the SWP includes up to the second nearest neighbors in the diamond structure. The three-body term has the same cutoff distance:

$$f_3(\boldsymbol{r}_{ij}/\sigma, \boldsymbol{r}_{ik}/\sigma) = \lambda \exp\left[\gamma(r_{ij}/\sigma - \alpha)^{-1} + \gamma(r_{ik}/\sigma - \alpha)^{-1}\right](\cos\theta_{jik} + 1/3)^2 , \tag{3.7}$$

where λ and γ are positive parameters and θ_{jik} is the angle between $\boldsymbol{r}_{ij}$ and $\boldsymbol{r}_{ik}$ (see Fig. 3.2).

The three-body term is expressed in terms of a separable product of radial exponential functions (see Fig. 3.5) and an angular function (see Fig. 3.3) and describes the forces against-bond bending. Both, the radial terms and the angular term, are positive everywhere. The radial parts smoothly vanish beyond a cutoff distance which is identical to that of the pairwise potential. The cutoff distance formally expresses our assumption that contributions to bond bending from atoms beyond the second nearest neighbors are negligible. The angular term has two minima at the tetrahedral

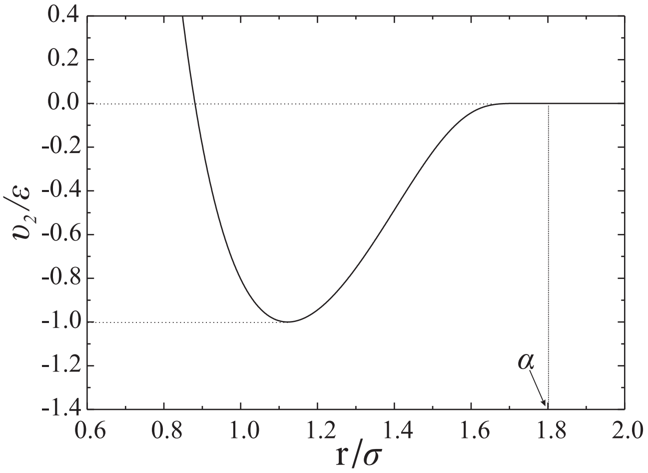


Figure 3.4: The reduced pair term of the SWP vs distance. It vanishes at the reduce cutoff distance $r/\sigma = 1.8$.

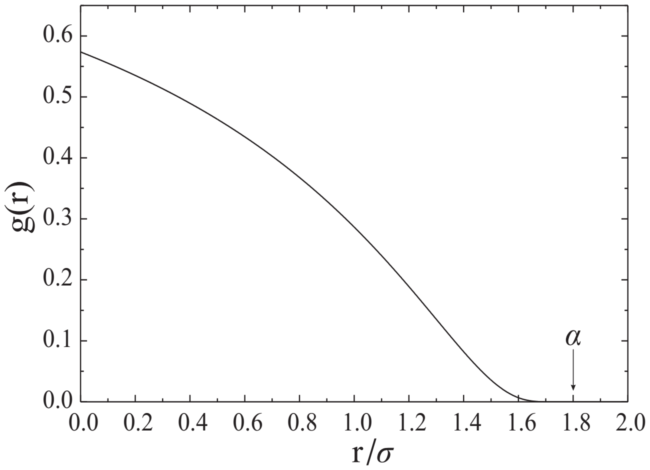


Figure 3.5: The radial exponential term $g(r) = \exp\left((r/\sigma - \alpha)^{-1}\right)$ vs separation distance. It is positive everywhere and vanishes ideally at the same cutoff reduced distance σ/r as the pairwise part (see Fig. 3.4).

angles $\cos\theta_{\text{tetr}} = \cos(2\pi - \theta_{\text{tetr}}) = -1/3$ in order to stabilize the formation of sp^3 bonds.

Although SWP have a simple form and lose their physical content in systems without sp^3 hybridization, they proved to be successful in describing many different structures/systems, such as point defects, liquid phases and certain surface structures [143]. The SWP in their original formulation as developed for silicon requires to fit seven parameters (A, B, p, q, α, λ, and γ). The energy unit ε has been fitted to the cohesive energy of the silicon diamond structure. In general it is also possible to use a single set of parameters for GaN, assuming only Ga-N bonds. Within this approach only systems with fully coordinated atoms (each cation is surrounded by four anions and vice versa) can be treated. In order to have a more transferable potential the three possible interactions must be treated explicitly. Thus, three different sets of parameters are required for the three different interactions (Ga-Ga, N-N, and Ga-N). The next section describes the parameterization used in this work.

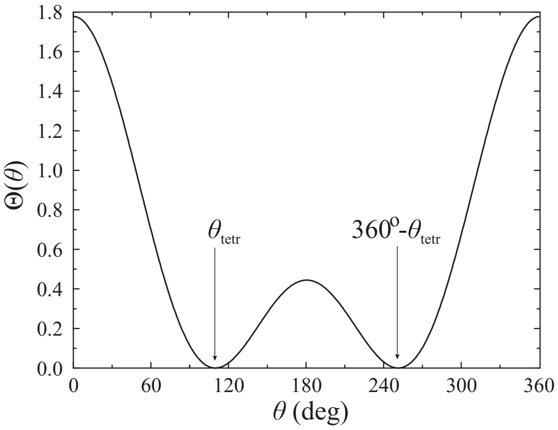


Figure 3.3: The angular function $\Theta(\theta) = (\cos\theta + 1/3)^2$ vs the angle. It is positive everywhere and becomes zero at two angles θ that coincide with the angle characteristic for tetrahedrally bonded systems.

Table 3.1: Parameters of the modified SWp. α, γ, A, B, and λ are dimensionless parameters.

Type of bond	α	γ	σ (Bohr)	ε (eV)	A	B	λ
Ga-N	1.8	1.2	3.204	2.170	7.917	0.720	32.50
Ga-Ga	1.8	1.2	3.250	0.655	7.917	0.720	26.76
N-N	1.8	1.2	2.460	0.655	7.917	0.720	26.76

3.3.1 Parametrization for GaN

As has been mentioned in Sec. 3.2 in the initial parametrization of the GaN Stillinger-Weber potential, a Gaussian term was added to the potential to describe interactions beyond the second nearest neighbors cite. The most widely used modification of the potential keeps the original analytic form proposed by Stillinger and Weber [137] for the two- and the three- body terms (Eqs. (3.6) and (3.7)) [139]. The parameters α, γ, p, and q were taken equal to those proposed for Si [137] independent of bond type: $\alpha = 1.8$, $\gamma = 1.2$, $p = 4$, and $q = 0$. The remaining parameters were fitted to closely reproduce lattice parameters, experimental elastic constants [144], and *ab-initio* calculated formation energies of inversion domain and stacking mismatch boundaries in GaN [145].

The modification of the SWP presented here is based on the parametrization proposed by Aïchoune *et al.* [139]. We use the same analytical formula proposed by Stillinger and Weber for Si and Aïchoune *et al.* for GaN (Eqs. (3.6), (3.7), with $p = 4$ and $q = 0$). Our intension is to construct a SWP that accurately describes the elastic long range interactions around edge type threading dislocations (see Sec. 8.6). Thus, we fitted the parameters to accurately reproduce the *ab-initio* derived lattice parameters and elastic constants for wurtzite GaN. The fitted parameters are given in Table 3.1. In Fig. 3.6 the two-body terms for the three different interaction types (namely Ga-N, Ga-Ga, and N-N) are shown. The two-body term for the Ga-N interaction has a deep minimum at the equilibrium bond length and smoothly vanishes just before the second nearest neighbor.

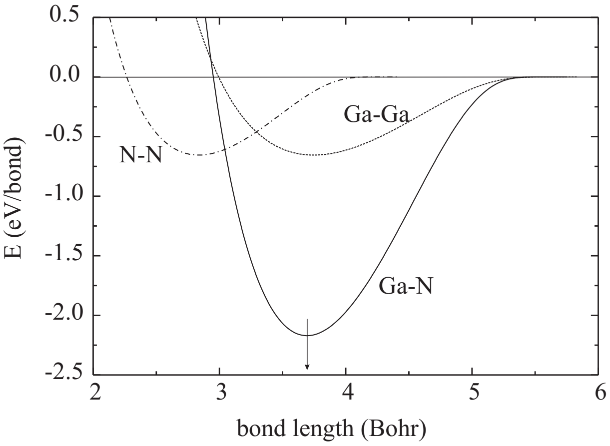


Figure 3.6: Two-body terms for Ga-N, Ga-Ga, and N-N interactions. The arrow indicates the equilibrium Ga-N first nearest neighbor distance.

In Table A.2 the *ab-initio* and SWP calculated and experimental elastic constants are given. As can been seen the SWP accurately reproduces the *ab-initio* calculated and experimental elastic constants. Thus, we expect to accurately describe the elastically strained region between threading

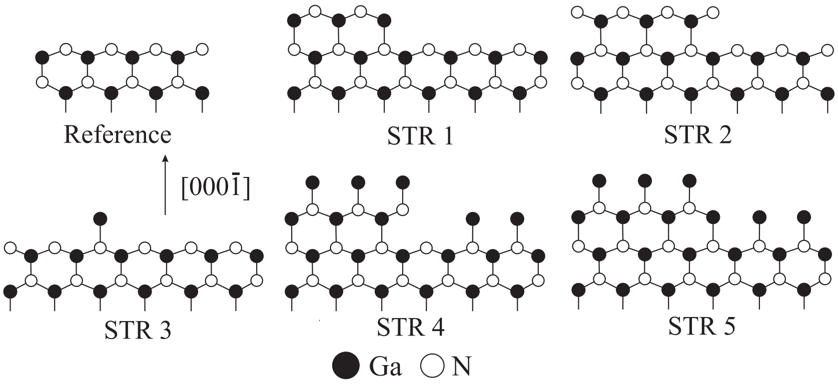


Figure 3.7: Side view of the five different structures used for the *ab-initio* and SWP calculations.

dislocations. Although the potential has been optimized with respect to the bulk properties we also checked its transferability: We performed total energy and force calculations for four different step configurations on the $\left(000\bar{1}\right)$ GaN surfaces using both SWP and DFT. The surfaces were modelled by a repeated 6×1 slab geometry consisting of two GaN double layers. The steps were introduced on the $\left(000\bar{1}\right)$ GaN surface (see Fig. 3.7). The atoms in the steps were fully relaxed. For the DFT calculations the dangling bonds at the bottom side of the slab have been passivated by partially charged pseudohydrogens that provide $\frac{5}{4}e^-$ ($^{1.25}$H). The settings for the DFT calculations are the same as those given in Sec. 4.4. For a detailed discussion on slab calculations see Sec. 4.2. The formation energies as function of the N chemical potential are shown in Figs. 3.8(a) and (b) for the DFT and SWP calculations respectively.

Let us first focus on the configurations STR 1, STR 2, and STR 3 (dashed lines in Figs. 3.8(a) and (b)). These structures are N-terminated and are distinguished by the Ga-terminated STR 4 and STR 5. As can be clearly seen there is a qualitative agreement between SWP and DFT calculated energies. However, agreement with the *ab-initio* results is unsatisfactory for the Ga-terminated step configurations (STR 5 and STR 6 indicated by the solid lines). Furthermore, SWP calculations have been performed for the step configurations on Ga-polar, Ga-terminated surfaces as discussed in detail in Chapter. 4 (see Figs. 4.11 and 4.12). The results are compared in Figs. 3.8(c) and (d). As can be seen there is no quantitative nor a qualitative agreement between DFT and SWP calculated energies. While the SWP (at least qualitatively) describes the N-terminated structures, it fails to describe the Ga-terminated surfaces. The reason is that SW is not well suited to describe metallic bonds (such as Ga-Ga bonds on that surface). Since, metallic-type Ga-Ga bonds may exist in the core regions of edge type dislocation in GaN (see, e.g., Fig. 8.5), SWP are not expected to correctly describe the energetics of the highly strained core region. However, their ability to accurately describe bulk properties (lattice parameters and and elastic constants) will allow us to develop a multiscale approach and thus to study systems with up to macroscopically relevant dislocation densities.

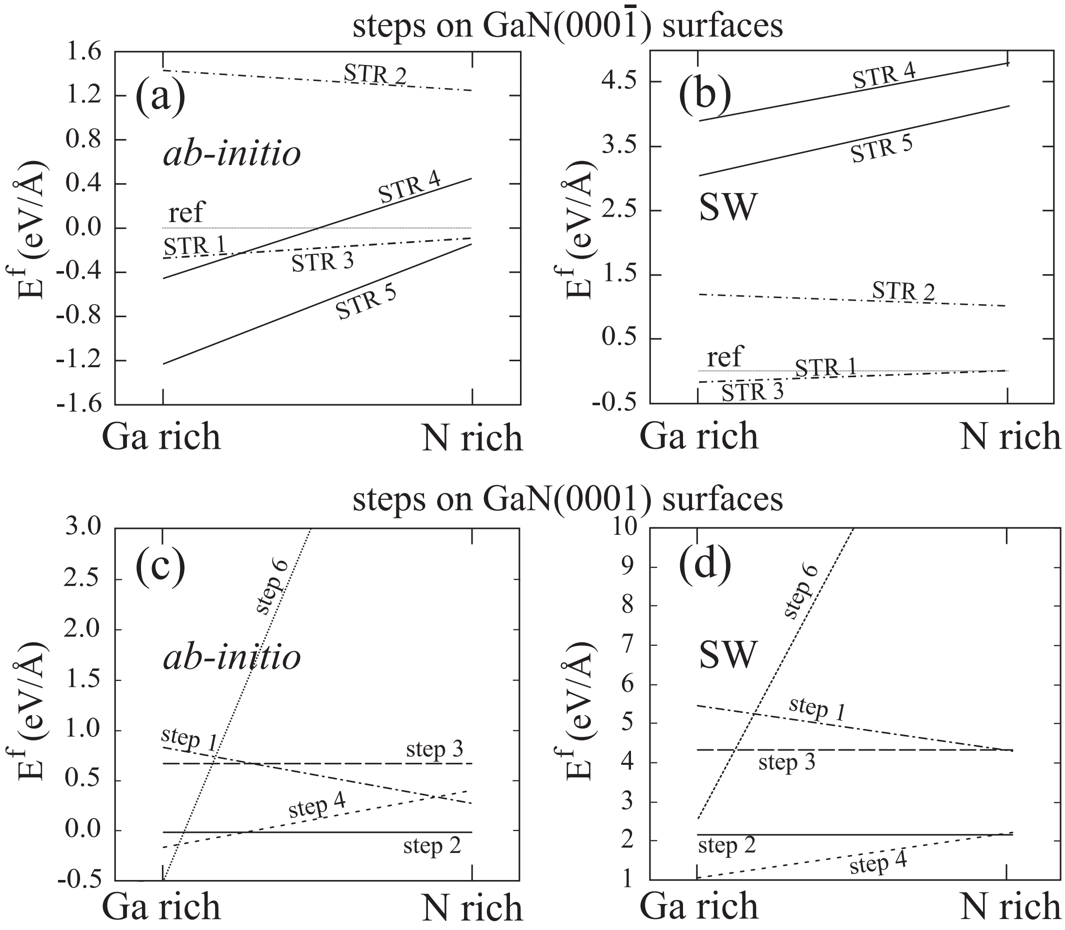


Figure 3.8: (a) & (b) Relative formation energies of four different step configurations and the Ga adatom on the N polar surface (see Fig. 3.7) derived by DFT (a) and SWP (b) calculations. The zero energy corresponds to the N terminated $(000\bar{1})$ GaN surface. solid lines indicate the Ga-rich configurations. (c) & (d) Relative formation energies of five different step configurations on the Ga polar surface (see Fig. 4.11) derived by DFT (c) and SWP (d) calculations. The zero energy corresponds to the Ga terminated (0001) GaN surface. The *ab-initio* calculations on the steps on the Ga polar surface are separately described in Sec. 4.4.

Part II

Results

Chapter 4

Steps on GaN{0001} surfaces: Implications on surface morphology and growth

The most common and technologically most relevant growth direction of wurtzite GaN is normal to the {0001}[1] basal plane [147]. The morphology of these surfaces is known to be extremely sensitive to the growth conditions. In general GaN films grown with a low III/V ratio are expected to display rough/faceted surface morphologies. Going to more metal-rich conditions a transformation of the morphology occurs accompanied by smoother surfaces and better film quality (see Secs 5.1 and 5.2).

Atomic steps on surfaces are often chemically more active than the flat surface and may act as nucleation and compensation centers. They thus play a crucial role in the growth of GaN epilayers. Here, we study the geometry, energetics, and the implications on surface roughening of possible step configurations on the (0001)GaN surfaces. In Sec. 4.1 previous experimental observations of steps and islands on (0001)GaN surfaces are briefly discussed. In Secs. 4.2 and 4.3 an introduction on the surface calculations using the slab approach and on the thermodynamics of the surface energies respectively is given. Finally, the *ab-initio* calculations on possible step configurations are presented (Sec. 4.4) and the implications of steps on the surface morphologies are discussed (Sec. 4.5).

---

[1] It is important to introduce here the different closures (i.e., brackets) used in crystal terminology to denote directions, planes,and symmetry equivalent planes: Square brackets ($[hkjl]$) denote a single direction, while angle brackets ($\langle hkjl \rangle$) designate a family of symmetry equivalent directions. Parenthesis ($(hkjl)$) denote a single plane or a set of parallel planes. Planes equivalent by symmetry may be denoted by curly brackets ($\{hkjl\}$). More details can be found in standard solid-state textbooks (e.g., see Ref. [146])

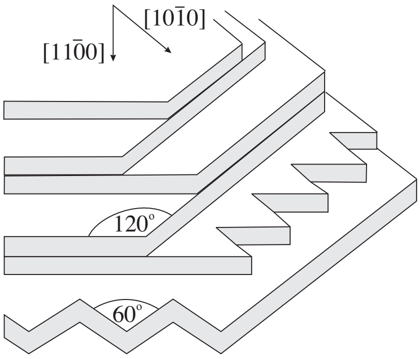


Figure 4.1: Schematic illustration of the anisotropic step-flow growth mode. The growth anisotropy leads to alternating smooth and jugged step edges and step bunching.

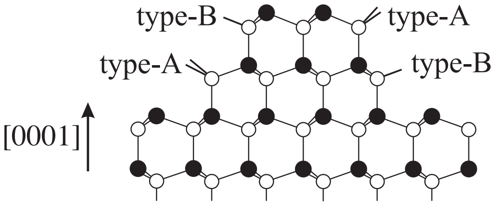


Figure 4.2: Side view of steps on the {0001} wurtzite GaN surfaces indicating the alternating character of the step edges. The unreconstructed type-A (-B) edge is characterized by two (one) dangling bonds per edge atom.

4.1 Introduction

In homoepitaxy three different growth modes can be in general encountered (see Sec. 5.2 and Refs therein). At low deposition rates and high growth temperatures one-dimensional step-flow growth is observed. In contrast, at low growth temperatures and high deposition rates the surface is expected to roughen and a three dimensional growth mode takes place. In an intermediate regime of growth temperatures and deposition rates the growth follows a two-dimensional mode where islands nucleate and coalesce on the surface. Atomic steps and island edges play an important role to understand/describe these regimes.

The first direct observations of the different growth modes during the homoepitaxy of GaN by MOCVD have been reported by Stephenson *et al.* [148]. They found that the growth mode depends on the growth temperature for constant III/V ratio: Going from high to lower temperatures the growth mode changes from step-flow to layer by layer and 3D growth. A similar, temperature dependence of the growth modes has been reported by Xie *et al.* during the MBE growth of (0001)GaN surfaces [149]. At higher temperatures, where the adatom diffusion length is long compared to step terrace width, they observed a 1D step-flow growth mode. For lower temperatures the adatom diffusion length becomes considerably smaller and a transition to 2D island growth is observed. The most interesting finding by Xie *et al.* is the growth anisotropy that characterizes both 1D and 2D growth modes [149]. In the step-flow growth mode, they observed that as one descends the terraces along a symmetry crystallographic direction (e.g. [10$\bar{1}$0]) the step edges alternate between jagged and smooth. They also observed along certain symmetry crystallographic directions (see Fig. 4.1) step bunching consisting of double bilayer height steps. On the other hand, in the 2D island growth mode the growth anisotropy resulted in the formation of triangular shaped islands. They proved that the growth anisotropy is due to the adatom/edge-atom binding anisotropy and not due to a possible anisotropic subsurface influence on the adatom mobility: In

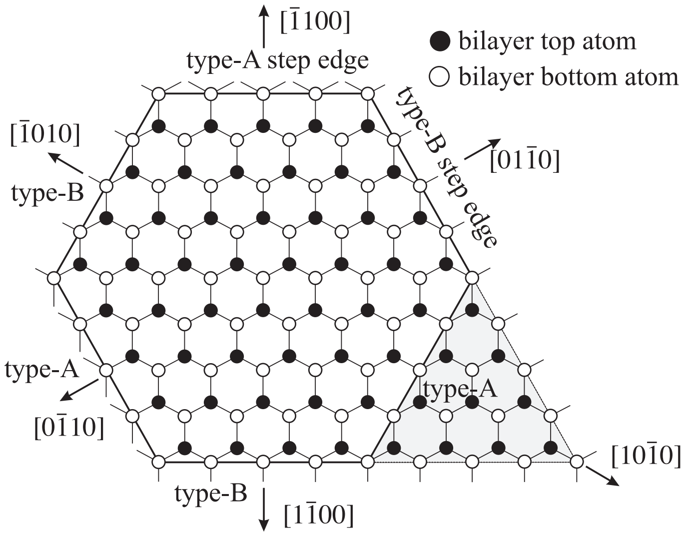


Figure 4.3: Top view of a bilayer showing the low index type A and B edges on (0001) and $(000\bar{1})$ wurtzite GaN surfaces. The unreconstructed type A step edges, which are normal to $\{10\bar{1}0\}$ directions (in this illustration), are characterized by two dangling bonds per edge atom. The normal to $\{1\bar{1}00\}$ directions (in this illustration) type B step edges have one dangling bond per edge atom. The shaded triangle indicates a jagged shaped step that would be formed assuming that type A steps grow faster than type B steps.

addition to the hcp, fcc islands have been observed in the island growth mode. Although a hcp and a fcc island on a given terrace have different subsurface stacking geometries have been found to have similar sizes and qualitatively similar triangular shapes suggesting that growth anisotropy in hcp and fcc GaN surfaces is governed by anisotropic edge accommodation. Similar results have been also reported by other groups for GaN films grown by MOCVD [150, 151, 152].

The growth anisotropy has been explained in terms of a simple model based on the dangling bond configurations at the steps [149, 150, 151]. As can be seen in Fig. 4.3 two low index step edges can be distinguished: type-A and type-B. The unreconstructed type-A edges are characterized by two dangling bonds per edge atom, while the type-B by one dangling bond. It is reasonable to assume that adatoms prefer to stick to a type-A edge rather than to a type-B edge, where the number of the dangling bonds is minimized. This behavior will eventually result in a faster growth of type-A steps. Thus it is expected that along directions normal to type-A edges, the steps will be triangular shaped, bounded by type-B edges. Along directions normal to type-B edges the steps will be smooth. In the 2D island mode the growth anisotropy results in triangular shaped islands while in the step-flow growth mode it results in a smooth-jagged sequence of step edges. Moreover, in the wurtzite structure, along a given crystallographic direction, the type of the step edges alternates (see Fig. 4.2). Thus in the step flow growth mode the faster growing type-A edges are expected to catch up with the underlying slower moving type-B edges. This mechanism leads to step bunches consisting of bilayer height steps.

Triangular shaped ”ghost” islands have been also observed during submonolayer deposition of GaN by MBE [153]. In the STM imaging the ”ghost” islands appeared hollow surrounded by bright frames, but continuous STM imaging or altering the scanning voltage, irreversibly converted the ”ghost” islands into normal islands. The ”ghost” islands were observed for growth only under excess Ga surface coverage. In a later work the authors explained the observation of ”ghost”

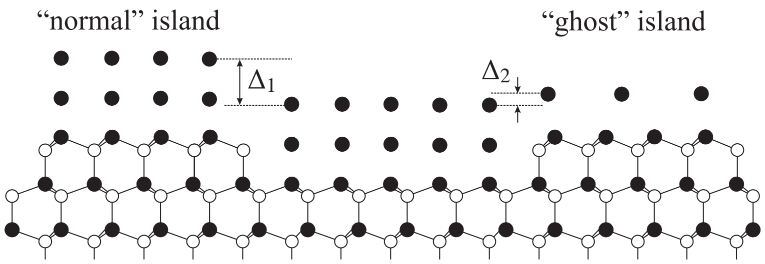


Figure 4.4: Schematic illustration of a model for GaN "ghost" islands. Filled and open circles correspond to Ga and N atoms respectively.

islands in terms of a model based on a surfactant mediated process, where the excess Ga present on the surface plays the role of the surfactant [154]. Recent *ab-initio* calculations revealed that the most stable (0001)GaN surface structure under Ga-rich conditions is the laterally contracted Ga bilayer structure (see Ref. [43] and Sec. 5.4.1)[2]. For the deposited N atoms to incorporate on the film an exchange process between the Ga atoms in the first layer and the N atoms should take place. The authors proposed that the "ghost" islands are nucleated via the aforementioned site-exchange action, where upon completion of the process the top surface becomes deficient of excess Ga. While the surrounding terraces remain Ga rich, the island surfaces will be bare or covered by less than a monolayer Ga (see Fig. 4.4). That will affect both the height of the "ghost island and the reconstruction of the island surface that subsequently affects the density of states (DOS). Hence STM imaging would give different contrast for the "ghost" and the normal islands. It has been speculated that the lack of excess Ga from the island surfaces upon interruption of the growth and quenching of the film is due to possible energy barriers for the Ga atoms to diffuse on the islands. Upon STM imaging Ga atoms can be transferred on the "ghost" islands via the tip or by the disturbance of the surface.

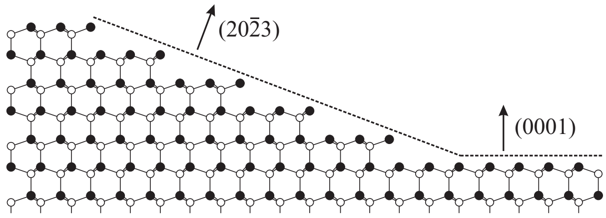


Figure 4.5: Schematic representation of the $(20\bar{2}3)$ vicinal surface on the (0001)GaN surface. Vicinal surfaces can be assumed as group of steps. Filled and open circles correspond to Ga and N atoms respectively.

Vicinal surfaces can be in general assumed as groups of steps on surfaces (see Fig. ??). It is of major interest to investigate the thermodynamic stability of steps and thus of vicinal surfaces on the (0001)GaN surfaces and their implications on surface morphology and growth. Furthermore, only speculations on the origins of the growth anisotropy and ghost island formation have been reported up todate. In order to address the above issues we have studied the geometry and thermodynamics of steps on the (0001)GaN surfaces. The results are cited in Sec. 4.4 and are discussed in Sec. 4.5.

---

[2] According to the laterally contracted bilayer model, the Ga polar GaN surface is covered by two Ga layers. The first layer sits atop the surface, while the top most is contracted and out of registry. Both layers give in total 2.33 ML Ga surface coverage.

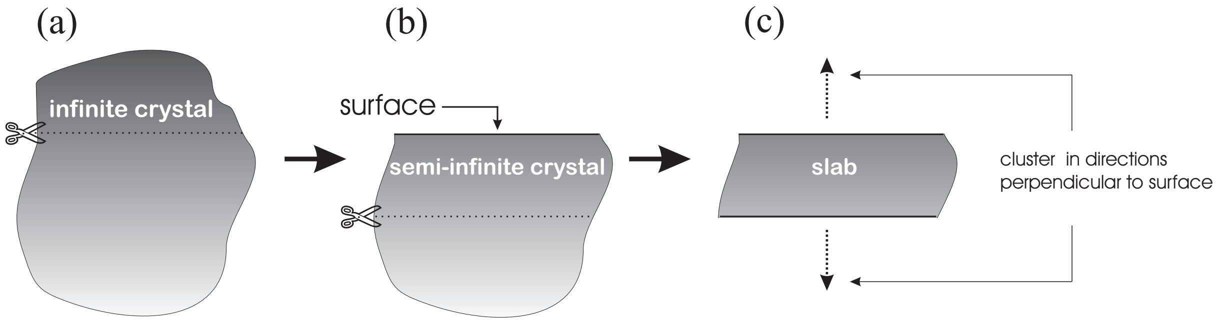


Figure 4.6: Illustration of creating a surface from an infinite crystal and the corresponding slab system.

4.2 Slab approach

Surfaces can be regarded as defects that destroy the periodicity of the bulk crystal structure. They can be created by cleaving and removing half of the crystal (see Fig. 4.6(1) and (2)). Within the supercell approach surfaces can be studied using a slab geometry. In order to create a slab we cut the semi-infinite crystal (that has a free surface) at a particular lattice plane parallel to the surface and throw away the rest (Fig. 4.6(b) and (c)).

Within the slab approach, the system can be assumed as cluster in the directions perpendicular to the surface (see Fig. 4.7). This is achieved by using a vacuum region in between the slab and its periodic images. The vacuum thickness should be optimized to (i) avoid artificial surface-surface interactions which arise if the vacuum is too small while (ii) minimize the computational cost (i.e. to make the vacuum thickness as thin as possible).

Obviously the slab has two surfaces. The ”top” side corresponds to the surface of interest, while special care must be taken for the treatment of the bottom surface of the slab. If both surfaces are inequivalent (e.g. GaN (0001) and $\left(000\bar{1}\right)$) or band structure calculations are to be performed, the dangling bonds of the bottom surface have to be passivated by fictitious partially charged hydrogens (pseudohydrogens) in order to saturate the dangling bonds of this surface and to avoid charge accumulation. The passivation is done in terms of the electron counting rule (ECR) [155]. According to the ECR each group-III cation contributes $\frac{3}{4}e^-$ to a bond, whereas each group-V anion contributes $\frac{5}{4}e^-$. Hence, the N atoms at the bottom surface should be saturated by pseudohydrogen that provide $\frac{3}{4}e^-$ ($^{0.75}$H) and the Ga atoms by pseudohydrogens that provide $\frac{5}{4}e^-$ ($^{1.25}$H). For fractionally charged hydrogens we use ionic potentials $-Z^*/r$ with Z^* being 0.75 and 1.25 respectively. Even in the case of equivalent slab-surfaces the saturation of the dangling bonds at the bottom side is desirable, since it provides a more efficient decoupling of the two surfaces and allows the use of relatively thin slabs [156].

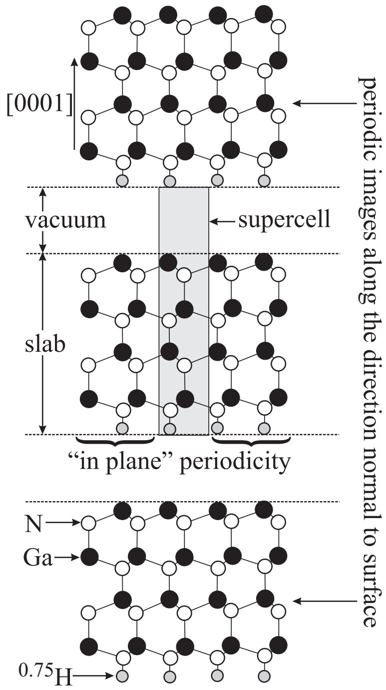


Figure 4.7: Schematic illustration of a slab configuration for calculations on the GaN (0001) surfaces. The shaded rectangle indicated the 1×1 supercell. The slab consist of 4 GaN double layers. The thickness of the slab is more than enough to accurately describe the surface and the underlying bulk material (see Sec. 6.4). The dangling bonds of the $(000\bar{1})$ bottom surface are saturated by partially charged hydrogens according to the ECR.

4.3 Surface energy

As has been mentioned in Sec. 4.2 the surfaces can be created by cleaving and removing half of the crystal (see Fig. 4.6(a) and (b)). The basic effect of this procedure is to break bonds between atoms on either side of the cleavage plane. By definition, the surface energy is the energy required to create two equivalent surfaces following the aforementioned procedure. A powerful way to study the energetics and the stability of the surfaces is to combine the microscopic first principles results with thermodynamic concepts. Since surfaces interact with the environment (adsorption/desorption, adatom diffusion, diffusion from the surface to the bulk and vice-versa) to apply a thermodynamic approach the surfaces must be close to thermodynamic equilibrium. This corresponds to high growth temperatures and/or negligible growth rates. The formalism given in this section is based on the assumption of thermodynamic equilibrium.

Let us assume a system consisting of n_{Sp} species and let n_i be the number of particles of the i_{th} species. At constant temperature T and pressure p the Gibbs free energy of the system is:

$$G = \sum_{i}^{n_{Sp}} n_i \mu_i \,. \tag{4.1}$$

Here μ_i is the *chemical potential* of the i_{th} species and describes the change in the free energy if the number of particles of the i_{th} species changes by dn_i under constant temperature, pressure, and the population of the other particles:

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_{j\neq i}} . \tag{4.2}$$

A system consisting of two or more phases is in thermodynamic equilibrium, if the chemical potential μ_i of any species is the same in all phases (see Fig. 4.8). Since the chemical potential of any species in any phase can be derived by first principles calculations, the concept of chemical potentials allows us to efficiently describe realistic systems. Furthermore, the chemical potentials can be experimentally controlled and the link between theory and experiment can be straightforwardly established.

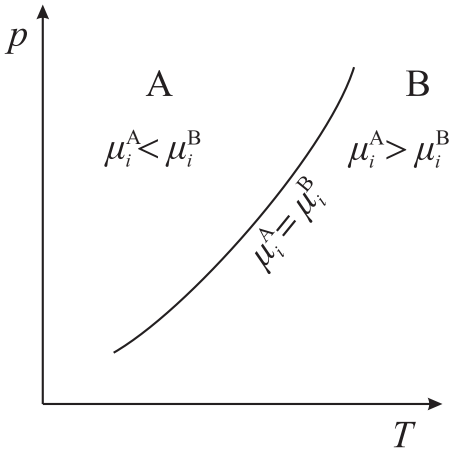


Figure 4.8: Schematic representation of a $p-T$ phase diagram. μ_i^A and μ_i^B indicate the chemical potentials of the i_{th} species in phases A and B respectively. The curve distinguishes the A- and B- stable regions and indicates the conditions for thermodynamic equilibrium between the two phases.

Commonly the chemical potentials are not independent variables, but have to obey certain boundary conditions. For the case discussed here (growth of undoped GaN) the Ga and N species should be in thermodynamic equilibrium with GaN:

$$\mu_{\text{GaN (bulk)}} = \mu_{\text{Ga}} + \mu_{\text{N}}. \tag{4.3}$$

Hence, only one of the two chemical potentials (μ_{Ga} or μ_{N}) can be chosen freely. Further restrictions stem from the requirement to avoid the formation of undesirable phases, i.e. the formation of the Ga bulk phase and the N_2 molecules. To prevent the formation of these phases, the Ga and N chemical potentials must be lower - and thus thermodynamically more stable - than the corresponding parasitic phases:

$$\begin{array}{rcl} \mu_{\text{Ga}} & < & \mu_{\text{Ga (bulk)}} \\ \mu_{\text{N}} & < & \mu_{\text{N (N}_2\text{ molecule)}} \end{array} . \tag{4.4}$$

Assuming that the entropy and pressure contributions are negligible, the chemical potential of a Ga atom in Ga bulk is:

$$\mu_{\text{Ga (bulk)}} = G_{\text{Ga (bulk)}} = E^{\text{tot}}_{\text{Ga (bulk)}} - TS + pV \approx E^{\text{tot}}_{\text{Ga (bulk)}} . \tag{4.5}$$

Here, $E^{\text{tot}}_{\text{Ga (bulk)}}$ is the total energy of Ga bulk per Ga atom. Similarly, for the N chemical potential will have:

$$\mu_{\text{N (N}_2\text{ molecule)}} = G_{\text{N (N}_2\text{ molecule)}} = E^{\text{tot}}_{\text{N (N}_2\text{ molecule)}} - TS + pV \approx E^{\text{tot}}_{\text{N (N}_2\text{ molecule)}} . \tag{4.6}$$

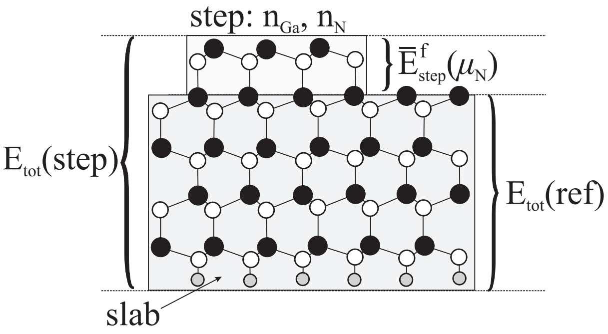


Figure 4.9: Schematic illustration of the model structure used to calculate the formation energies of steps on (0001) surfaces. The step geometry consists of n_{Ga} and n_{N} Ga and N atoms respectively. $E^{\mathrm{tot}}_{\mathrm{step}}$ is the *ab-initio* calculated total energy of the system and $E^{\mathrm{tot}}_{\mathrm{ref}}$ indicates the total energy of the reference structure (bare (0001) GaN surface). $E^{f}_{\mathrm{step}}\left(\mu_{\mathrm{N}}\right)$ is the energy required to form the step configuration on the (0001) surface and is given by Eq. (4.12).

Here, $E^{\mathrm{tot}}_{\mathrm{N\,(N_2\,molecule)}}$ is the total energy of the N_2 molecule at $p = 0$ and $T = 0$ K per N atom. While for solid and liquids the entropy and pressure contributions to the Gibbs free energy can be assumed to be negligible, for gases, such as a N_2-gas they may be significant. For realistic growth conditions the entropy contributions are larger than the pressure contributions and hence the total energy $E^{\mathrm{tot}}_{\mathrm{N\,(N_2\,molecule)}}$ can be assumed as an upper limit for the N chemical potential (see Ref. [157]). In order to get a lower limit for the Ga and N chemical potentials, we introduce the GaN heat of formation:

$$\Delta H^{f}_{\mathrm{GaN}} = \mu_{\mathrm{GaN\,(bulk)}} - \mu_{\mathrm{Ga\,(bulk)}} - \mu_{\mathrm{N\,(N_2\,molecule)}}\,. \tag{4.7}$$

Based on Eqs. (4.4) and (4.7) the thermodynamically allowed range of values for the Ga and N chemical potentials are:

$$\mu_{\mathrm{Ga\,(bulk)}} + \Delta H^{f}_{\mathrm{GaN}} \quad < \mu_{\mathrm{Ga}} < \quad \mu_{\mathrm{Ga\,(bulk)}} \tag{4.8}$$

$$\mu_{\mathrm{N\,(N_2\,molecule)}} + \Delta H^{f}_{\mathrm{GaN}} \quad < \mu_{\mathrm{N}} < \quad \mu_{\mathrm{N\,(N_2\,molecule)}} \tag{4.9}$$

. The lower limit in Eq. (4.8) (Eq. (4.9)) corresponds to N-rich (Ga-rich) conditions, while the upper one to Ga-rich (N-rich) conditions.

The energy required to create a surface (γ) is [158]:

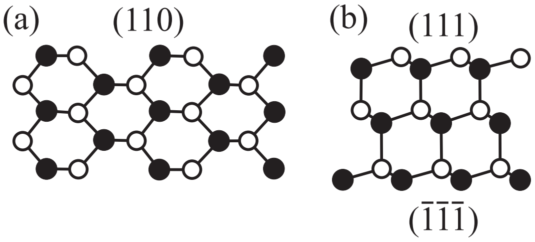
$$\gamma \mathrm{A} = F_{\mathrm{surface}} - \sum_{i}^{n_{Sp}} n_i \mu_i\,. \tag{4.10}$$

Here, n_i is the number atoms of the i_{th} species with chemical potential μ_i, A is the area of the surface, and $F_{\mathrm{surface}} = E_{\mathrm{surface}} - TS_{\mathrm{surface}}$ is the Helmholtz surface free energy. Thus for a system consisting of two species (Ga and N) the surface energy at zero temperature would be:

$$\gamma = \frac{1}{A}\left(E_{\mathrm{surface}} - n_{\mathrm{Ga}}\mu_{\mathrm{Ga}} - n_{\mathrm{N}}\mu_{\mathrm{N}}\right)\,. \tag{4.11}$$

If the calculations are performed on slabs with two equivalent surfaces the surface internal

Figure 4.10: Side view of the (a) (110) and (b) $(111)/(\overline{1}\overline{1}\overline{1})$ surfaces of zincblende GaN. Cutting along $\{110\}$ two identical surfaces will be created, while cutting along $\{111\}$ will result into inequivalent surfaces: (111) and $(\overline{1}\overline{1}\overline{1})$



energy E_{surface} will be the half of the first-principles calculated total energy and the formation energy γ will correspond to the absolute surface formation energy. An example of a slab with two equivalent surfaces is the zincblende (110)GaN surface (see Fig. 4.10(a)). On the other hand there exist directions where this is not possible: Cutting along $\{111\}$ surface planes will result in two inequivalent surfaces, (111) and $(\overline{1}\overline{1}\overline{1})$ (see Fig. 4.10(b)). However, four symmetry equivalent $\{111\}$ facets exist for the zincblende crystal. Thus, a finite crystal limited only by $\{111\}$ facets can be created. Taking the crystal sufficiently large, edge effects can be neglected and absolute surface formation energies can be derived. In practice, this approach requires large cells and it is computationally expensive. Alternative more sophisticated approaches based on the concept of *"energy-density"* have been developed [159, 160, 161, 162]. The symmetry equivalent of the (111) zincblende surface, for the wurtzite crystal is the (0001) surface. The low crystal symmetry of the wurtzite structure does not allow us to construct a finite crystal limited only of (0001) facets (see Fig. 4.7) [163]. Thus, absolute formation energies cannot be derived and a reference system is required. Usually depending on the polarity of the surfaces under consideration, the bare (0001) or $(000\overline{1})$ are used as reference systems correspondingly. In this case the internal energy E_{surface} in Eq. (4.11) should be substituted by the energy difference $E^{\text{tot}}_{\text{system}} - E^{\text{tot}}_{\text{ref}}$, where $E^{\text{tot}}_{\text{system}}$ and $E^{\text{tot}}_{\text{ref}}$ are the first-principles calculated total energies of the system under consideration (e.g. a step on the (0001) GaN surface) and of the reference structure (i.e., the bare (0001) GaN surface) respectively. Based on Eqs. (4.11) and (4.3), we get the following formula for the relative formation energy as a function of the N chemical potential[3]:

$$E^{f}_{\text{system}}\left(\mu_{\text{N}}\right) = E^{\text{tot}}_{\text{system}} - E^{\text{tot}}_{\text{ref}} - n_{\text{Ga}}\mu_{\text{GaN}} - \left(n_{\text{N}} - n_{\text{Ga}}\right)\mu_{\text{N}}\,. \tag{4.12}$$

In Fig. 4.9 a schematic illustration of the *ab-initio* total energies needed to derive the step formation energies is given. All the parameters involved in Eqs. (4.9) and (4.12) are known in advance or can be derived by first principles calculations. For non-stoichiometric systems ($n_{\text{Ga}} \neq n_{\text{N}}$) the N chemical potential is an independent variable. However, this variable cannot be chosen fully freely but it is restricted in the range predicted for thermodynamic equilibrium (Eq. (4.9)). Thus, the formation energy of the system is calculated as function of the N (or Ga) chemical potential for the thermodynamically allowed range of chemical potentials. The link between theory and

---

[3] Based on Eq. (4.3) equivalent formula can be derived for the Ga chemical potential.

experiment (Ga- or N-rich growth conditions) is straightforward. In practice, the formation energies are plotted as function of the N chemical potential. This diagram is called "phase diagram" and provides information regarding the thermodynamic stability of the structures under consideration for growth under various growth conditions (III/V ration during growth, see e.g., Fig. 4.12). The method outlined above will be used in the next Section to analyze the first principles calculations on various step configurations.

4.4 First principles calculations

In order to theoretically analyze the formation of steps on Ga polar GaN surfaces and their implication on the surface roughening, first principle calculations have been performed for a number of different step configurations. The surfaces are modelled by a repeated 6×1 slab geometry consisting of four GaN double layers (see Fig. 4.7). The steps consist of 2 surface unit cells and are introduced on the (0001) GaN surface (see Fig. 4.11). The atoms in the steps and in the three top double GaN layers of the slab were fully relaxed. The dangling bonds of the N atoms in the bottom surface of the slab (N polar surface) has been passivated by partially charged pseudohydrogen (see Sec. 4.2). The calculations have been performed employing density functional theory (DFT), using soft Troullier-Martins pseudopotentials [101] and the Perdew-Burke-Ernzerhof (PBE) generalized-gradient approximation (GGA) to describe exchange/correlation [81]. The Ga 3d semicore states are described in the frozen core approximation (nlcc) [96]. The calculations have been performed using a plane waves basis set (energy cutoff 50 Ry). An equivalent of 6×6 Monkhorst-Pack k-point sampling for the unit cell has been applied to model the Brillouin zone. The DFT program packages FHI98MD [96, 164] and sfhingx [97] were used for the calculations. The aforementioned settings have been checked to be more than enough to accurately describe the geometry and the energetics of steps on the (0001)GaN surfaces (see Sec. 6.4 and Ref. [165]).

Each step structure consists of one type-A and one type-B step edge (see Sec. 4.1). Careful convergence checks have been performed in order to ensure that edge-edge and step-step interactions are negligible: Calculation on steps consisting of 2 and 3 surface unit cells and terrace lengths consisting of 3 and 5 surface unit cells revealed differences in the formation energies of less than 0.001 eV/Å and 0.01 eV/Å. As reference structure for the relative step formation energies the bare 1×1 (0001)GaN surface has been used.

Fig. 4.11 shows the six different step configurations that have been calculated. The relative formation energies of the steps have been calculated as a function for the N chemical potential (μ_{N}) using Eq. (4.12) within the the thermodynamically allowed (Eq. (4.9)). Our calculations indicate that the formation enthalpy of GaN ($\Delta \mathrm{H}^{f}_{\mathrm{GaN}}$) is -1.13 eV. This value is in good agreement with the value of -1.28 eV calculated by previous *ab-initio* calculations within the PBE-GGA and (nlcc) [157] and with the experimental value of -1.14 eV [166].

The relative step formation energies as function of the N chemical potential are shown in

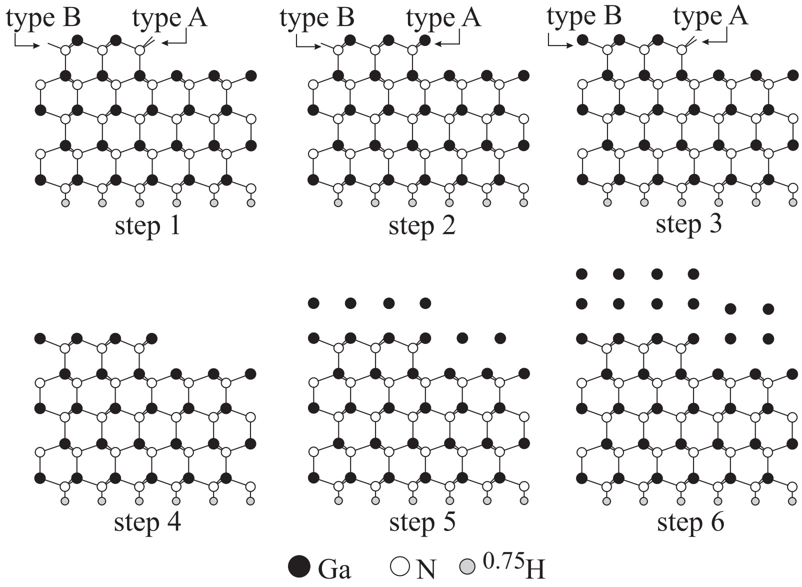


Figure 4.11: Side view of the six different step configurations used for the *ab-initio* calculations. All the steps consist of a pair of type-A and type-B edges. Step 1 has both edges N-terminated, step 2 and step 3 have the type-A and type-B edges terminated by Ga respectively, while step 4 has both edges Ga-terminated. Step 5 and Step 6 are formed starting from step 4 and adding 1 and 2 ML Ga respectively. As reference structure for the calculation of the relative step formation energies, the bare Ga polar surface has been used.

Fig. 4.12. Let us first have a closer look on the stoichiometric surface configurations (step 2 and step 3). The difference between these configurations is the termination: In configuration 2 the A-step is terminated by Ga and the B-step by Ga. In configuration 2 the two terminations are exchanged. The step configuration having only the type-B edge terminated by Ga atoms (step 3) is highly unfavorable compared to the step configuration having only the type-A edge terminated by Ga atoms (step 2). This can be explained in terms of the local energy balance at the step edges. The Ga atoms bound at the type-A edge saturate two dangling bonds, while those bound at the type-B edge saturate one dangling bond. Thus, for stoichiometric step structures, the system minimizes its energy by preferably incorporating Ga atoms at type-A edges. During the 2D island growth mode, this favors the formation of triangular shape islands bounded by type-B edges, while in the 1D step-flow growth mode this behavior favors the jagged edge morphology and step bunching (see Fig. 4.1). Both conclusions nicely explain/confirm recent experiments and provide a microscopic basis to understand the growth anisotropy (see Sec. 4.1 and Refs. therein).

Under more N-rich conditions the Ga terminated surface becomes energetically degenerate with the step 2 structure (the stoichiometric step configuration having the type-A edge Ga terminated). The step structure having both edges N-terminated (step 1), is highly unfavorable over the entire thermodynamically allowed range of N chemical potentials. Only under extreme N-rich conditions (beyond the thermodynamic equilibrium) it may become energetically favorable. On the other hand, moving towards metal-rich conditions the step configuration having both edges Ga saturated becomes energetically favorable, while at the limit of Ga-rich conditions the step 6 structure covered by 2 ML of Ga is more stable. Based on these results one might conclude that step formation is exothermic throughout the thermodynamically allowed range of N chemical potentials, i.e., the surface would be unstable against step formation. Thus, the surface would be expected to be unstable

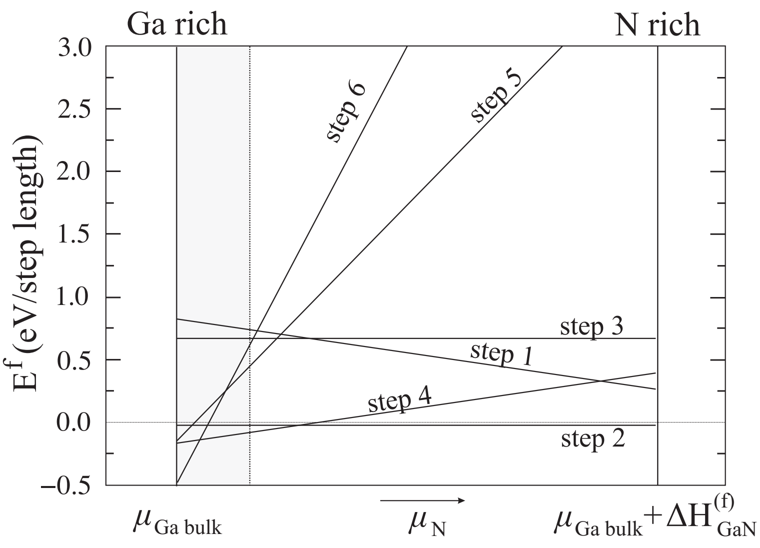


Figure 4.12: Relative formation energies of the six different step configurations (see Fig. 4.11). The zero energy corresponds to the Ga terminated (0001)GaN surface. The shaded rectangle indicates the region of the N chemical potentials where the Ga bilayer atop the Ga polar surface is more stable than the Ga terminated (0001)GaN surface.

against step formation and to roughen even under Ga-rich conditions. While, surface roughening for growth under N-rich conditions has been reported by a number of previous experiments, it is not expected for growth under Ga-rich conditions (see Sec. 5.2 and Refs. therein). To replace this discrepancy it is important to include that for Ga-rich conditions a rather different surface structure - the Ga rich laterally contracted bilayer structure - is the most stable surface configuration [43]. Thus, for Ga-rich conditions, we have to use as reference structure not the (0001)GaN surface but the laterally contracted bilayer structure.

Since the treatment of steps on the compressed Ga-bilayer leads to large surface unit cells (to comply with the underlying registry) and different slab sizes a model structure that closely resembles the original Ga-bilayer structure has been used. Specifically, a 1 × 1 Ga bilayer atop the (0001) surface as a reference structure for the Ga-rich region has been used. We find that, for growth under Ga-rich conditions, the formation of the metal-rich step 6 configuration on the Ga bilayer surface is highly endothermic: 0.54 eV/step length. The Ga-bilayer structure is less favorable than the laterally contracted Ga-bilayer under Ga-rich conditions [43]. Thus, the step 6 configuration will be at least 0.54 eV/step length higher in energy that the laterally contracted Ga-bilayer surface. Hence, we find the formation of steps/vicinal surfaces under metal-rich conditions on the laterally contracted bilayer surface to be highly endothermic.

4.5 Summary

Based on *ab-initio* based slab calculations a phase diagram has been derived that predicts the relative stability of an extensive set of possible step configurations (see Fig. 4.12). An analysis of the aforementioned phase diagram reveals that under low III/V ratio (N-rich conditions) steps may spontaneously form on the surface. That will eventually result in surface roughening and faceting, in agreement with previous *ab-initio* based adatom diffusion studies and experimental observations (see Sec. 5.2). Going towards metal-rich conditions the step terraces are predicted to be covered by 2 ML of Ga. This conclusion is consistent with the model proposed by Adelmann *et al.* to explain the elimination of faceting observed under high III/V ratios [48]. The topmost layer of the Ga bilayer is expected to be liquid and thus weakly experiences the underneath surface. Thus, the surface energy can be minimized by minimizing the surface area. Since faceting increases the surface area, it is expected that it will be unfavorable for growth under Ga-rich conditions. Indeed, here we find that the Ga bilayer structure is energetically favorable over all the various step configurations studied here at the metal-rich limit.

The step configuration covered by a Ga bilayer fits also the assumptions done by Xie *et al.* in order to explain the formation of ghost islands [154]. However, while the calculations presented here confirm the Ga-rich character of the stable step configuration for growth under metal-rich conditions, we can only speculate about the possible energy barriers the Ga adatoms experience and which kinetically hinder the transport of Ga adatoms onto the terraces upon a growth interruption.

It is important to note here that at the limit of extreme Ga-rich conditions, the energetically most stable step configurations are Ga terminated. This has interesting consequences on the sticking coefficients and incorporation rates of both Ga and N adatoms. For the N adatoms, which are the minority species under Ga-rich conditions, this implies that their sticking coefficient at the step edges will be close to one. On the other hand the sticking coefficient of Ga adatoms will be negligible: Thus, Ga adatoms are distributed almost uniformly on the surface terraces. Both conclusions will be used in Sec. 5.5 as key to derive a growth model that describes the adsorption of Ga on GaN surfaces.

Chapter 5

Gallium Adsorption on (0001) GaN Surfaces.

Despite the intensive work on epitaxy of group III-Nitrides an understanding of the basic processes during growth is still in its infancy. For example, while it is well established that MBE growth of GaN has to be carried out under Ga-rich conditions in order to obtain optimized surface and material properties, studies on GaN growth, report contradicting results regarding activation barriers and prefactors for the transitions between different growth regimes.

To address this issue, first principles calculations fundamental growth mechanisms such as Ga adsorption and nucleation (i.e. formation of small Ga islands) on top the Ga polar GaN surface have been performed. These results combined with experimental RHEED data of Ga adsorption on (0001)GaN surface and a lattice gas growth model, allow for the first time a consistent description of previous experimental results. Specifically this model is able to solve the experimental within a unified picture. The results have been published in Ref. [47].

5.1 Surface reconstruction of (0001) and $\left(000\bar{1}\right)$ GaN surfaces.

The common growth direction of wurtzite GaN is normal to the hexagonal $\{0001\}$ basal plane. The *c*-axis in wurtzite structure ($\langle 0001\rangle$ direction) is a polar axis (lack of inversion symmetry, see e.g Ref. [2]). In the $\{0001\}$ basal plane the atoms are arranged in bilayers consisting of two closely spaced hexagonal layers, one with cations and the other with anions. Thus the bilayers have polar faces and a basal surface should be either Ga- or N- faced (see Fig. 5.2). The term Ga-face (or (0001) surface) indicates Ga on the top position of the $\{0001\}$ bilayer. Similarly, N-face (or $\left(000\bar{1}\right)$ surface) indicates N on the bottom position of the $\{0001\}$ bilayer[1].

Two different groups of reconstructions have been identified for the $\left(000\bar{1}\right)$ and (0001) GaN

---

[1] Ga (N) -face or Ga (N) -polar surface do not indicate the termination of the surface. For example, a N-polar Ga-terminated surface (and actually does - see Ref. [38]) exists.

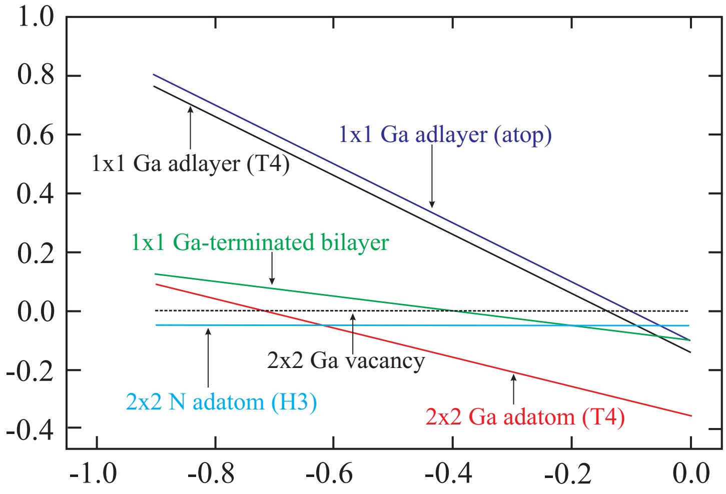


Figure 5.1: Relative formation energies of possible 1×1 and 2×2 models of GaN(0001) surface as a function of the Ga chemical potential. The energy zero corresponds to the 2×2 Ga vacancy reconstruction (it is constructed by removing a Ga atom from the 2×2 surface unit cell of the ideal GaN(0001) surface) After Smith *et al.* [38]

surfaces. Combining STM, LEED, Auger, and *ab-initio* calculations Smith *et al.* [38, 39] four dominant reconstructions for GaN$(000\bar{1})$, which in order of increasing surface Ga/N ratio are the following: 1×1, 3×3, 6×6, and $c\,(6 \times 12)$. For the technologically more relevant GaN(0001) surface the following reconstructions have been reported: 1×2, 2×2, 4×4, 5×5, 6×4, and pseudo-"1×1" [40, 41, 42, 167, 168].

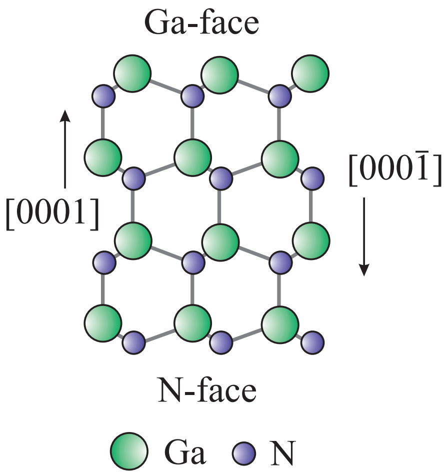


Figure 5.2: The two different polarities of wurtzite GaN along the *c*-axis.

Theoretical studies of the GaN (0001) surface have been performed by several groups employing pseudopotential plane wave [38, 169, 170] or tight binding methods [171, 172]. All studies focused on 1×1 and 2×2 reconstructions. The 2×2 reconstruction is the experimentally most commonly observed on the Ga polar surfaces. Since the 2×2 symmetry has not been observed on the GaN$(000\bar{1})$ surface, it can be assumed as a fingerprint of the Ga-polar surface. In Fig. 5.1 the relative formation energies as calculated by Smith *et al.* [38] are shown. Under N-rich conditions the 2×2-H3 N-adatom is the energetically most favorable structure with the 2×2 Ga-vacancy being only slightly higher in energy. Going to more Ga-rich conditions the 2×2-T4 Ga adatom is favored. These results are in agreement with calculations by other groups [169, 170, 172]. Only the tight binding calculations by Elsner *et al.* [171] found the Ga-monolayer and Ga-trimer models to be more stable under Ga-rich conditions. This can be attributed to the insufficient description of the metallic bonds by the tight binding approach.

Smith *et al.* reported also on the Ga-polar surface the existence of a pseudo-"1×1" reconstruction under very Ga-rich conditions [42]. The term pseudo-"1×1" relates to the fact that

RHEED shows a 1 × 1 pattern, while LEED measurements suggests an incommensurate surface structure. This discrepancy has been explained in terms of a discommensurate fluid phase, similar to that observed for Au(111) and Pt(111) [173]. The "1×1" GaN(0001) surface is the most Ga-rich reconstruction. Modelling of Auger spectral intensities indicates that it contains between 2 and 3 ML of Ga above the last GaN bilayer, while STM images reveal the height of the "1×1" Ga layer to be 3.8 Å indicating $\sim$ 2 ML of Ga atoms [40, 42]. Since the melting point of bulk Ga (29.8^o) is very near room temperature, it is reasonable to assume a fluid character of this Ga-rich surface. Thus the 1×1 corrugation pattern seen in the RHEED patterns is the result of the time averaging inherent in the STM measurements. Based on first principle calculations Northrup *et al.* confirmed that a surface consisting of a laterally contracted bilayer of Ga is the most stable structure under Ga-rich conditions (see Sec. 5.4.1).

5.2 Growth diagram of GaN(0001) surfaces.

Metallorganic Chemical Vapor Deposition (MOCVD) is nowadays the most commonly used technique for fabricating GaN-based devices. On the other hand, molecular beam epitaxy (MBE) compared to MOCVD provides the advantage to accurately control growth at the monolayer scale by monitoring in site surface reconstructions by using reflection high energy electron diffraction (RHEED). Thus, *growth diagrams* (referred also as *surface phase diagrams*) for the growth of GaN are considerably helpful for the MBE growth of high quality materials. In general, a growth diagram provides information regarding the surface morphology (and thus the grown material quality) as function of the growth conditions (Ga/N ratio fluxes and growth temperature).

In Fig. 5.3 a typical MBE GaN growth diagram is shown [46]. One of the fundamental findings of such a surface phase diagram is that there the surface morphology depends sensitively on the III/V ratio during growth [46]. In general, three different regimes can be distinguished:

- growth under N-rich conditions (N-stable regime),
- growth under slightly metal-rich conditions (intermediate regime), and
- growth under more Ga-rich conditions and at lower temperatures (Ga-droplets regime).

Films grown under N-rich conditions have in general spotty a RHEED patterns (which characterizes a rough surface morphology) and a poor crystalline quality. On the other hand, growth under Ga-stable conditions produces films having streaky RHEED patterns (which are characteristic for smooth two-dimensional growth morphologies). At low substrate temperatures or extreme Ga-rich conditions Ga-droplets formation is observed [46, 48, 147, 174, 175]. Besides the surface morphology, the growth conditions affect also the electrical and optical properties. Hall measurements on MBE grown GaN epilayers indicate that higher carrier mobilities are obtained for material grown under slightly Ga-rich conditions (intermediate regime in the growth diagram) [176]. Furthermore,

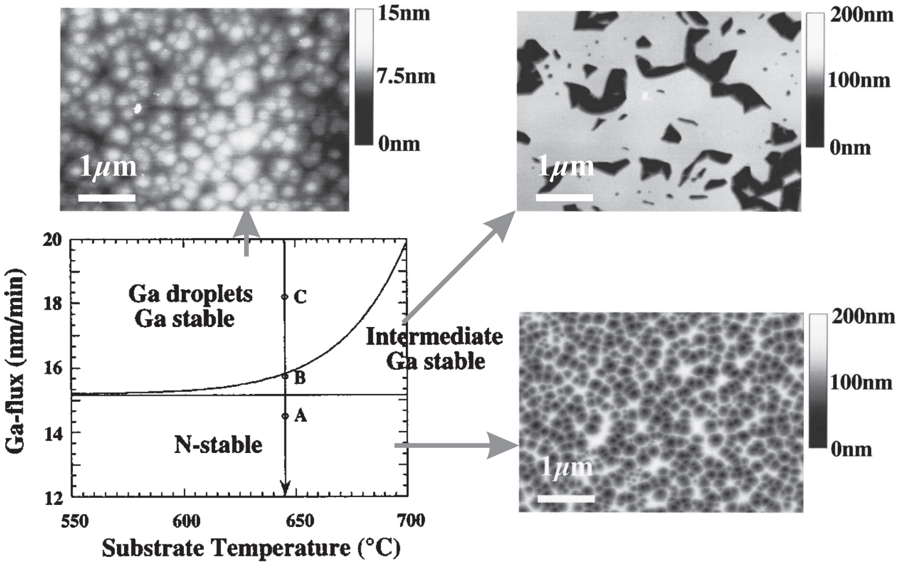


Figure 5.3: Phase growth diagram indicating the growth conditions (Ga - flux and substrate temperature) at constant N-flux, for the three regimes: Ga - droplets, intermediate, and N - stable. The arrows point to AFM images (after the removal of excess Ga by HCl etching), characteristic for each of the three regimes. After Heying *et al.* [46].

photoluminescence (PL) measurements showed that N stable growth deteriorates the optical properties of the film: growth under those conditions was found to result in material with increased yellow luminescence and a decreased near-band edge emission [147]. At the same time, an increased yellow luminescence and a decrease of the near-band-edge emission were observed for regions grown under a Ga droplet [177].

A conclusion of these observations is that an optimal quality of the material (i.e., a smooth surface morphology, optimum electronic and optical properties) can be achieved if the growth takes place under slightly Ga-rich conditions (i.e., in the intermediate regime). This can be explained in terms of surface kinetics. Zywietz *et al.* [178] based on *ab-initio* calculations found that N adatoms on Ga-terminated surfaces experience significantly higher diffusion barriers than Ga adatoms. Thus, under N-rich conditions, N-terminated surfaces can be kinetically stabilized. They also found that N-terminated surfaces exhibit roughly five times higher diffusion barriers than Ga-terminated surfaces, i.e., N on GaN surfaces significantly reduces the Ga diffusion lengths. This effect will eventually lead to a statistical roughening of the surface. Also, it is possible that adatoms are trapped at positions not corresponding to the ideal bulk positions: This may be the critical stage of stacking fault formation. On the other hand, Ga-rich conditions will result in a smaller amount of N present on the surface and in a higher mobility of the Ga adatoms. Both effects are expected to lead to a 2D step-flow growth mode and to a reduced density of stacking faults. For a detailed analysis of surface structures and growth kinetics of group III-Nitrides see Ref. [179].

As can be seen in Fig. 5.3 the Ga fluxes at which transition between the different growth regimes

Table 5.1: Experimental activation energies $E_{\mathrm{A}}^{\mathrm{exp}}$ and prefactors $\nu_{\mathrm{des}}^{\mathrm{exp}}$ for the transition fluxes between different Ga coverage regimes as obtained from the adsorption ($\upsilon = 0$) and growth ($\upsilon > 0$) phase diagrams (from Refs. as indicated). Here, υ denotes the growth rate (N flux). The $0 \rightarrow 1$ transition corresponds to the adlayer formation atop the bare surface for both surface polarities. For the Ga-polar surface the $1 \rightarrow 2$ and $2 \rightarrow 3$ transitions correspond to the formation of the laterally contracted layer atop the pseudomorphic adlayer and the formation of Ga droplets atop the bilayer respectively. For the N-polar surface the $1 \rightarrow 2$ transition corresponds to the formation of Ga-droplets on top of the Ga adlayer. α the linear temperature coefficient, and $\nu_{\mathrm{des}}^{\mathrm{ren}}$ the renormalized prefactor (see Sec. 5.6).

Reference	Transition	$E_{\mathrm{A}}^{\mathrm{exp}}$ (eV)	$\nu_{\mathrm{des}}^{\mathrm{exp}}$ (Hz)	υ (ML/s)	α (meV/K)	$\nu_{\mathrm{des}}^{\mathrm{ren}}$ (Hz)
(0001)GaN surface:						
Ref. [47]	$1 \rightarrow 2$	5.2	3.0×10^{25}	0	-2.4	3.0×10^{25}
Ref. [47]	$2 \rightarrow 3$	5.1	2.0×10^{25}	0	-2.3	9.8×10^{24}
Ref. [49]	$0 \rightarrow 1$	4.9	6.6×10^{24}	0	-2.1	1.0×10^{24}
Ref. [49]	$1 \rightarrow 2$	3.7	6.2×10^{18}	0	-0.9	1.4×10^{18}
Ref. [49]	$2 \rightarrow 3$	3.2	1.9×10^{16}	0	-0.4	4.9×10^{15}
Ref. [48]	$1 \rightarrow 2$	3.7	5.0×10^{17}	0.28	-0.9	1.4×10^{18}
Ref. [48]	$2 \rightarrow 3$	4.8	1.0×10^{24}	0.28	-2.0	3.3×10^{23}
Ref. [46]	$2 \rightarrow 3$	2.8	1.0×10^{14}	1.1	> -0.1	5.4×10^{13}
$(000\bar{1})$GaN surface:						
Ref. [49]	$0 \rightarrow 1$	3.7	4.5×10^{18}	0	-0.9	1.4×10^{18}
Ref. [49]	$1 \rightarrow 2$	3.1	4.1×10^{15}	0	-0.3	1.6×10^{15}

occurs (transition fluxes) depend on the growth temperature. To obtain a quantitative description of the growth diagram we assume an Arrhenius behavior for the transition fluxes:

$$\Phi = \Phi_i + \nu^{\mathrm{exp}} \exp\left(-E_{\mathrm{A}}^{\mathrm{exp}}/k_B T_S\right) . \tag{5.1}$$

Here, $E_{\mathrm{A}}^{\mathrm{exp}}$ and ν^{exp} are the activation energy and prefactor respectively which are characteristic for each transition. The offset Φ_i must be the same for all transitions. In Table 5.2 the activation energies and prefactors for the transition fluxes between different growth regimes (as indicated) as derived from experimental phase diagrams are given. A closer inspection of the values poses a number of questions:

- In Table 5.2 the values for both growth ($\upsilon > 0$) and adsorption ($\upsilon = 0$) phase diagram experiments are cited[2]. Why are the activation energies and prefactors in the adsorption and growth phase diagrams so different?

- Why are the values for the same transition ($2 \rightarrow 3$ for growth) in Refs. [48] and [46] so different.

---

[2] The growth phase diagram experiments investigate the Ga transition fluxes during growth (finite N flux). In the adsorption experiments the Ga transition fluxes are investigated after interruption of N flux and adsorption of finite amount of Ga (see Sec. 5.3)

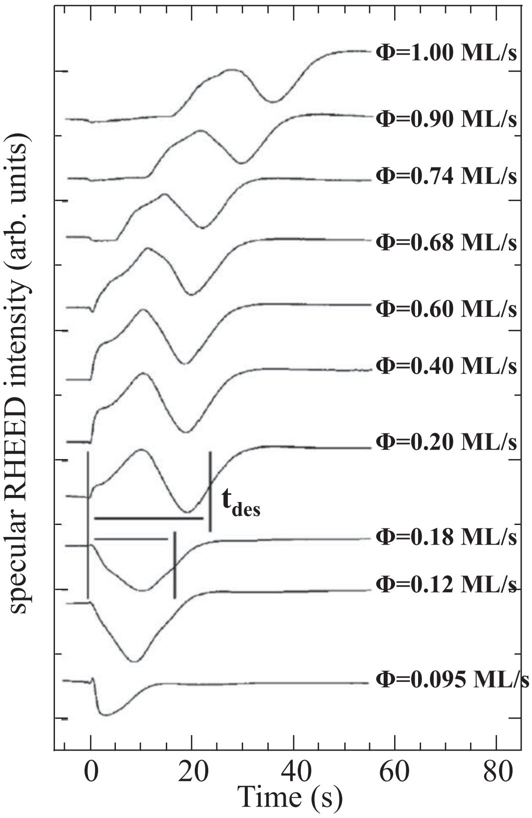


Figure 5.4: Specular RHEED intensity during Ga desorption from a (0001) GaN surface. Beforehand, Ga adsorption was carried out for 1 min at indicated Ga fluxes (substrate temperature $T_S = 740^oC$). The desorption time is defined/measured as the time interval between the Ga flux interruption and the last inflection point in the RHEED intensity. After Adelmann *et al.* [47].

- The activation energy of 2.8 eV for the 2 → 3 transition reported in the growth phase diagram in Ref. [46] is close to the cohesive energy of bulk Ga (2.81 eV [146]). It is thus tempting to interpret is as a Ga desorption barrier. However, what is then the physical origin of the 5.1 eV activation barrier reported in Ref. [47]?

- What is the physical meaning of a prefactor of 10^{25} Hz (which would correspond to a vibrational energy in the order of $\sim$ 10 GeV)? Typically observed prefactors for diffusion or desorption processes are in the order of $\sim 10^{13}$ Hz.

In order to address the above questions experimental data are analyzed in terms of a simple lattice-gas growth model. We will first briefly describe the experimental RHEED studies which allow to accurately monitor the desorption of Ga from GaN(0001) surfaces and that give a quantitative description of the growth phase diagram.

5.3 RHEED measurements

The experimental procedure to monitor Ga desorption is as follows: First, a well determined amount of Ga is adsorbed on the surface. This is done by switching off the N-flux and varying the Ga-flux. After a fixed adsorption time the Ga flux is interrupted and the specular RHEED

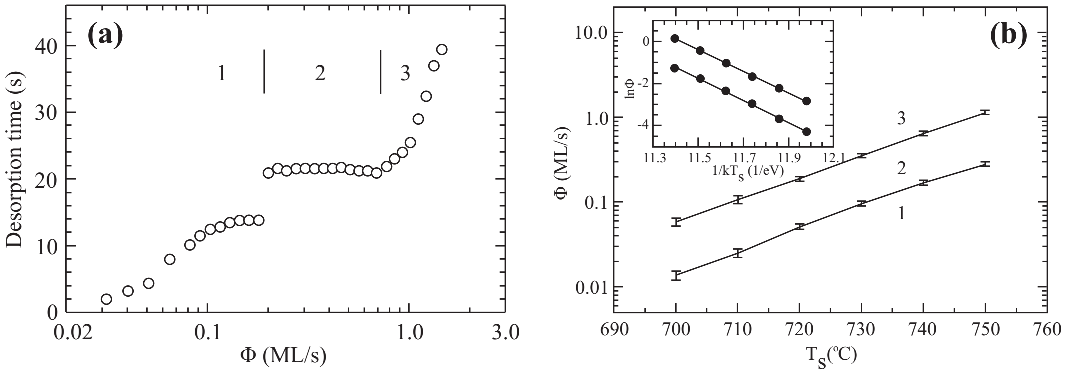


Figure 5.5: (a) Ga desorption time as a function of the impinging Ga flux after equilibrium has been attained for regions 1 and 2 and after 1 min of Ga adsorption for region 3. The substrate temperature is 740 °C. (b) The impinging Ga fluxes at transitions between the three regimes as a function of the substrate temperature. The inset corresponds to an Arrhenius plot of the data. After Adelmann *et al.* [47].

intensity during the subsequent Ga evaporation is recorded.

Fig. 5.4 shows the specular RHEED transients (the specular RHEED intensity corresponds to the electron beam intensity whici is not diffracted but specularly reflected by the surface) during Ga desorption after 1 min of Ga adsorption for indicated Ga fluxes and subsequent interruption of the Ga flux (at $t = 0$). The duration of the oscillatory transients observed during the desorption ($t_{\rm des}$) depends on the amount of Ga present on the surface after adsorption. The fundamental finding is that the desorption time $t_{\rm des}$ for Ga fluxes $\Phi \leq 0.72$ ML/s is independent of the previous adsorption time. A closer analysis of the desorption time as function of the nominally impinged amount of Ga ($\theta = \Phi t_{\rm des}$) shows that for Ga fluxes $\Phi \leq 0.72$ ML/s finite equilibrium Ga surface coverages exist. For higher values of the Ga flux no finite equilibrium exists and thus Ga infinitely grows and forms macroscopic droplets on the surface.

The results of Fig. 5.4 are summarized in Fig. 5.5(a) where the desorption time - and qualitatively the Ga surface coverage - are given as function of the impinging Ga flux. The substrate temperature was kept constant (T_S = 740 °C). This diagram can be assumed as Ga adsorption isotherm. Three different regimes can be discriminated: (i) an S-shaped increase of $t_{\rm des}$ for $\Phi < 0.20$ ML/s, (ii) a constant time $t_{\rm des}$ up to $\Phi = 0.72$ ML/s, and (iii) no finite equilibrium and continuous Ga accumulation. In Fig. 5.5(b) the variation of the adsorption isotherm as function of the substrate temperature is given. The transition fluxes vary exponentially with the temperature. Assuming an Arrhenius behavior with an offset equation $\Phi = \Phi_i + \nu_{\rm des}^{\rm exp} \exp\left(-E_{\rm A}^{\rm exp}/k_B T_S\right)$ (see Eq. (refeq:Arrehnius)) the activation energies and prefactors are obtained by a least square fit (see inset of Fig. 5.5). The corresponding values are: $E_A^{(12)} = 5.2 \pm 0.1$ eV and $\nu_{\rm des}^{(12)} = 3 \times 10^{25}$ Hz for the transition from the first to the second regime and $E_A^{(23)} = 5.1 \pm 0.05$ eV and $\nu_{\rm des}^{(23)} = 2 \times 10^{25}$ Hz for the transition from the second to the third regime.

Restricting the analysis to regimes 1 and 2, i.e., to regimes where equilibrium surface coverages are obtained, a quantitative description of the Ga adsorption isotherm can be achieved. This is shown in Fig. 5.6. We see that the increase of the desorption time in the first region (see Fig. 5.5(a)) corresponds to an increase in the Ga surface coverage from almost zero to a value close to 1 ML. The transition from the first to the second region is accompanied by an abrupt increase of the coverage to ∼ 2.5 ML.

Since 1 ML coverage corresponds to a Ga adatom on each GaN site, it can be assumed that in region 1 Ga adsorbs as a coherent (pseudomorphic) adlayer. The 2.5 ML surface coverage in region 2 and the abrupt increase by almost 1.5 ML at the $1 \rightarrow 2$ transition are consistent with the Ga surface coverage of the laterally contracted bilayer model of 2.33 ML (see Sec 5.4.1 and Ref. [43]). Further evidence that the adsorption in the second region is related to the formation of the laterally contracted bilayer comes from a closer analysis of the RHEED patterns: For rapidly quenched samples a 2.73 ± 0.03 Å lattice parameter is found. This value is close to the value of 2.75 Å calculated for the in lattice parameter of the laterally contracted bilayer.

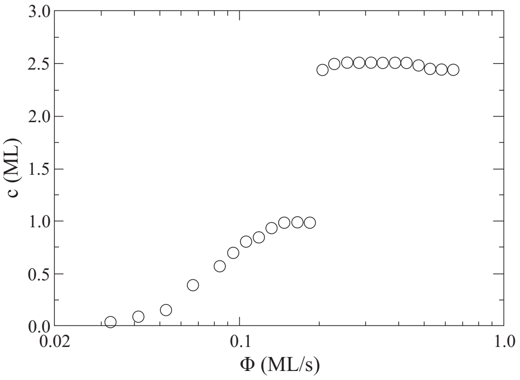


Figure 5.6: Quantitative description of the Ga adsorption isotherm shown in Fig. 5.5(a). After Adelmann *et al.* [47].

The laterally contracted bilayer model is further corroborated by recent Ga adsorption studies by Koblmüller *et al.* [49]. They showed that an adlayer consisting of 2.4 ML Ga can be stabilized atop the bare Ga-polar surface. Additionally they found that this adlayer consists of two layers: a top layer with 1.4 ML and a bottom layer with 1 ML Ga surface coverage, with the bottom layer being more strongly bound than the top one.

The three different regimes of the isotherm given in Fig. 5.5(a) can be summarized as follows: (1) Successive formation of a coherent Ga monolayer, with Ga coverage $c \leq 1$ ML, (2) formation of a laterally contracted Ga bilayer structure with Ga coverage $c = 2.5$ ML, and (3) Ga droplets formation on top of the Ga bilayer. The activation energies for the $1 \rightarrow 2$ and $2 \rightarrow 3$ transitions were found identical ($E_A^{(12)} = 5.2 \pm 0.1$ eV and $E_A^{(23)} = 5.1 \pm 0.05$ eV).

5.4 First principles analysis.

In order to identify the origin of the aforementioned discrepancies, *ab-initio* calculations of the formation of small Ga islands on top of the Ga laterally contracted bilayer structure have been performed. Since the activation barriers for the $1 \rightarrow 2$ and $2 \rightarrow 3$ transitions in the adsorption phase diagram are identical, it is expected that the mechanisms/energetics behind these transitions are rather similar. Thus, without losing generality, the first principle calculations have been focused on the Ga-bilayer structure and the $2 \rightarrow 3$ transition. Due the importance of the laterally contracted bilayer model, a short description of it will be given before the discussion of the first principles calculations.

5.4.1 The laterally contracted bilayer model.

The laterally contracted bilayer model as proposed by Northrup *et al.* [43], has been shown by *ab-initio* calculations to be the most stable (0001)GaN surface structure under Ga-rich conditions. A schematic representation of the laterally contracted bilayer structure is shown in Fig. 5.7. It consists of two Ga layers atop the bare (0001)GaN surface. The bottom layer (layer 2 in Fig. 5.7) is coherent (pseudomorphic) to the Ga-polar surface and contains 3 Ga atoms per $\sqrt{3} \times \sqrt{3}$ surface cell. The top layer (layer 1) contains 4 Ga atoms per $\sqrt{3} \times \sqrt{3}$ cell. Both layers give total Ga surface coverage of 2.33 ML.

The aforementioned findings compare favorably with the pseudo-"1×1" reconstruction of the (0001)GaN surface that is believed to prevail during Ga-rich growth. Auger spectra indicated that this reconstruction contains between 2 and 3 ML of Ga above the last GaN bilayer [40, 41, 42]. Additionally it has been proposed that it corresponds to an incommensurate structure whose 1×1 corrugation pattern observed in the STM arises from the the dynamical nature of the topmost Ga layer and the time averaging inherent in the STM measurement [42].

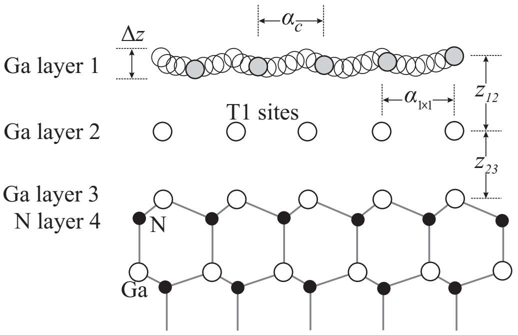


Figure 5.7: Side view of the laterally contracted bilayer structure proposed by Northrup *et al.* [43]. The filled circles in the top layer represent Ga atoms at a particular time. The open circles represent a time averaged image of the Ga atoms. The time averaged vertical corrugation of the top layer is approximately 0.16 Å.

5.4.2 First principles calculations.

The *ab-initio* calculations were performed employing density functional theory (DFT) using soft Troullier-Martins pseudopotentials [101] and the Perdew-Burke-Ernzerhof (PBE) generalized-gradient approximation (GGA) to describe exchange/correlation [81]. The Ga 3d semicore states are described in the frozen core approximation (nlcc) [96]. The DFT program package FHI98MD [96, 164] was used for the calculations. The calculations have been performed using a plane waves basis set (energy cutoff 50 Ry). The surfaces were modelled using slab geometries with $\sqrt{3} \times \sqrt{3}$ and $2\sqrt{3} \times 2\sqrt{3}$ unit cells and two GaN double layers (see Fig. 5.8 for a top view and Fig. 5.9 for a side view). The slab thickness has been checked to describe accurately the energetics of the system. As has been shown (see Sec. 6.4) an increase of the slab thickness to four double layers changes the surface energy by less than 0.5 meV/1×1. The Brillouin zone has been sampled by a $4 \times 4 \times 1$ and $2 \times 2 \times 1$ Monkhorst-Pack mesh for the $\sqrt{3} \times \sqrt{3}$ and $2\sqrt{3} \times 2\sqrt{3}$ unit cells respectively [95].

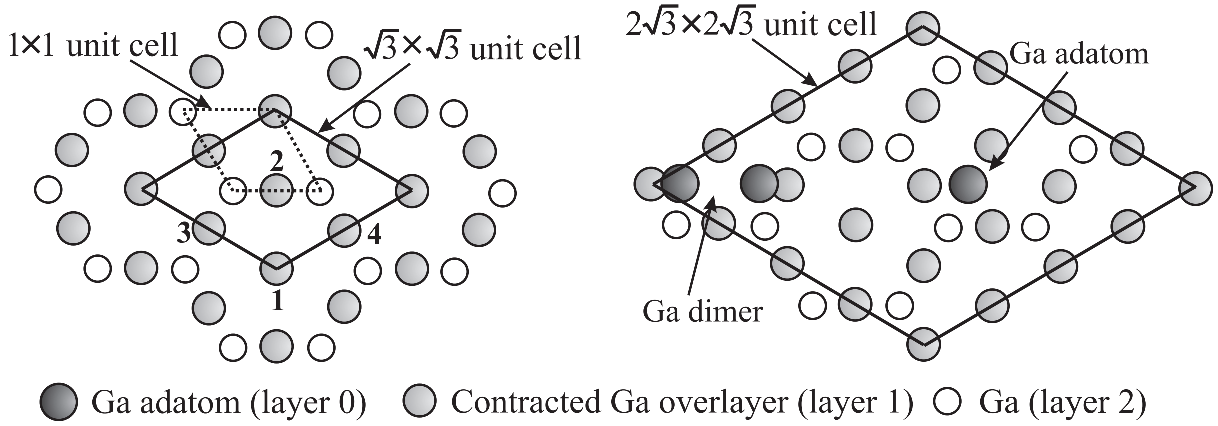


Figure 5.8: Schematic top view of the Ga bilayer structure (a) and an adatom and a 2 atom island (b) on the Ga bilayer structure. The 1×1, $\sqrt{3} \times \sqrt{3}$, and $2\sqrt{3} \times 2\sqrt{3}$ surface unit cells are also shown. The white balls mark the positions of the Ga atoms in the second layer, the grey balls the positions of Ga atoms in the contracted top layer, and the black balls indicate the positions of the Ga adatoms on top the contracted epilayer.

The lower surface has been passivated by partially charged pseudo-hydrogen in order to remove the electrically active surface states (see Sec. 4.2). The adatoms and the first two surface layers have been fully relaxed. Since the onset of Ga droplets on top the Ga bilayer structure takes place under Ga rich conditions, the formation energies discussed in the following have been calculated for the Ga rich limit.

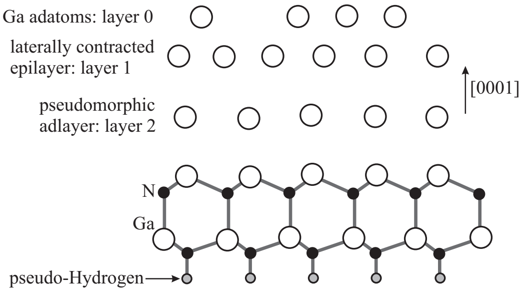


Figure 5.9: Schematic side view of an adatom and a 3 atom island on top the Ga bilayer structure. The slab consists of two GaN double layers. The lower surface is passivated by partial charged pseudo-Hydrogens in order to remove the electrically active surface states.

The lattice parameter of bulk GaN is $\alpha_{1\times1} = 3.19$ Å (6.03 Bohr). Hence the in lattice parameters of the contracted layer is $\alpha_c = \left(\sqrt{3}/2\right)\alpha_{1\times1} = 2.76$ Å. The configuration of the bilayer structure corresponds to the registry B reported in Ref. [43]: one of the Ga atoms (the atom 1 in Fig. 5.8(a)) of the contracted layer (layer 1) sits on top of one of the Ga atoms of the pseudomorphic adlayer (layer 2). After performing atomic relaxation it is found that while the three Ga atoms in layer 1 (atoms 2, 3, and 4, see Fig. 5.8(a)) are coplanar (within less than 0.05 Å vertical corrugation), atom 1 is located approximately 0.25 Å higher. The pseudomorphic adlayer is located approximately 2.53 Å above the (0001)GaN bare surface and the contracted layer approximately 2.34 Å above layer 2. The same vertical separations (within a tolerance of 0.01 Å) have been calculated also for the structures with Ga adatom/islands atop the contracted layer.

In a first step the desorption energy of an adatom on top of a Ga bilayer has been calculated. The desorption energy (the energy needed to remove an adatom from the surface) is defined by

$$E_{\text{des}} = -\left(E_{\text{tot}}^{\text{adatom}} - E_{\text{tot}}^{\text{slab}} - E_{\text{tot}}^{\text{atom}}\right) \tag{5.2}$$

where $E_{\text{tot}}^{\text{adatom}}$ is the total energy of the surface including the adatom, $E_{\text{tot}}^{\text{slab}}$ that of the free laterally contracted bilayer surface, and $E_{\text{tot}}^{\text{atom}}$ that of the free (spin polarized) Ga atom[3]. In the equilibrium configuration, the adatom sits on a three fold coordinated hollow site (see Fig. 5.8). Using Eq. (5.2), the adatom binding is found to be 2.52 eV for the $\sqrt{3} \times \sqrt{3}$ unit cell and 2.46 eV for the $2\sqrt{3} \times 2\sqrt{3}$ structure. The vertical corrugation of the contracted layer is approximately 0.50 Å. The Ga-Ga bond length between the adatom and the surface is 2.68 Å. This distance is close to the nearest neighbor distance in α-Ga of 2.71 Å [181].

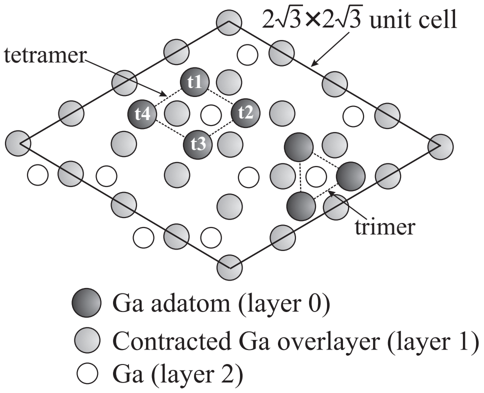


Figure 5.10: Schematic top view of a three (trimer) and a four (tetramer) atom island. The white balls mark the positions of the Ga atoms in the pseudomorphic adlayer, and the grey balls the positions of Ga atoms in the contracted top layer. The black balls indicate the positions of the Ga adatoms in the islands.

In a second step the binding energy of an adatom to an island has been calculated. For an island consisting of n_{ad} atoms the binding energy is:

$$E_{\text{isl}} = -\left[\frac{1}{n_{\text{ad}}}\left(E_{\text{tot}}^{\text{isl}} - E_{\text{tot}}^{\text{slab}} - n_{\text{ad}} E_{\text{tot}}^{\text{atom}}\right) - E_{\text{des}}\right]. \tag{5.3}$$

Here $E_{\text{tot}}^{\text{isl}}$ is the total energy of the surface including the island and E_{des} the desorption energy of an adatom given by Eq. (5.2). For an island consisting of two adatoms we find binding energies of 0.16 eV and 0.21 eV for the $\sqrt{3} \times \sqrt{3}$ and $2\sqrt{3} \times 2\sqrt{3}$ unit cells respectively. The Ga-Ga bond length between adatoms and surface is 2.67 Å and the bond length of the dimer is 2.71 Å. Both results are very close to the nearest neighbor distance in α-Ga.

For larger islands, consisting of three adatoms, $E_{\text{isl}} = 0.30$ eV. For a four atom island $E_{\text{isl}} = 0.34$ eV. The positions of the adatoms in the the three and four atom islands are schematically shown in Fig. 5.10. The Ga-Ga bond length between the island atoms and the surface is on average 2.60 Å, while the distance between the island atoms is 2.81 ± 0.04 Å. The Ga-Ga bond length between the atoms in a four adatom island is 2.81 ± 0.07 Å (except the distance between the atoms t2 and t4 shown in Fig 5.10). The t1, t2, and t3 adatoms of the tetramer have bond lengths of 2.60 Å with the laterally contracted surface, while the t4 atom has a bond length of 2.80 Å.

Finally, we have calculated the formation energy of Ga atom in an island, i.e., the energy required for a Ga atom to adsorb on the surface and then bind to an island. Since for all islands this energy is smaller than the cohesive energy of Ga bulk ($E_{\text{des}} + E_{\text{isl}} = 2.75$ eV compared to

---

[3] The value of 0.148 eV has been used for the spin-polarized correction of the Ga atom [180].

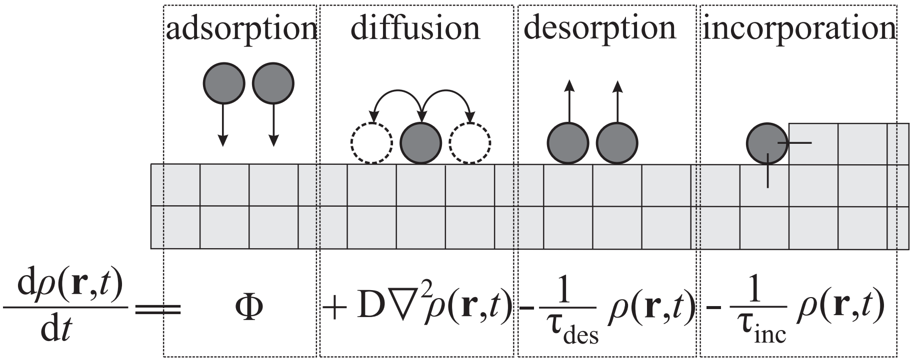


Figure 5.11: Schematic representation of the adatom density time evolution $\mathrm{d}\rho\left(\mathbf{r},t\right)/\mathrm{d}t$.

$E_{\mathrm{coh}}^{\mathrm{Ga\,bulk}} = 2.8$ eV), all islands are unstable against droplet formation. Thus, the islands are precursors for Ga droplet formation. Indeed, the formation of Ga islands minimizes the total energy of the system by realizing a configuration that allows the Ga atoms to form bonds with lengths close to the first nearest neighbors in α-Ga. This further corroborates the statement that the Ga islands on top the bilayer structure act as nucleation centers for Ga droplets formation.

5.5 Growth and adsorption lattice gas model.

In Sec. 5.4 the adsorption and formation of small Ga clusters on the Ga bilayer structure has been described. Here, the dependence of the Ga adatoms density ρ on parameters such as temperature T, Ga flux Φ, and growth rate v will be derived[4]. As schematically shown in Fig. 5.11, the time derivative of the adatom density depends on the amount of Ga that arrives on the surface (impinging Ga flux Φ), the surface diffusion ($D\nabla^2\rho\left(\mathbf{r},t\right)$), the amount of Ga that desorbs ($-(1/\tau_{\mathrm{des}})\rho\left(\mathbf{r},t\right)$), and the amount of Ga that incorporates ($-(1/\tau_{\mathrm{inc}})\rho\left(\mathbf{r},t\right)$):

$$\frac{\mathrm{d}\rho\left(\mathbf{r},t\right)}{\mathrm{d}t} = \Phi + D\nabla^2\rho\left(\mathbf{r},t\right) - \frac{1}{\tau_{\mathrm{des}}}\rho\left(\mathbf{r},t\right) - \frac{1}{\tau_{\mathrm{inc}}}\rho\left(\mathbf{r},t\right). \tag{5.4}$$

Here, D is the surface diffusion constant and $\mathbf{r}$ is the lateral position on the surface. The desorption time t_{des} is given by

$$\tau_{\mathrm{des}}^{-1} = \nu_{\mathrm{des}} e^{-E_{\mathrm{des}}/k_{\mathrm{B}}T}, \tag{5.5}$$

where E_{des} is the desorption barrier and k_{B} the Boltzmann constant.

Under Ga rich conditions growth is realized via a 2D step flow mode (see Sec. 5.2). For step flow growth, incorporation takes place mainly at the step edges. Since the steps are moving, the

---

[4] Here, the density ρ indicates the adatom density on top of the last completed Ga layer. That is a monolayer for the $1 \rightarrow 2$ transition and a bilayer for the $2 \rightarrow 3$ transition. The adatom density must be distinguished from the Ga surface coverage c, that denotes the total amount of Ga present on the bare surface (including completed Ga monolayers or bilayers).

incorporation rate $\tau_{\text{inc}}^{-1}(\mathbf{r},t)$ is in general inhomogeneous and time dependent. For ideal step flow it will be zero on the terraces and > 0 only at the step edges.

The growth rate υ and the incorporation rate are directly related by

$$\upsilon(t) = \frac{1}{A}\int \frac{1}{\tau_{\text{inc}}(\mathbf{r},t)}\rho(\mathbf{r},t)\,\mathrm{d}\mathbf{r}, \tag{5.6}$$

where the integration is performed over the total surface area A.

As has been mentioned in Sec. 5.3, in regimes 1 and 2 of the phase diagram stationary conditions are realized. Furthermore the transitions we are interested in occur exclusively under Ga rich conditions. Under stationary conditions the growth rate is time independent (i.e. $\upsilon(t) = \upsilon$) and the time derivative of the adatom density vanishes (i.e. $\mathrm{d}\rho(\mathbf{r},t)/\mathrm{d}t = 0$). Additionally, as shown in Sec. 4.4 under Ga rich conditions the steps on the surface are Ga terminated. In a conventional step flow growth mode the sticking probability of an adatom at the step edge is expected to be close to one. In contrast to the aforementioned conventional picture, here we have two species with very different concentrations. For the nitrogen adatoms (which are the minority species under very Ga rich conditions) the sticking coefficient at the Ga-terminated step will be close to one. The steps act as sinks to the nitrogen concentration, which becomes highly inhomogeneous. On the other hand, for the Ga adatoms the sticking coefficient at the Ga-terminated step edges will be very low. This has two consequences: First, the Ga incorporation rate τ_{inc}^{-1} will be very small. Second, the effect of steps on the Ga concentration will be negligible hence the Ga-adatom density $\rho(\mathbf{r},t)$ and the incorporation rate $\tau_{\text{inc}}^{-1}(\mathbf{r},t)$ will be virtually homogeneous. Thus, $D\nabla^2\rho(\mathbf{r}) = 0$ and Eq. (5.6) entails $(1/\tau_{\text{inc}})\,\rho = \upsilon$.

Based on the above discussion, Eq. (5.4) can be rewritten as:

$$0 = \Phi - \upsilon - \frac{1}{\tau_{\text{des}}}\rho_0. \tag{5.7}$$

Here, ρ_0 is the equilibrium Ga adatom density. It worths to note here that the above equation applies both for growth ($\upsilon > 0$) and adsorption ($\upsilon = 0$). Eq. (5.7) can be rewritten as:

$$\rho_0 = (\Phi - \upsilon)\,\tau_{\text{des}}. \tag{5.8}$$

Nucleation occurs if the (stable) nuclei are in thermodynamic equilibrium with the lattice gas (which is described by the adatom density) on the surface. For low adatom densities ($\rho_0 \ll 1$), interactions in the lattice gas itself can be neglected and we obtain

$$\frac{N_{\text{ad}}}{N_{\text{sites}}} = \rho_{\text{crit}} = \exp\left(-\frac{\Delta E_{\text{nuc}}}{k_{\text{B}}T}\right), \tag{5.9}$$

where N_{ad} the number of adatoms in the lattice gas, N_{sites} is the total number of surface sites which can be occupied by the adatoms, and ΔE_{nuc} is the energy an adatom gains if it is attached

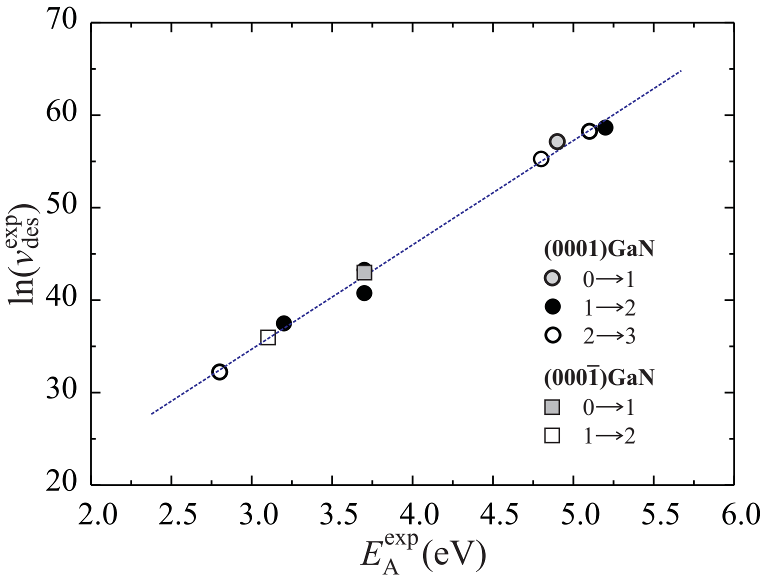


Figure 5.12: Natural logarithm of the experimental prefactors $\ln(\nu_{\mathrm{des}}^{\mathrm{exp}})$ as function of the activation energy $E_{\mathrm{A}}^{\mathrm{exp}}$ as taken from Refs. [46, 47, 48], and [49]. The transitions correspond to those mentioned in Table (5.2). The line represents a linear fit of the data.

to a subcritical nucleus making the latter stable.

Combining Eqs. (5.8) and (5.9), we obtain the critical Ga flux Φ_{crit} at which the phase transitions occur:

$$\Phi_{\mathrm{crit}} = \upsilon + \tau_{\mathrm{des}}^{-1} \exp\left(-\frac{\Delta E_{\mathrm{nuc}}}{k_{\mathrm{B}}T}\right). \tag{5.10}$$

Using Eq. (5.5), this equation can be rewritten as

$$\Phi_{\mathrm{crit}} = \upsilon + \nu_{\mathrm{des}} \exp\left(-\frac{E_{\mathrm{des}} + \Delta E_{\mathrm{nuc}}}{k_{\mathrm{B}}T}\right). \tag{5.11}$$

This equation is valid both for the adsorption and growth phase diagram.

5.6 Interpretation of the results.

The first principles calculations described in Sec. 5.4 revealed an activation energy for the formation of small Ga islands atop the Ga bilayer structure of 2.75 eV. Comparing this number with the reported experimental data as shown in Table 5.2, we see that the only activation energy which is close to the *ab-initio* calculated energy is for experiments performed at high growth rates (υ = 1.1 ML/s). For lower growth rates or adsorption the experimentally reported activation energies are higher. In Fig. 5.12 the natural logarithm of the prefactors $\ln\left(\nu_{\mathrm{des}}^{\mathrm{exp}}\right)$ is plotted as a function of the activation energy $E_{\mathrm{A}}^{\mathrm{exp}}$. As revealed by Fig. 5.12, there is a clear relation between the prefactors and the activation energies:

$$\ln\left(\nu_{\mathrm{des}}^{\mathrm{exp}}\right) = \alpha_1 E_{\mathrm{A}}^{\mathrm{exp}} + \alpha_0. \tag{5.12}$$

Here the subscript "exp" indicates experimental data. For the data in Table 5.2 we get $\alpha_1 = 11.27\ \mathrm{eV}^{-1}$ and $\alpha_0 = 0.897$[5].

In order to explain this unusual and unexpected dependence between prefactor and activation energies we have developed and tested a growth model with a temperature depended activation energy. Since the experimentally accessible range of temperatures is rather small ($\simeq 50$ K), a linear temperature dependence of the activation energy E_A is assumed:

$$E_\mathrm{A} = E_0 + \alpha\,(T - T_o)\,. \tag{5.13}$$

Here, E_0 is a temperature independent contribution, T_0 is a temperature offset, and α is a linear temperature coefficient. Substituting the activation energy given by Eq. (5.13) in Eq. (5.11), the critical Ga flux at phase transition can be written as:

$$\Phi_\mathrm{crit} = \upsilon + \nu_\mathrm{des} e^{-\alpha/k_\mathrm{B}} e^{-(E_0 - \alpha T_0)/k_\mathrm{B} T}. \tag{5.14}$$

Based on the previous equation, we can assume that the experimentally observed activation energies have the following temperature dependence:

$$E_\mathrm{A}^\mathrm{exp} = E_0 - \alpha T_0. \tag{5.15}$$

Similarly, comparing Eqs (5.11) and (5.14) we see that the experimentally attempt prefactors $\nu_\mathrm{des}^\mathrm{exp}$ should be temperature dependent. The temperature dependence of the prefactors is expressed by assuming that they have the following renormalized form:

$$\nu_\mathrm{des}^\mathrm{ren} = \nu_\mathrm{des} e^{-\alpha/k_\mathrm{B}}. \tag{5.16}$$

Here, ν_des corresponds to the actual attempt frequencies. From Eq. (5.15) the linear temperature coefficient α can be written as

$$\alpha = \frac{E_0 - E_\mathrm{A}^\mathrm{exp}}{T_0}. \tag{5.17}$$

Taking the natural logarithm on both sides of Eq. (5.16) and substituting the temperature coefficient from Eq. (5.17) we get:

$$\ln\left(\nu_\mathrm{des}^\mathrm{ren}\right) = \left[\frac{1}{k_\mathrm{B} T_0}\right] E_\mathrm{A}^\mathrm{exp} + \left[\ln\left(\nu_\mathrm{des}\right) - \frac{E_0}{k_\mathrm{B} T_0}\right]. \tag{5.18}$$

---

[5] The small differences to the values reported in Ref. [47] (where most of the work discussed in this chapter has been published) are due to including here additional experimental data from Ref. [49]. This publication was not available at the time of the publication of Ref. [47]. For the following discussion these differences are negligible.

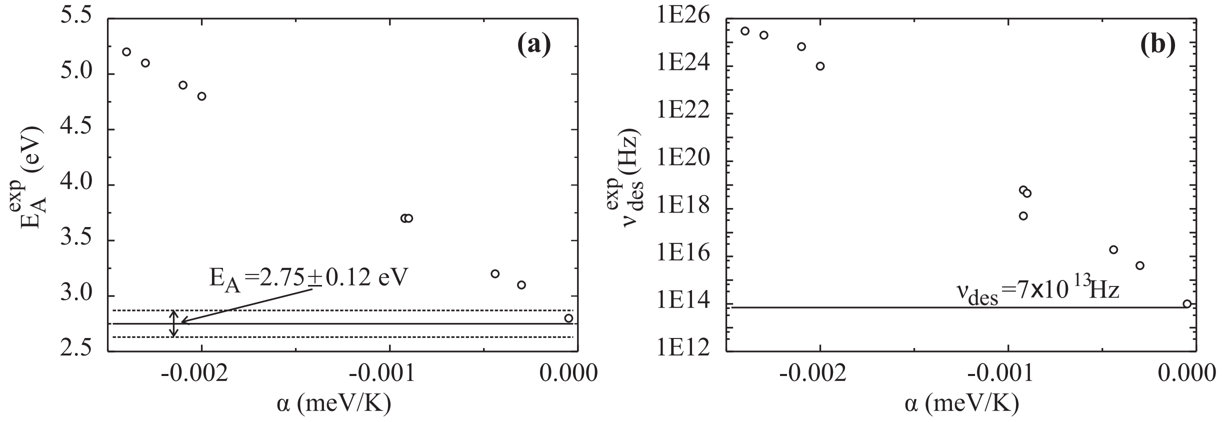


Figure 5.13: Experimentally derived (a) activation energies $E_{\mathrm{A}}^{\mathrm{exp}}$ and (b) prefactors $\nu_{\mathrm{des}}^{\mathrm{exp}}$ as function of the linear temperature coefficient α. The activation energies $E_{\mathrm{A}}^{\mathrm{exp}}$ differ by almost a factor of 2 and the prefactors $\nu_{\mathrm{des}}^{\mathrm{exp}}$ by up to 12 orders of magnitude. For comparison (a) the variation of the effective energies E_{A} and (b) the actual (i.e., non-renormalized) attempt frequency ν_{des} are plotted. The data are taken from Table 5.2.

Comparing Eqs. (5.12) and (5.18) we have:

$$\alpha_1 = \frac{1}{k_{\mathrm{B}}T_0} \quad \text{and} \quad \alpha_0 = \ln\left(\nu_{\mathrm{des}}\right) - \frac{E_0}{k_{\mathrm{B}}T_0} \tag{5.19}$$

Using these relations and the value of 2.75 eV for the temperature independent activation energy E_0 (as found from the first principles calculations) we obtain $T_0 = 757$ °C, which is close to the experimentally accessed temperatures (700 °C – 750 °C). Thus, the assumption of the linear temperature dependence of the activation energies is well justified. Also we find for the actual (i.e. non-renormalized) attempt frequency $\nu_{\mathrm{des}} \simeq 7 \times 10^{13}$ Hz, which is close to standard attempt frequencies.

In Table 5.2 temperature coefficients α and renormalized prefactors $\nu_{\mathrm{des}}^{\mathrm{ren}}$ as calculated using Eqs. (5.16) and (5.17) with $\nu_{\mathrm{des}} = 7 \times 10^{13}$ are listed. Having a closer look at the data of Table 5.2 the following conclusions can be drawn:

- There is a good agreement between the renormalized and the experimentally measured prefactors. That shows that the assumption of a linear temperature dependent activation energy with our lattice gas growth and adsorption model consistently describes all experimental studies to determine activation barrier and prefactor. In particular, for the first time it is possible to describe in the same model both adsorption and growth experiments.

- The large variation in frequencies and activation energies is accompanied by a large variation in the temperature coefficients α. In general the temperature dependence is larger if no N flux is present ($\upsilon = 0$) and it becomes minimum for the highest growth rate (for $\upsilon = 1.1$ ML/s, $\alpha \rightarrow 0$). This indicates that *the N flux affects the temperature dependence of the frequencies* *and the activation barriers.*

- The temperature dependence has a huge effect on the experimentally derived apparent activation energies $E_{\mathrm{A}}^{\mathrm{exp}} = E_0 - \alpha T_0$: The values found differ by almost a factor of 2 (see Fig. 5.13(a)). A similar effect has been found for the experimentally measured prefactor $\nu_{\mathrm{des}}^{\mathrm{exp}}$ (by up to 12 orders of magnitude!-see Fig. 5.13(b)). On the other hand, the actual change in the effective desorption energy $E_{\mathrm{A}} = E_0 - \alpha\,(T - T_0)$ is small. For $\Delta T = 50$ K and assuming the highest calculated temperature coefficient the effective energies are within a small energy window of 2.75 ± 0.12 eV. For the case of high growth rates, where the temperature dependence is considerably smaller, this deviation is even smaller. This can be understood by considering that *the experimentally measured activation energy represents a linear projection to zero temperature, even when its real temperature dependence deviates strongly from linear behavior outside the experimental temperature window.*

5.7 Summary

The work presented here was motivated by a number of inconsistences arising from experimental studies on the GaN growth diagram. Different activation energies and unphysical prefactors (which were up to 10^{12} orders of magnitude to large) have been reported for the activation of Ga droplet formation during Ga rich GaN growth [46, 47, 48]. The results presented in this section have been published in Ref. [47]. Experimental data which had been published only after this publication has been added and included in the present analysis.

Performing *ab-initio* calculations the Ga adsorption and formation of small Ga islands atop the Ga bilayer structure have been studied. We find that the formation energy of islands on top the bilayer structure is smaller than the cohesive energy of bulk Ga and hence the islands can be assumed as precursors of Ga droplets formation. This has interesting consequences in the growth of GaN under metal-rich conditions. In conventional III-V semiconductors growth under metal-rich conditions is expected to result to metal accumulation and droplet formation on the surface, which is detrimental to the quality of the grown material. In Chapter 4 we shown that, for growth under Ga-rich conditions, steps and vicinal surfaces are energetically highly unfavorable against the formation of the laterally contracted Ga bilayer configuration. Further growth of thicker Ga wetting layer on top the (0001)GaN surface would block the growth of GaN. However this is prevented by the formation of Ga droplets on top the Ga bilayer structure.

In combination with RHEED experiments, an *ab-initio* based growth and adsorption model has been derived. This model describes consistently the adsorption of Ga on GaN surfaces as well the as the accumulation of Ga during Ga-rich growth and explains the aforementioned discrepancies in terms of a temperature dependent activation energy. The activation barrier was found to reduce with temperature and only speculations can be made for the origin of this behavior. The in lattice parameter and thus the geometry of the top contracted Ga layer is expected to change with temperature. In turn, the change in surface geometry is expected to affect the island formation energy. However, further first principles calculations are required in order to confirm this assumption.

Chapter 6

Morphology and surface reconstruction of m-plane GaN.

In Section 5.1 we have seen that the common growth direction of wurtzite group III-Nitrides is the normal to the hexagonal basal plane $\{0001\}$. Thus, the majority of theoretical and experimental growth studies has been directed to the polar (0001) and $\left(000\bar{1}\right)$ surfaces (see e.g. the reviews by Strite and Morkoç [1], Jain *et al.* [31], and Ambacher [2]). One of the main problems in growth of group III-Nitrides is the high density of threading dislocations and domain boundaries defects (see e.g. Sec. 8.1). These boundaries often coincide with non-polar surfaces of the material. Non-polar surfaces have been also observed to form the inner surface of extended defects such as nanotubes and dislocation cores. Early theoretical studies on non-polar $\left(10\bar{1}0\right)$ and $\left(11\bar{2}0\right)$ GaN surfaces have been mainly motivated by the need to get a deeper understanding of the electronic structure and stability of extended defects [182, 183, 184]. Experimentally, growth of m-plane GaN$\left(10\bar{1}0\right)$ films has been motivated by the fact that huge internal electrostatic fields are generated by the spontaneous and piezoelectric polarization along the polar [0001]-axis [61]. Although these fields are used in some applications (such as e.g. confining carriers in a two-dimensional electron gas), in many applications they are unwanted (such as e.g. in the fabrication of high-efficiency GaN-based light emitting diodes). Since the c-axis is a singular polar axis in wurtzite structure, it is expected that growth at any direction orthogonal to it (e.g along the $\left[10\bar{1}0\right]$ or $\left[11\bar{2}0\right]$ directions) will result in films free of pyroelectric or piezoelectric (if shear stresses in the growth plane are absent) properties [61].

Here we present and discuss the results of *ab-initio* calculations on the GaN m-plane. Particularly, we propose a model in which a surface reconstruction with an approximate symmetry of "4×5" exists on this surface. This work has been motivated by recent STM experiments on m-plane GaN films grown by MBE. These experiments revealed the existence of a surface reconstruction with metallic character and "4×5" symmetry. The outline of this chapter is as follows: First the pyro- and piezo-electric properties of wurtzite group III-Nitrides are discussed and a brief review of the few growth studies on the non-polar GaN surfaces are given. Next, recent STM experiments on the *m*-plane GaN films will be presented and in Section (6.3) we focus on previous theoretical works on non-polar GaN surfaces. Finally an *ab-initio* based model of the *m*-plane GaN surface with an approximate symmetry of "4×5" will be discussed in detail. The analysis/discussion given in Sec. 6.4 together with the STM experiments and the *ab-initio* calculations on the 1×1 Ga adlayer configurations (given in Section 6.3), have been published in Refs. [185] and [186].

6.1 Polarization properties of wurtzite group III-Nitrides

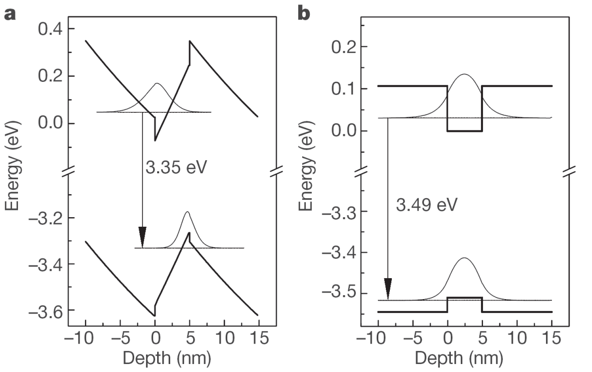


Figure 6.1: Calculated band profiles in (5 nm GaN)/(10 nm $Al_{0.1}Ga_{0.9}N$). (a) The very large electrostatic fields induced by polarization along the *c*-axis result in quantum confined Stark effect and poor electron-hole overlap. (b) The $[1\bar{1}00]$ orientation is free of electric fields. After Waltereit *et al.* [61].

Materials with a singular polar axis can exhibit spontaneous polarization (SP), i.e., polarization even in zero strain. Since in wurtzite structure the [0001] axis is a singular polar axis it is expected that materials with this symmetry exhibit SP. Indeed, it has been predicted recently by theory and confirmed by experiment that group III-Nitrides with wurtzite crystal structure are pyroelectric materials. On the other hand, zincblende crystals cannot be polarized, since they have four symmetry-equivalent polar axes, whose contributions cancel each other. Wurtzite is the structure with the highest symmetry compatible with SP [187, 188]. Bernardini *et al.* predicted by *ab-initio* calculations that the group III-Nitrides in wurtzite phase are characterized by large values of SP and that the magnitude of this effect sensitively depends on the structural parameters (i.e. the degree of divergence from the ideal wurtzite structure[1]) [189].

The direction of the SP is determined by the polarity. For group III-Nitrides it is along the (0001)-axis. The pyroelectric crystals are always piezoelectric and, in the absence of an electric field, the total macroscopic polarization is the sum of the SP and the strain induced piezoelectric polarization. As is mentioned in Appendix A.2, in epitaxial growth of wurtzite group III-Nitrides the epilayers are under biaxial strain. Under these conditions and assuming the shear stresses to be negligible, the piezoelectric polarization has only one non-vanishing component which is along the growth direction (for a detailed proof see the recent work by Ambacher *et al.* [190]):

$$P_3^{\mathrm{pz}} = 2\epsilon_1 \left(e_{31} - e_{33}\frac{C_{13}}{C_{33}} \right) . \tag{6.1}$$

Here, e_{31} and e_{33} are two of the three non-vanishing piezoelectric constants of the wurtzite crystal,

---

[1] For the ideal wurtzite structure see Appendix A.1.

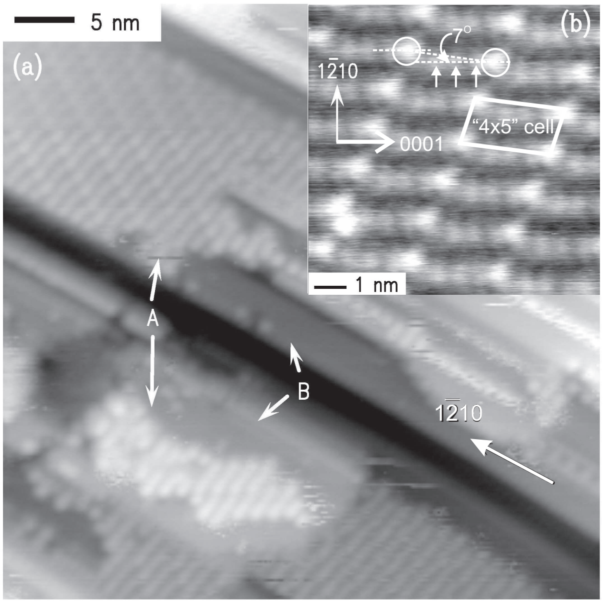


Figure 6.2: (a) STM image of Ga-rich GaN$(1\bar{1}00)$ surface. Two different reconstructions are observed, labelled "A" and "B". (b) Expanded view of the region with the "A"-type reconstruction. The arrows indicate the corrugation maxima. The brighter maxima are indicated by the circles. The "4×5" surface unit cell is also shown. After Lee *et al.* [185].

that link polarization with strain. The third piezoelectric constant e_{15} is related to polarization induced by shear strain. It is small and thus commonly neglected. For a definition of the strain ϵ_1 and of the elastic constants C_{13} and C_{33} see Appendix A.2.

The polarization induces charges ($\rho_P = -\nabla P$) and strong electric fields along the c-axis. These electric fields often have a substantial influence on the optical and electronic properties of strained group III-Nitride heterostructures. The are also have useful for device applications: examples are the realization of a two dimensional electron gas (2DEG) near a heterointerface [2, 190, 191]. However, they also produce a shift in the wavelength of emission lines and reduce the quantum efficiency due to the spatial separation of electrons and holes in the quantum wells (see for example Fig. (6.1)).

An interesting option to avoid polarization in wurtzite GaN epilayers is to grow along alternative directions. Since the [0001]-axis is the only polar axis in the wurtzite structure, any direction orthogonal to it, such as the $[1\bar{1}00]$ direction, will be free of SP components [192]. Consequently, growth studies on $(1\bar{1}00)$ oriented (m-plane) GaN films have been performed. Along this orientation, the crystal symmetry forbids pyroelectric and piezoelectric field (at least in the absence of shear stress in the growth plane) [61, 192, 193, 194, 195]. Waltereit *et al.* were the first to report growth of m-plane GaN films on $\gamma - \mathrm{LiAlO_2}\,(100)$ [61, 192, 193]. They observed slate-like surface morphologies and they showed that in contrast to conventional GaN(0001) growth, GaN$(1\bar{1}00)$ films are free of electrical polarization and indeed exhibit improved quantum efficiencies. No surface reconstructions have been reported. Recently, MBE growth of m-plane GaN$(1\bar{1}00)$ on $\gamma - \mathrm{LiAlO_2}\,(100)$ has been recently studied by Sun *et al.* [194, 195]. They reported that growth conditions play an important role on morphology and quality of the grown surfaces. More specific, they observed that growth under N-rich conditions induces a phase mixture of *c*- and *m*-planes, while under Ga-rich conditions and at low temperatures, smooth *m*-plane surfaces are formed.

6.2 RHEED and STM observations.

The *m*-plane GaN films were grown on ZnO$\left(1\bar{1}00\right)$ substrates by plasma-assisted MBE. A low temperature (∼ 500°C) GaN buffer layer was used in order to obtain smooth morphology. Growth under Ga-rich conditions was found to result in smooth surfaces and streaky RHEED patterns. On the other hand growth under N-rich conditions resulted in spotty RHEED patterns indicative of three-dimensional growth.

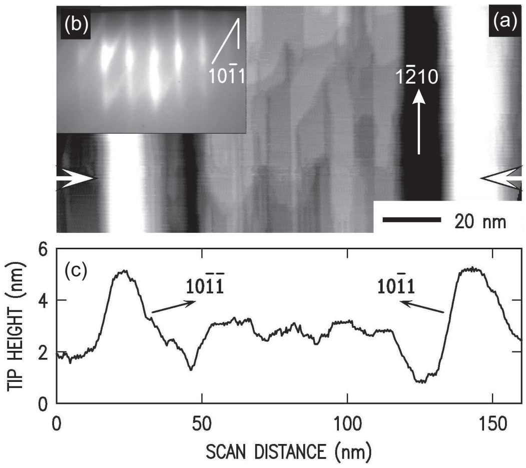


Figure 6.3: (a) STM image of a 1.0 μm thick GaN film. (b) RHEED pattern with electron beam along $\left[11\bar{2}0\right]$. (c) Line cut taken at the position of the arrows in (a). After Lee *et al.* [185].

In Fig. 6.3(a) the STM image and in Fig 6.3(b) the corresponding RHEED pattern of a 1.0 μm thick film are shown. The RHEED pattern displays oblique streaks indicative of $\left(1\bar{1}01\right)$ or $\left(1\bar{1}0\bar{1}\right)$ facets. These facets are also seen in the STM image as revealed in the accompanying line cut shown in Fig. 6.3(c). Assuming that faceting is not kinetically induced, it can be deduced that the $\left(1\bar{1}00\right)$ and $\left(1\bar{1}01\right)$ surfaces are having comparable formation energies. Hence, faceting is a potential problem for the growth of flat $\left(1\bar{1}00\right)$ films.

The surface morphology of the films is characterized by large GaN$\left(1\bar{1}00\right)$ terraces, as e.g. those between the ridges in Fig. 6.3(a). These terraces elongate in the $\left[11\bar{2}0\right]$ direction and have slate-like morphologies. Several surface reconstructions have been observed with the clearest observations being for the films grown under Ga-rich conditions. In Fig. (6.2a) the STM image of a Ga-rich GaN$\left(1\bar{1}00\right)$ surface is shown. Two different reconstructions can be distinguished on that surface labelled "A" and "B". The "B"-type reconstruction is featureless (that allows to speculate a small unit cell such as 1 × 1) and its atomic configuration is unknown yet.

The "A"-type reconstruction is the most commonly observed and consists of corrugated rows extended in directions ±7° away from the [0001] direction (see Fig. 6.2(a)). In the expanded view shown in Fig. 6.2(b) it can be seen that the corrugation maxima form rows with short segments extending in the [0001] direction. At every fourth or fifth maxima a brighter maximum occurs that is shifted in the $\left[11\bar{2}0\right]$ direction compared to its neighbors. The unit cell of this reconstruction is indicated in the image. For a better visualization a ball and stick model of the top two layers of the *m*-plane GaN surface is shown in Fig. 6.4. The primitive vectors of the 1 × 1 unit cell are

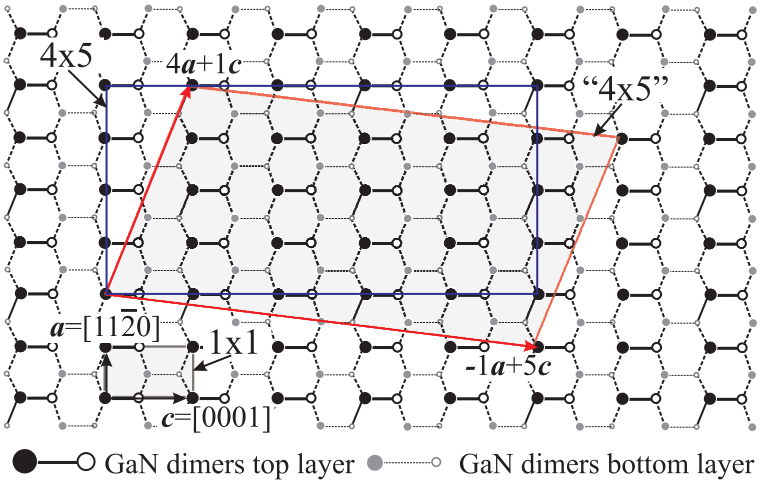


Figure 6.4: Top view of the top two layers of the GaN$(1\bar{1}00)$ surface. The 1×1, 4×5, and "4×5" are shown. As it can been seen, the "4×5" cell can be formed by slightly rotating and shearing the 4×5 cell.

the $\boldsymbol{a}$ vector in the $[11\bar{2}0]$ direction with length equal the wurtzite lattice parameter α and the $\boldsymbol{c}$ vector in the [0001] direction and length equal the c lattice parameter of the wurtzite structure. The experimentally observed unit cell of the "A"-type reconstruction matches that of one with primitive vectors $4\mathbf{a} + 1\mathbf{c}$ and $-1\mathbf{a} + 5\mathbf{c}$, which contains 21 1×1 unit cells. The shorthand notation "4×5" (including the quotation marks to denote an approximate symmetry) is used for the "A"-type reconstruction because it can be constructed by slightly rotating and shearing the $4\boldsymbol{a} \times 5\boldsymbol{c}$ cell[2]. Finally it should be noted that STM imaging even at low voltages of ±0.1 V is posible and that the images are independent on the bias. This behavior is characteristic for to *metallic* surfaces.

6.3 Earlier theoretical works.

The m-plane GaN surfaces have been studied theoretically employing DFT calculations by Northrup and Neugebauer [183] and Filippetti *et al.* [182] and Hartree-Fock calculations by Jaffe *et al.* [184]. In Fig. (6.6) a schematic top view of the GaN$(1\bar{1}00)$ surface is shown. It consists of equal numbers of three fold coordinated Ga and N atoms that form an array of Ga-N dimers. Thus, charge neutrality can be obtained without changes in stoichiometry or reconstruction.

In general, surface relaxation on heteropolar covalent or ionic semiconductors is driven by electron-electron Coulomb repulsion, by quantum mechanical hybridization effects, and by classical Coulomb attraction between anions and cations (see e.g. the review by Pollmann and Krüger in Ref. [196]). Even in the case of covalent systems, following the electron counting rule, charge

---

[2] In matrix notation it can be written as $\begin{pmatrix} 1 & 4 \\ -1 & 5 \end{pmatrix}$. For reasons of simplicity the shorthanded "4×5" notation is used.

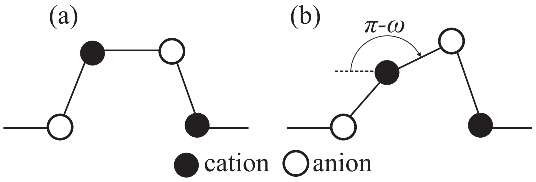


Figure 6.5: Side view of a III-V semiconductor (110) or $\left(1\bar{1}00\right)$ surface, for (a) unrelaxed and (b) relaxed configuration.

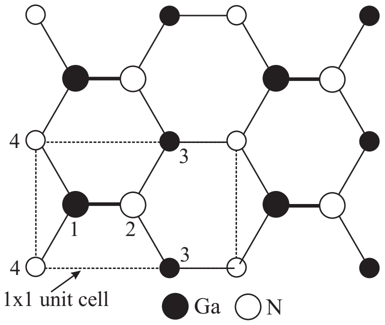


Figure 6.6: Schematic top view of the m-plane surface of wurtzite GaN. The 1×1 unit cell is indicated by the dash line. Atoms 1 and 2 are in the surface, while atoms 3 and 4 and are in the second layer.

transfer between surface atoms can occur and hence, some of these will behave as anions and others as cations. In order to reduce the hybridization energy the cations tend to move below the surface and to establish planar sp^2-like bonds with their three nearest neighborhoods. On the other hand, the anions tend to move upwards and to establish a p^3-like configuration. The re-hybridization is forwarded by a rotation ω of the dimer (see Fig. 6.5). This rotation is accompanied by a lowering of the energy of the occupied anion dangling bond and by an increase of the energy of the cation empty state. Finally, the relevant cation-anion displacement costs electrostatic energy and in order to minimize it, the anion-cation distance tends to shorten.

Depending on the ionicity and the symmetry of the system one of the aforementioned relaxation mechanisms might dominate over the others. For more covalent systems the quantum mechanical re-hybridization effects dominate, giving rise to relatively large bond rotation angles, while the anion-cation bond length is nearly preserved. On the contrary, for more ionic systems the anion-cation Coulomb attractions dominate the relaxation, resulting in a relatively small buckling angle and relatively large bond contraction. Since the group III-Nitrides are fairly ionic, it is expected that the m-plane GaN surface relaxation should be mainly governed by the ionic Coulomb attractions. The bond-angle rotation is expected to be relatively small (compared e.g. to the bond-angle rotation in the equivalent (110) III-V semiconductor surfaces) and the surface bonds are expected to contract.

Indeed, for the GaAs(110) surface Alves *et al.* [198] based on DFT calculations reported relatively large bond-rotation angle ($\sim 30^o$) and almost negligible bond contraction ($\sim 1\%$). Regarding the electronic structure, they found surface states in the fundamental bandgap (see Fig. 6.7 (a) & (b)), that arise from the empty cation-derived C_3 state and the occupied anion-derived dangling bond A_3 state (see Fig. 6.7(c) & (d)). Upon relaxation the empty C_3 state moves upwards and the occupied A_5 state shifts to lower energies, and a reduction of surface energy is obtained. On the contrary, as reported by Northrup and Neugebauer [183], the more ionic character of GaN,

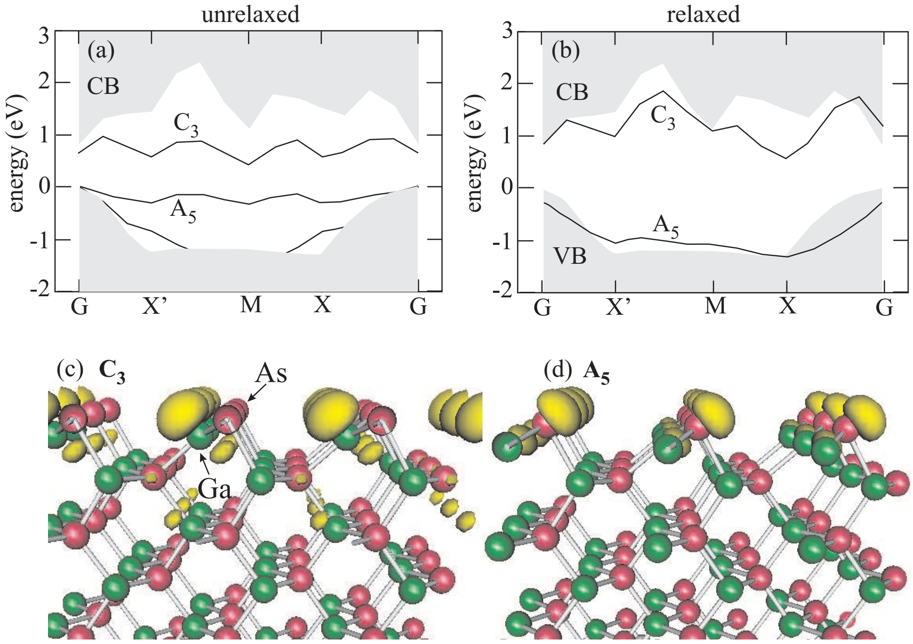


Figure 6.7: Band structure for the GaAs(110) (a) unrelaxed and (b) relaxed surface. The projected bulk band structures are indicated by the shaded areas. Charge densities of the (c) C_3 and (d) A_5 states at the M point of the surface Brillouin zone, shown in (b). The calculations have been performed within the local density approximation with basis-set cutoff energy of 10 Ry (see Refs. [197]and [198]).

compared to GaAs, results to smaller buckling angles ($\sim 7^o$) and larger bond contractions ($\sim 6\%$). They found that, similar to the GaAs(110) surfaces, the Ga atoms move downwards and relax to an sp^2 configuration. On the other hand, the N atoms remain close to their original position, tending to adopt a more p^3-like configuration. A similar surface relaxation has been reported by Fillipetti *et al.* [182] for the other group III-Nitride non-polar surfaces. In general the non-polar surfaces of group III-Nitrides exhibit only quantitative differences compared to the III-V counterparts.

The stoichiometric dimer surface (SD) (see Fig. 6.8(a)) would be expected to be present after cleavage or cracking. However, recent STM experiments (see Sec. 6.2) revealed the existence of non-stoichiometric *m*-plane GaN surfaces. The first studies on non-stoichiometric $\left(1\bar{1}00\right)$GaN surfaces were reported by Northrup and Neugebauer [183]. Two different non-stoichiometric configurations were studied that formed by replacing the surface Ga or N atoms by N or Ga atoms and are called N-N-dimer and Ga-Ga dimer or GD respectively. Although the formation of the N-N dimer configuration might be expected to be favorable under N-rich conditions, calculations revealed that the N-N dimer surface is higher in energy than the others for all possible thermodynamically allowed Ga chemical potentials. The SD surface was found to be energetically favorable over a large range of the thermodynamically allowed chemical potentials while, the GD surface becomes energetically favorable under more Ga-rich conditions, indicating that this surface could be formed under Ga-rich growth conditions.

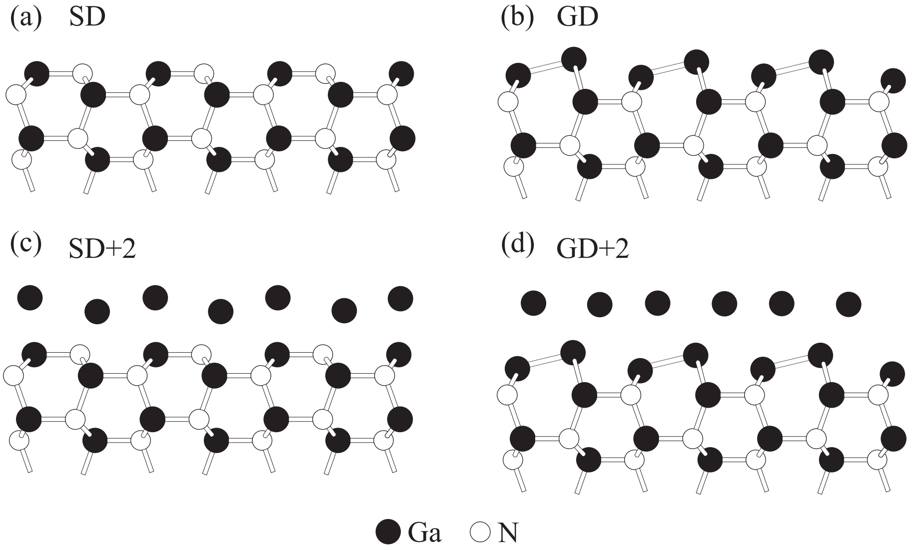


Figure 6.8: Side views of various 1 × 1 structural models, shown in a $[11\bar{2}0]$ projection. SD stands for the single dimer and corresponds to the stoichiometric dimer surface. GD stands for the Gallium dimer structure and it can be formed by replacing the N surface atoms by a Ga atoms. The SD+2(GD+2) models can be formed by the addition of two monolayers of Ga to the SD(GD) surface.

The stability of the GD configuration under more Ga-rich conditions and the metallic character of the the m-plane surface grown under Ga-rich conditions, which have been revealed by the STM experiments, motivated calculations for two additional possible configurations. These configurations were formed by adding two monolayers of Ga on top the SD (called SD+2) and the GD structures (called GD+2) (shown in Figs. 6.8(c) and 6.8(d) respectively)[3]. One monolayer indicates one atom per 1 × 1 unit cell. The calculations have been performed within the local density approximation using Troullier-Martins pseudopotentials with the Ga 3d orbitals in the valence and a plane waves basis set with 60 Ry cutoff energy. The surfaces were modeled by a slab geometry with 4 layers thickness. The relative stability of the configurations shown in Fig. 6.8(a)-(d) is given in Fig. 6.12. The SD+2.14 configuration shown there will be discussed in details in Sec. 6.4. As can been seen the SD+2 and GD+2 are more stable than the SD configuration in the very Ga-rich conditions.

The aforementioned configurations do not directly correspond to the experimentally observed ”4 × 5” surface reconstruction, since they all have a 1 × 1 symmetry. Nevertheless, the relative stability of the SD+2 and GD+2 models under very Ga-rich conditions, provides strong evidence that metallic surfaces are thermodynamically allowed and favored in the Ga-rich growth regime.

---

[3] These results have been published in Ref. [185], along with the first-principle calculations presented in Sec. 6.4.

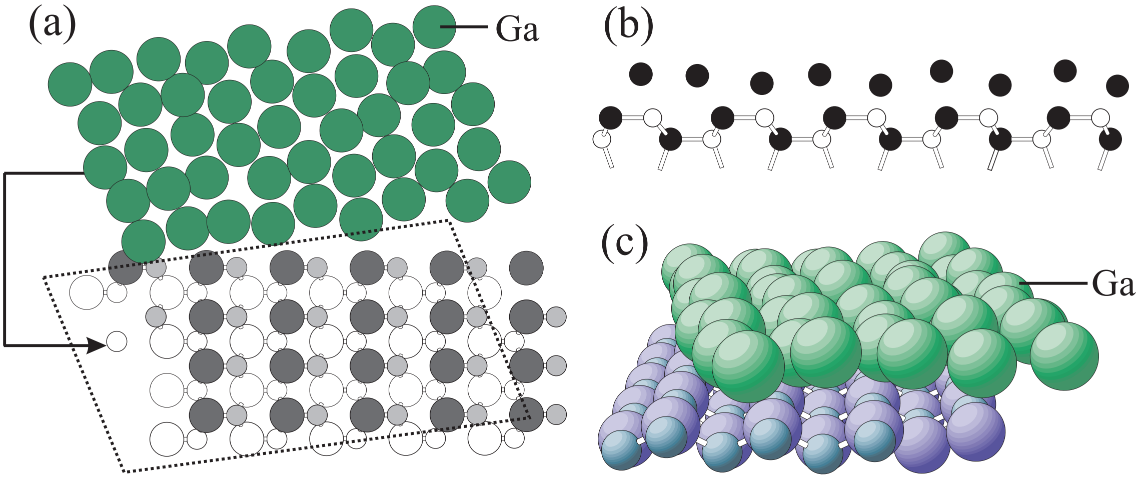


Figure 6.9: Schematic representation of the reconstructed and compressed Ga adlayer GaN$(1\bar{1}00)$ surface: (a) top view with the "4 × 5" cell shown by the dashed line. (b) Side view, and (c) perspective view. In (a) large black or small grey filled circles indicate the Ga or N atoms, respectively, in the top surface layer. In (b) black (white) balls indicate the Ga (N) atoms. In (c) the large blue balls mark the Ga atoms and the smaller grey balls the N atoms.

6.4 The contracted "4 × 5" model.

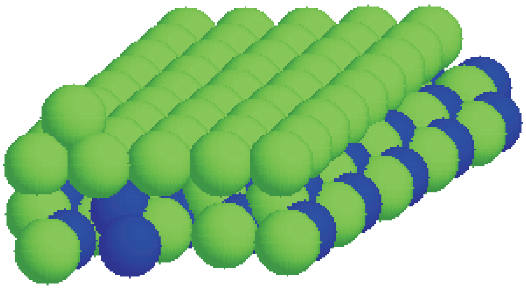


Figure 6.10: Perspective view of the initial geometry used to model the SD+2.14 model. A single Ga adatom is adsorbed in order to destroy the the initial high symmetry of the surface which after performing a few atomic steps is removed.

The Ga-Ga spacing in the adlayers of the SD+2 and GD+2 configurations mentioned in Sec. 6.3 is slightly larger than in Ga bulk. Thus, it is expected that an additional lowering of the surface energy can be achieved by compressing the Ga atoms in the top layer such that their density approaches something close to that of Ga bulk. To set up the new model with higher Ga adatom density we used the structural information from the STM data regarding size, symmetry and orientation of the surface unit cell (see Sec. 6.2 and Fig. 6.2). Specifically, models have been considered that have the experimentally observed "4 × 5" unit cell rotated by 7° away from the [0001] direction (see Fig. 6.4). The relaxed structure is shown in Fig. 6.9. It consists of the stoichiometric m-plane surface containing 21 Ga-N dimers. On top of this surface a compressed Ga adlayer consisting of 45 atoms is placed (see Fig. 6.9(a)). The adlayer corresponds to a Ga coverage of $45/21 = 2.14$ ML. Thus, hereafter this structure will be referred as SD+2.14 model.

A direct calculation of the SD+2.14 model using the same settings as for the SD+2 and GD+2 models (e.g. calculations within the LDA with 60 Ry plane wave cutoff energy and a 4 layers thick slab that makes in total 381 atoms) turned out to be computationally too expensive. In order to derive a set of parameters that minimizes the computational effort but still allows to obtain accurate surface geometries and energies, careful convergence checks have been performed. It turned out that accurate surface energies of the SD and SD+2 configurations can be calculated within the PBE-GGA and using Troullier-Martins pseudopotentials where the 3*d* orbitals are treated in the non-linear core correction (nlcc). A plane wave cutoff energy of 30 Ry was used and surfaces were modelled using slab geometries with 2 and 4 GaN layers thickness.

layers	SD	SD+2
2	0.090	0.086
4	0.107	0.081

Table 6.1: Convergence of the surface energy with respect to slab thickness for the stoichiometric (SD) and Ga adlayer (SD+2) GaN$(1\bar{1}00)$ surface. All energies are in eV/Å. The energies for the SD+2 surface are for the Ga-rich limit.

The convergence of the surface energy with respect to the slab thickness is shown in Table 6.1. We find that the surface energy of the stoichiometric SD surface significantly increases by more than 0.018 eV/Å^2 when going from the two to the four layers slab. The corresponding energy difference for the SD+2 structure is less than 0.005 eV/Å^2 while the surface geometry is described with an accuracy better than 0.03 Å. Hence, the use of slabs consisting of only two GaN layers in combination with the use of planewaves with 30 Ry cutoff energy within the PBE-GGA treating the Ga 3*d* states in the nlcc, is more than enough to accurately describe energetics and geometry of the Ga adlayer configurations in the Ga-rich limit.

The question that arises here is, why a less thicker slab is enough to describe the metallic SD+2 surface better than the stoichiometric SD surface? As has been mentioned in Sec. 6.3 the relaxation mechanism of the non-polar GaN surfaces consists of a charge transfer from the Ga to the N dangling bond, re-hybridization, and eventually a strong contraction of the bonds between the surface Ga and N atoms. In the stoichiometric SD surface the presence of N dangling bonds causes considerable strain in the lattice which decays only slowly with increasing slab thickness. On the other hand, for the SD+2 surface the presence of the Ga adlayer prevents the formation of N dangling bonds. Thus, the N atoms stay close to their bulk positions and the strain effects in the GaN layer turn out to be negligible.

Using the aforementioned parameter set, a geometry optimization of the SD+2.14 surface has been performed. A single Ga adatom has been adsorbed on the surface in order to destroy the initially high symmetry of the surface and help the system to overcome any possible local minima of the surface energy (see Fig. 6.10). After a few atomic steps the adatom has been removed. The resulting equilibrium surface is shown in Figs. 6.9(a-c). As can been seen the Ga adlayer buckles and forms a low symmetry structure. This can be further corroborated by the distributions shown in Figs. 6.11(a & b). In Fig. 6.11(a) the distribution of the in plane Ga-Ga bond lengths of the first four neighboring shells in the adlayer structure and in Fig. 6.11(b) the distribution of the heights of the Ga atoms in the adlayer are shown. A rather wide distribution is observed in both cases. The large buckling revealed from the height distribution (≈ 0.8 Å) is related to the fact that the

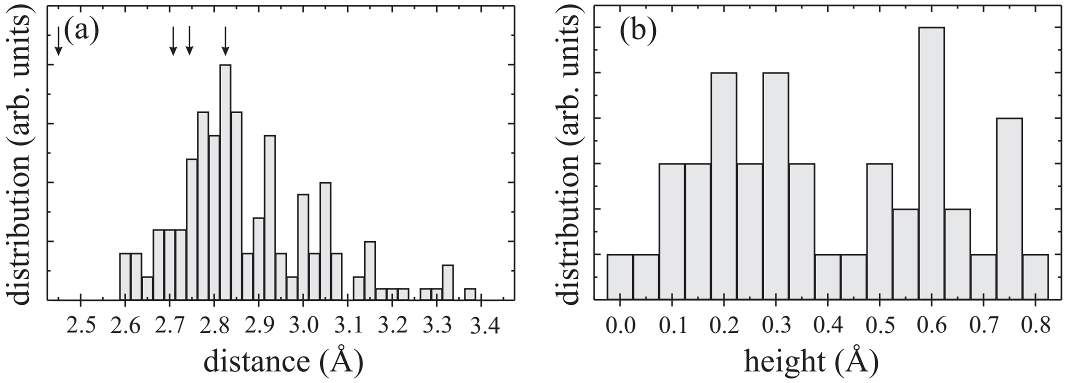


Figure 6.11: (a) Distribution of the in plane bond lengths for the first four neighboring shells of the Ga adlayer. The arrows indicate the bond lengths of the first four neighboring shells in bulk α-Ga. (b) Height distribution for the atoms in the Ga adlayer. The lower atom of the adlayer is taken as the reference height.

Ga atoms in the Ga adlayer are out of registry to the underlying Ga-N surface.

The relative formation energy of the SD+2.14 configuration along with the other configurations is shown in Fig. 6.12. As can be seen the SD+2 and SD+2.14 configurations almost degenerate in the Ga-rich limit. Initially it was assumed that by compressing the Ga adlayer such that Ga density becomes closer to that of Ga bulk will further decrease the surface energy. Indeed, the compression of the Ga adlayer is expected to strengthen the Ga-Ga bonds. On the other hand, the out of registry of the adlayer to the underlying Ga-N surface, weakens the Ga-N bonds. The total energy calculations show that the two competing effects almost cancel each other.

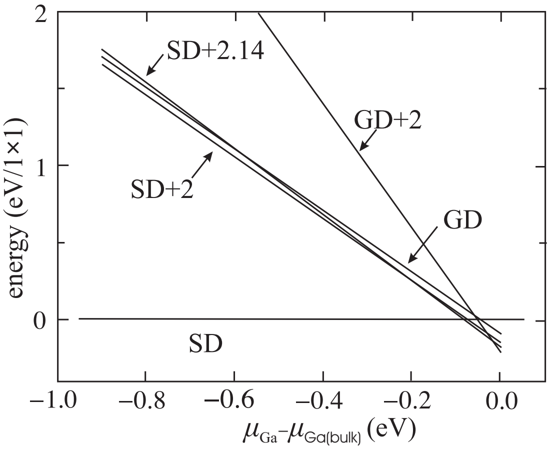


Figure 6.12: Relative formation energies of the various structural model as indicated. In the Ga rich limit the formation energy of the SD+2.14 configuration almost degenerate with the SD+2 configuration.

As can been seen in Fig. 6.11(a) the Ga adlayer structure can be further compressed in order to bring the Ga-Ga bond lengths closer to those of bulk Ga and subsequently to enhance the bond strength. Hence, it can be considered that a different density and/or different arrangement of the adlayer may produce a reduction in the energy. In any case, judging from the experimental STM images (see Fig. 6.2), the arrangement of the adlayer must indeed rather complicated. The exact arrangement of the "4 × 5" reconstruction cannot be fully determined by the model presented here and further work is required. However the SD+2.14 model clearly shows that structures exceeding 2 MLs of Ga are feasible and thermodynamically stable under extreme Ga-rich conditions.

6.5 Summary

In this chapter the possibility of growing GaN films free of polarization effects has been considered. Such Ga-rich surfaces should be characterized by a higher sticking coefficient for the N adatoms and a higher mobility for the Ga adatoms under metal-rich conditions. In analogy to the Ga-rich polar GaN surfaces, these effects are expected to result in two dimensional growth and smoother surface morphologies. Thus, it would be advantageous to employ Ga-rich surfaces as an intermediate step in atomic layer epitaxy. The existence of $\left(1\bar{1}01\right)$ facets revealed by recent RHEED and STM experiments imposes a potential problem for the growth of high quality $\left(1\bar{1}00\right)$ films. While the $\left(1\bar{1}01\right)$ surfaces have been studied recently employing first-principles calculations [43], no information regarding the relative stability between these and the m-plane surfaces, can be derived. Thus, further work is required in order to make a step forward to better mastering of growth of non-polar GaN surfaces.

Chapter 7

kinetically stabilized ordering in $\mathrm{Al_xGa_{1-x}N}$ alloys

Alloys of group III-Nitrides have been under intensive investigations and are viewed as the materials system of choice for optoelectronic devices in the blue/ultraviolet region of the optical spectra. In this chapter first a brief review of recent studies on growth and properties of ternary group III-Nitrides will be given. Next, experiments which provide evidence for ordering and which give insight view into the effect of ordering on the electronic structure will be reported. Finally, in Secs. 7.4 and 7.5, results of *ab-initio* calculations of ordered and disordered $\mathrm{Al_{0.5}Ga_{0.5}N}$ alloys will be discussed.

The work presented in this chapter is based on a joined experimental and theoretical investigation of ordering in $\mathrm{Al_xGa_{1-x}N}$ alloys. The theoretical part addresses energetics and electronic properties but also possible kinetic pathways to create ordered $\mathrm{Al_xGa_{1-x}N}$. Here we will focus on ordered and disordered phases of $\mathrm{Al_xGa_{1-x}N}$ ($\mathrm{x} = \frac{1}{2}$) alloys. The results have been published in Ref. [199].

7.1 Ordering in III-Nitride alloys

In order to introduce the terms of order and disorder, let us first assume the occupation statistics of the metal atoms (Al and Ga) in the metal sublattice of the wurtzite $\mathrm{Al_xGa_{1-x}N}$ alloy. In the state of complete disorder all lattice points are occupied at random by either of the two kinds of atoms. The X-rays analysis of the crystal will reveal a spectra as if all the points of the sublattice were occupied by the same average manner. On the other hand, if on set of points are occupied predominantly by Al and the other set by Ga atoms the state of the alloy would be identified as ordered or partially ordered. Since the X-rays distinguish between the two sets of points new lines will appear in the spectra (superstructure or superlattice lines) [146, 200]. Since the X-rays take averages over rather large numbers of atoms, the order state revealed by X-rays experiments should

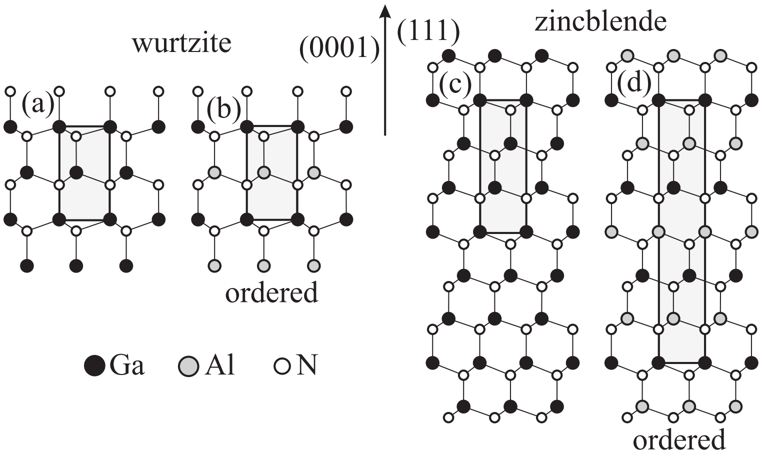


Figure 7.1: Side view of binary GaN ((a) and (c)) and ordered ternary $Al_{0.5}Ga_{0.5}N$ ((b) and (d)) alloys in wurtzite ((a) and (b)) and zincblende ((c) and (d)) phases. The ordered phases correspond to 1:1 superlattices along the [0001] and [111] directions in the wurtzite (b) and zincblende (d) crystals respectively. The rectangles indicate the unit cells. In wurtzite structure ordering lowers the crystal symmetry but conserves the size of the unit cell. In zincblende structure ordering doubles the size of the unit cell.

be defined as *long range order*. This should be confronted with *short range order*. In the absence of long range order, short range order appears when atoms of the one species are preferentially surrounded by atoms of the other species and vice versa for the first nearest neighbors.

Let us assume the case of the 1:1 ACBC superlattice[1]. If r_{A} is the probability to find an atom of type A on an A site, the degree of long range order (which will refer to as order for shortness) is quantified by the order parameter $S = 2r_{\mathrm{A}} - 1$ [201]. The values $S = \pm 1$ correspond to perfect order states. Intermediate values of S correspond to intermediate degrees of order, while the value of $S = 0$ corresponds to the state of complete disorder.

The first atomic long range ordering in a GaN-AlN pseudobinary system was reported by Korakakis *et al.* [58] in films grown by MBE. They found atomic ordering to occur along the [0001] direction (which is the direction of growth). The superlattice structure observed in that study consists of alternating Ga-rich and Al-rich monolayers (ML). The degree of ordering was found to depend on the Al content with maximum ordering to occur at 50% AlN mole fraction that corresponds to a 1:1 ordered phase. The ordering was found to reduce the lattice symmetry from $P6_3mc$ (wurtzite structure) to P3mc but to conserve the size of the unit cell (see Fig. 7.1). The two places of the metal sublattice in the wurtzite unit cell are occupied by one Ga and one Al atom. The two places of the group-V sublattice are occupied by two N atoms. For a more detailed discussion of the wurtzite structure see Appendix A.1 on page 137.

In a following work Iliopoulos *et al.* [59] reported superlattices with 7 and 12 ML periodicity along the growth direction ([0001]) in films deposited under different growth conditions (see Fig (7.2)). In contrast to the previously reported 1:1 long range ordering, those structures do no consist of a simple sequence of AlN and GaN layers but they involve mixed AlN-GaN monolayers. Recently Benamara *et al.* [62] found in addition to the 1:1 ordering, a pyramidal-plane ordering that

---

[1] For the case of $Al_xGa_{1-x}N$ alloys discussed here, the stacking $i : j$ represents a periodicity of i GaN layers followed by j AlN layers along the [0001] direction.

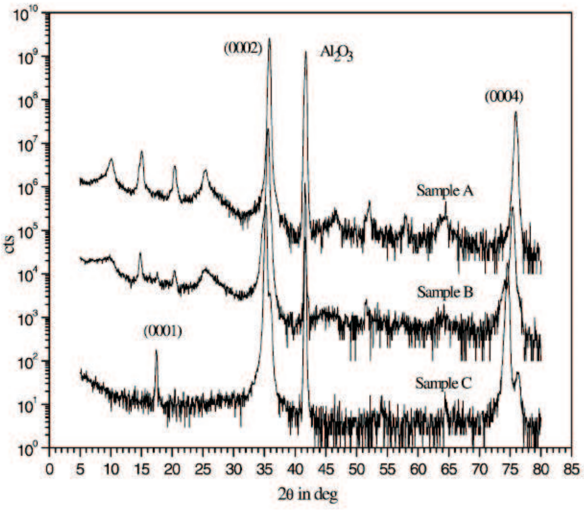


Figure 7.2: On-axis XRD spectra of three AlGaN films deposited under different III/V ratio during growth. Additional to the characteristic of the wurtzite structure (0002) and (0004) peaks, peaks forbidden for a random alloy appear. The (0001) peak in the spectra of Sample C (50% Al content) corresponds to 1:1 superlattice along the [0001] direction. For the other samples (higher Al contents) the peaks corresponding to superlattices with 7 and 12 ML periodicity. After Iliopoulos *et al.* [59].

consists of alternate stacking planes of GaN and AlN layers along the $[1\bar{1}01]$ direction in AlGaN samples grown by MBE.

Evidence of ordering in AlGaN films deposited by MOCVD has been reported by Ruterana *et al.* [60]. They observed the coexistence of 1:1, 3:1, and 10:2 ordered phases. The 10:2 ordering corresponds to a series of 5 GaN cells and one AlN along the [0001] direction (i.e. 10 GaN ML and 2 AlN ML). The 3:1 ordered phase corresponds to a superlattice along the [0001] direction with successive occupation of 3 Ga atoms and one Al atom of the metal sublattice. A superlattice with 3:1 periodicity has been observed also in InGaN films grown by MOCVD by Behbehani *et al.* [202]. The first results of ordering in $In_xGa_{1-x}N$ grown by MOCVD on (0001) sapphire has been reported by Ruterana *et al.* [203].

The phenomenon of ordering in group III-Nitride ternaries has been explained in terms of surface energetics and surface kinetics. Jenny *et al.* [204] studied the influence of Al over the Ga desorption kinetics in AlGaN alloys grown by gas source MBE. They found Al to preferentially incorporate into the AlGaN film. Two mechanisms were proposed to explain the preferential incorporation of Al. The fist mechanism is based on the higher cracking efficiency of ammonia by Al. The second mechanism is based on the observation that if ammonia flux is not sufficient to incorporate both Al and Ga, Al participates in an Al-for-Ga exchange that kicks out previously incorporated Ga.

Similar observations have been reported by Iliopoulos *et al.* [59] for AlGaN films grown by MBE with a constant ratio of Al to Ga fluxes. The growth kinetics was found to depend strongly on the ratio of the III/V fluxes. Under metal rich conditions Al preferentially incorporates into the film. On the other hand, under group V rich conditions, where there is an abundance of reactive N at the growing surface, both Al and Ga atoms are completely adsorbed at the surface and incorporate into bulk. This was explained in terms of local energy balance at the growing surface: Since the Al-N bonds are stronger that the Ga-N bonds (e.g., the formation energy of AlN is found both experimentally and theoretically to be greater than the formation energy of GaN - see Ref. [82]

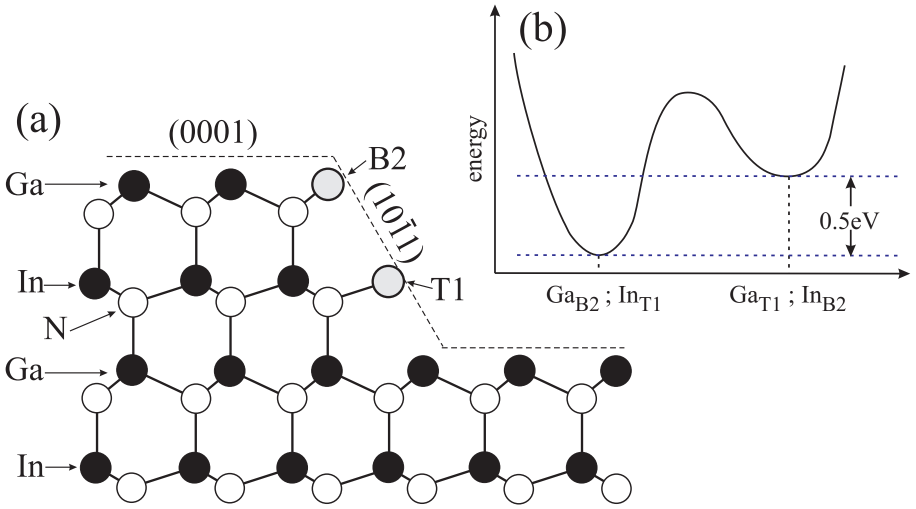


Figure 7.3: (a) Ball and stick model of a bilayer step of height on (0001) surface. At such a step metal atoms can attach at B2 (T1) sites and form two (one) bonds with the N atoms in the layer beneath. (b) Schematic representation of the energy required to exchange a Ga atom attached at B2 site ($\mathrm{Ga_{B2}}$) with an In atom at T1 site ($\mathrm{In_{T1}}$). The energy difference of 0.5 eV was calculated by Northrup *et al* [37] with ab-initio calculations. This model provides a machanism that explains ordering in InGaN alloys. 1:1 superlattice along the c-axis is formed as Ga atoms preferentially incorporate at B2 sites, In atoms at T1 sites and the steps flow across the surface.

and Refs. therein) the cohesive energy), under metal rich conditions the available active nitrogen preferentially bonds with Al, leading to a higher Al sticking coefficient. Under N rich conditions the sticking coefficient of Ga increases and leads to films with 50% AlN mole fraction.

Benamara *et al.* [62] explained the pyramidal-plane ordering in AlGaN alloys using a model initially proposed by Northrup *et al.* [37] in order to explain chemical ordering in InGaN. According to this model which is based on DFT calculations, it was shown that 1:1 ordering along the c-axis is thermodynamically unstable with respect to mixing of In and Ga within a given (0001) plane. Within their ab-initio based model, growth takes place at step edges or small $(10\bar{1}1)$ facets (see Fig. (7.3)). A bilayer step of height c has the same local atomic structure as a $(10\bar{1}1)$ microfacet. Since the In-Ga bond is longer and weaker than the Ga-N bond, Ga atoms would preferentially incorporate at B2 sites in Fig (7.3), where two N bonds can be saturated. Exchanging a Ga atom from a B2 site with an In atom at T1 site requires 0.5 eV. This energy is compared to the thermal energy($k_b\mathrm{T}_{\mathrm{growth}} \approx 0.09\,eV$ at a typical growth temperature of $\mathrm{T} = 1000$ K). If the sufficient surface mobility for In atoms is sufficient to sample the two inequivalent step types (one type of step edge offers only B2 sites, the other only T1 sites for the metal atoms), the same ordering mechanisms can be applied for single steps of height c/2. In the case of AlGaN alloys, the Al-N bonds are stronger than the Ga-N bonds. Thus, it is expected that Al atoms will preferentially incorporate at the higher coordinated B2 sites. DFT calculations for the $(1\bar{1}01)$ surface that confirm the above model for $\mathrm{Al_xGa_{1-x}N}$ have been reported in a recent publication by M. Albrecht *et al.* [199] along

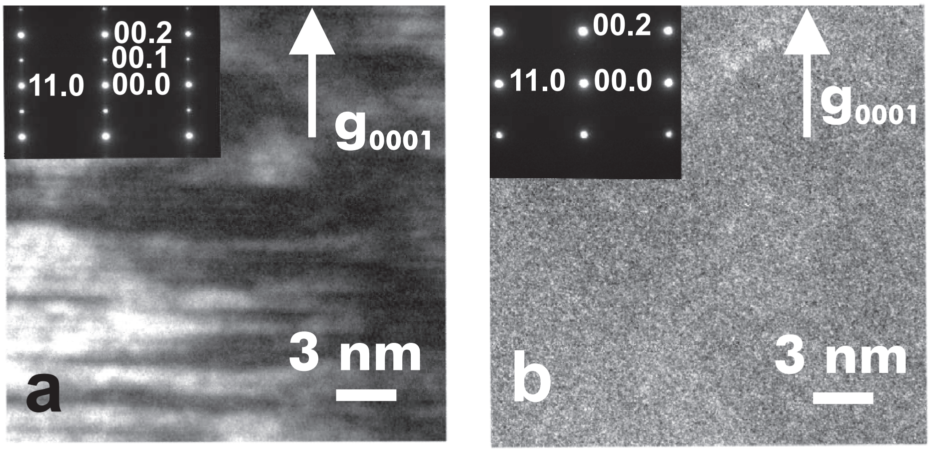


Figure 7.4: Cross-sectional dark field images using the order sensitive [0001] reflection and transmission electron diffraction patterns taken along the [$1\bar{1}00$] zone axis (insets) of an Al$_x$Ga$_{1-x}$N sample (x = 0.58). (a) As grown sample. In the electron diffractogramm [0001] superlattice reflections can been seen. Order domains (bright regions) in the dark filed image are present. (b) Annealed sample. No order is present either in the electron diffractogramm or in the TEM image. After Albrecht *et al.* [199].

with the results presented here.

A number of experimental works that deal with the optical properties of Al$_x$Ga$_{1-x}$N alloys have been published over the past years [205, 206, 207, 208, 209]. A correlation of the bandgap bowing of the Al$_x$Ga$_{1-x}$N alloys and the growth method used, based on bandgap measurements as reported in 20 previous studies, has been reported by Lee *et al.* [205]. They propose that the intrinsic bandgap bowing parameter of unstrained Al$_x$Ga$_{1-x}$N alloys should be $b = +0.62\,(\pm 0.45)$ eV. Misra *et al.* [206] speculated that in analogy with III-V cubic ordered alloys, ordering should reduce the size of the bandgap also in Al$_x$Ga$_{1-x}$N alloys, but experimental data supporting this argument is missing. Recently Dudiy and Zunger [36], based on DFT-LDA calculations, found that ordering reduces ther bandgap by 90 meV, i.e. leads to a redshift in the bandgap of Al$_{0.5}$Ga$_{0.5}$N.

So far it has been shown that the existence of ordering in AlGaN alloys is well established. Also, a number of works concerning the optical characterization of AlGaN alloys have been published. However, studies that link thermodynamics, growth kinetics and optical properties of Al$_x$Ga$_{1-x}$N alloys are missing. In the following Section the experimental evidence of ordering in AlGaN and the influence on the optical properties of the grown material will be briefly discussed.

7.2 Experimental evidences of ordering in Al$_x$Ga$_{1-x}$N alloys.

First, the experimental part of the joint experimental-theoretical study as published by Albrecht *et al.* [199] will be discussed. The experiments have been performed on 1 μm thick Al$_x$Ga$_{1-x}$N layers grown on sapphire (0001) substrates by radio frequency plasma enhanced MBE. The characteri-

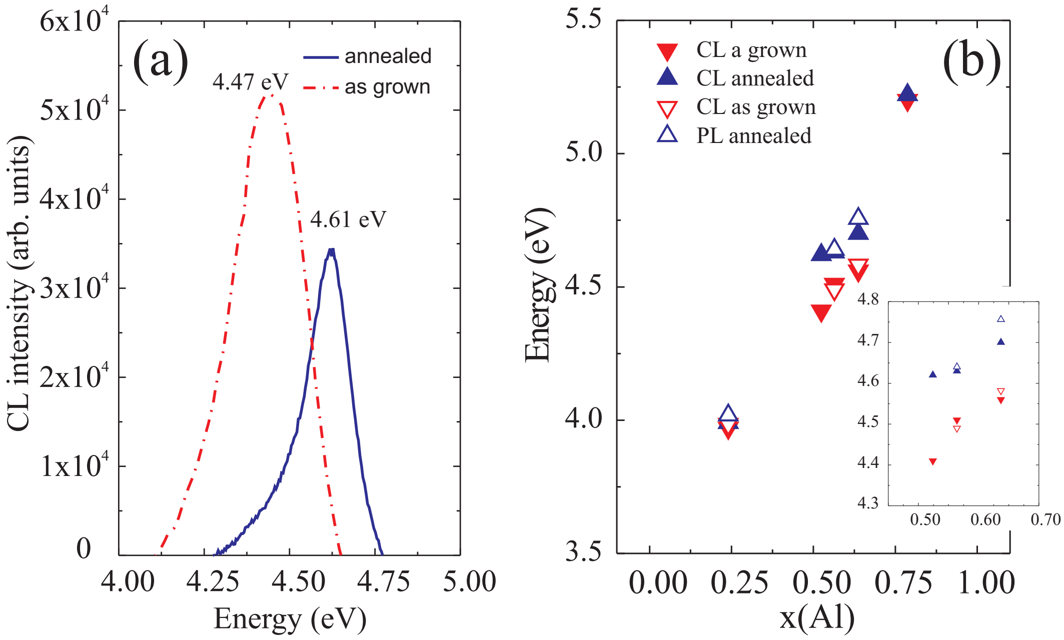


Figure 7.5: (a) TEM-CL spectra taken at 90 K of ordered (as grown) and disordered (annealed) samples with x=0.58 under identical excitation conditions. The CL peak of the ordered sample shows 140 meV redshift, accompanied by increase in the FWHM and in the intensity of the peak. (b) Summary of the PL and CL peak positions of the ordered and disordered samples throughout the whole range of Al concentrations. The ordered induced redshift is maximum for AlN mole fractions around 50% (inset). After Albrecht *et al.* [199].

zation of the samples has been performed by X-ray diffraction, Rutherford backscattering (RBS), and cathodoluminescence (CL) and photoluminescence (PL) spectroscopy.

The presence of ordering in the samples is revealed by the presence of the order sensitive (0001) spots in the electron diffraction pattern of the as grown samples (inset of Fig. (7.4a). Ordering is additionally revealed by the dark field images along the [0001] direction (Fig. (7.4)): the bright intensity regions are indicative of highly ordered domains. The ordered domains are of 100 nm to 250 nm lateral extension and of 3 nm to 10 nm thickness. The presence of order was found in the samples through the whole compositional range (x=0.240, 0.524, 0.564, 0.637, 0.787). For AlN mole fraction close to 50% the maximum value of the degree of ordering S was found to be 0.35 and corresponds to an average over ordered and disordered regions, due to the limited resolution of the diffraction method (100 nm in diameter by 100 nm in thickness).

The thermodynamic stability of ordering was examined by annealing at high temperature (for 5h at 1350 °C) and high pressure (nitrogen pressure of 9.5 kbar). In the electron diffraction patterns of the annealed samples the (0001) spots disappeared and no indication of long-range ordering can be found in electron diffraction data (see Fig. (7.4)). Additionally, X-ray data showed no indication of ordering, while the Al concentration of as grown and annealed samples was found to be the same. The disappearance of ordering upon annealing was found in all samples. Thus, *ordering is not* *thermodynamically stable for* $\mathrm{Al_xGa_{1-x}N}$ *alloys.*

Fig. (7.5a) shows the CL spectra of ordered (as grown) and disordered (annealed) samples with x=0.56 under identical excitation conditions. The position of the as grown peak is redshifted by 140 meV with respect to the annealed sample. The shift is accompanied by an increased FWHM and an increased peak intensity. Fig. (7.5b) summarizes the PL and CL peak positions of the the ordered and disordered samples throughout the whole range of Al concentrations. In all cases *ordering causes a decrease in the bandgap.* This decrease is most pronounced for samples with AlN mole fraction of around 50%.

So far evidence and properties of ordered $\mathrm{Al_xGa_{1-x}N}$ alloys have been based on experimental studies. In the next section the following issues will be addressed:

- Confirm the thermodynamic stability of the disordered configuration and the order induced redshift on a first theoretical basis.
- Identify the origin of the bandgap reduction as caused by ordering.
- Identify the mechanism which stabilizes the thermodynamically unstable alloy.

7.3 Modelling of random alloys

The first question which has to be addressed when modelling alloys, is how to overcome the problem of modelling a random alloy using finite size supercells with periodic boundaries.

Early electronic structure theories described semiconductor $\mathrm{A_xB_{1-x}}$ alloys in terms of weak, symmetry-preserving perturbations around the end-point constituents [210]. Examples are the virtual crystal approximation (VCA) and the site-coherent potential approximation (SCPA). In the VCA the alloy is assumed to have a single average type of atom $< AB >$, while in SCPA all A's and separately all B's are assumed to be equivalent and each is embedded in a structureless uniform medium. In both methods atomic relaxations are excluded and are said to be *non-structural*, since they consider only the average occupation of a site (i.e. how much A or B character does the site have) but have no information associated with the geometrical arrangement around the sites.

The validity of non-structural approaches is limited to those properties that do not depend but weakly on the local environment of atoms. For example, in tetrahedral pseudobinary alloys $\mathrm{(AC)_{1-x}(BC)_x} = \mathrm{A_{1-x}B_xC}$ each C is assumed to have four identical medium nearest neighbors. Since this configuration can be denoted as $\mathrm{CX_4}$, the point symmetry around C is T_d. Thus, lower point symmetries arising from different possible nearest neighbor configurations, like $\mathrm{CA_3B}$ and $\mathrm{CAB_3}$ (C_{3v} symmetry) as well as $\mathrm{CA_2B_2}$ (D_{2d} symmetry), do not contribute to the property in question. Thus, any effect in the properties due to symmetry reductions is excluded. For example, positional relaxation of the centered atom C and A-B charge transfer are not allowed due to symmetry restrictions. Both, experimental techniques capable of probing the average local properties of alloys and theoretical calculations give a clear evidence of the importance of the local-atomic arrangements (see Ref. [210] and references therein).

On the other hand *structural* techniques explicitly specify the alloy structure and can incorporate atomic relaxation. The difficulty with these theories arises from the rather large computational cost that is required to study a random alloy. For a binary (or pseudobinary) alloy with N sites, there are 2^{N} possible atomic configurations denoted as σ. A measurable property of the random alloy, as for example the total energy, is the ensemble average over all the 2^{N} possible configurations σ:

$$\langle E \rangle = \sum_{\sigma}^{2^{\mathrm{N}}} \rho\left(\sigma\right) E\left(\sigma\right), \tag{7.1}$$

where $\rho\left(\sigma\right)$ denotes the probability to find the configuration σ in an ensemble of systems. Each of these configurations needs to be structurally relaxed and then averaged in order to obtain the total energy of the system. In practice, a smaller number of representative configurations is chosen (Monte Carlo techniques [211]), or a single supercell with a random occupation of the sites by the A and B atoms [211, 212]. The main obstacles with these approaches are the large number of configuration that one has to take into account ($\sim 10^6$ in Monte Carlo studies) or the large supercell sizes ($\sim 10^3$ atoms) that one has to use.

Zunger *et al.* proposed a technique to model random alloys, using supercell that have sizes capable for *ab-initio* calculations ($\sim$ 100 atoms) [210, 213]. This technique is based on the concept of *special quasirandom structures* (SQS's) and is briefly described in Sec. (7.3.1).

7.3.1 Special Quasirandom Structure

Within the concept of special quasirandom structures, one attempts to mimic an infinite random alloy, by selectively occupying the finite N sites of a supercell by A and B atoms. The usage of periodic supercells will introduce spurious correlations beyond a certain distance. However, many physical properties, as for example the total energy of the system, depend mainly on the local environment around the atoms (see Ref. [210] and references therein). Thus the construction of SQS's is driven by the principle of close reproduction of the structural correlation functions (see below) of the random alloy for the first few shells around a given site. Any possible periodicity errors are put over to more distant neighbors. The characterization of the SQS's relies on the *structural correlation functions*. A detailed analysis can be found in Refs [210, 213], here only a brief description will be given.

A given configuration σ can be discretized in terms of its component *figures* f. A figure is defined by the number of atoms located on its vertices, and the order of neighbor distances separating them. Thus, a figure can be denoted as: $f = \{k, m\}$, where k is the number of atoms (or number of vertices) and m the order of neighbor distances. For example $f = \{2, 1\}$ corresponds to pairs separated by 1st neighbor distance. Following the terminology of Ising models each site

Table 7.1: Correlation functions of N-atom fcc SQS's at x = $\frac{1}{2}$. In brackets next to $\overline{\Pi}_{\{k,m\}}$ is the degeneracy factor $D_{\{k,m\}}$. For the corresponding perfectly random alloy all the correlation functions vanishes. After Wei *et al.* [210].

	$\overline{\Pi}_{\{2,1\}}$ [6]	$\overline{\Pi}_{\{2,2\}}$ [3]	$\overline{\Pi}_{\{2,3\}}$ [12]	$\overline{\Pi}_{\{2,4\}}$ [6]	$\overline{\Pi}_{\{2,5\}}$ [12]	$\overline{\Pi}_{\{2,6\}}$ [4]	$\overline{\Pi}_{\{4,1\}}$ [2]	$\overline{\Pi}_{\{4,2\}}$ [12]
SQS-8	0	0	$\frac{1}{24}$	$-\frac{1}{12}$	$\frac{1}{12}$	0	0	$-\frac{1}{6}$
CuPt	0	-1	0	1	0	-1	-1	0

i of the figure is assigned by a spin variable $\widehat{S}_i$ which takes the value -1 if the site is occupied by an atom A, or +1 if it is occupied by an atom B. A figure $\{k,m\}$ located at position $\boldsymbol{l}$ in the configuration σ is characterized by the product of the spin variables $\widehat{S}_i$:

$$\Pi_{\{k,m\}}\left(\boldsymbol{l},\sigma\right)=\prod\hat{S}_i. \tag{7.2}$$

Then the lattice average over all locations $\boldsymbol{l}$ of a figure $f=\{k,m\}$ is given as:

$$\overline{\Pi}_{\{k,m\}}\left(\sigma\right)=\frac{1}{ND_{\{k,m\}}}\sum_{\boldsymbol{l}}\Pi_{\{k,m\}}\left(\boldsymbol{l},\sigma\right), \tag{7.3}$$

where N is the number of sites, and $D_{\{k,m\}}$ is the number of equivalent figures per site. The value of a physical property E for the configuration σ is given by:

$$E\left(\sigma\right)=N\sum_{\{k,m\}}D_{\{k,m\}}\overline{\Pi}_{\{k,m\}}\varepsilon_{\{k,m\}}, \tag{7.4}$$

where $\varepsilon_{\{k,m\}}\left(\boldsymbol{l}\right)$ is the contribution of the figure $f=\{k,m\}$ located at $\boldsymbol{l}$, to the property $E\left(\sigma\right)$. The ensemble average of the property E is

$$\langle E\rangle=N\sum_{\{k,m\}}D_{\{k,m\}}\langle\overline{\Pi}_{\{k,m\}}\rangle\varepsilon_{\{k,m\}}, \tag{7.5}$$

where $\langle\overline{\Pi}_{\{k,m\}}\rangle$ is the ensemble average of the correlation function $\overline{\Pi}_{\{k,m\}}$. For a perfectly random infinite binary alloy $\mathrm{A_xB_{1-x}}$ the ensemble average of its correlation function can be proven to be [210]:

$$\langle\overline{\Pi}_{\{k,m\}}\rangle_{\mathrm{R}}=\left(2\mathrm{x}-1\right)^k. \tag{7.6}$$

For $x=\frac{1}{2}$ all the ensemble averages of the correlation functions vanish to all orders, except for $\{k,m\}=\{0,1\}$, that can be proven to be $\langle\overline{\Pi}_{\{0,1\}}\rangle_{\mathrm{R}}=1$. In a perfectly random alloy, a given site (occupied, for example, by A) has an average $\langle O_m\rangle_{\mathrm{R}}$ mth order neighbors of opposite (i.e., B), that can be proven to be

$$\langle O_m \rangle_{\mathrm{R}} = D_m \pm \sqrt{\frac{D_m}{2}}. \tag{7.7}$$

Here $D_m = Z_m/2$ and Z_m is the number of mth order neighbors for a given site.

Within the SQS's method one constructs special N atom periodic supercells, whose correlation functions $\overline{\Pi}_{\{k,m\}}(\sigma)$ best match the corresponding functions of the perfectly random infinite alloy $\langle \overline{\Pi}_{\{k,m\}} \rangle_{\mathrm{R}}$. The difference of the ensemble average of the property E and its value of the special supercell having the configuration σ is

$$\langle E \rangle - E(\sigma) = \sum_{\{k,m\},\, k \neq 0,1} D_{\{k,m\}} \left((2\mathrm{x}-1)^k - \overline{\Pi}_{\{k,m\}}(\sigma) \right) \varepsilon_{\{k,m\}} \tag{7.8}$$

The contributions $\varepsilon_{\{k,m\}}$ to the property E are expected to decrease with the size of the figure. Thus, the configuration σ (or selective occupation by A and B atoms) of the SQS can be chosen so that Eq. (7.8) is minimized in a hierarchical manner. In Table 7.1 the correlation functions of an 8 atom fcc SQS proposed by Wei *et al.* [210] (SQS-8) and the CuPt structure, are given as an example. The coordinates of an $\mathrm{A_xB_{1-x}C}$ alloy with the SQS-8 structure are given in Appendix B (p. 147). It is worth to mention that here the SQS-8 structure corresponds to a $\mathrm{A_2B_3A_2B_1}$ superlattice along [113]. Nevertheless, the first few correlation functions of the SQS-8 configuration coincide with those of the perfectly random alloy.

The wurtzite structure differs from the zincblende in that not all of the four tetrahedral bonds are equivalent. The bond along the c-axis is inequivalent to the other three bonds, since the $\langle 0001 \rangle$ direction is a polar axis and inversion symmetry is lacking[2]. The question that arises is whether for the calculation of the correlation functions the four tetrahedral bonds will be assumed equivalent, or the bond along the c-axis will be considered to be inequivalent. Therefor, three different SQS's have been used to study the random $\mathrm{Al_xGa_{1-x}N}$ ($\mathrm{x} = \frac{1}{2}$) alloys which will be abbreviated as SQS1, SQS2, and SQS3. These structures have been constructed in such a way that *they mimic different nitrogen local environments*. The unit vectors, atomic positions and ball and stick models of the corresponding structures are given in Appendix B. The SQS1 and SQS2 utilize 2 × 2 × 1 supercells and the SQS3 a 2 × 1 × 2 supercell.

Table 7.2 gives, for up to the 6th order pairs, the correlation functions $\overline{\Pi}_{\{2,m\}}$, the degeneracy factors $D_{\{2,m\}}$, and the average number of neighbors of opposite sign $\langle O_m \rangle$, of the SQS1, SQS2, and SQS3. The SQS3 is a good candidate to model a random alloy for the case that the four tetrahedral bonds are assumed equivalent. It reproduces the correlation functions of the perfectly random alloy for neighboring shells up to the third order. On the other hand if the tetrahedral bond along the

---

[2] For example, in the ideal (unrelaxed) wurtzite structure the c/α ratio is $\sqrt{8/3}$ and the four tetrahedral bonds have the same length (see Appendix A.1). If volume relaxation is taken into account, the value of the ratio differs from the ideal one and the length of the bond along the c-axis differs from the other three equivalent and of equal length bonds. The most pronounced example from the current calculations comes from the case of AlN, where the ratio of the relaxed structure is 1.60, i.e., by ∼ 1% smaller.

Table 7.2: The three SQS's used to model random $\mathrm{Al_xGa_{1-x}N}$ alloys at $\mathrm{x} = \frac{1}{2}$. The upper half is dedicated to the case where the four tetrahedral bonds are assumed equivalent, while for the lower half the bond along the *c*-axis is assumed inequivalent to the other three. The correlation functions $\overline{\Pi}_{\{k,m\}}$ and the degeneracy factors $D_{\{k,m\}}$ up to the sixth order neighboring shells are given for the pair figures. For the perfectly random pseudobinary alloy with $\mathrm{x} = \frac{1}{2}$ the correlation functions vanish to all orders, except $\langle\overline{\Pi}_{\{0,1\}}\rangle_{\mathrm{R}}$. For structures shown here, the correlation functions for the triplets, $\langle\overline{\Pi}_{\{3,m\}}\rangle$, vanish. The average number of neighbors $\langle O_m\rangle$ of opposite type of the mth order shell, and the corresponding values in the perfectly random alloy, calculated by Eq. (7.7), are also given.

	$\overline{\Pi}_{\{2,1\}}$	$\overline{\Pi}_{\{2,2\}}$	$\overline{\Pi}_{\{2,3\}}$	$\overline{\Pi}_{\{2,4\}}$	$\overline{\Pi}_{\{2,5\}}$	$\overline{\Pi}_{\{2,6\}}$
THE FOUR TETRAHEDRAL BONDS ARE EQUIVALENT.						
$D_{\{k,m\}}$	12	6	2	18	12	6
SQS1	0	-1	1	0	0	1
SQS2	$-\frac{1}{3}$	1	1	$-\frac{1}{3}$	$-\frac{1}{3}$	1
SQS3	0	0	0	$-\frac{1}{9}$	0	1
Number of neighbors of opposite sign $\langle O_m\rangle$						
Random	6 ± 1.7	3 ± 1.2	1 ± 0.7	9 ± 2.1	6 ± 1.7	3 ± 1.2
SQS1	6 ± 0.0	6 ± 0.0	0	9 ± 1.9	6 ± 3.7	0
SQS2	8 ± 0.0	0	0	12 ± 0.0	8 ± 0.0	0
SQS3	6 ± 2.0	3 ± 2.3	1 ± 1.1	10 ± 2.6	6 ± 4.8	0
THE FOUR TETRAHEDRAL BONDS ARE NOT EQUIVALENT.						
$D_{\{k,m\}}$	6	6	6	2	12	6
SQS1	0	0	-1	1	0	0
SQS2	$-\frac{1}{3}$	$-\frac{1}{3}$	1	$-\frac{1}{3}$	$-\frac{1}{3}$	1
SQS3	$-\frac{1}{3}$	$\frac{1}{3}$	0	0	$-\frac{1}{3}$	$\frac{1}{3}$
Number of neighbors of opposite sign $\langle O_m\rangle$						
Random	3 ± 1.2	3 ± 1.2	3 ± 1.2	1 ± 0.7	6 ± 1.7	3 ± 1.2
SQS1	3 ± 1.9	3 ± 1.9	6 ± 0.0	0	6 ± 3.7	3 ± 1.9
SQS2	4 ± 0.0	4 ± 0.0	0	0	8 ± 0.0	4 ± 0.0
SQS3	4 ± 1.3	2 ± 2.1	3 ± 2.3	1 ± 1.1	8 ± 2.6	2 ± 2.1

c-axis is assumed to be inequivalent to the other three, the SQS1 will be a good candidate: For this SQS the correlation functions of the first two shells vanish. Additionally, it closely reproduce in the same range the average number of opposite neighbors of the perfectly random alloy.

In order to explore the electronic properties and the energetics of ordered and disordered AlGaN alloys, ab-initio calculations has been performed. In the following sections the details of the calculations and the results will be analytically presented.

7.4 Energetics of ordering.

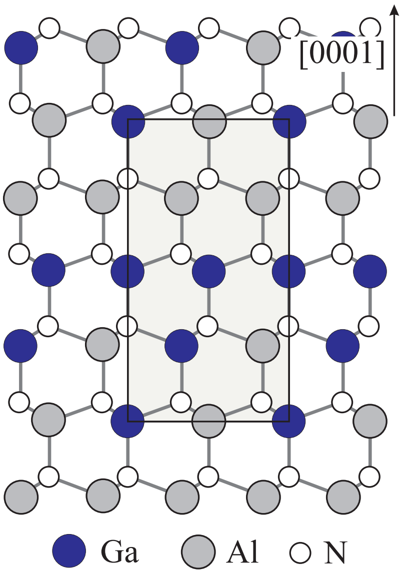


Figure 7.6: Side view of the SQS3 structure. The rectangle corresponds to the $2 \times 1 \times 2$ supercell.

In order to examine the thermodynamics of ordering in $Al_{0.5}Ga_{0.5}N$ alloys, first principle calculations on ordered and disordered configurations have been performed. The ordered structures have been described in a $1 \times 1 \times 1$ supercell where GaN and AlN layers alternate along (0001). This structure corresponds to 1:1 supercell along the c-axis (see Fig. 7.1(b)). The disordered alloys have been presented by the SQS1, SQS2, and SQS3 configurations (see Figs 7.7 and 7.6) as described in Sec. (7.3.1). The primitive vectors and the coordinates of the atoms of these structures are given in Appendix B (Eqs. (B.1), (B.2), and (B.3)).

Additional calculations have been performed for ordered and disordered $Al_{0.5}Ga_{0.5}N$ alloys having the zincblende structure. The zincblende ordered structure has been presented by a 1:1 superlattice along the [111] direction (the equivalent to the [0001] direction in the wurtzite structure). The coordinates are given in Eq. (B.4), p. 148. Unlike for the wurtzite structure, in zincblende ordering along the [111] axis cannot be modelled by a $1 \times 1 \times 1$ supercell: Due to symmetry the supercell has to be three times larger than the unit cell along the [111] direction. The disordered structures are modelled using the SQS-8 configuration (see Table B.4 and Eq. (B.5)), which has been proposed by Wei *et al.* [210] to be a good candidate for disordered binary alloys with cubic symmetry.

The calculations have been performed employing DFT in the PBE-GGA. Troullier-Martins pseudopotentials [101] have been used and a 70 Ry plane wave cutoff energy has been applied. An equivalent of $4 \times 4 \times 1$ Monkhorst-Pack k-point mesh has been used for sampling the Brillouin zone. For all structures, the shape/size of the supercell (lattice parameters) and the internal degrees of freedom (internal coordinates) have been fully relaxed. The accuracy of the electronic minimization and the structural relaxation is less than 10^{-3} eV. The formation energy of the

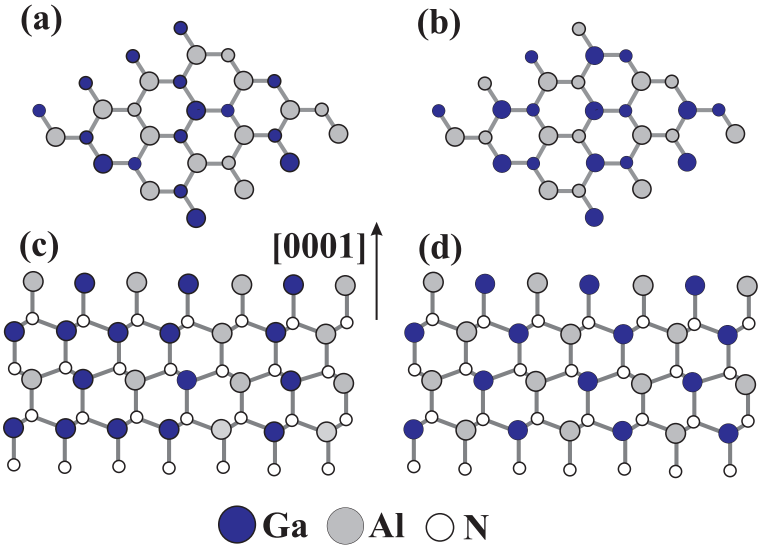


Figure 7.7: Top ((a) and (b)) and side ((c) and (d)) views of the SQS1 ((a) and (c)) and SQS2 ((b) and (d)) structures. Both structures utilize 2 × 2 × 1 supercells.

ordered and disordered configuration has been calculated as

$$E_f = E_{\text{tot}}\left(\text{Al}_\text{x}\text{Ga}_{1-\text{x}}\text{N}\right) - \text{x}E_{\text{tot}}\left(\text{AlN}\right) - \left(1-\text{x}\right)E_{\text{tot}}\left(\text{GaN}\right), \tag{7.9}$$

where $E_{\text{tot}}\left(\text{AlN}\right)$ and $E_{\text{tot}}\left(\text{GaN}\right)$ are the total energies of bulk AlN and GaN, respectively and $\text{x} = \frac{1}{2}$. The calculated lattice constant, c/α ratio, formation energy, and bandgap are given in Table 7.3. Our results show that all the structures give almost degenerate results. The only exception is found for the ordered configuration which has a somewhat higher formation energy (by ~ 3 meV/cation). Additionally, there is a reduction in free energy in the disordered structure arising from increased configurational entropy. For a binary (or pseudobinary) alloy the entropy contribution is

$$S = k_B T\left(\text{x}\ln\text{x} - \left(1-\text{x}\right)\ln\left(1-\text{x}\right)\right). \tag{7.10}$$

For $\text{x} = \frac{1}{2}$ the configurational entropy is $\ln\left(2\right)k_B T \approx 0.67 k_B T$. This effect is significant for growth temperatures and favors the disordered structure. Thus, the theoretical results are in full accordance with the annealing experiments described in Sec. (7.2): **the ordered structure is thermodynamically unstable**.

In Sec. 7.1 a possible mechanism that explains ordering in InGaN alloys has been described. The question is whether a similar mechanism may be operational also in $\text{Al}_\text{x}\text{Ga}_{1-\text{x}}\text{N}$ alloys. In $\text{Al}_\text{x}\text{Ga}_{1-\text{x}}\text{N}$ the Al-N bond is stronger than the Ga-N bond. Thus, the system can minimize its energy during growth by incorporating Al atoms preferably at B2 sites (see Fig. (7.3) where two N-bonds can be saturated. Ab-initio calculations showed that the energy reduction obtained upon exchanging an Al atom bound to a B2 site with a Ga atom bound to a T1 site is 0.67 eV. This

Table 7.3: Calculated lattice constant, c/α ratio, and bandgap for wurtzite (wz) and zincblende (zb) GaN, AlN, and $Al_{0.5}Ga_{0.5}N$ ordered and disordered alloys. Also the formation energies of the alloys calculated using Eq. (7.9). For the zincblende AlN the E_{gap} corresponds to the bandgap energy at the Γ-point.

	phase	α (Bohr)	c/α	E_f (meV/cation)	E_{gap} (eV)
GaN	wz	6.04	1.63	-	2.06
AlN	wz	5.90	1.60	-	4.13
$Al_{0.5}Ga_{0.5}N$ ordered	wz	5.97	1.62	8.5	2.95
SQS1	wz	5.97	1.62	5.4	3.03
SQS2	wz	5.97	1.62	2.1	3.04
SQS3	wz	5.97	1.62	3.0	2.97
GaN	zb	8.53	-	-	1.86
AlN	zb	8.31	-	-	4.06
$Al_{0.5}Ga_{0.5}N$ ordered	zb	8.42	-	4.8	2.77
$Al_{0.5}Ga_{0.5}N$ disordered	zb	8.42	-	-1.5	2.83

energy is large compared to k_BT at growth temperature. Therefore, this mechanism provides a significant driving force for ordering. The model explains also why ordering occurs only in a limited temperature range: The Al and Ga adatoms on the surface should have enough time to diffuse to the step edges or vicinal facets, where they get incorporated at their preferred sites. This can be achieved at high growth temperatures (compared to growth temperatures usually obtained during MBE) and low growth rates. This is in agreement with the experiment by Ebling *et al.* [207], that the degree of ordering in MBE-$Al_xGa_{1-x}N$ alloys improves with increasing the growth temperature.

7.5 Effects of ordering on the electronic structure.

CL and PL measurements on samples with and without chemical ordering showed that chemical ordering leads to a redshift in the spectra, i.e. chemical ordering reduces the bandgap (see Fig. 7.5). This reduction is most pronounced for alloys with a AlN mole fraction close to 50% ($\sim$ 140 meV). In order to investigate the effect ordering has on the electronic properties of the $Al_xGa_{1-x}N$ alloys, the bandgaps of bulk GaN and AlN and of ordered and disordered $Al_{0.5}Ga_{0.5}N$ alloys have been calculated.

In Table 7.3 the bandgaps of both ordered and disordered alloys are given. As can be clearly seen ordering reduces the bandgap. This is consistent the experimental observations. The calculated bandgaps of the SQS1 and SQS2 are approximately 90 meV larger than those of the ordered configuration. These values are lower than the experimental value of 140 meV, but in excellent agreement with the value of 90 meV obtained by Dubiy and Zunger [36] using a similar approach (DFT-LDA). On the other hand, the SQS3 shows only a $\sim$ 20 meV larger bandgap. The calculations for cubic ordered and disordered $Al_{0.5}Ga_{0.5}N$ alloys reveal that ordering gives rise to a 60 meV bandgap reduction. While the discrepancies between the experiment and the SQS1 and SQS2 band calculations can be attributed to intrinsic limitations of DFT (e.g underestimation of the bandgap), two questions remain:

- Why is the reduction of the calculated bandgap for the SQS3 small compared to both experiment and the theory when using the SQS1 and SQS2 structures or those published in Ref [36]?

- *What is the origin of the redshift?* In zincblende semiconductors, the reduction of the bandgap is explained in terms of level repulsion due to Brillouin zone folding. Ordering in zincblende semiconductors increases the unit cell along the ordering direction (e.g. constructing a a 1:1 superlattice along the $\{111\}$ direction results a unit cell which is twice larger than the original cell - see Figs. 7.1(c) and (d)). This increase will fold non-Γ points onto the Brillouin zone center. In contrast, in the wurtzite structure Brillouin zone folding does not occur, since ordering causes changes only in the occupation of the $1 \times 1 \times 1$ unit cell, the periodicity (and thus the size of the unit cell) along the c-axis remains unchanged.

In order to address these questions the wavefunctions of the states at the valance band maximum (VBM) and conduction band minimum (CBM) have been examined in detail. In Fig. 7.8 the wavefunctions of the states at the VBM and the CBM are shown. As can be clearly seen, the p-like states of the VBM and the s-like states of the CBM are strongly localized at those N atoms that are bound to three Ga atoms (i.e., in the GaN-layer of the 1:1 superlattice). This essentially resembles localization of electrons and holes in a *mono-atomic GaN quantum well embedded in mono-atomic AlN barriers*. On the other hand for the SQS1 and SQS2 configurations this type of localization is absent (see Figs 7.9 and 7.10). A similar localized/delocalized behavior is found for the VBM and CBM of ordered/disordered zincblende $\mathrm{Al_xGa_{1-x}N}$ alloys (see Fig. 7.11). We can thus, conclude that *the origin of the bandgap reduction is carrier localization.*

So far, the electronic properties of the SQS3 structure have not been addressed. As demonstrated in Sec. 7.3.1 that SQS3 configuration is a good candidate to model random binary alloys in the wurtzite structure provided that the four tetrahedral bonds are equivalent. However, as pointed already due to the hexagonal symmetry of the wurtzite structure this is not true. Assuming the bond along the c-axis inequivalent to the other three, the correlation functions of the SQS3 structure are clearly worse than of the SQS1 and SQS2 structures (see Table 7.2). However, comparing the corresponding data in Table 7.2, we find a correlation function that is comparable to that of the SQS2 configuration. The question is therefore why do the SQS1 and SQS2 structures show almost identical results for the bandgap reduction, while the SQS3 gives a reduction which is lower by a factor of four?

To understand this behavior in Fig. 7.12 the wavefunctions of the VBM and CBM states for the SQS3 structure are shown. Unlike the SQS1 and SQS2 holes and less pronounced electrons which are localized can be seen. The localization of the wavefunctions takes place at N atoms bound to three Ga atoms. It must be noted here that the SQS3 is only a single configuration

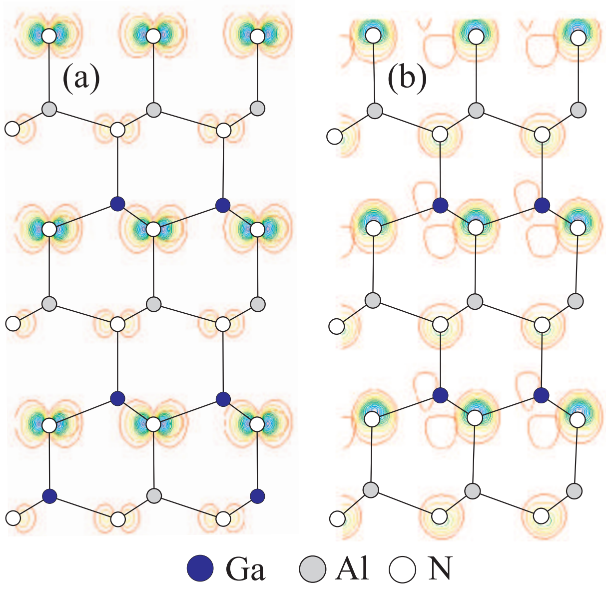


Figure 7.8: Contour plots of the states at the VBM (a) and CBM (b) of the ordered wurtzite $\mathrm{Al_{0.5}Ga_{0.5}N}$. Strong localization of both states at N bound to three Ga atoms, (i.e. in the GaN-layer) can been seen.

out of a large set of possible configurations ($> 10^3$), which can be realized with a cell which is four times the volume of the wurtzite unit cell. The fact, that half of the supercell along the c-axis consists of a 1:1 superlattice-like structure, has to considered as accidental. An important conclusion from this observation is that constructing an ”optimum” disordered alloy solely based on the correlation function is not sufficient. While the SQS2 and SQS3 configurations have a similar correlation function for the first two neighboring shells, they considerably differ in the nearest neighbor environment of the N atoms.

In order to explain the strange behavior of SQS3, let us first examine the band structure of GaN, AlN and $\mathrm{Al_xGa_{1-x}N}$ at the Γ point. Both GaN and AlN in the wurtzite phase are direct bandgap semiconductors. As can be seen in Fig. 7.13 the top of the VB consists almost entirely of N $2p$ states, while the bottom of the CB originates from s states of N and Ga (Al) atoms. Even although spin-orbit interaction is not included in our approach, the triple degenerated VBM (the Γ_{15}^{v} state in Fig. (7.13)) is split into a double degenerate Γ_6^v state with $p_{x,y}$ symmetry and a single degenerate Γ_1^v state with p_z symmetry. This energy splitting is called crystal field splitting Δ_{cr} (it arises from the crystal symmetry at the Γ-point). If the spin-orbit interaction Δ_{so} is included the Γ_6^v split into Γ_9^v and Γ_7^v, while the single degenerate Γ_1^v state is labelled also Γ_7^v. The sign of the Δ_{cr} that determines the order of the Γ_1^v and Γ_6^v states and its magnitude, depends strongly on the kind of material and the structural parameters (the c/α ratio of the lattice constants and the internal lattice parameter u). For, GaN Δ_{cr} is positive and according to first-principles calculations it is rather large compared to Δ_{so}, although Δ_{cr} itself is not large. Experimental measurements give a value of the crystal field splitting that is positive but comparable in value to that of the spin-orbit splitting. On the other hand, for AlN the crystal field is negative and large. For more details see Ref. [214]. The negative value of AlN for Δ_{cr} is related to the strong ionicity of this material: Due

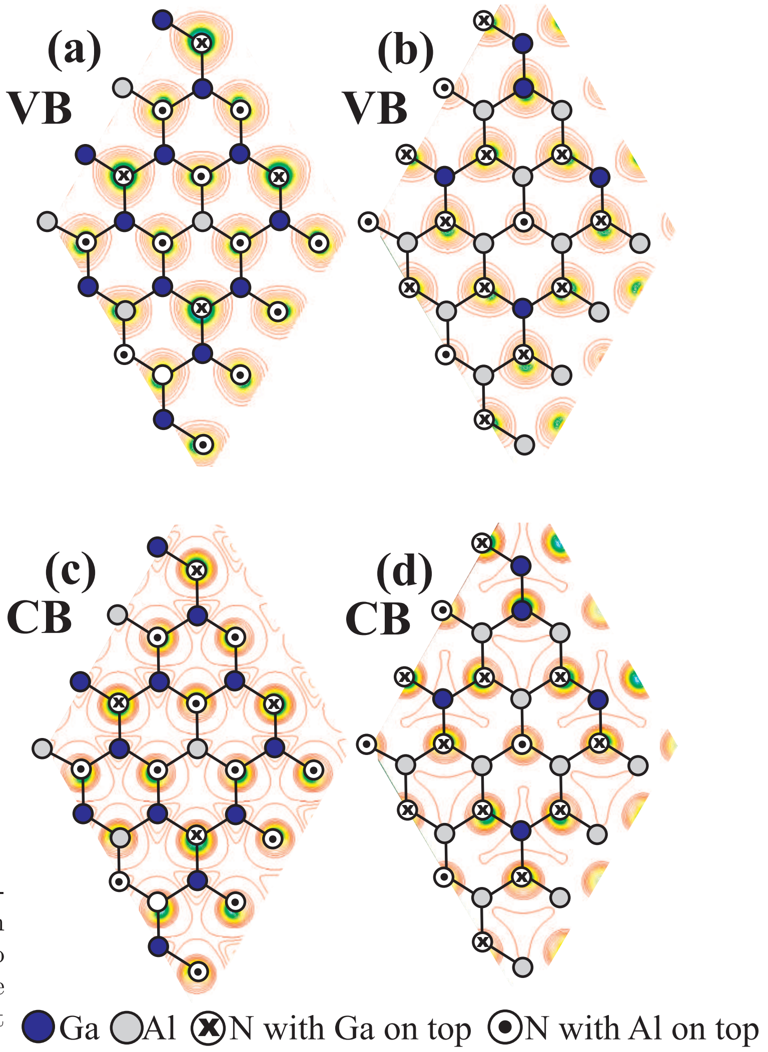


Figure 7.9: Contour plots of the valence ((a) and (b)) and conduction ((c) and (d)) band edges at the two N layers in the SQS1 supercell. The Ga and Al atoms shown are located at the layer below the N atoms.

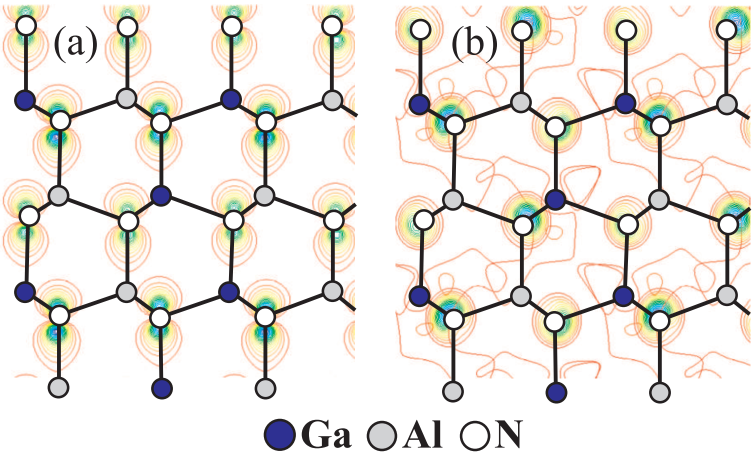


Figure 7.10: Contour plots of the states at the VBM (a) and CBM (b) of the SQS2. The wavefunctions of both are delocalized.

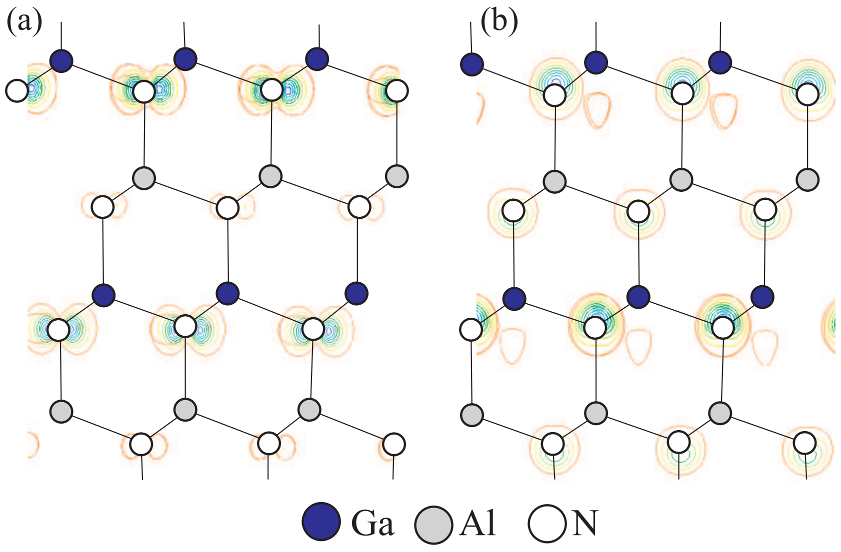


Figure 7.11: Contour plots of the states at the VBM (a) and CBM (b) of ordered cubic $Al_{0.5}Ga_{0.5}N$. Clear hole and electron localization at N bound to three Ga atoms can been seen.

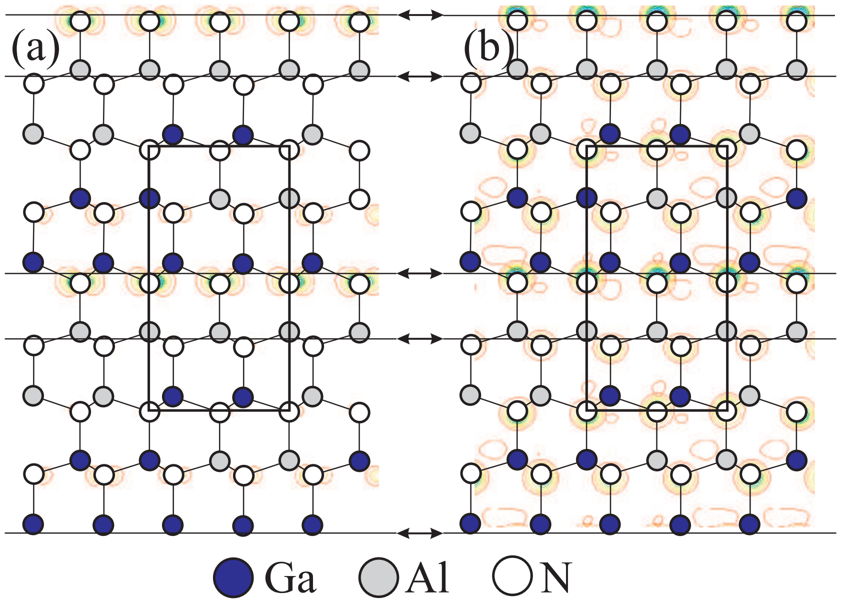


Figure 7.12: Contour plots of the states at the VBM (a) and CBM (b) of the SQS3. Strong hole localization (a) and less pronounced electron localization (b) is obtained. The rectangle indicates the supercell. The horizontal lines and the arrows indicate the vertical positions of the pure GaN and pure AlN layers.

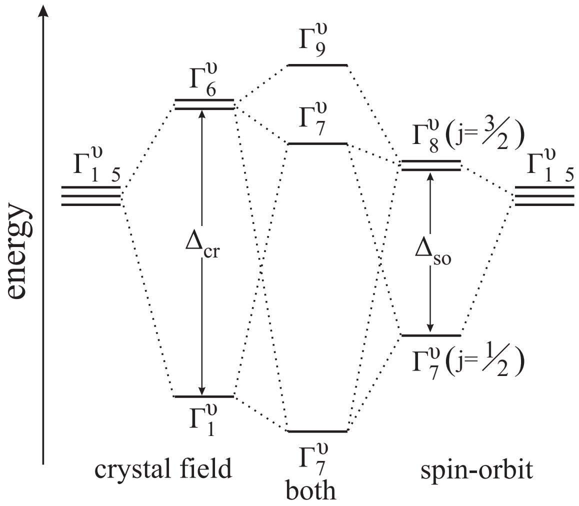


Figure 7.13: Schematic diagram showing the energy splitting at the VB edge of the wurtzite GaN. Left side: splitting induced only by the crystal field. Right side: splitting induced only by spit-orbit interaction. The combined case is shown in the middle. The scale of the diagram is artificial.

to it AlN shows a considerable deviation from the ideal wurtzite structure ($c/\alpha = 1.60 < \sqrt{8/3}$).

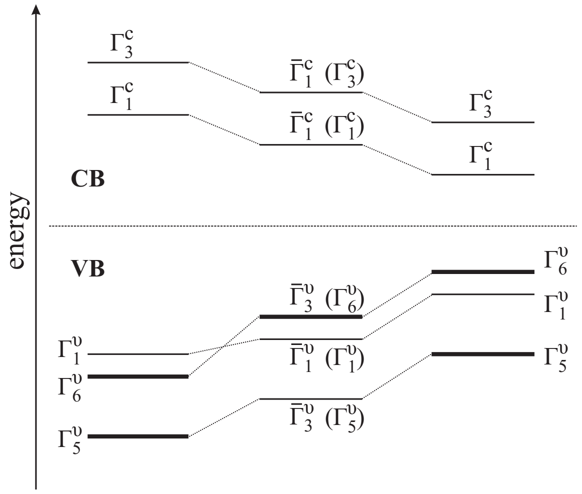


Figure 7.14: Schematic diagram of the VB and CB states GaN, AlN, and ordered $Al_{0.5}Ga_{0.5}N$. The scale of the diagram is artificial.

The symmetry of the states at the valence and conduction band edges for ordered $Al_{0.5}Ga_{0.5}N$ alloys as described by Dudiy and Zunger [36] is shown in Fig. 7.14. In an ordered structure the Γ^{v}_{1} state of the constituents AlN and GaN forms the $\overline{\Gamma}^{v}_{1}(\Gamma^{v}_{1})$ state[3]. Similarly, the Γ^{v}_{6} state forms the $\overline{\Gamma}^{v}_{3}(\Gamma^{v}_{6})$ state. The Γ^{v}_{5} state which is deeper in the valence band forms the $\overline{\Gamma}^{v}_{3}(\Gamma^{v}_{5})$. While the Γ^{v}_{6} and Γ^{v}_{5} states for AlN and GaN have a different symmetry, the resulting $\overline{\Gamma}^{v}_{3}(\Gamma^{v}_{6})$ and $\overline{\Gamma}^{v}_{3}(\Gamma^{v}_{5})$ states have the same symmetry. In order to avoid crossing, these states repel each other, thus pulling the VBM upwards[4]. Similarly, upon ordering, the Γ^{c}_{1} and Γ^{c}_{3} states of the CB will give the $\overline{\Gamma}^{c}_{1}(\Gamma^{c}_{1})$ and the $\overline{\Gamma}^{c}_{1}(\Gamma^{c}_{3})$ states which have the same symmetry. Thus they repel each other decreasing further the bandgap. As shown in Ref. [36] the top of the VB in ordered $Al_{0.5}Ga_{0.5}N$ has $p_{x,y}$ character and the crystal field splitting $\Delta_{\text{cr}} = \overline{\Gamma}^{v}_{3}(\Gamma^{v}_{6}) - \overline{\Gamma}^{v}_{1}(\Gamma^{v}_{1})$ is positive (113 meV). In the random $Al_{0.5}Ga_{0.5}N$, on the other hand, the crystal field becomes negative (-77 meV) and the topmost valence state has p_z character.

It is therefore important to stress that a criterion which can be used to identify ”optimum” quasirandom configurations of an $Al_{0.5}Ga_{0.5}N$ alloy, has to take into account (besides the correlation functions) also the local environment of the N atoms. If the number of N atoms with Al-N bonds (which are considerably more ionic than the Ga-N bonds) at adjacent sites exceeds a critical value,

---

[3] The resulting states of the ordered alloy are denoted by a bar and in parenthesis their parenthood states are given.

[4] For example, in the band structure of GaN and AlN the parenthood Γ^{v}_{6} and Γ^{v}_{5} states cross at the A point.

the crystal field becomes "locally" positive. In these regions the electronic properties of the system exhibit a more "ordered" character. Exactly this mechanism is operational for the SQS3 structure, where part of the supercell contains a 1:1 superlattice structure. The absence of such a local arrangement in the SQS2 configuration makes this structure a better random alloy, although it has a similar correlation function as the SQS3 structure.

7.6 Conclusions

Based on first principles calculations on the energetics and electronic structure we were able to get a very detailed insight into the intricate interplay between chemical ordering, thermodynamic stability, and electronic structure. For example, consistent with experiment ordered structures are found to be thermodynamically unstable and to have a reduced bandgap (reduction by $\simeq 90\,meV$). A detailed analysis showed that unlike for III-V semiconductors, the bandgap reduction is caused by a strong hole and electron localization in mono-atomic GaN quantum wells embedded in mono-atomic AlN barriers: Thus, ordered nitrides behave as a natural multi-quantum well structure. Our analysis showed also that the local environment around N atoms significantly affects the electronic properties of the alloy. This has been explained in terms of the ionicity differences between Al-N and Ga-N bonds and their effect on the order of the states at the edge of the VB. Adjacency of several N atoms with more Al first neighbors, can locally change the sign of the crystal field and alter the sequence of the states at the VBM, leading to carriers localization and bandgap reduction. Finally, a possible kinetic mechanism for growth of ordered AlGaN alloys and provides a growth window for ordering is proposed. According to this mechanism ordering is introduced due to growth kinetics at surface steps and is caused by the energy difference of the bonds between of Al-N and Ga-N. Thus, growth of ordering requires high growth temperatures, compared to those obtained during MBE, and low growth rates, in order the adatoms to be able to diffuse on the surface to the step edges.

Since the degree of order can be controlled by growth temperature or growth rate, lattice matched heterostructures of ordered and disordered $\mathrm{Al_xGa_{1-x}N}$ alloys having the same composition become possible. We have seen (Sec. 7.5) that the topmost valence states of AlN and GaN have (due to the stronger ionic character of the AlN) a different symmetry (Γ_6^v with $p_{x,y}$ character for GaN and Γ_1^v with p_z character for AlN). In the disordered $\mathrm{Al_xGa_{1-x}N}$ alloys the VBM state stems from the Γ_1^v state and thus has p_z character. Therefore the principle optical transition is z-polarized and strongly reduces the light emission along the c-axis in $\mathrm{Al_xGa_{1-x}N}$ structures. However, since ordering alters the sequence of the valence bands back to that found in GaN, employing chemically ordered $\mathrm{Al_xGa_{1-x}N}$ alloys may be an interesting option. An example would be vertical cavity surface emitting lasers (VCELS) in the deep UV region of the optical spectrum.

Chapter 8

Edge type dislocations in GaN

During the 1990s there has been a remarkable progress in group III-Nitride semiconductors. Nevertheless, due to the lack of lattice and thermal matched substrates, growth of wurtzite GaN films results typically in high threading dislocation (TD) densities. Since the work by Lester *et al.* [215] there is a large controversial debate regarding the effect dislocations have on the electronic and optical properties of group III-Nitride based devices. These authors measured the dislocation density in high efficiency LEDs fabricated by Nakamura's group [216]. They found the dislocation density in the active region of the devices to be of the order of $\sim 10^{10}\,\mathrm{cm}^{-2}$. Until recently the density of dislocations in the GaN films has been $10^{8} \dots 10^{10}$ [31, 217]. However, different techniques have been developed that allow the reduction of TD in the GaN epilayers. The reduction of extended defects has been demonstrated for the homoepitaxial growth GaN on GaN bulk crystals or bulk-like substrates [218, 219, 220]. On the other hand, in the field of heteroepitaxy, the ELOG (Epitaxial Lateral Overgrowth) technique provides epilayers characterized by relatively low density of threading defects. The reduction of the threading defects is achieved by masking part of the dislocated initial GaN material and subsequent growth on top (see Ref. [221] and Refs therein).

Transmission electron microscopy studies reveal that in wurtzite GaN edge, screw, and mixed type dislocations are present (see Refs [222, 223] and Refs therein). Edge dislocations are characterized by a Burgers vector $\boldsymbol{b} = \boldsymbol{\alpha} = \frac{1}{3}\left\langle 11\bar{2}0 \right\rangle$ (α-type), screw dislocations by $\boldsymbol{b} = \boldsymbol{c} = \langle 0001 \rangle$ (c-type), and mixed by $\boldsymbol{b} = \boldsymbol{\alpha} + \boldsymbol{c} = \frac{1}{3}\left\langle 11\bar{2}3 \right\rangle$ ($\alpha + c$-type). α and c are the wurtzite lattice constants (see Fig. 8.1). The relative concentration of the various types of TD depends on the growth parameters. In general the majority of dislocations are of edge type and have been proposed to be associated with the coalescence of slightly misoriented islands [222]. Electronic and atomic structure calculations on edge type TD in GaN have been performed on four candidate structures. The full (or 8-atoms) core, the open (or 5/7-atoms) core, the Ga-vacancy and the N-vacancy cores have been assumed (see Fig. 8.5). The vacancy core geometries have been constructed by removing a certain fraction of low coordinated Ga or N atoms present in the core regime.

In the present chapter we investigate the electronic structure, energetics and geometry of the

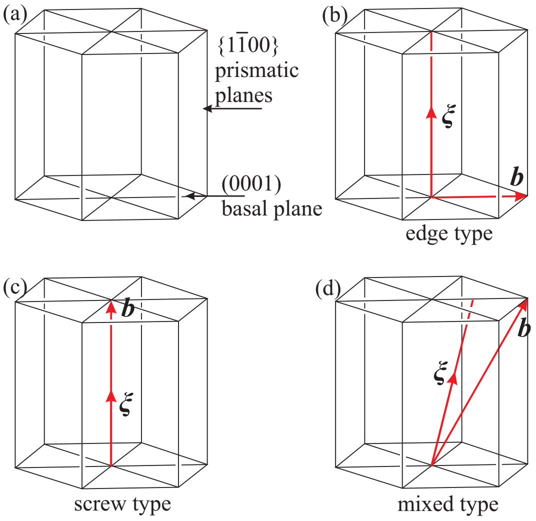


Figure 8.1: Schematic representation of threading dislocations in wurtzite GaN. (a) The GaN unit cell. (b) An edge type dislocation with Burgers vector $\boldsymbol{b} = \frac{1}{3}\langle 11\bar{2}0\rangle$. (c) A screw type dislocation with Burgers vector $\boldsymbol{b} = \langle 0001\rangle$. (d) Mixed type dislocation with Burgers vector $\boldsymbol{b} = \frac{1}{3}\langle 11\bar{2}3\rangle$. $\boldsymbol{\xi}$ represents the dislocation's sense.

α-type TD in GaN. Based on a multiscale theoretical approach we demonstrate the existence of a new type of edge dislocation core (called 4-atoms core). This structure has been first studied by empirical potential calculations and had been discarded since it was found to be energetically unfavorable [56, 57]. A unique feature of the new structure compared to all previously identified structures is that all core atoms are fully coordinated, i.e., they are as in bulk GaN surrounded by four anions/cations). Thus, while the 8- and the 5/7-atom core structures have broken bonds in the core region and are thus expected to induce states in the bandgap, the 4-atom core is not expected to be electrically active. However, this is not valid for edge dislocations in GaN: As will be shown the local strain field around the dislocation core in combination with the small lattice constant of GaN induces in this material deep defect states. Clear evidence of the 4-atom core edge dislocation in GaN comes from recent HRTEM experiments and image simulations. Part of the theoretical calculations presented here along with the aforementioned HRTEM experiments has been published in Ref. [224].

8.1 Previous studies on TD in GaN.

In conventional III-V semiconductor devices threading dislocations act as non-radiative centers and reduce the optical output. The recombination of the minority carrier centers results in heat generation and subsequently to the motion and multiplication of the defects. At the end the devices burn out or stop operating. While in conventional III-V semiconductor devices a dislocation density of the order of $10^6\,\mathrm{cm}^{-2}$ would be fatal, GaN-based optoelectronic devices operate efficiently even at large dislocation concentrations as large as $10^{10}\,\mathrm{cm}^{-2}$.

Initially Lester *et al.* proposed that dislocations do not act as recombination centers for the minority carriers [215]. More recently Xin *et al.* based on atomic resolution Z-contrast imaging experiments reported the first direct observation of the atomic structure of TD cores in GaN [225]. Based on the observation that the central column of the edge type dislocation core (column of low coordinated atoms of the 8-atom core shown in Fig. 8.5) has a similar atomic configuration as the $\{10\bar{1}0\}$ surface they proposed that this core should not impose any deep states in the bandgap. Indeed, as discussed in Sec. 6.3, the re-hybridization that takes place in the $(10\bar{1}0)$ surfaces cleans the bandgap from any induced states. More recently, based on electron energy loss spectra (EELS) experiments, Xin *et al.* reported that α-type dislocations do not exhibit any states in the bandgap [226]. Furthermore Arslan and Browning, using multiple - scattering simulations in combination with Z-contrast imaging and EELS experiments, proposed that edge type TD in GaN are not electrically active. Rather, they speculated that point defects are related with any electrical activity at dislocations and limit the device performance [227]. However, more recently Bangert *et al.* studied the effects dislocations have on low (lowEELS) and core electron loss spectra [228]. They demonstrated that, while the core excitation EELS are very sensitive to changes in the bonding, lowEELS spectra are very sensitive to empty states in the bandgaps. Based on these observations they showed that edge and mixed type dislocations induce gap states and hence are electrically active.

This observation is nowadays confirmed a large number of experimental studies which indicate that threading dislocations in GaN act as recombination centers for the minority carriers (see Refs [228, 229, 230, 231, 232, 233] and references therein). Rosner *et al.* reported the first work based on a combination of AFM, TEM, and CL experiments, that links TD with a reduction in the minority carrier density [229]. As has been proposed by them, the high performance of devices based on materials with high density of electrically active dislocations can be explained in terms of the diffusion length of the minority carriers. The efficiency of the device will not degrade as long as the dislocation spacing is significantly longer than the carrier diffusion length.

The first electronic structure calculations on α- and c- type TD in GaN have been performed by Elsner *et al.* [50]. They used both *ab-initio* local density functional and density functional based tight binding methods on clusters with a single dislocation and on supercells with a dislocation dipole. The authors found induced states in the bandgap only for the closed core screw dislocation. Removing the hexagonal core of the screw dislocation they created the open core screw configuration which was found to be energetically more favorable than the close core structure. The open screw core consists of an internal surface having $(10\bar{1}0)$ facets and was found to induce only shallow gap states arising from the high local strains related to the dislocation displacement field. For the edge type TD they reported that the 8-atom core structure is the energetically favorable one with no deep states in the bandgap. In subsequent papers they reported that oxygen impurities, Ga vacancies (V_{Ga})), and defect complexes of these two species, may be trapped by the dislocation stress field [51, 54]. They proposed that segregation of oxygen that sits on a nitrogen site (O_N) in the inner $(10\bar{1}0)$ facet of the open screw core, and the formation of $V_{Ga} - (O_N)_3$ chemically inert defects is energetically favorable and may result to the formation of nanopipes. Regarding the edge type TD they found that V_{Ga} and oxygen-related defect complexes may be trapped by the dislocation stress field, with the most favorable position being at the core, and induce deep states in the gap that contribute to the YL intensity.

In a later DFT study Wright and Grossner examined the possibility of edge type TD in GaN to be charged [55]. They focused on four candidate core structures: 8-atom core, 5/7 atom core, Ga vacancy (V_{Ga}), and N vacancy (V_{Ga})[1] They studied the formation energy as a function of the Fermi level for different growth conditions. They showed that the edge type dislocations may well be charged giving rice to deep states in the bandgap. For neutral dislocations, the full-core structure was found to be energetically favorable over the other cores for growth under N-rich conditions. For Ga-rich conditions the neutral full core and N vacancy configurations energetically degenerate and are lower in energy than the other neutral core configurations. The same core structures have been studied by Lee *el al.* performing ab-initio based tight binding calculations [234]. They found that the full, open and N-vacancy core structures, which are energetically favorable under Ga-rich conditions, impose deep states in the bandgap. For the Ga vacancy structure they found only shallow levels $\sim$ 0.2 eV below the CBM and above the VBM. Thus, they proposed that edge dislocations under Ga-rich conditions can act as trap centers for electrons and contribute to the YL emission. Two recent first principles works on screw type TD have been published by Northrup [52, 53]. Different core geometries have been examined. It was found that for growth under Ga-rich conditions the core prefers a configuration where N atoms are removed from the core, while going to more N-rich conditions it becomes energetically favorable for N atoms to be present in the core. As a result different growth conditions are expected to lead to different electrical activity. While for Ga-rich conditions the Ga-rich core was found to impose highly dispersive states in the gap indicating that the dislocation may be electrically highly conductive, for N-rich conditions the core imposed states show a much shorter bandwidth. Empty states in the upper half gap have been also reported for the edge type dislocations by Fall *et al.* [235]. The corresponding wavefunctions of the defect states were found to be centered at the dislocation cores, while for lower values of the isosurfaces or equivalently for lower values of the corresponding charge density, the states were seen to extend into the strained regions around the dislocations.

The relevant stability of dislocations in GaN has been studied also employing SW empirical potentials [56, 57]. Besides the full and open core structures, a novel, fully reconstructed core consisting of a 4-atom ring has been assumed. In both works cluster calculations have been performed and the open core structure was found to be the energetically most favorable. However, while the SW potentials are suitable to describe the elastic distortions within the elastic limit, it is not expected to accurately describe the large distortions (beyond the elastic limit) and the broken bonds that may exist in the core regions. The accuracy of the above results regarding the core energetics are, due to the approximate nature of the applied method, hard to access. For a more

---

[1] The geometries of the 8- and 5/7 atom cores are shown in Fig. 8.5. The Ga and N vacancy structures are derived from the 8-atom core by removing the low coordinated Ga and N atoms present in the core respectively.

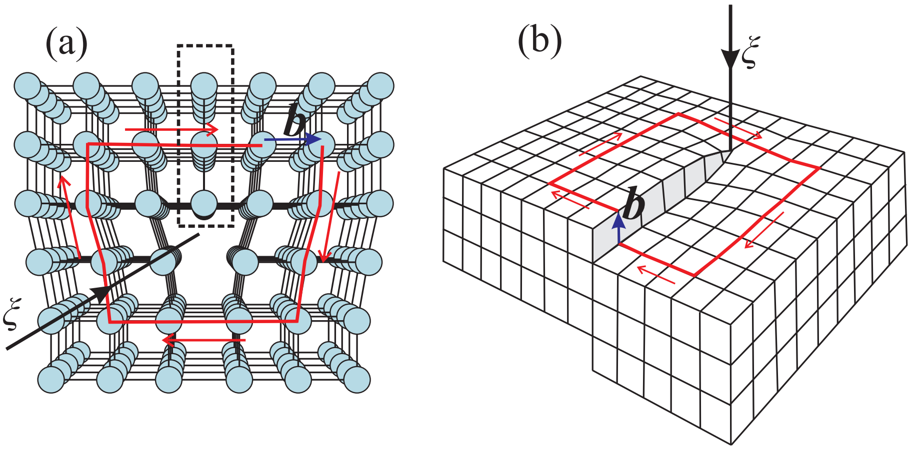


Figure 8.2: Edge (a) and screw (b) dislocations in a simple cubic crystal. $\boldsymbol{\xi}$ is the sense of the dislocations, while the red lines and arrows indicate the Burgers circuits. The closure failure of the circuit defines the Burgers vector $\boldsymbol{b}$. The dotted rectangle in (a) indicates the additional semi-infinite plane incorporated into the crystal to create the dislocation.

detailed discussion on the methods used to study theoretically the dislocations see Sec. 8.4.

8.2 A Brief Introduction to Dislocation Theory

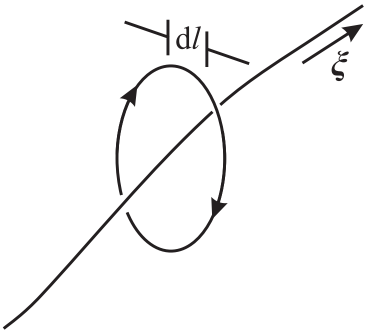


Figure 8.3: Integration (Burgers) circuit around a dislocation of sense $\boldsymbol{\xi}$.

The concept of dislocations has been first introduced in order to describe plastic deformation of crystals, namely the non-reversible deformation that occurs for stresses beyond Hooke's law [236, 237]. Let us first assume the crystal as a continuum media. If the crystal was ideal then any integration of the displacement field $\boldsymbol{u}$ around a closed circuit would vanish. If, on the other hand the circuit enclose a dislocation (see Fig. 8.3) then the above integral would result:

$$\boldsymbol{b} = \oint \frac{\partial \boldsymbol{u}}{\partial s} \mathrm{d}s. \tag{8.1}$$

Here, $\boldsymbol{b}$ is the Burgers vector of the dislocation.

In general, we can define dislocations as line defects with non-vanishing Burgers vectors. In order to get a better visualization of dislocations, let us assume a cubic crystal and an additional semi-infinite plane pushed in the crystal (see Fig. 8.2(a)). The edge line of this semi-infinite plane is called edge dislocation. As schematically shown in Fig. 8.2(a), the disturbance of the lattice is large in the vicinity of the dislocation, while as we go some lattice periods away the crystal gets again the correct registry. However, a deformation (deviation from the ideal bulk structure) is present even at large distances away from the dislocation. This becomes more clear by drawing a closed circuit going from lattice

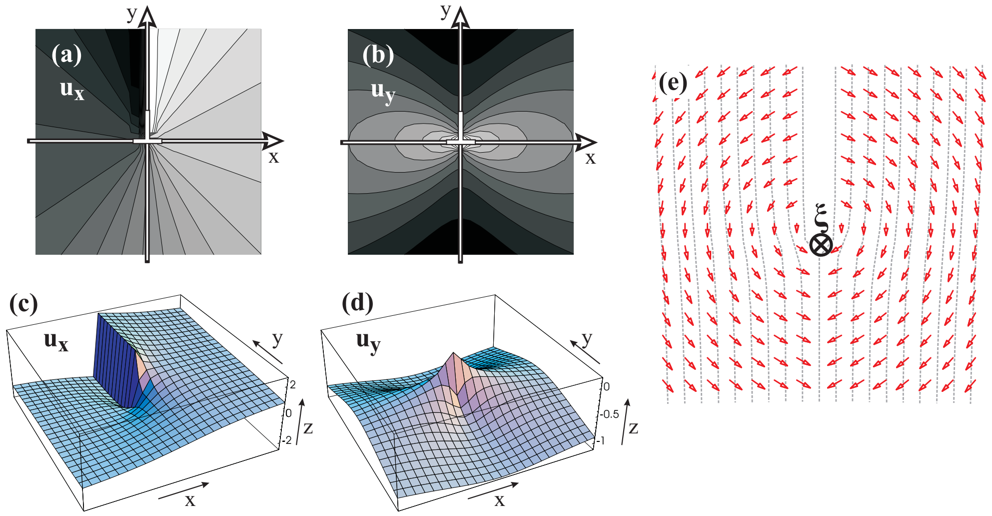


Figure 8.4: Contour ((a) & (b)), 3D ((c) & (d)), and (e) vector field plots of the components u_x and u_y of the displacement field $\boldsymbol{u}$ imposed by an edge type dislocation (see Eq. (8.5)). The dotted lines in (e) indicate the distortion of the lattice planes imposed by the displacement field. The dislocation line is normal to the paper and is placed at the axes origins.

point to lattice point in the ideal bulk crystal. Equal number of steps must be taken along the parallel sides of the circuit. If the same circuit is drawn around in the dislocated crystal around the core then an additional vector, that is normal to the dislocation line and equals a lattice period, would be required in order to close the circuit. The additional vector is the Burgers vector $\boldsymbol{b}$ that characterizes the strength of the dislocation.

A different situation is illustrated in Fig. 8.2(b). Here, the result of a cut by half a lattice plane and the mutual shift of both parts is shown. The edge of the cut corresponds to an edge type dislocation. The screw dislocation differs from the edge on the fact that the Burgers vectors of the former are parallel to the dislocation line. In general a dislocation is characterized by the Burgers vector $\boldsymbol{b}$ and the sense $\boldsymbol{\xi}$. The sense is the tangent to the dislocation line vector. For an edge dislocation the following condition holds:

$$\boldsymbol{b}_\mathrm{edge} \cdot \boldsymbol{\xi} = 0, \tag{8.2}$$

For a screw dislocation, instead of the dot product the cross product has to be taken:

$$\boldsymbol{b}_\mathrm{screw} \times \boldsymbol{\xi} = 0 \tag{8.3}$$

Since in this work we focus on α-type dislocations in GaN, the following analysis focuses only on edge type dislocations.

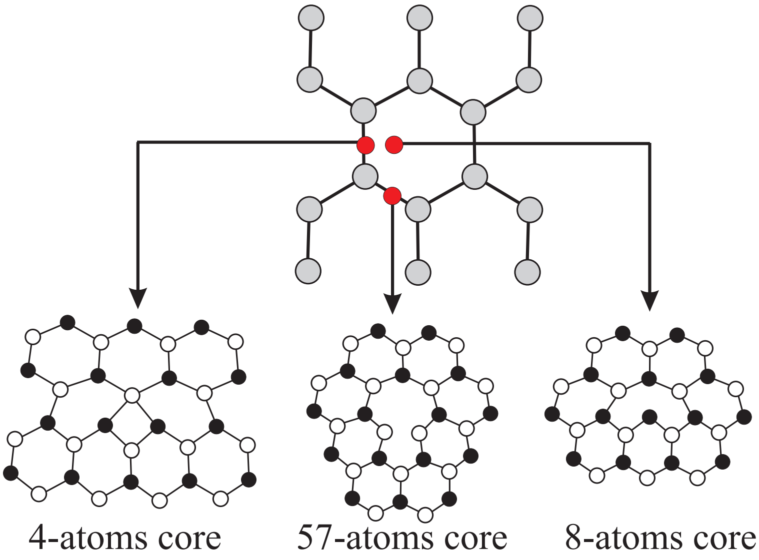


Figure 8.5: Three different core structures that can be created by mapping the displacement field predicted by elasticity theory (see Eq. (8.5)) onto the bulk lattice at different origins. The red points in the bulk lattice indicate the origins that produce the three different dislocation cores.

8.2.1 Displacement field associated with edge type dislocations

Based on the equilibrium equations of elasticity theory (Eq. (A.2)) it can be proved that the displacement field around a straight edge dislocation is [238]:

$$\nabla^2 \boldsymbol{u} + \frac{1}{1-2\nu} \nabla \mathrm{div} \boldsymbol{u} = -b\delta\left(r\right)\widehat{\mathbf{j}}. \tag{8.4}$$

Here, the dislocation line coincides with the z-axis, and the Burgers vector is along the positive x-axis ($b_{\mathrm{x}} = b$, $b_{\mathrm{y}} = b_{\mathrm{z}} = 0$), ν is Poisson's ratio and $\widehat{\boldsymbol{j}}$ is the unit vector in y-direction. Due to symmetry reasons the displacement field is restricted in the xy-plane ($\boldsymbol{u} = u_{\mathrm{x}}\widehat{\mathbf{i}} + u_{\mathrm{y}}\widehat{\mathbf{j}}$, $u_{\mathrm{z}} = 0$)

Eq. (8.4) can be solved analytically and yields the components of the displacement field imposed by a single straight edge dislocation [239]:

$$\begin{aligned} u_{\mathrm{x}} &= \frac{b}{2\pi}\left[\tan^{-1}\frac{y}{x} + \frac{\pi}{2}\mathrm{sign}\left(y\right)\left[1-\mathrm{sign}\left(x\right)\right] + \frac{1}{2\left(1-\nu\right)}\frac{x\,y}{x^2+y^2}\right] \\ u_{\mathrm{y}} &= -\frac{b}{2\pi}\left[\frac{1-2\nu}{2\left(1-\nu\right)}\ln\sqrt{x^2+y^2} + \frac{1}{2\left(1-\nu\right)}\frac{x^2}{x^2+y^2}\right]. \end{aligned} \tag{8.5}$$

Here $\mathrm{sign}\left(x\right)$ is a function of x that gives -1, 0 or 1 depending on whether x is negative, zero, or positive respectively. In Fig. 8.6(b) the displacement field $\boldsymbol{u} = u_{\mathrm{x}}\widehat{\mathbf{i}} + u_{\mathrm{y}}\widehat{\mathbf{j}}$ produced by Eqs. (8.5) is shown, while in Fig. 8.4 contour and 3D plots of the components u_{x} and u_{y} are shown. As can be seen the u_{x} is discontinuous for positive values of y at $x = 0$ due to the factor $\frac{\pi}{2}\mathrm{sign}\left(y\right)\left[1-\mathrm{sign}\left(x\right)\right]$ in Eqs. (8.5). The discontinuity gives the additional half plane required for the dislocation to be formed and has a thickness

$$\Delta u_{\mathrm{x}} = \left[u_{\mathrm{x}\,(x\to 0^+)} - u_{\mathrm{x}\,(x\to 0^-)}\right]_{y>0} - \left[u_{\mathrm{x}\,(x\to 0^+)} - u_{\mathrm{x}\,(x\to 0^-)}\right]_{y<0} = b. \tag{8.6}$$

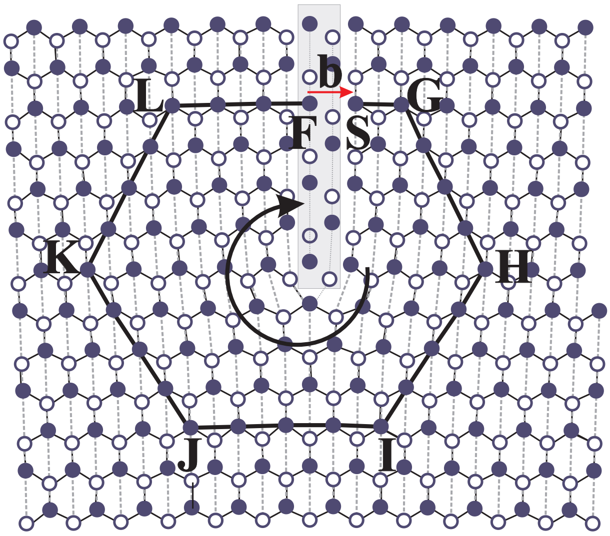


Figure 8.6: The 4-atom core geometry created by mapping the displacement field shown in Fig. 8.4(e) onto the bulk lattice. The crooked line **SGHIJKIF** indicates the Burgers circuit around the dislocation. The closure failure **FS** identifies an edge dislocation with Burgers vector $\boldsymbol{b} = \frac{1}{3}\left[11\bar{2}0\right]$. The shaded area indicates the additional material required for the dislocation to be formed. It consists of a semi-infinite $\left(11\bar{2}0\right)$ GaN double layer. The dotted lines indicate the distorted planes around the dislocation, while, the open and the filled circles the Ga and the N atoms. The sense of the dislocation is normal to the paper plane and points into the paper.

The core geometries are created by mapping the displacement field given by Eq. (8.5) onto the bulk lattice. As proposed by Béré and Serra [56], placing the dislocation line at different points of the bulk lattice three different core structure are formed (see Fig. 8.5). These are called 4-atom core, 5/7-atom (or open) core, and 8-atom (or close) core. Chen *et al.* reported that the 5/7-atom core can be also constructed by removing the column of 3-fold coordinated atoms from the 8-atom core [57]. Indeed, atomic relaxation of the 5/7-atom core structure created by elasticity theory and the structure produced by substracting the low coordinated atoms from the 8-atom core, yields identical positions for the core atoms.

In Fig. 8.6(a) the Burgers circuit (**SGHIJKIF**) around a 4-atom core dislocation is drawn. The closure failure **FS** (**S**: start, **F**: finish), identifies an edge dislocation with Burgers vector $\boldsymbol{b} = \frac{1}{3}\left[11\bar{2}0\right]$. The modulus of the Burgers vector $|\boldsymbol{b}|$ coincides with the lattice parameter α of the wurtzite structure. Hence, a semi-infinite plane of thickness α should be added normal to the Burgers vector. That corresponds to a double $\left(11\bar{2}0\right)$ GaN layer, as it is indicated by the shaded area in Fig. 8.6(a). Identical results arise from the analysis of the 8-atom and the 5/7-atom core dislocations (e.g. see Fig. 8.11 for the 8-atom core and Fig. 8.15 for the 5/7-atom core structures).

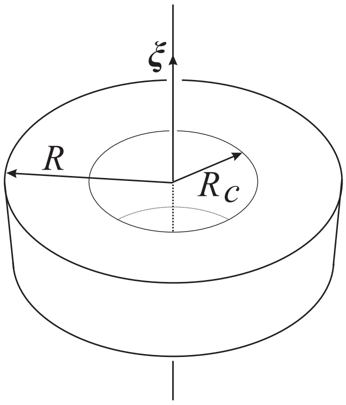


Figure 8.7: Schematic representation of a dislocation line with the core region.

The energy of an infinite straight dislocation comes from the elastic part that is contained in the elastically strained bonds and can be calculated analytically using linear elasticity theory and from the energy stored in the core:

$$E_{\mathrm{disl}} = E_{\mathrm{el}} + E_{\mathrm{core}}. \tag{8.7}$$

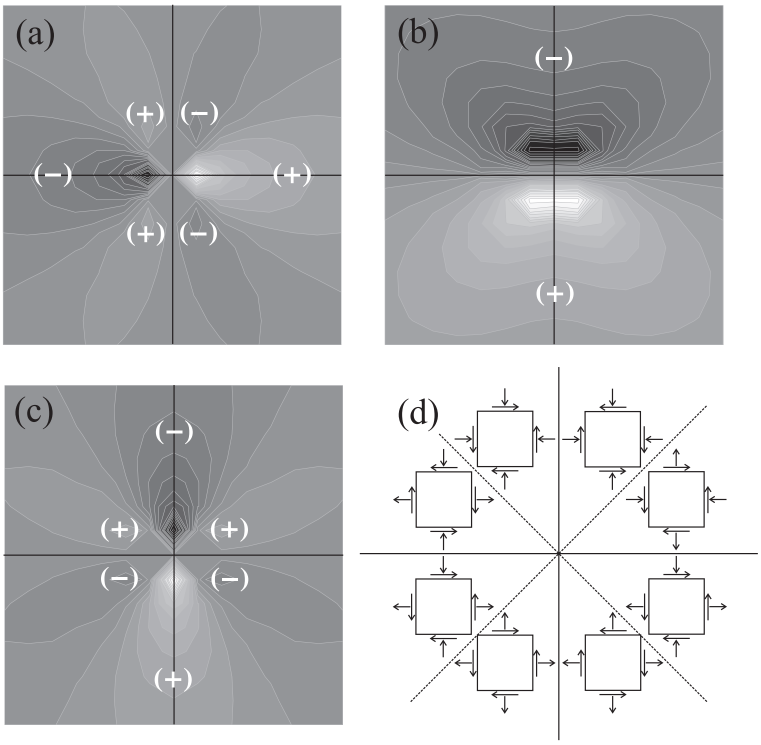


Figure 8.8: Contour plots of the stresses (a) σ_{xy}, (b) σ_{xx}, (c) σ_{yy}, and (d) schematic representation of the stress field around an edge type dislocation. The stress field around an edge dislocation is inversion antisymmetric.

The elastic energy per unit length of a straight edge type dislocation stored in a region bounded by cylinders of radius R_{C} and R around the dislocation line (Fig 8.7) in an isotropic medium is [240]:

$$\frac{E_{\mathrm{el}}\left(R\right)}{L} = \frac{\mu b^2}{\left(1-\nu\right)4\pi} \ln\left(\frac{R}{R_{\mathrm{C}}}\right). \tag{8.8}$$

Here, R_{C} is the core radius, ν is the Poisson's ratio, and μ is the shear modulus. The logarithmic term in Eq. (8.8) diverges as $R \to +\infty$ or as $R_{\mathrm{C}} \to 0$. The divergence with R_{C} arises from the inadequacy of the linear elasticity theory to describe large lattice distortions occurring near the dislocation line. On the other hand the divergence with R denotes that a characteristic energy cannot be ascribed to a dislocation, since it depends on the size of the crystal. The core can be defined as the minimum region around the dislocation line that cannot be described by elasticity theory.

Eq. (8.8) is applicable in isotropic medium. If the crystalline anisotropy is taken into account the elastic energy will be [240, 241]:

$$\frac{E_{\mathrm{el}}\left(R\right)}{L} = \frac{K\,b^2}{4\pi} \ln\left(\frac{R}{R_{\mathrm{C}}}\right), \tag{8.9}$$

where K is the energy factor that depends on the geometry (direction of the dislocation line and the Burgers vector) and the elastic constants of the material. For a straight prismatic α-type dislocation with $\boldsymbol{\xi} = [0001]$ and $\boldsymbol{b} = \frac{1}{3}\left[11\bar{2}0\right]$ the energy factor is [240]:

$$K = \frac{C_{11}^2 - C_{12}^2}{2C_{11}}. \tag{8.10}$$

The elastic constants C_{11} and C_{12} are those defined in Eq. (A.4).

8.2.2 Infinite array of parallel edge dislocations

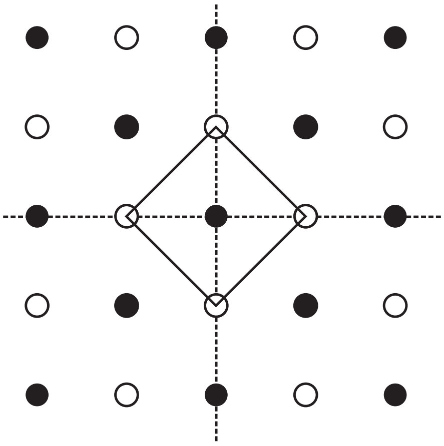


Figure 8.9: Schematic representation of an infinite array of dislocations in the quadrupolar configuration. Open and filled circles correspond to opposite sign dislocations. The rhombus indicate a potential unit cell with a dislocation dipole.

Alternatively to a single dislocation, a distribution of parallel straight dislocations normal to the xy-plane can be assumed:

$$\sum_i (\pm 1)_i \, b \, \delta\left(\boldsymbol{r} - \boldsymbol{r_i}\right). \tag{8.11}$$

Here, $\boldsymbol{r_i}$ correspond to the intersections of the dislocation lines with the xy-plane and the Burgers vectors are assumed to be along the positive (+1) or the negative (-1) x-direction. In order to embed this dislocation distribution in a unit cell with periodic boundary conditions (that will allow us later to follow the supercell approach), the sum of the Burgers vector must vanish. Otherwise, the elastic energy of the system would be infinite due to the logarithmic dependence of Eq. (8.9). Thus, the following condition should hold:

$$\sum_i (\pm 1)_i = 0. \tag{8.12}$$

Let us assume an infinite array of parallel dislocations that satisfies Eq. (8.12). Several different dislocation configurations may be regarded. In order to find the optimum configuration that minimizes the defect-defect interactions, we investigate the stress field arising from an edge-type dislocation. In Fig 8.8 the contour plots and a schematic representation of the stress field are given. It becomes clear that the stress field is inversion antisymmetric. Thus if the dislocations are aligned according the inverse symmetric quadrupolar configuration (Fig. 8.9), the net forces acting on each dislocation cancel and the dislocation-dislocation interactions are minimized [242, 243]. In order to describe dislocations in a supercell geometry, a unit cell with two dislocations of opposite sign is required. Ideally this unit cell can be created, assuming the superposition of the displacement fields (Eq. (8.5)) of an infinite array of dislocation in the quadrupolar configuration as shown in Fig. 8.9. The infinite size of the array ensures that periodicity can be applied to a cell subtracted from the center. In practice a large but finite size array turns out to be sufficient to ensure periodicity. In Fig 8.10 the maximum displacement produced by arrays of different size is shown. As it becomes obvious an array of 12 × 12 is more than

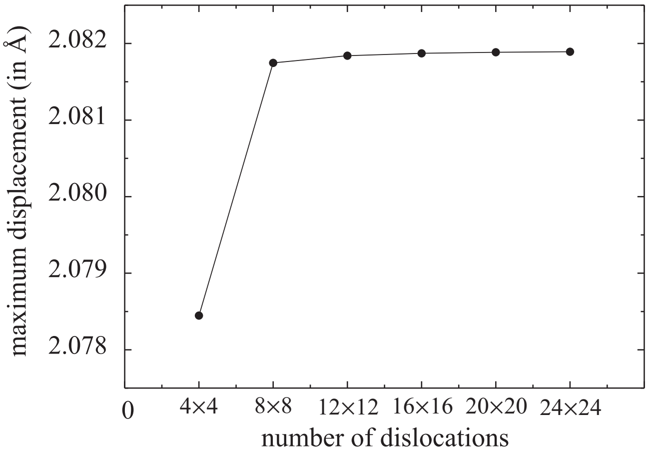


Figure 8.10: Maximum displacement resulting from the superposition of the displacement fields of $n \times n$ parallel dislocations in a quadrupolar configuration as function of the array size.

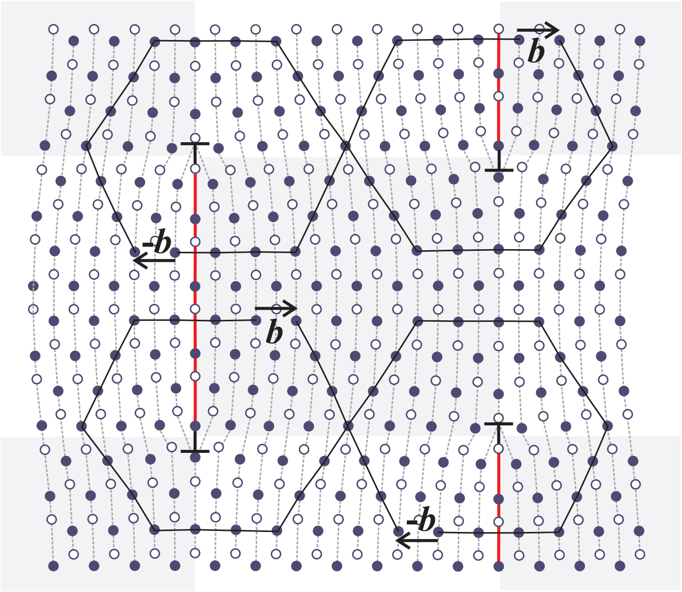


Figure 8.11: A system consisting of 960 atoms with four 8-atom core dislocations in the quadrupolar configuration is shown. Around each dislocation the corresponding Burgers circuit is drawn. The shaded regions correspond to positive values of the component u_x of the displacement field and are shown. The dotted lines indicate the deformation of the lattice due to the presence of the planes induced by the dislocations (thick red lines). Open and filled circles indicate the Ga and the N atoms respectively.

enough to correctly describe the periodicity of the supercell with an accuracy better than 10^{-4} Å.

The next step in creating the supercell is to map the displacements onto the bulk structure. Similarly to the case of a single dislocation (see Figs. 8.5 and 8.6) the core geometry depends on the origin of the dislocations within the bulk lattice. Choosing again different origins the three core geometries shown in Fig. 8.5 will be produced. In Fig. 8.11 the atomic geometry of an array of 8-atom core dislocations is shown. Around each dislocation line the corresponding Burgers circuits are drawn in order to demonstrate the Burgers vectors.

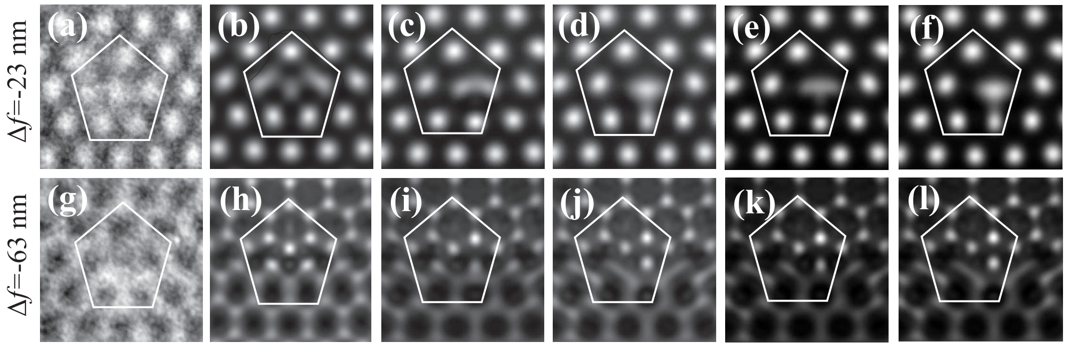


Figure 8.12: Enlargement of the two micrographs shown in Fig. 8.13 (a,g). Image simulations for the 4-core structure (b,h), full core structure (c,i), the open structure (d,j), the N-vacancy structure (e,k), and the Ga-vacancy structure (f,l). The upper (lower) panel shows defocus values $\Delta f = -23$ nm ($\Delta f = -63$ nm).

8.3 Experimental Evidence of the 4-atom Core Structure.

So far only the 8-atom core structure has been identified experimentally. Here we briefly present High Resolution Transmission Electron Microscopy (HR-TEM) experiments that provided together with out theoretical studies direct evidence of the 4-atom core structure. The HR-TEM experiments were performed on 200 μm thick GaN layers, grown by hydride vapor phase epitaxy (HVPE) onto 6H-SiC substrates at 1050°C. For each dislocation a defocus series with defocus steps of 5 nm was taken including at least one contrast inversion. Fig. 8.13 shows the HR-TEM pictures of the same dislocation at two defocus values ($\Delta f = -23$ nm and -63 nm). An enlargement of the core region is shown in Figs 8.12(a) & (g).

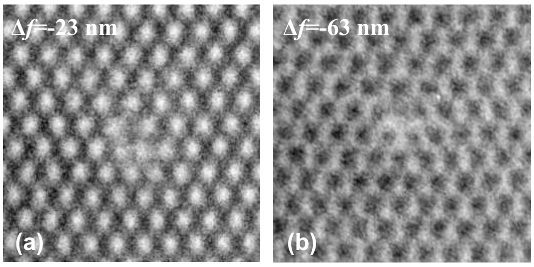


Figure 8.13: Unfiltered HR-TEM image of the same threading dislocation in GaN at defocus values $\Delta f = -23$ and $\Delta f = -63$ nm.

In order to identify the core structure image simulations have been performed. The thickness of the sample and the defocus values of the corresponding experimental images are the necessary parameters for the contrast calculations. Initially, as input the core geometries, as calculated in previous DFT studies (see Sec. 8.1) have been used: the 5/7-atom core (see Fig. 8.23(f) and respective contrast simulations in Figs 8.12(d) & (j)), the 8-atom core (Fig. 8.23(d) and respective contrast simulations in Figs 8.12(c) & (i)), and the Ga-and N-vacancy structure (where under-coordinated Ga and N atoms have been removed from the full core structure). Comparison with the experimental image in Fig. 8.12(a) & (g), the simulated contrasts in Fig. 8.12(c,i) & (d,j) shows no correspondence. Thus the ”classical” core structures fail to reproduce the experimental images observed in our samples and can be ruled out as structural models. If, instead of the 8- and the 5/7-atom cores, the 4 atom core is used as input, the simulated contrasts (Figs 8.12(b) & (h)) agree very well with the experimental images.

8.4 Multiscale Approach.

It becomes clear from Eqs (8.8) and (8.9), that atomic scale calculations of the dislocation cores will be complicated by the long-ranged elastic interactions. In practice atomic scale calculations on dislocations are based on two different approaches. The first approach consist of using a cluster system, while the in second one a supercell with a dislocation dipole and periodic boundary conditions are realized.

In the cluster method a single dislocation is imposed in a finite system and periodic boundary conditions are applied only along the direction of the dislocation line (see Fig. 8.14). The advantage of the cluster method is that it allows the treatment of an isolated defect. Thus, there is no need to take care of defect-defect interactions. On the other hand, the problem of treating the free surface of the cluster arises. If one uses short-range empirical potentials (such as the SW potentials) this problem can be circumvented as follows: The positions of the atoms within a cylindrical shell around the cluster are kept fixed to the positions predicted by elasticity theory. The thickness of the shell can be restricted to the range of the potential. The rest atoms are allowed to fully relax their positions. As becomes obvious from the strain energy distribution shown in Fig. 8.15, as one moves far away from the dislocation line, the lattice distortions smoothly vanish. Thus in order to resemble the correct asymptotic behavior a large enough cylindrical is required. The energy factor and the core radius depend on the elastic constants and on the elastic distortions of the crystal (see Eqs. (8.10) and (8.9)). Thus they can be accurately described by empirical potential calculations. However the SW empirical potentials fail to describe broken bonds and/or bond distortions beyond the elastic limit that may be present in the core region. Furthermore, no information regarding the electronic structure of the core can be derived.

On the other hand, DFT-cluster calculations can overcome the drawbacks regarding the treatment of the core region, but impose a number of problems concerning the treatment of the dangling bonds at the cluster surface. Following a similar procedure as commonly applied for slab calculations, the dangling bonds can be passivated by partially charged pseudo-hydrogens (see Sec. 4.2). As in the slab systems, the passivation of the cluster surface restricts us in the calculating relative formation energies. In the slab systems this is straightforward, since the geometry of the hydrogen passivated surfaces is the same for all slabs under comparison. However, in the dislocation systems, each core structure requires different cluster and thus different surface geometry, even for clusters of identical size. Thus, the passivation with pseudohydrogens cannot be used as well defined and fixed reference system but would give rise to uncontrollable contributions in the electronic structure and total energies. Hence, no comparison between them could be made.

Alternatively a supercell with a dipole of opposite sign dislocations and periodic boundary conditions overcomes the difficulties imposed by the existence of a cluster surface. The defect-cluster surface interactions now are replaced by defect-defect interactions between the infinite number of dislocation images produced by the periodicity. The use of supercells leads to a fairly puzzling situation: First the core structure itself is rather localized. On the other hand, the strain field is significant even far away from the dislocation distance. While DFT calculations provide an accurate description they are restricted to rather small systems (a few hundred atoms), thus excluding part of the strain effects. The SW empirical potentials are suitable in describing the elastically strained regions and can treat relatively large systems (more than 10^5 atoms), but their accuracy is limited to the core region. Finally, the correct asymptotic limits that will allow us to describe the macroscopic relevant dislocation densities, are given by continuum elasticity theory, that does not allow an atomic description. In order to overcome the aforementioned problems we introduce a multiscale approach:

Multiscale Approach Here, we combine elasticity theory with *ab-initio* and SW empirical potentials calculations in order to accurately describe the dislocations in a wide range of scales. In a first step linear elasticity theory is used to create the initial geometry. Following the method introduced in Sec. 8.2.1 the displacement field of a $n \times n$ dislocation array in the quandrupolar configuration is mapped onto the bulk crystal. The supercells with a dislocation dipole are derived from the center of the array. The dimensions of the array should be large enough to ensure periodicity. A systematic set of supercells with various dislocation distances are constructed. The smallest cells are of the scale of the dislocation cores, while the largest approach the macroscopically observed dislocation distances (or equivalently dislocation densities). Next SW empirical potentials are used to relax the atomic geometries. For the smaller cells equivalent *ab-initio* calculations are performed. The *ab-initio* calculated cells provide information regarding the electronic structure of the dislocation cores. The larger cells provide information regarding the long range elastic interactions and bridge the atomic scale *ab-initio* described zone with the macroscopic relevant zone (dislocation densities lower than 10^{10} cm^{-2} for the edge type dislocations in GaN).

8.5 Cluster approach

In order to calculate the energy factors K and the core radius R_C for the three different edge dislocations, we have performed cluster calculations in large cylindrical systems containing more than 15,000 atoms (Fig. 8.14) using the SW empirical potentials. The radius of the cylinder was more than 100 Å and periodic boundary conditions were applied only along the dislocation line (z-direction). The atomic positions were fully relaxed, except those included in a cylindrical shell of ≈ 6 Å thickness.

The SW potential calculations compared to DFT calculations offer the advantage of easily assigning the excess energy of each individual atom relevant to the energy it has in the ideal bulk position. Thus, the strain energy included in a cylindrical region of radius R around the dislocation

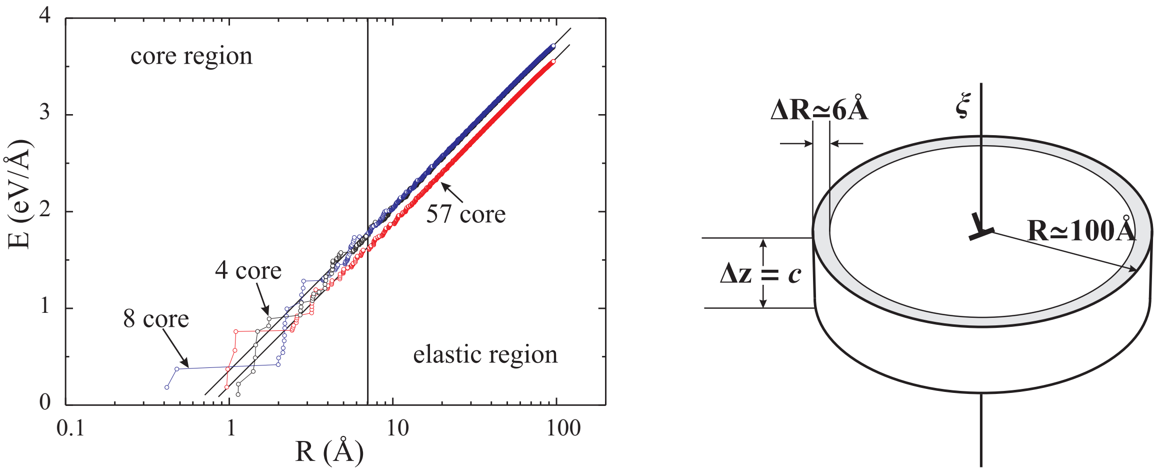


Figure 8.14: *Left*: Total energy stored in a cylinder of radius R as a function of R for the α-type dislocations in GaN. *Right*: Schematic representation of the cell used in the SW-cluster calculations. The shaded cylindrical shell indicates the atoms that are kept fixed during the atomic relaxation.

line can be easily derived by integration. The results for the three cores are shown in Fig. 8.14. As expected from Eq. (8.9) the energies follow a logarithmic behavior for radius R greater than a certain radius R_{C}, which is estimated ≈ 7 Å. For $R < R_{\mathrm{C}}$ the energies exhibit a behavior that cannot be described by a logarithmic dependence.

Since the energy factor depends only on the elastic constants of the material (Eq. (8.10)), the three dislocation cores exhibit the same prelograrithmic factor. The slope of the lines in Fig. 8.14 corresponds to an energy factor of 146 GPa. Based on Eq. (8.10) and and the elastic constants calculated by the SW potential given in Table A.2, a similar energy factor of 153 GPa is found.

As has been mentioned previously the core region cannot be described by the elasticity theory. The core energy depends strongly on the atomic geometry. Hence, the relevant shift of the curves in Fig. 8.14 is associated with the different core geometries. The core energy can be calculated as the energy stored in a cylinder of radius R_{C}. As can be deduced from Fig. 8.14 the 5/7-atom core would be expected to be the energetically most favorable, while the 4- and 8-atom core to energetically degenerate. Similar results based on different parametrizations of the SW potentials have been reported in Ref. [56], while in Ref. [57] the 4-atom core was found to be higher in energy.

In Fig. 8.15 a mapping of the strain energy around a 5/7-core dislocation is shown. The different strain state of the atoms is indicated by the different color code. As can be seen the strain energy distribution is not symmetric around the dislocation line. It is more extended along the direction parallel to the Burgers vector. The highest values of the strain energy are observed near the dislocation line. With increasing distance from the dislocation core the strain energy vanishes. In the inset of Fig. 8.15 a zoom of the core region with different color scale is shown. The highest values of the strain energy can be found for the atoms closest to the dislocation line. This can be explained in terms of the fairly large bond stretching and bond bending that this atoms have

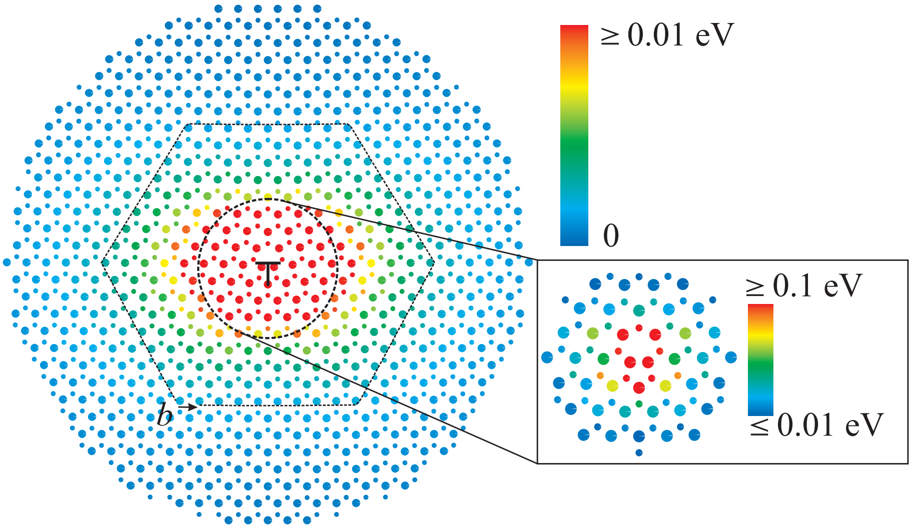


Figure 8.15: Strain energy around a single 5/7-core dislocation. The hexagon around the core indicates the Burgers circuit. *Inset*: Zoom in the core region with different scaling for the elastic energy.

and in terms of the existence of dangling bonds and/or wrong bonds between Ga-Ga and/or N-N atoms. Both reasons increase the energy assigned to each of these atoms.

The SW empirical potentials have been parameterized to describe distortions of the tetrahedral bonds within the elastic limit. It has been shown in Sec. 3.3.1 that the SW potentials fail to correctly describe the metallic Ga-Ga bonds. Since in the core regions distortions beyond the elastic limit and wrong bonds may exist, the SW potentials are not expected to correctly describe this region. Thus, the aforementioned assumptions regarding the relevant stability of the different cores are not fully reliable. In the next Section the *multiscale* approach that combines DFT with SW empirical potential calculations and elasticity theory is used to study the different dislocation cores and to obtain accurate results.

8.6 Geometry, energetics, and electronics of α-type dislocations.

Here we follow the multiscale approach introduced in Sec. 8.4 in order to investigate the geometry, energetics, and electronic structure of charge neutral edge type dislocations in GaN. The displacement field of a 12×12 dislocations array is used to create the initial geometries. It becomes clear from Fig. 8.10 that the size of the array is more than enough to ensure periodicity with an accuracy higher than 10^{-4} Å.

The set of supercells derived from the center of the dislocation array, has been used to simulate dislocation densities in the range between 10^{14} cm^{-2} and 10^{11} cm^{-2}. This corresponds to dislocation distances ranging between ≈ 11 Å and ≈ 180 Å. The smallest cells consist of 108 atoms while the largest consist of $\approx 60,000$ atoms. In the next step using SW empirical potentials, the atomic geometries were relaxed until the forces on all atoms are $< 10^{-3}$ eV/Å. Based on cluster calculations the core radius was roughly estimated to be 7 Å. Hence, for the smallest cells, the distance between the dislocations may not be long enough to avoid core - core overlap. Indeed, allowing all atoms in the supercell to relax and using a quite efficient optimization routine that overcomes shallow local minima, the energetic ground state-ideal wurtzite crystal results. Thus, for the smallest cells a number of atoms between the cores are kept fixed (Fig. 8.16).

For the smallest cells, consisting of up to 144 atoms, equivalent *ab-initio* calculations were performed. The calculations were performed within the DFT, using the PBE-GGA, a plane wave basis set (energy cutoff 50 Ry), and soft Troullier-Martins pseudopotentials. The Brillouin zone was sampled by a 1 × 1 × 2 Monkhorst-Pack mesh. As shown in Sec. 6.4 these settings ensure converged results for the geometry, energetics, and electronics of the systems studied here.

8.6.1 Energetics

Let us first focus on the energetics of the different cores. In order to identify the energetically most favorable core structure we have calculated for all core configurations the dislocation formation energy:

$$E_f = \frac{1}{2L_z}\left(E_{\text{tot}}\left(\text{disl.}\right) - E_{\text{tot}}\left(\text{bulk}\right)\right) \qquad (8.13)$$

Here, E_{tot} (disl.) is the total energy of the supercell containing the two dislocations after atomic relaxation, and E_{tot} (bulk) is the total energy of GaN bulk consisting of the same number of atoms as the dislocation supercell. L_z is the length of the supercell along the dislocation core (i.e., along [0001]), and the factor 1/2 takes into account that there are two equivalent dislocations per supercell.

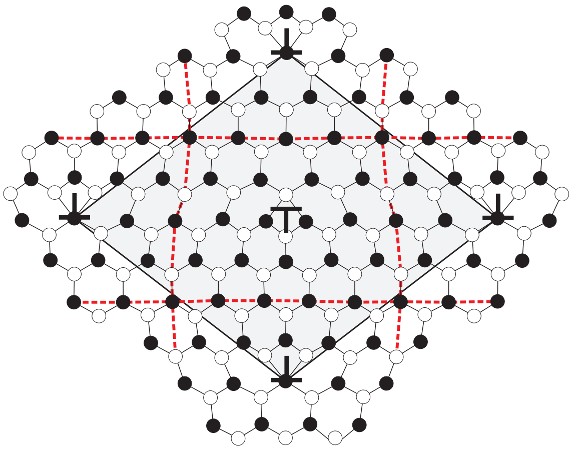


Figure 8.16: Ball and stick model of a small supercell (shaded rhombus) consisting of 144 atoms with two 4-atom core dislocations of opposite sign. One dislocation is located at the center of cell. The other is located at the four corners. Open and filled circles correspond to Ga and N atoms. The dashed lines indicate atoms that are kept fixed during the geometry optimization with SW empirical potentials.

The relevant stability as function of dislocation - dislocation separation for all three edge dislocation structures (8-, 5/7- and 4-atom core) is shown in Fig. 8.17. The 8-atom core structure is used as a reference. The figure contains two major information: (i) The slope of the energy dependence (which describes the defect-defect interaction) and (ii) the absolute value (which represents the energy of the inner core region). Since (i) is mainly realized by elastic interaction it should be well described by the empirical potentials. The inner core region, however, relevant for (ii) is characterized by broken bonds, a situation for which empirical potentials are commonly not well suited and ab-initio calculations become mandatory. We therefore use the results for the largest ab-initio calculated supercell (the 144 atom cell) to shift for each dislocation type the empirical po-

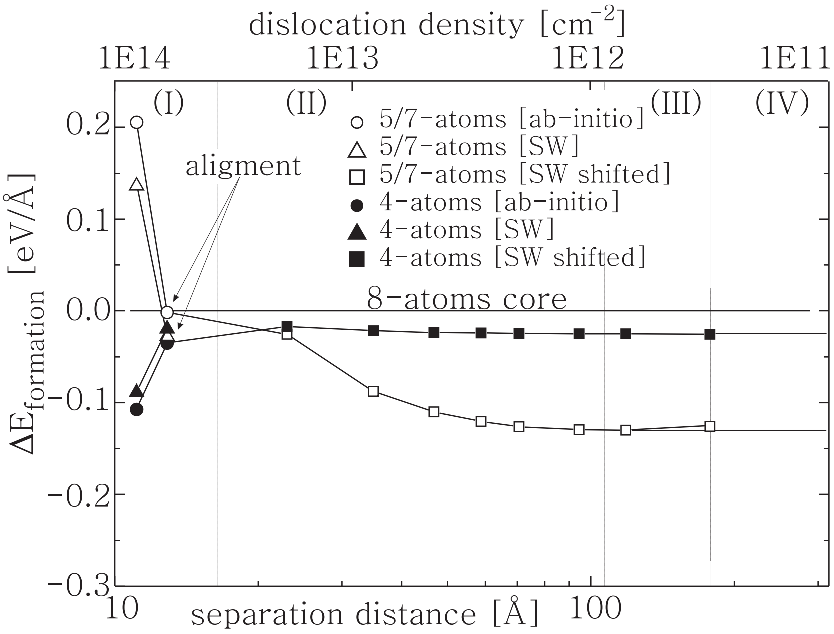


Figure 8.17: Relative formation energies per dislocation as function of dislocation separation (lower axis) and dislocation density (upper axis). The formation energy of the 8-atom core structure has been used as reference energy. The symbols mark the calculated energies: ab-initio calculations (cycles), shifted (squares) and unshifted empirical potentials (triangles). The vertical dashed lines separate the four zones discussed in text. Alignment shows the cell size for which the ab-initio and the empirical potential calculations have been aligned.

tentials (see Fig. 8.18). The shift is small (2 (24) meV/Å for the 4-atom (5/7-atom) core structure) indicating the high quality of the empirical potential.

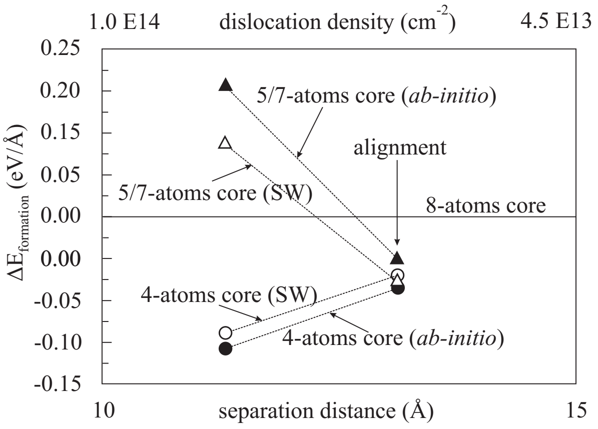


Figure 8.18: Enlargement of Region I of Fig. 8.17. The shift required to align *ab-initio* and SW empirical calculated cells is rather small: 2 (24) meV/Å for the 4-atom (5/7-atom) core structure.

In order to verify the accuracy of the multiscale approach Fig. 8.17 can be divided into four different regions. The first region (I) allows a direct comparison between ab-initio and empirical potential calculations. As can be seen the error is negligible (0.069 and -0.018 eV/Å for the 5/7- and 4-atom core structure, respectively) compared to the actual energy difference between the core structures. In the third region (III) the energy dependence becomes independent from the dislocation distance which is the expected behavior when linear elastic theory applies. This region connects empirical potentials with elastic theory and thus can be used to extrapolate the formation energy to experimentally relevant dislocation densities. Finally, the second region (II), accessible by empirical potential calculations only, bridges between the first (ab initio microscopic) and the third (experimentally relevant macroscopic) zone. The multiscale analysis (Fig. 8.17) clearly shows that for stoichiometric and neutral dislocations the 5/7-atom core is the most stable structure. This result is in apparent contrast to the assignment based on the image simulations described in Sec. 8.3 where the 4-atom core structure had been identified. We have therefore tested the effect strain may have on the relative stability of the three investigated core structures.

8.6.2 Effect of strain on the stability of the dislocation cores

It is well known that HVPE grown material to be under significant tensile strain [244, 245, 246]. In general the strain in the GaN layers is the sum of the extrinsic compressive and the intrinsic tensile strain:

$$\epsilon_{\text{tot}} = \epsilon_{\text{extr}} + \epsilon_{\text{intr}} \tag{8.14}$$

The extrinsic strain arises from the mismatch of the thermal expansion coefficients between the epilayers and the substrate and is induced by the cooling process after growth. For growth on e.g. sapphire substrates, the cooling process will result to faster shrinking of the substrate and the GaN epilayers will become compressively strained. Experimental evidences of the intrinsic strain in epitaxially grown GaN films has been reported in Ref. [244]. Böttcher *et al.* showed that the origin of the intrinsic tensile strain is the high temperature island coalescence during epilayer growth (see Fig. 8.19). As long as the gap closure between two randomly placed islands by elastic deformation, which reduces the surface energy at the expense of elastic energy, is energetically favorable, tensile strain will develop in the epilayers.

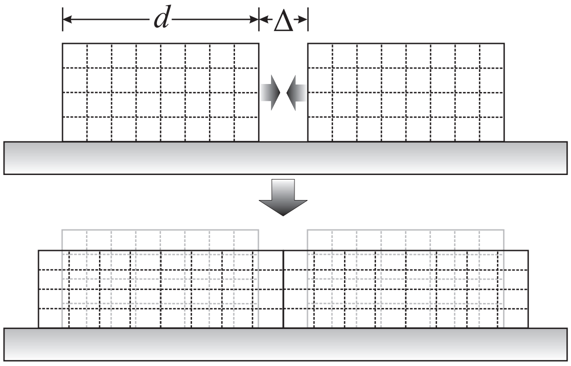


Figure 8.19: Schematic representation of the formation of intrinsic tensile strain due to island coalescence. The islands are assumed to be mobile on the substrate and the resulting strain is homogenous. If the islands are assumed anchored to the substrate the strain will be heterogeneous, with the highest values occurring at the boundaries.

In order to test the dependence of the dislocation core structure on strain we calculated the formation energy of the dislocations in large supercells (dislocation separation $\sim$ 177 Å; $\approx 10^5$ atoms). Various compressive and tensile biaxial strains (see Eq. (A.18)) were applied. The calculations have been performed using the SW empirical potentials. The results are shown in Fig. 8.20. The relative energy of the 5/7-atom core structure with respect to the 8-atom and 4-atom core structure dramatically changes when going from compressive to tensile strain: While the 5/7-atom core structure is the most stable structure for compressive and zero strain conditions, the 4-atom core becomes energetically most favorable for a tensile strain larger than ≈ 0.015. As has been mentioned, HVPE grown material is well known to be under significant tensile strain and hence

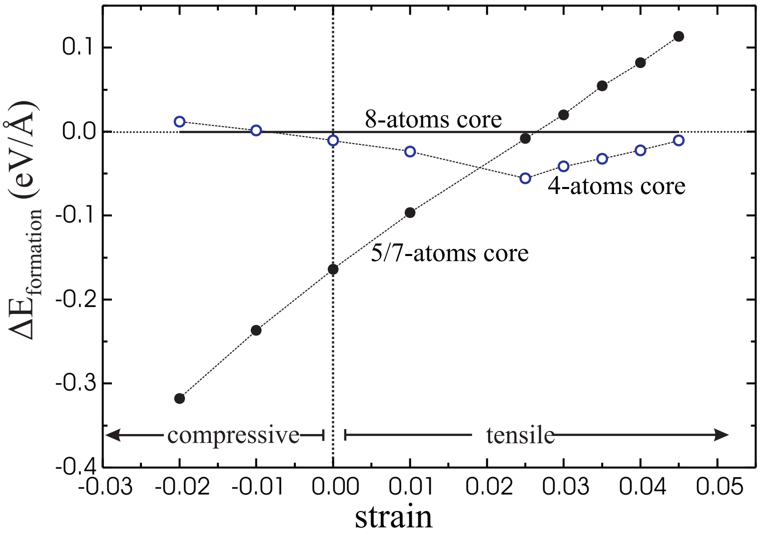


Figure 8.20: Relative formation energies as a function of the applied biaxial strain ϵ. The formation energy of the full core structure has been used as reference energy.

the stabilization of the 4-atom core structure can be assumed as a consequence of the significant intrinsic-tensile strain in the sample.

8.6.3 Effect of vacancies on the structure and stability of the 4-core structure

As has been discussed in Sec. 8.1 non-stoichiometric cores are present for highly doped samples (n- as well as p-type doping) and for Ga- or N- rich conditions. Although, in the present studies the focus is on undoped material under stoichiometric growth conditions, it is worth to investigate the effect of vacancies on the structure and stability of the dislocations. To do this, the formation energy of isolated Ga and N vacancies at various positions in the strain field around the dislocation line were calculated.

The calculations have been performed in large supercells consisting of $\approx 25,000$ atoms using SW empirical potentials. The atomic positions were fully relaxed until the forces on all atoms were $< 10^{-3}$ eV/Å. Since the strain is not isotropic around the dislocation (see Fig. 8.15) the vacancy formation energy as function of vacancy-dislocation distance has been calculated along various high symmetry directions. This is shown in Fig. 8.21(a) which also shows the schematic supercell (solid lines) with the arrangement of the dislocations. The dots symbolize the three inequivalent directions chosen (vertical up - VU, vertical down - VD, and horizontal - H). The corresponding positions at the dislocation core are shown in Fig. 8.21(b).

Placing the vacancies as described above, the formation energy for both Ga and N vacancies has been calculated as function of vacancy-core distance. The result is shown in Fig. 8.22(a) for the example of the Ga vacancies and in Fig. 8.22(b) for the N vacancies. As can be seen in Fig. 8.22 the dislocation core strongly attracts vacancies (the reason is that vacancies lead to a significant relaxation of the strain field). The capture zone around the dislocations (i.e. the maximum separation at which the vacancies still "feel" an attraction) has a radius of ~ 15 Å for

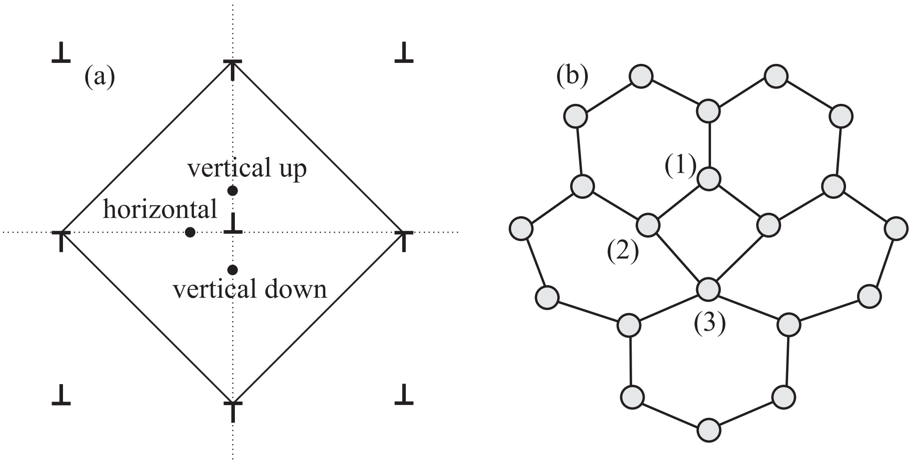


Figure 8.21: (a)Schematic arrangement of the dislocation and the vacancies in the supercell. The solid lines mark the supercell and the points indicate the different directions at which the vacancy has been positioned. (b) Schematic arrangement of the vacancies at the dislocation core. The three directions introduced in Fig. 1 correspond here to: (1) - vertical up, (2) - horizontal, and (3) - vertical down.

the Ga vacancies and ~ 10 Å for the N vacancies. Beyond that region the formation energy of the vacancy coincides with that of a vacancy in unstrained GaN bulk.

An interesting and fully unexpected feature occurs if the vacancy gets into the core: If the vacancy is placed in position (2) (see Fig. 8.21(b)) the 4-atom core structure becomes unstable and forms spontaneously (i.e. without having to overcome an activation barrier) the 8-atom core vacancy structure. Indeed, also the formation energy is degenerate and the structures resemble each other. This allows us to align out results for the vacancy structures with the *ab-initio* based calculations for the charge neutral dislocations reported by Wright and Grossner [55]. A comparison with the energy of the 8-atom core structure (see again Fig. 8.22) shows that the vacancy structures are always less stable than the 8-atom core structures. Only under very Ga-rich conditions the energetically most favorable vacancy structure (which is the 8-atom core N vacancy) gets energetically degenerate. For all other conditions the vacancies are energetically unfavorable and thus thermodynamically unstable.

We can thus conclude that vacancies may and will become important under extreme Ga-rich conditions (which are characteristic for MBE grown GaN but not for HVPE grown GaN) and/or for highly doped GaN (where vacancies may become charged and act as compensating centers). However, the samples investigated in the present work are highly compensated (as characteristic for HVPE where compensating hydrogen atoms are highly abundant) and vacancies will remain neutral. Thus, while vacancies in general may have a dramatic effect on both dislocation structure and energetics for HVPE grown GaN (the situation of the HRTEM experiments described in Sec. 8.3) the formation of vacancies is highly endothermic. Thus, as shown above the vacancy concentration is too small to affect dislocation energetics or structure.

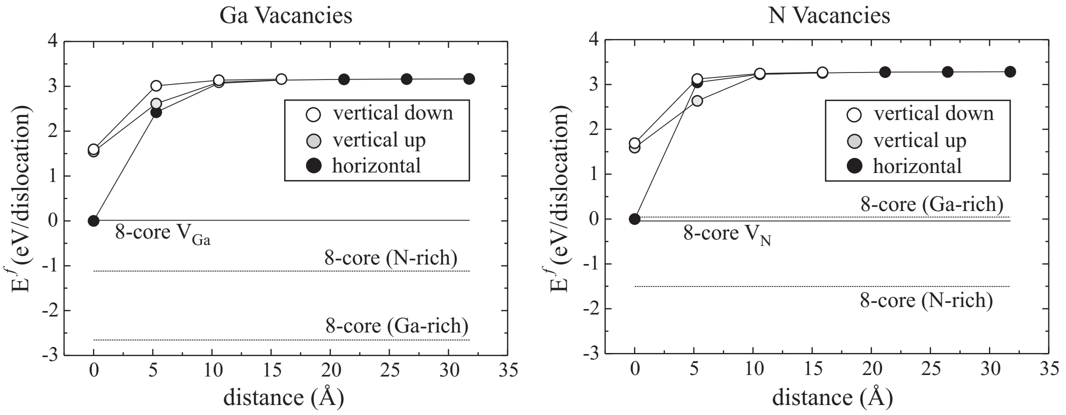


Figure 8.22: (a) & (b) Formation energy (in eV per dislocation and per unit cell length) to create a Ga(N)-vacancy as function of distance to the core and for placing the vacancies in various positions around the 4-core dislocation line respectively (see Fig. 8.21). The energy reference (energy zero) is the 8-core structure with a Ga(N) vacancy respectively. Further included are formation energies of the 8-core structure without vacancies for Ga and N-rich conditions as calculated by Wright & Grossner [55].

8.6.4 Electronic Structure

Let us now focus on the electronic character of the dislocation cores. Using the results from the *ab-initio* calculations first the density of states (DOS) for the three different dislocation cores has been calculated. The results are shown in Figs 8.23(a), (b), and (c) for the 4-, 8-, and 5/7-atom core structures respectively. The DOS clearly shows that all cores have states in the bandgap and are thus electrically active. This is the expected behavior for the 5/7- and 8-atom core structure since both have broken bonds in the core region. Indeed, the existence of these states has been reported also in previous density-functional theory studies by Lee *et al.* [234] and by Fall *et al.* [235]. The 4-atom core structure, however, is fully reconstructed. Each core cation (anion) is surrounded by four anions (cations) and no broken bonds exist in the core. Thus the 4-atom core would be expected to be electrically inactive in apparent conflict to the calculated DOS.

In order to identify the mechanism inducing these states the corresponding defect wave functions have been analyzed (Figs 8.23(b), (d), and (f) for the 4- 8-, and 5/7-atom core respectively). As can be seen the states are characterized by orbitals between two Ga atoms and are quantitatively similar to the states forming the metallic bonds in bulk Ga. For lower values of the isosurfaces the wavefunctions are seen to extend into the strained regions away from the core. A similar behavior has been reported by Fall *et al.* for the 8-atom core structure [235]. The origin of these states is the large local strain-field in the vicinity of the dislocation line and (ii) the small lattice constant characteristic for GaN (the lattice constant of GaAs, e.g., is larger by 20%). Both effects result in a Ga-Ga separation of 2.74 Å which is close to the bond length in bulk Ga (2.44 Å).

It is interesting to note that the strain-induced formation of empty bond states is not restricted to the 4-atom core structure but appears to be a general feature of all α-type dislocations in GaN

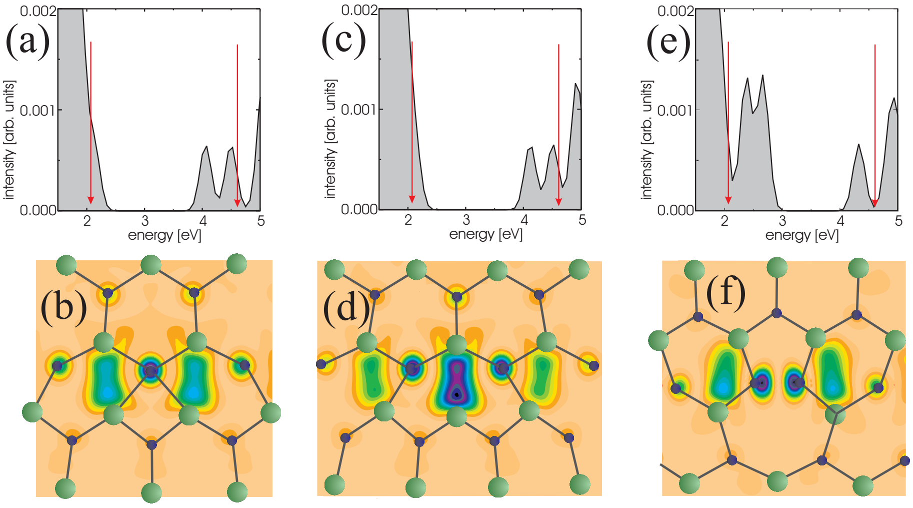


Figure 8.23: Density of states and contour plots of dislocation induced deep electronic states for the (a), (b) 4-atom core, (c), (d) 8-atom core, and (e), (f) 5/7-atom core structure respectively. Large (small) balls correspond to Ga (N) atoms.

(Fig. 8.23(b),(d), and (f)). Indeed, recent EELS studies revealed the appearance of states at an energy of 3.2 eV above the valence band edge, which is in excellent agreement with the energetic position we find from our calculation [228]. The main difference between the fully coordinated 4-atom core structure and the other core structures is that in the latter case also the dangling bond orbitals clearly contribute to this state.

8.7 Summary

Here we proposed a *multiscale* approach that combines *ab-initio* and empirical potential calculations with elasticity theory and provides an accurate tool in the study of the geometry, energetics, and electronic structure of dislocations.

Based on the *multiscale* approach a core structure for dislocations in GaN that has never been assumed in DFT calculations, neither has been experimentally identified, could be revealed. This core exists of a four atom ring and is fully reconstructed. Direct experimental evidence of the 4-atom core has been given by recent HR-TEM experiments on HVPE grown GaN. Despite the fact that the 4-atom core is fully reconstructed the electronic structure clearly shows the existence of deep dislocation related states. This result is in contrast to the classical picture of dislocations according to which the existence of deep states is related to the presence of broken bonds (and thus under-coordinated atoms) in the dislocation core. An analysis of the deep electronic states showed that they are related to the giant local strain field around a threading dislocation in GaN which is strong enough to induce a partial metallization in form of metallic-like Ga-Ga bonds. It is crucial to note that in contrast to the classical picture, where deep electronic states can be removed by reconstructing the core, strain induced deep states are rather insensitive to the specific core reconstruction and prevail even for fully reconstructed core structures.

Chapter 9

Synopsis

The goal of this thesis has been to provide a microscopic understanding of selected extended defects in and on group III-Nitrides and their effect on growth process and on the properties of the grown epilayers.

The main challenge in describing extended defects was the presence of two very different length scales: Short range phenomena on the atomic scale in the vicinity of the defect core and long range phenomena due to the elastic field around the core. We therefore had to develop a multiscale approach which connects the length scales. The multiscale approach enables to describe the physical systems in a wide variety of scales: From the atomic level (i.e., from the dislocation core which consists of a few atoms) all the way up to experimentally accessible scales (as needed e.g. to describe the macroscopically relevant dislocation densities which require system sizes consisting of $\sim 10^5$ atoms). The multiscale approach developed here has been applied to study edge type threading dislocations in GaN. However, with appropriate modifications (e.g., by constructing/using appropriate empirical potentials) the same concept can be applied to describe systems as different as edge, screw, or mixed type dislocations and in various materials. The other key approach used here is a combination of *ab-initio* calculations with the thermodynamic concept of chemical potentials. Based on this combination a first link between growth experiments and theory could be established: Particularly, phase diagrams indicating the thermodynamic stability as function of the growth conditions could be identified.

In a first application of these methods, extended defects on GaN surfaces (e.g., steps/vicinal surfaces) have been considered. Our results are in line with previous studies and allow to derive general trends of group III-Nitride surfaces: Under Ga-rich conditions they are stabilized by metal-rich surfaces (consisting of a few Ga-layers). An interesting observation which emerged from these studies is that surface steps/vicinal surfaces may spontaneously form under N-rich conditions and will thus stimulate surface roughening. A second application addressed the formation of small Ga islands on top GaN surfaces. These islands can be regarded as precursors of Ga droplet formation: The formation of droplets prevents the growth of a thicker Ga wetting layer and modifies the growth from 2D to 3D.

In a further study the kinetically stabilized ordering in AlGaN alloys which form monoatomic GaN-AlN superlattice has been considered. Our results show that ordering induces a redshift in the bandgap which is consistent with experiment. A detailed analysis showed that in contrast to zincblende semiconductors the dominant effect in bandgap reduction is not due to level repulsion by Brillouin zone folding but caused by strong carrier localization. The effect could be technologically used: Ordering changes the symmetry of the valance band maximum from z-polarized to xy-polarized and will thus enhance the light emission along the c-axis. As a last example of multiscale simulations the structure, electrical activity and stability of edge type threading dislocations in GaN has been addressed. Based on this approach we could demonstrate that the electrical activity is **not** a consequence of the core structure but originates from the giant strain field around edge dislocations. An important consequence of this result is that electrical activity is an intrinsic property of any edge dislocation in GaN irrespective of the specific core structure.

The wide variety of phenomena ranging from surfaces, alloys, nucleation, to defects which could be successfully solved using *ab-initio* based multiscale techniques demonstrate the power and the potential of this technique. The key achievement is that it enables to describe systems from the microscopic to the macroscopic scale with an accuracy that is in principle limited only by the accuracy of the *ab-initio* calculations. So far, through out this work we assumed the system to be in thermodynamic equilibrium with its environment. When kinetic effects become important, e.g, to simulate the growth of ordering in AlGaN alloys, other approaches, such as kinetic Monte Carlo simulations, are needed and expected to provide a better microscopic understanding. This could be the ansatz for a future work regarding both methodological aspects and the study of group III-Nitrides.

Part III

Appendices

Appendix A

Bulk Properties of GaN

The calculations on structural (lattice parameters) properties and elastic constants of wurtzite GaN are presented here.

The calculation of equilibrium bulk properties of GaN is needed in the surface and defect calculations. Using the equilibrium lattice constants in surface calculations ensures that surfaces are strain free. Moreover, it is well known that HVPE grown GaN films are under significant intrinsic biaxial tensile stress (see Sec. 8.6.2). In order to study the effect biaxial strain has on the energetics and structure of edge type threading dislocations, the relation between the strain components for biaxial stress (see Eqs. (A.16) and (A.18)) and thus the elastic constants is needed. SW empirical potential are used in this thesis to describe the long range elastic interaction between dislocations (see Sec. 8.6). Thus, the elastic constants are needed in order to fit the parameters of the potentials and test their reliability to describe elastically strained material.

A.1 Structural properties

Unlike the cubic (fcc) III-V semiconductors such as GaAs, the thermodynamically most favorable phase of GaN is the hexagonal wurtzite structure (Space group: P6_3mc). In Fig. A.1 the GaN wurtzite structure (α-phase) and the zincblende counterpart (β-phase) are shown. The first to report that the thermodynamically most stable phase of GaN is the wurtzite structure were Juza and Hahn [247]. GaN in the zincblende structure has been reported for growth on cubic substrates: Growth on (001) GaAs, on cubic SiC, on MgO, and on (001) Si (see e.g. the review by Strike and Morkoç [1] and references therein).

The primitive vectors of the zincblende cell are $\boldsymbol{\alpha}_1 = \frac{\alpha_{\text{zinc}}}{2}(1,1,0)$, $\boldsymbol{\alpha}_2 = \frac{\alpha_{\text{zinc}}}{2}(1,0,1)$, and $\boldsymbol{\alpha}_3 = \frac{\alpha_{\text{zinc}}}{2}(0,1,1)$, where α_{zinc} is the zincblende lattice constant. The primitive unit cell has two atoms, one Ga at $(0,0,0)$ and one N at $\left(\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)$ in units of $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, and $\boldsymbol{\alpha}_3$. The determination of the equilibrium geometry is straightforward since there is only one lattice constant.

The primitive wurtzite cell has 2 Ga atoms at $(0,0,0)$, $\left(\frac{2}{3},\frac{1}{3},\frac{1}{2}\right)$ and 2 N atoms at $(0,0,u)$,

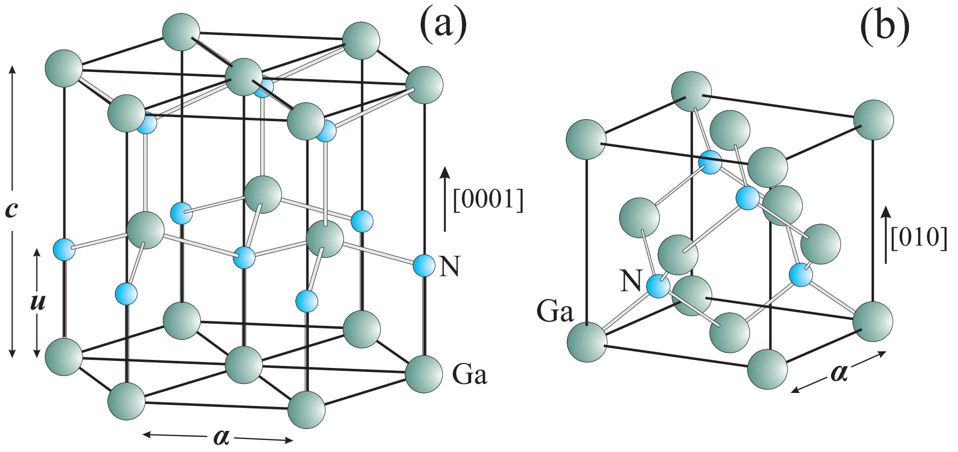


Figure A.1: Ball and stick models of the (a) α- (wurtzite) and (b) β- (zincblende) GaN structures.

$\left(\frac{2}{3}, \frac{1}{3}, \frac{1}{2} + u\right)$ in units of the primitive vectors: $\boldsymbol{\alpha}_1 = \alpha_{\text{wur}}\left(\frac{1}{2}, \frac{\sqrt{3}}{2}, 0\right)$, $\boldsymbol{\alpha}_2 = \alpha_{\text{wur}}\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}, 0\right)$, and $\boldsymbol{\alpha}_3 = \alpha_{\text{wur}}\left(0, 0, \frac{c}{\alpha_{\text{wur}}}\right)$. Here α_{wur} and c are the lattice constants of the wurtzite structure and u is a parameter characterizing the internal structure (i.e., the basis in the supercell). For the ideal wurtzite structure all Ga-N bonds are identical which gives: $u = 3/8$ and $c/a = \sqrt{8/3}$. For the determination of the equilibrium geometry, a three step procedure is used. There are three independent parameters to be optimized, the lattice constant α_{wur}, the ratio c/α_{wur}, and the internal parameter u. First, an ideal wurtzite structure is assumed (i.e, u and c/α are fixed at their ideal values) and the equilibrium volume is determined by varying the lattice constant α_{wur}. Next, the volume is kept fixed and the c/α_{wur}-ratio is optimized. At the end the new c/α_{wur}-ratio is kept fixed and once more the volume is determined. At each step the internal parameter u was optimized by fully relaxing the atomic positions.

In both cases (zincblende and wurtzite structures) the final optimum values were obtained by fitting the total energy versus volume, using the Murnaghan equation of state [250] (see Sec. A.3). The calculated values of the lattice constants, the bulk modulus and the bulk modulus derivatives are given in Table A.1. The calculations have been performed employing DFT in the PBE-GGA. Troullier-Martins pseudopotentials [101] have been used and a 50 Ry cutoff energy has been applied. The Brillouin zone was sampled by $4 \times 4 \times 1$ and $3 \times 3 \times 3$ Monkhorst-Pack k-point samplings for the wurtzite and the zincblende structures respectively. For a detailed discussion and convergence checks with respect to exchange-correlation functionals see e.g. [82, 251].

A.2 Elastic Constants

The elastic properties and the elastic constants of the wurtzite GaN crystal as needed to describe GaN in a continuum picture will be derived here.

Table A.1: Lattice parameters and bulk modulus of $\alpha-$ and $\beta-$ GaN. $d_{\text{III}-\text{V}}$ is the nearest neighbor cation-anion distance. The experimental lattice constants for the wurtzite phase and correspond to bulk crystals grown at extreme conditions of temperature and pressure (1800 K and 15 kbar) while for the zincblende phase correspond to film grown on GaAs [248]. The values for both wurtzite and zincblende correspond to 25 °C temperature. For the experimental wurtzite bulk modulus and bulk modulus derivative we give the range of values published over the past years (see Ref. [4] and Refs. therein), while for the zincblende structure the values are from [249].

	α (Å)	c/α	u	V_0/dimer $\left(\text{Å}^3\right)$	$d_{\text{III}-\text{V}}$ (Å)	B (Mbar)	B'
wurtzite							
this work	3.182	1.6407	0.377	22.89	1.968	1.84	4.62
experiment	3.189	1.626	-	22.83	1.955	1.73-2.45	3.2-4.3
zincblende							
this work	4.516	-	-	23.02	1.955	1.71	4.41
experiment	4.511	-	-	22.95	1.953	1.73	3.70

Let x_i, $(i = 1,\ 2,\ \text{or}\ 3)$ be orthogonal cartesian coordinates and σ_{ij} the stresses acting upon a cube of the material under consideration. σ_{ij} is the i_{th} component of the force per unit area acting on a plane whose normal is parallel to the positive x_j direction (see Fig. A.2). At rest each infinitesimal volume must be in mechanical equilibrium[1]:

$$\sigma_{ij} = \sigma_{ji}. \tag{A.1}$$

Moreover, no net force act on the element:

$$\frac{\partial \sigma_{1i}}{\partial x_1} + \frac{\partial \sigma_{2i}}{\partial x_2} + \frac{\partial \sigma_{3i}}{\partial x_3} + f_i = 0 \qquad i = 1,\ 2,\ 3 \tag{A.2}$$

where f_i is the i_{th} component of the body force per unit volume. Eqs (A.2) are the *equilibrium equations* of classical elasticity theory.

When stresses are acting upon a body, the body deforms. We denote by u_i the displacements at the point $\boldsymbol{r}$ due to the applied stresses. The strains are defined as:

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right). \tag{A.3}$$

The strains define the deformation and are dimensionless.

Hooke's law states that in a linear elastic material the strain is directly proportional to the stress. It is valid only in the limit that the stresses and the displacement are approaching zero. Otherwise we say that we are in the *nonlinear* region. The generalized Hooke's law has the following

---

[1] No net torque is present in the element. Additionally internal torques are assumed to contribute negligible to the strain energy and thus are ignored.

form[2]:

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl}. \tag{A.4}$$

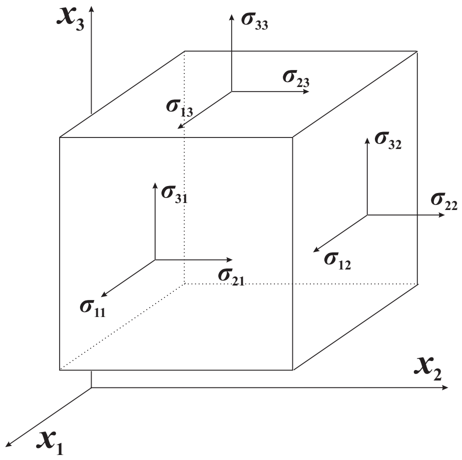


Figure A.2: Stress distribution on an infinitesimal volume element. σ_{ij} is the i_{th} component of the force per unit area acting on a plane whose normal is parallel to the positive x_j direction.

Here the coefficients C_{ijkl} are the elastic constants of the material under consideration. The above equation can be written in a matrix form:

$$\{\sigma_{ij}\} = \begin{array}{cc} & (kl) \rightarrow \\ (ij) & \{C_{ijkl}\}\,\{\epsilon_{kl}\} \\ \downarrow & \end{array} \tag{A.5}$$

The 9×9 matrix $\{C_{ijkl}\}$ relates the nine stress elements σ_{ij} with the 9 strain elements ϵ_{kl}. The elastic constants can be written in a contracted notation as C_{mn}, where m and n correspond to a pair of ij or kl indices according to the following rule:

$$\begin{array}{ccccccccccccc} ij & \text{or} & kl & 11 & 22 & 33 & 23 & 31 & 12 & 32 & 13 & 21 \\ m & \text{or} & n & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array}. \tag{A.6}$$

Due to symmetry reasons the 9×9 matrix representation in Eq. (A.5) can be reduced to a 6×6 representation, where the matrix of the elastic constants is symmetric:

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{31} \\ \sigma_{12} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{12} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{13} & C_{23} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{14} & C_{24} & C_{34} & C_{44} & C_{45} & C_{46} \\ C_{15} & C_{25} & C_{35} & C_{45} & C_{55} & C_{56} \\ C_{16} & C_{26} & C_{36} & C_{46} & C_{56} & C_{66} \end{bmatrix} \begin{bmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \epsilon_{33} \\ \gamma_{23} \\ \gamma_{31} \\ \gamma_{12} \end{bmatrix}. \tag{A.7}$$

Here $\gamma_{ij} = 2\epsilon_{ij}$ with $i \neq j$ correspond to the shear strains. The above 21 elastic constants C_{ij} can be further reduced by taking into account the symmetry elements of the crystal under consideration. For hexagonal crystals, with the third axis perpendicular to the basal plane (as in wurtzite crystals),

---

[2] The Einstein convention (summation over repeated indices) is employed.

the matrix of the elastic-constants has the following form:

$$\{C\} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{bmatrix}. \tag{A.8}$$

Here

$$2C_{66} = C_{11} - C_{12}. \tag{A.9}$$

Thus there are 5 independent elastic constants for the wurtzite structure to be calculated. Eq. (A.9) indicates that hexagonal crystals are isotropic in the basal plane, namely $\{C\}$ is invariant to a rotation about the z-axis.

When a unit element volume deforms by differential strain increments $\mathrm{d}\epsilon_{ij}$, the stresses do work on the element by an amount[3]:

$$\mathrm{d}U = \sigma_{ij}\mathrm{d}\epsilon_{ij} = C_{ijkl}\epsilon_{kl}\mathrm{d}\epsilon_{ij}. \tag{A.10}$$

By integration of the above equation (using the notation indicated in Eq. (A.6) for the strain elements) we get:

$$U = \frac{1}{2}\sum_{i=1}^{6}\sum_{j=1}^{6} C_{ij}\epsilon_i\epsilon_j. \tag{A.11}$$

Here U is the elastic energy density.

In general applying strain onto a crystal will change the angles between and the length of the primitive vectors of the crystal. To express this mathematically the strain components can be represented by a symmetric strain matrix $\boldsymbol{\epsilon}$:

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 & \epsilon_6/2 & \epsilon_5/2 \\ \epsilon_6/2 & \epsilon_2 & \epsilon_4/2 \\ \epsilon_5/2 & \epsilon_4/2 & \epsilon_3 \end{bmatrix}. \tag{A.12}$$

Then, the primitive vectors $\boldsymbol{\alpha_i}$ $(i = 1,\, 2,\, 3)$ are transformed to the new vectors under strain by

$$\begin{bmatrix} \boldsymbol{\alpha'}_1 \\ \boldsymbol{\alpha'}_2 \\ \boldsymbol{\alpha'}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \boldsymbol{\alpha}_3 \end{bmatrix} \cdot (\boldsymbol{I} + \boldsymbol{\epsilon}) \tag{A.13}$$

---

[3] The Einstein convention is employed again.

Table A.2: Wurtzite GaN elastic constants (in GPa).

	C_{11}	C_{12}	C_{13}	C_{33}	C_{44}	C_{66}	$2C_{13}/C_{33}$	B_0
Calculations by Wright [33]	367	135	103	405	95	116	0.509	$202^{(a)}$
Calculations by Kim *et al.* [252, 253]	$396^{(c)}$	$144^{(c)}$	100	392	$91^{(c)}$	126	0.510	$207^{(a)}$
Measurement by Savastenko and Sheleg [254]	296	130	158	267	24 ± 2	83	1.184	$195^{(a)}$
Measurement by Polian *et al.* [144]	390	145	106	398	105	123	0.533	$210^{(a)}$
Measurements by Tagaki *et al.* [255]	374	106	70	379	101	134	0.369	$180^{(a)}$
Present calculations (PBE-GGA)	378	100	104	404	104	139	0.515	$197^{(a)}/186^{(b)}$
Present calculations (SWp)	361	142	114	353	99	109	0.646	$201^{(a)}/208^{(b)}$

(a) Bulk modulus calculated using Eq. (A.24).

(b) Bulk modulus calculated using the Murnaghan equation of state (Eq. (A.23)).

(c) Values derived from the zincblende elastic constants using Martin's transformation [256]. Martin's transformation provides an analytic expression between the elastic constants of cubic-zincblende and hexagonal-wurtzite crystal structures.

with $\boldsymbol{I}$ being the unit matrix.

The five independent elastic constants of the wurtzite structure (Eq. (A.8)) can be determined from Eq. (A.11) by calculating equilibrium structure and total energy as function of the strain variable δ for five different strained configurations. To classify the structures the form $D = [\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6]$ is used. Here we follow the selection used by A. F. Wright [33] and use the following five modes:

- $D_1 = [\delta, \delta, 0, 0, 0, 0] \rightarrow U_1 = (C_{11} + C_{12})\,\delta^2$,
- $D_2 = [\delta, \delta, -2\delta, 0, 0, 0] \rightarrow U_2 = (C_{11} + C_{12} - 4C_{13} + 2C_{33})\,\delta^2$,
- $D_3 = [0, 0, \delta, 0, 0, 0] \rightarrow U_3 = \frac{1}{2} C_{33}\delta^2$,
- $D_4 = [0, 0, 0, 0, 0, \delta] \rightarrow U_4 = \frac{1}{4}\,(C_{11} - C_{12})\,\delta^2$,
- $D_5 = [0, 0, 0, \delta, \delta, 0] \rightarrow U_5 = C_{44}\delta^2$.

The elastic constants are determined by calculating the energies E_i ($U_i = E_i/\Omega_0$, Ω_0: the volume of the equilibrium structure) for small values of δ between -0.02 and $+0.02$. The calculated values using the two different methods (DFT within the PBE-GGA and the SWP), are given in Table A.2. Also included are calculated and experimentally measured elastic constants from previous studies. The calculations by Wright [33] were performed using DFT within the LDA using the plane waves pseudopotential method. The results by Kim *et al.* [252, 253] are also from DFT calculations but implemented via the full-potential muffin-tin orbital method. The measured values were derived from temperature-dependent broadening of powder x-ray diffraction spectra by Savastenko and Sheleg [254], and Brillouin scattering by Polian *et al.* [144] and Takagi *et al.* [255].

Another useful way to describe the mechanical properties of crystals are the elastic compliance constants S_{ij}. Following Eq. (A.5), the elastic compliance relates the strains with the stresses as follows:

$$\{\epsilon_{ij}\} = \begin{array}{cc} & (kl) \rightarrow \\ (ij) & \{S_{ijkl}\}\,\{\sigma_{kl}\} \\ \downarrow & \end{array} . \tag{A.14}$$

The elastic compliance constants can be calculated by inversion of the elastic constants matrix (Eq. (A.7)):

$$\begin{aligned} S_{11} &= \frac{C_{11}C_{33} - C_{13}^2}{(C_{11} - C_{12})\left[C_{33}\,(C_{11} + C_{12}) - 2C_{13}^2\right]} \\ S_{12} &= \frac{C_{12}C_{33} - C_{13}^2}{(C_{11} - C_{12})\left[C_{33}\,(C_{11} + C_{12}) - 2C_{13}^2\right]} \\ S_{13} &= \frac{C_{13}}{C_{33}\,(C_{11} + C_{12}) - 2C_{13}^2} \\ S_{33} &= \frac{C_{11} + C_{12}}{C_{33}\,(C_{11} + C_{12}) - 2C_{13}^2} \\ S_{44} &= \frac{1}{C_{44}}\ . \end{aligned}$$

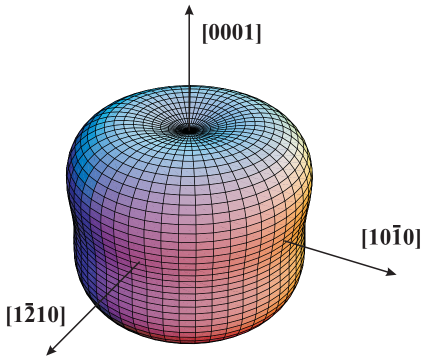


Figure A.3: Reciprocal Young's modulus as function of the crystallographic direction for wurtzite GaN.

The reciprocal Young's modulus Y^{-1} describes the hardness of the material. For hexagonal materials the reciprocal Young's modulus along an arbitrary direction at an angle θ with respect to the c-axis is given by [188, 257]

$$Y^{-1}(\theta) = S_{11}\sin^4(\theta) + S_{33}\cos^4(\theta) + (S_{44} + 2S_{13})\sin^2(\theta)\cos^2(\theta)\,. \tag{A.15}$$

Fig. A.3 shows a three dimensional plot of the hardness as function of the crystallographic direction. In Fig. A.4 polar plots of the reciprocal Young's modulus in the basal plane (a) and in the $\left(01\bar{1}0\right)$ plane (b) are shown. The hardness in the basal plane is isotropic for all the group III-Nitrides (see Fig. A.4(a))[4]. In epitaxial growth along the c-axis of the III-Nitrides the strain is caused by the lattice and thermal mismatch between the layer and the substrate. The grown material can freely relax along the c-direction. The isotropic hardness in the basal plane results in a biaxial stress ($\sigma_1 = \sigma_2$, $\sigma_3 = 0$ - the crystal is laterally strained within the (0001) plane but freely relaxes along the [0001] direction). Thus, it is of interest to study the strain in a wurtzite crystal

---

[4] The fact that the Young's modulus of GaN and AlN coincide in the basal (see Fig. A.4(a)) can be attributed to the different approaches used for the calculation of the corresponding elastic constants (for GaN the PBE-GGA elastic constants calculated here were used while for AlN and InN the LDA calculated elastic constants by Wright [33] were used). If elastic constants calculated with the same approach are used, AlN will appear slightly softer in the basal plane [190]. In Fig. A.4(b) using elastic constants calculated with the same approach has no effect on the picture.

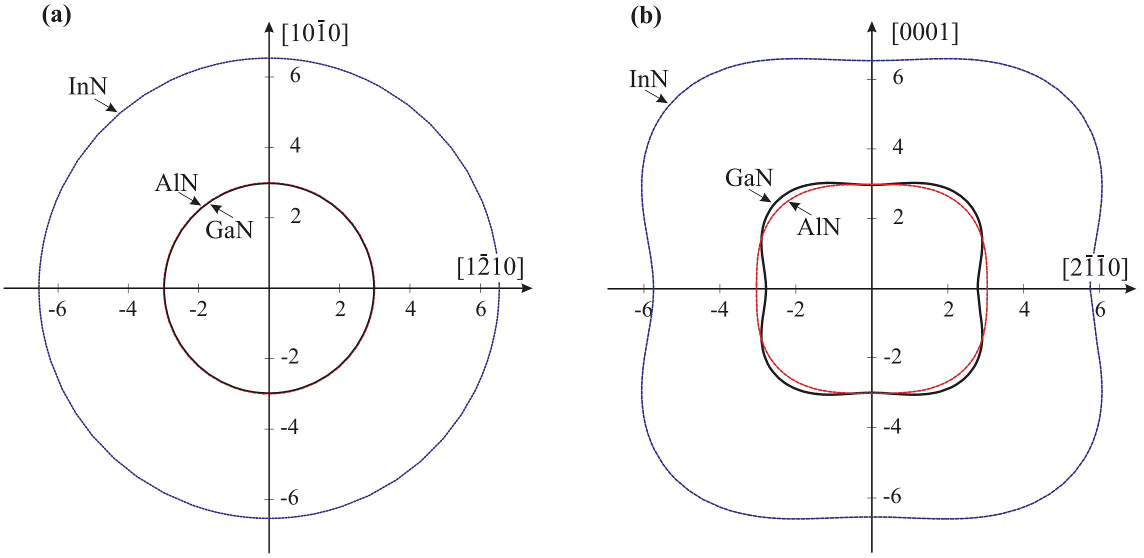


Figure A.4: Polar plots of the reciprocal Young' modulus Y^{-1} (in 10^{-3} GPa^{-1}) for wurtzite GaN, AlN, and InN in (a) the basal plane and (b) the $\left(01\bar{1}0\right)$ plane. Due to the hexagonal symmetry (see Fig. A.3), any other plane that includes the [0001]-axis would be identical with (b). For GaN the PBE-GGA calculated elastics constants were used (see Table A.2). For AlN and InN the elastics constants calculated by Wright [33] were used.

resulting from biaxial stress. In this case there are only three non-vanishing strain components:

$$\epsilon_1 = \epsilon_2 = \frac{\alpha - \alpha_0}{\alpha_0}, \quad \text{and} \quad \epsilon_3 = \frac{c - c_0}{c_0}. \tag{A.16}$$

Here $(\alpha_0,\ c_0)$ and $(\alpha,\ c)$ are the lattice constants of the unstrained and the strained crystal respectively.

For a pseudomorphic epilayer the in-plane lattice parameter is strained to be equal to the lattice constant of the substrate, $\alpha = \alpha_{\text{sub}}$. The ϵ_3 strain component can be calculated by minimizing the elastic energy density (Eq. (A.11)), that now takes the following form:

$$U = \frac{1}{2}\left(2C_{11}\epsilon_1^2 + 2C_{12}\epsilon_1^2 + 4C_{13}\epsilon_1\epsilon_3 + C_{33}\epsilon_3^2\right). \tag{A.17}$$

The ϵ_3 strain component that results from the minimization of U in Eq. (A.17) is:

$$\epsilon_3 = -2\frac{C_{13}}{C_{33}}\epsilon_1. \tag{A.18}$$

Then, the strained lattice constant c can be calculated using Eq. (A.13):

$$c = c_0\left(1 - 2\frac{C_{13}}{C_{33}}\left(\frac{\alpha_{\text{sub}} - \alpha_0}{\alpha_0}\right)\right). \tag{A.19}$$

From Table A.2 it can be seen that the calculated values (with internal relaxation) agree with the previous calculated and experimental results, except the measurement by Savastenko and Sheleg [254]. The small differences that exist with the earlier theoretical results can be explained in terms of the the different methods used. As Wright [33] suggested, the large differences between the results by Savastenko and Sheleg and the other theoretical and experimental values by the other groups, may show that their values are not accurate. This may attributed to the fact that the experiments by Savastenko and Sheleg are rather old (1978) and the material quality at that time might be rather poor. This statement can be supported by examining the ratio:

$$2\frac{C_{13}}{C_{33}} = \left|\frac{\epsilon_3}{\epsilon_2}\right|, \tag{A.20}$$

This ratio for all the calculated and measured elastic constants are close to 0.5 (except the values by Tagaki [255] who found the C_{13} to low), while for Savastenko and Sheleg is close to 1.2 and predicts that the perpendicular strain will be larger than the in-plane strain.

A.3 Bulk Modulus B

The bulk modulus B_0 describes the resistance of the material to the change of its volume and is defined as the ratio of the change in hydrostatic pressure acting on a volume to the dilatation($\epsilon_1 = \epsilon_2 = \epsilon_3 = \frac{1}{3}\delta$):

$$B_0\left(T,P\right) = -V\left(\frac{\partial P}{\partial V}\right)_T. \tag{A.21}$$

The bulk modulus derivative B_0' is given by:

$$B_0' = -\frac{\partial}{\partial P}\left(V\frac{\partial P\left(T,V\right)}{\partial V}\right)_{T,P=0}. \tag{A.22}$$

The energy as function of volume is given by the Murnaghan equation of state [250] in terms of B_0 and B_0':

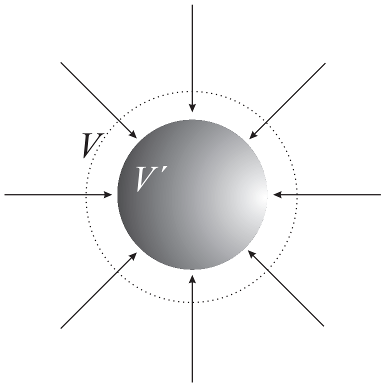


Figure A.5: The bulk modulus describes the change in volume as the pressure on it changes: $B_0\left(V,T\right) = -V\left(\frac{\partial P}{\partial V}\right)_T$.

$$E\left(V\right) = E_0 + \frac{B_0 V}{B_0'\left(B_0'-1\right)}\left[B_0'\left(1-\frac{V_0}{V}\right) + \left(\frac{V_0}{V}\right)^{B_0'} - 1\right]. \tag{A.23}$$

Eq. (A.23) is often used to fit the calculated total energy versus volume data, in order to obtain the equilibrium volume V_0, the total energy $E\left(V_0\right)$, the bulk modulus B_0, and its derivative B_0'.

A further relation for calculating the wurtzite bulk modulus in terms of the wurtzite elastic constants C_{ij} is the following [253]:

$$B = \frac{C_{33}\left(C_{11}+C_{12}\right) - 2\left(C_{13}\right)^2}{C_{11}+C_{12}-4C_{13}+2C_{33}}. \tag{A.24}$$

It can be used as an internal check to test the accuracy/consistency of of the measured or calculated elastic constants.

The calculated bulk modulus and bulk modulus derivatives for zincblende and wurtzite GaN are shown in Table A.1. In Table A.2 the same values are accompanied by the values calculated using Eq. (A.24) and earlier theoretical and experimental results. The differences between the predicted values from the two methods (via the Murnaghan equation of state and via the elastic constants) are small, $\sim 5\%$ indicating the correctness of the elastic constants calculations.

Appendix B

Coordinates of random configurations

B.1 Wurtzite $\mathrm{Al_xGa_{1-x}N}$

Here the primitive vectors and the reduced coordinates[1] of the atoms of the three SQS's used to model random wurtzite $\mathrm{Al_xGa_{1-x}N}$ alloys for $\mathrm{x} = \frac{1}{2}$ are given.

- **SQS1**

$$
\begin{array}{rcl}
\boldsymbol{\alpha_1} & = & \alpha\left(1,\ \sqrt{3},\ 0\right) \\
\boldsymbol{\alpha_2} & = & \alpha\left(1,\ -\sqrt{3},\ 0\right) \\
\boldsymbol{\alpha_3} & = & \alpha\left(0,\ 0,\ \frac{c}{\alpha}\right)
\end{array}
\qquad
\begin{array}{rclrcl}
\mathrm{Ga}_1 & = & \left(\frac{1}{2},\ 0,\ 0\right) & \mathrm{Al}_1 & = & \left(0,\ 0,\ 0\right) \\
\mathrm{Ga}_2 & = & \left(0,\ \frac{1}{2},\ 0\right) & \mathrm{Al}_2 & = & \left(\frac{1}{6},\ \frac{1}{3},\ \frac{1}{2}\right) \\
\mathrm{Ga}_3 & = & \left(\frac{1}{2},\ \frac{1}{2},\ 0\right) & \mathrm{Al}_3 & = & \left(\frac{1}{6},\ \frac{5}{6},\ \frac{1}{2}\right) \\
\mathrm{Ga}_4 & = & \left(\frac{2}{3},\ \frac{1}{3},\ \frac{1}{2}\right) & \mathrm{Al}_4 & = & \left(\frac{2}{3},\ \frac{5}{6},\ \frac{1}{2}\right) \\
 & & & & & \\
\mathrm{N}_1 & = & \left(0,\ 0,\ u\right) & \mathrm{N}_2 & = & \left(\frac{1}{6},\ \frac{1}{3},\ u+\frac{1}{2}\right) \\
\mathrm{N}_3 & = & \left(\frac{1}{2},\ 0,\ u\right) & \mathrm{N}_4 & = & \left(\frac{1}{6},\ \frac{5}{6},\ u+\frac{1}{2}\right) \\
\mathrm{N}_5 & = & \left(0,\ \frac{1}{2},\ u\right) & \mathrm{N}_6 & = & \left(\frac{2}{3},\ \frac{1}{3},\ u+\frac{1}{2}\right) \\
\mathrm{N}_7 & = & \left(\frac{1}{2},\ \frac{1}{2},\ u\right) & \mathrm{N}_8 & = & \left(\frac{2}{3},\ \frac{5}{6},\ u+\frac{1}{2}\right)
\end{array}
\tag{B.1}
$$

- **SQS2**

---

[1] The reduced coordinates are given in units of the primitive lattice vectors $\boldsymbol{\alpha_1}$, $\boldsymbol{\alpha_2}$, and $\boldsymbol{\alpha_3}$.

$$
\begin{array}{rcl}
\boldsymbol{\alpha_1} & = & \alpha\left(1,\ \sqrt{3},\ 0\right) \\
\boldsymbol{\alpha_2} & = & \alpha\left(1,\ -\sqrt{3},\ 0\right) \\
\boldsymbol{\alpha_3} & = & \alpha\left(0,\ 0,\ \frac{c}{\alpha}\right)
\end{array}
\quad
\begin{array}{rclrcl}
\mathrm{Ga}_1 & = & (0,\ 0,\ 0) & \mathrm{Al}_1 & = & \left(\frac{1}{2},\ 0,\ 0\right) \\
\mathrm{Ga}_2 & = & \left(\frac{1}{2},\ \frac{1}{2},\ 0\right) & \mathrm{Al}_2 & = & \left(\frac{1}{6},\ \frac{1}{3},\ \frac{1}{2}\right) \\
\mathrm{Ga}_3 & = & \left(\frac{2}{3},\ \frac{1}{3},\ \frac{1}{2}\right) & \mathrm{Al}_3 & = & \left(0,\ \frac{1}{2},\ 0\right) \\
\mathrm{Ga}_4 & = & \left(\frac{1}{6},\ \frac{5}{6},\ \frac{1}{2}\right) & \mathrm{Al}_4 & = & \left(\frac{2}{3},\ \frac{5}{6},\ \frac{1}{2}\right) \\
 & & & & & \\
\mathrm{N}_1 & = & (0,\ 0,\ u) & \mathrm{N}_2 & = & \left(\frac{1}{6},\ \frac{1}{3},\ u+\frac{1}{2}\right) \\
\mathrm{N}_3 & = & \left(\frac{1}{2},\ 0,\ u\right) & \mathrm{N}_4 & = & \left(\frac{1}{6},\ \frac{5}{6},\ u+\frac{1}{2}\right) \\
\mathrm{N}_5 & = & \left(0,\ \frac{1}{2},\ u\right) & \mathrm{N}_6 & = & \left(\frac{2}{3},\ \frac{1}{3},\ u+\frac{1}{2}\right) \\
\mathrm{N}_7 & = & \left(\frac{1}{2},\ \frac{1}{2},\ u\right) & \mathrm{N}_8 & = & \left(\frac{2}{3},\ \frac{5}{6},\ u+\frac{1}{2}\right)
\end{array}
\tag{B.2}
$$

- **SQS3**

$$
\begin{array}{rcl}
\boldsymbol{\alpha_1} & = & \alpha\left(1,\ -\sqrt{3},\ 0\right) \\
\boldsymbol{\alpha_2} & = & \alpha\left(\frac{1}{2},\ \frac{\sqrt{3}}{2},\ 0\right) \\
\boldsymbol{\alpha_3} & = & \alpha\left(0,\ 0,\ 2\frac{c}{\alpha}\right)
\end{array}
\quad
\begin{array}{rclrcl}
\mathrm{Ga}_1 & = & (0,\ 0,\ 0) & \mathrm{Al}_1 & = & \left(\frac{1}{2},\ 0,\ \frac{1}{2}\right) \\
\mathrm{Ga}_2 & = & \left(0,\ 0,\ \frac{1}{2}\right) & \mathrm{Al}_2 & = & \left(\frac{1}{3},\ \frac{1}{3},\ \frac{3}{4}\right) \\
\mathrm{Ga}_3 & = & \left(\frac{1}{2},\ 0,\ 0\right) & \mathrm{Al}_3 & = & \left(\frac{5}{6},\ \frac{1}{3},\ \frac{1}{4}\right) \\
\mathrm{Ga}_4 & = & \left(\frac{1}{3},\ \frac{1}{3},\ \frac{1}{4}\right) & \mathrm{Al}_4 & = & \left(\frac{5}{6},\ \frac{1}{3},\ \frac{3}{4}\right) \\
 & & & & & \\
\mathrm{N}_1 & = & \left(0,\ 0,\ \frac{u}{2}\right) & \mathrm{N}_2 & = & \left(\frac{1}{2},\ 0,\ \frac{u}{2}\right) \\
\mathrm{N}_3 & = & \left(0,\ 0,\ \frac{1}{2}(u+1)\right) & \mathrm{N}_4 & = & \left(\frac{1}{3},\ \frac{1}{3},\ \frac{1}{2}(u+\frac{1}{2})\right) \\
\mathrm{N}_5 & = & \left(\frac{1}{3},\ \frac{1}{3},\ \frac{1}{2}(u+\frac{3}{2})\right) & \mathrm{N}_6 & = & \left(\frac{5}{6},\ \frac{1}{3},\ \frac{1}{2}(u+\frac{1}{2})\right) \\
\mathrm{N}_7 & = & \left(\frac{1}{2},\ 0,\ \frac{1}{2}(u+1)\right) & \mathrm{N}_8 & = & \left(\frac{5}{6},\ \frac{1}{3},\ \frac{1}{2}(u+\frac{3}{2})\right)
\end{array}
\tag{B.3}
$$

B.2 Zincblende alloys

In analogy to the wurtzite structure, the primitive vectors and the arrangements of the atoms in the supercell of the zincblende structure, are cited here. The coordinates of the ordered configuration are taken from Ref. [210].

- **ordered**

$$
\begin{array}{rcl}
\boldsymbol{\alpha_1} & = & \alpha\left(\frac{\sqrt{2}}{4},\ -\frac{\sqrt{6}}{4},\ 0\right) \\
\boldsymbol{\alpha_2} & = & \alpha\left(\frac{\sqrt{2}}{4},\ \frac{\sqrt{6}}{4},\ 0\right) \\
\boldsymbol{\alpha_3} & = & \alpha\left(0,\ 0,\ 2\sqrt{3}\right)
\end{array}
\quad
\begin{array}{rclrcl}
\mathrm{Ga}_1 & = & \left(\frac{2}{3},\ \frac{1}{3},\ \frac{1}{3}\right) & \mathrm{Al}_1 & = & \left(0,\ 0,\ \frac{1}{2}\right) \\
\mathrm{Ga}_2 & = & \left(\frac{1}{3},\ \frac{2}{3},\ \frac{2}{3}\right) & \mathrm{Al}_2 & = & \left(\frac{2}{3},\ \frac{1}{3},\ \frac{5}{6}\right) \\
\mathrm{Ga}_3 & = & (0,\ 0,\ 1) & \mathrm{Al}_3 & = & \left(\frac{1}{3},\ \frac{2}{3},\ \frac{7}{6}\right) \\
 & & & & & \\
\mathrm{N}_1 & = & \left(\frac{1}{3},\ \frac{2}{3},\ \frac{7}{24}\right) & \mathrm{N}_2 & = & \left(\frac{2}{3},\ \frac{1}{3},\ \frac{11}{24}\right) \\
\mathrm{N}_3 & = & \left(0,\ 0,\ \frac{5}{8}\right) & \mathrm{N}_4 & = & \left(\frac{1}{3},\ \frac{2}{3},\ \frac{19}{24}\right) \\
\mathrm{N}_5 & = & \left(\frac{2}{3},\ \frac{1}{3},\ \frac{23}{24}\right) & \mathrm{N}_6 & = & \left(0,\ 0,\ \frac{9}{8}\right)
\end{array}
\tag{B.4}
$$

- **SQS8**

$$
\begin{array}{lcllcllcl}
 & & & \mathrm{Ga}_1 & = & (0,\,0,\,0) & \mathrm{Al}_1 & = & \left(0,\,\frac{1}{2},\,\frac{1}{4}\right) \\
 & & & \mathrm{Ga}_2 & = & \left(\frac{1}{2},\,-\frac{1}{4},\,\frac{1}{8}\right) & \mathrm{Al}_2 & = & \left(-\frac{1}{2},\,\frac{1}{4},\,\frac{3}{8}\right) \\
 & & & \mathrm{Ga}_3 & = & \left(-\frac{1}{2},\,\frac{3}{4},\,\frac{5}{8}\right) & \mathrm{Al}_3 & = & \left(0,\,0,\,\frac{1}{2}\right) \\
\boldsymbol{\alpha_1} & = & \alpha\left(1,\,\frac{1}{2},\,-\frac{1}{2}\right) & \mathrm{Ga}_4 & = & \left(-1,\,\frac{1}{2},\,\frac{3}{4}\right) & \mathrm{Al}_4 & = & \left(-\frac{1}{2},\,\frac{1}{4},\,\frac{7}{8}\right) \\
\boldsymbol{\alpha_2} & = & \alpha\left(\frac{1}{2},\,-\frac{1}{2},\,0\right) & & & & & & \\
\boldsymbol{\alpha_3} & = & \alpha\,(1,\,1,\,2) & \mathrm{N}_1 & = & \left(-\frac{1}{8},\,-\frac{1}{16},\,\frac{5}{32}\right) & \mathrm{N}_2 & = & \left(\frac{5}{2},\,-\frac{5}{16},\,\frac{9}{32}\right) \\
 & & & \mathrm{N}_3 & = & \left(\frac{1}{8},\,\frac{7}{16},\,\frac{13}{32}\right) & \mathrm{N}_4 & = & \left(-\frac{3}{8},\,\frac{3}{16},\,\frac{17}{32}\right) \\
 & & & \mathrm{N}_5 & = & \left(\frac{1}{8},\,-\frac{1}{16},\,\frac{21}{32}\right) & \mathrm{N}_6 & = & \left(\frac{3}{4},\,\frac{19}{8},\,\frac{17}{16}\right) \\
 & & & \mathrm{N}_7 & = & \left(-\frac{7}{8},\,\frac{7}{16},\,\frac{29}{32}\right) & \mathrm{N}_8 & = & \left(-\frac{3}{8},\,\frac{3}{16},\,\frac{33}{22}\right)
\end{array}
\tag{B.5}
$$

Bibliography

[1] S. Strite and H. Morkoç, J. Vac. Sci. Technol. B **10(4)**, 1237 (1992).

[2] O. Ambacher, J. Phys. D: Appl, Phys. **31**, 2653 (1998).

[3] T. L. Tansley and C. P. Foley, J. Appl. Phys. **59**, 3241 (1986).

[4] J. H. Edgar, S. Strite, I. Akasaki, H. Amano, and C. Wetzel, eds., *Properties, Processing and Applications of Gallium Nitride and Ralated Materials*, vol. 23 of *EMIS Datareview Series* (INSPEC, London, 1999).

[5] V. Y. Davydov, A. A. Klochikhin, R. P. Seisyan, V. V. Emtsev, S. V. Ivanov, F. Bechstedt, J. Furthmüller, H. Harima, A. V. Mudryi, J. Aderhold, O. Semchinova, and J. Graul, phys. status Solidi (b) **229**, r1 (2002).

[6] T. Matsuoka, H. Okamoto, M. Nakao, H. Harima, and E. Kurimoto, Appl. Phys. Lett. **81**, 1246 (2002).

[7] J. I. Pankove, Mater. Sc. Engin. **B61-62**, 305 (1999).

[8] O. Brandt, H. Yang, B. Jenichen, Y. Suzuki, L. Däweritz, and K. H. Ploog, Phys. Rev. B **52**, R2253 (1995).

[9] D. Schikora, M. Hankeln, D. J. As, K. Lischka, T. Litz, A. Waag, T. Buhrow, and F. Henneberger, Phys. Rev. B **54**, R8381 (1996).

[10] D. J. As, T. Frey, D. Schikora, K. Lischka, V. Cimalla, J. Pezoldt, R. Goldhahn, S. Kaiser, and W. Gebhardt, Appl. Phys. Lett. **76**, 1686 (2000).

[11] G. Feuillet, H. Hamaguchi, K. Ohta, P. Hacke, H. Okumura, and S. Yoshida, Appl. Phys. Lett. **70**, 1025 (1997).

[12] K. Lischka, phys. stat. sol. (a) **177**, 135 (2000).

[13] S. Nakamura and G. Fasol, *The Blue Laser Diode* (Springer-Verlag, Berlin Heidelberg, 1997).

[14] R. Juda and H. Hahn, Anorg. Allgem. Chem. **234**, 282 (1938).

[15] R. Juda and H. Hahn, Anorg. Allgem. Chem. **244**, 133 (1940).

[16] H. Grimmeiss and Z. H-Koelmans, Nature **14a**, 264 (1959).

[17] H. P. Maruska and J. J. Tietjen, Appl. Phys. Lett. **15**, 367 (1969).

[18] J. Neugebauer and C.G. Van de Walle, Proc. ICPS-22, World Scientific, Singapore, 1995, p. 2327.

[19] L. I. Pankove, E. A. Miller, and J. E. Berkeyheiser, J. Lumin. **5**, 84 (1972).

[20] S. Yoshida, S. Misawa, and S. Gonda, Appl. Phys. Lett. **42**, 427 (1983).

[21] H. Amano, N. Sawaki, I. Akasaki, and Y. Toyoda, Appl. Phys. Lett. **48**, 353 (1986).

[22] S. Nakamura, Jpn. J. Appl. Phys. **30**, L1705 (1991).

[23] H. Amano, I. Akasaki, T. Kozawa, K. Hiramatsu, N. Sawaki, K. Ikeda, and Y. Ishii, J. Lumin. **40-41**, 121 (1988).

[24] S. Nakamura, Jpn. J. Appl. Phys. **31**, L139 (1992).

[25] J. A. V. Vechten, J. D. Zook, R. D. Hornig, and B. Goldenberg, Jpn. J. Appl. Phys. **31**, 3662 (1992).

[26] J. Neugebauer and C. G. V. de Walle, Phys. Rev. Lett. **75**, 4452 (1995).

[27] J. Neugebauer and C. G. V. de Walle, Appl. Phys. Lett. **68**, 1829 (1996).

[28] F. A. Ponce and D. P. Bour, Nature **386**, 351 (1997).

[29] See: `http://compoundsemiconductor.net/articles/news/7/6/8/1`.

[30] J. I. Pankove and T. D. Moustakas, eds., *Gallium Nitride (GaN) I*, vol. 50 of *Semiconductors and semimetals* (Academic Press, New York, 1998).

[31] S. C. Jain, M. Willander, J. Narayan, and R. V. Overstraeten, J. Appl. Phys. **87**, 965 (2000).

[32] S. J. Pearton, J. C. Zolper, R. J. Shul, and F. Ren, J. Appl. Phys. **86**, 1 (1999).

[33] A. F. Wright, J. Appl. Phys. **82**, 2833 (1997).

[34] G. V. de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).

[35] L. Bellaiche and A. Zunger, Phys. Rev. B **57**, 4425 (1998).

[36] S. V. Dudiy and A. Zunger, Phys. Rev. B **68**, 041302(R) (2003).

[37] J. E. Northrup, L. T. Romano, and J. Neugebauer, Appl. Phys. Lett. **74**, 2319 (1999).

[38] A. R. Smith, R. M. Feenstra, D. W. Greve, J. Neugebauer, and J. E. Northrup, Phys. Rev. Lett. **79**, 3934 (1997).

[39] A. R. Smith, R. M. Feenstra, D. W. Greve, J. Neugebauer, and J. E. Northrup, Appl. Phys. A **66**, S947 (1998).

[40] A. R. Smith, R. M. Feenstra, D. W. Greve, M. S. Shin, M. Skowronski, J. Neugebauer, and J. E. Northrup, Appl. Phys. Lett. **72**, 2114 (1998).

[41] A. R. Smith, R. M. Feenstra, D. W. Greve, M. S. Shin, M. Skowronski, J. Neugebauer, and J. E. Northrup, Surf. Sci. **423**, 70 (1998).

[42] A. R. Smith, R. M. Feenstra, D. W. Greve, M. S. Shin, M. Skowronski, J. Neugebauer, and J. E. Northrup, J. Van. Sci. Technol. B **16**, 2242 (1998).

[43] J. E. Northrup, J. Neugebauer, R. M. Feenstra, and A. R. Smith, Phys. Rev. B **61**, 9932 (2000).

[44] A. Kley and J. Neugebauer, Phys. Rev. B **50**, 8616 (1994).

[45] K. Ohno, K. Esfarjani, and Y. Kawazoe, *Computational Materials Science, From Ab Initio to Monte Carlo*, Solid-State Sciences (Springer, Berlin, Heidelberg, New York, 1999).

[46] B. Heying, R. Averbeck, L. F. Chen, E. Haus, H. Riechert, and J. S. Speck, J. Appl. Phys. **88**, 1855 (2000).

[47] C. Adelmann, J. Brault, G. Mula, B. Daudin, L. Lymperakis, and J. Neugebauer, Phys. Rev. B **67**, 165419 (2003).

[48] C. Adelmann, J. Brault, D. Jalabert, P. Gentile, H. Mariette, G. Mula, and B. Daudin, J. Appl. Phys. **91**, 9638 (2002).

[49] G. Koblmüller, R. Averbeck, H. Riechert, and P. Pongratz, Phys. Rev. B **69**, 035325 (2004).

[50] J. Elsner, R. Jones, P. K. Sitch, V. D. Porezag, M. Elstner, T. Frauenheim, M. I. Heggie, S. Öberg, and P. R. Briddon, Phys. Rev. Lett. **79**, 3672 (1997).

[51] J. Elsner, R. Jones, M. I. Heggie, P. K. Sitch, M. Haugk, T. Frauenheim, S. Öberg, and P. R. Briddon, Phys. Rev. B **58**, 12571 (1998).

[52] J. E. Northrup, Appl. Phys. Lett. **78**, 2288 (2001).

[53] J. E. Northrup, Phys. Rev. B **66**, 045204 (2002).

[54] C. J. Fall, A. T. Blumenau, T. Frauenheim, R. Jones, and M. I. Heggie, MRS Internet J. Nitride Semicod. Res. **5S1**, W.9.3 (2000).

[55] A. F. Wright and U. Grossner, Appl. Phys. Lett. **73**, 2751 (1998).

[56] A. Béré and A. Serra, Phys. Rev. B **65**, 205323 (2002).

[57] J. Chen, P. Ruterana, and G. Nouet, Mater. Sci. Eng. **B82**, 117 (2001).

[58] D. Korakakis, K. Ludwig-Jr., and T. Moustakas, Appl. Phys. Lett. **71**, 72 (1997).

[59] E. Iliopoulos, K. Ludwig-Jr., T. D. Moustakas, and S. Chu., Appl. Phys. Lett. **78**, 463 (2001).

[60] P. Ruterana, G. D. S. Jores, M. Laügt, F. Omnes, and E. Bellet-Amalric, Appl. Phys. Lett. **78**, 344 (2001).

[61] P. Waltereit, O. Brandt, A. Trampert, H. T. Grahn, J. Menniger, M. Ramsteiner, M. Reiche, and K. H. Ploog, Nature (London) **406**, 865 (2000).

[62] M. Benamara, L. Kirste, M. Albrecht, K. W. Benz, and H. P. Strunk, Appl. Phys. Lett. **82**, 547 (2003).

[63] R. G. Parr and W. Yang, *Density–Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).

[64] F. Carlos, N. Fernando, and M. A. L. Miguel, eds., *A Primer in Density Functional Theory*, no. 620 in Lecture Notes in Physics (Springer-Verlag, Heidelberg, 2003).

[65] L. H. Thomas, Proc. Camb. Phil. Sic. **23**, 542 (1927).

[66] E. Fermi, Rend. Accad. Lincei **6**, 602 (1927).

[67] E. Fermi, Z. Phys. **48**, 73 (1928).

[68] E. Fermi, Rend. Accad. Lincei **7**, 342 (1928).

[69] P. Hohenberg and W. Kohn, Phys. Rev **136**, B864 (1964).

[70] W. Kohn and L. J. Sham, Phys. Rev **140**, A1133 (1965).

[71] P. A. M. Dirac, Proc. Cambridge Phil. Soc. **26**, 376 (1930).

[72] E. Wigner, Phys. Rev. **46**, 1002 (1934).

[73] M. Gell-Mann and K. A. Brueckner, Phys. Rev. **106**, 364 (1957).

[74] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[75] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[76] U. von Barth and L. Hedin, J. Phys. C: Solid State Phys. **5**, 1629 (1972).

[77] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[78] J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).

[79] C. L. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[80] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).

[81] J. P. Perdew, K. Burke, and M. Ernzehof, Phys. Rev. Lett. **77**, 3865 (1996).

[82] M. Fuchs, J. L. F. D. Silva, C. Stampfl, J. Neugebauer, and M. Scheffler, Phys. Rev. B **65**, 245212 (2002).

[83] J. Baker, M. Muir, and J. Andzelm, J. Chem. Phys. **102**, 2063 (1995).

[84] L. Fan and T. Ziegler, J. Am. Chem. Soc. **114**, 10 (1992).

[85] C. Filippi, D. Singh, and C. Umrigar, Phys. Rev. B **50**, 14947 (1994).

[86] G. Ortiz, Phys. Rev. B **45**, 11328 (1992).

[87] M. Städele, M. Moukara, J. A. Majewski, P. Vogl, and A. Görling, Phys. Rev. B **59**, 10031 (1999).

[88] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).

[89] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).

[90] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).

[91] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[92] A. Rubio, J. L. Corkill, M. L. Cohen, E. L. Shirley, and S. G. Louie, Phys. Rev. B **48**, 11810 (1993).

[93] E. Gross and R. M. Dreizler, *Density Functional Methods in Physics* (Plenum Press, New York, 1985), chap. Density functional approach to time-dependent and to relativistic systems, pp. 81–140, Nato ASI Series.

[94] D. J. Chadi and M. L. Cohen, Phys. Rev. B **8**, 5747 (1973).

[95] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[96] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, Comput. Phys. Commun. **107**, 187 (1997).

[97] S. Boeck, A. Dick, Ch. Freysoldt, F. Grzegorzewski, T. Hammerschmidt, L. Ismer, L. Lymperakis, and M. Wahn, ***SFHIngX*** *– User's Guide*, S.l Boeck and J. Neugebauer, editors, Berlin, 2003 Available on line: `http://www.sfhingx.de`,.

[98] V. Heine, *The Pseudopotential Concept*, vol. 24 of *Solid State Physics* (Academic Press, New York, 1970).

[99] J. C. Phillips and L. Kleinman, Phys. Rev. **116**, 287 (1959).

[100] M. Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1999).

[101] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[102] A. Redondo, W. A. G. III, and T. C. McGill, Phys. Rev. B **41**, 5038 (1977).

[103] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[104] A. Zunger and M. L. Cohen, Phys. Rev. B **20**, 4082 (1979).

[105] G. P. Kerker, L. Phys. C **13**, L189 (1980).

[106] G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[107] D. R. Hamman, Phys. Rev. B **40**, 2980 (1989).

[108] E. L. Shirley, D. C. Allan, R. M. Martin, and J. D. Joannopoulos, Phys. Rev. B **40**, 3652 (1989).

[109] M. Fuchs, M. Bockstedte, E. Pehlke, and M. Scheffler, Phys. Rev. B **57**, 2134 (1998).

[110] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[111] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[112] X. Gonze, R. Stumpf, and M. Scheffler, Phys. Rev. B **44**, 8503 (1991).

[113] S. Erkoc, Phys. Rep. **278**, 79 (1997).

[114] R. M. Nieminen, M. J. Puska, and M. J. Manninen, eds., *Many-Atom Interactions in Solids*, vol. 48 of *Proceedings in Physics* (Springer, Berlin, Heidelberg, 1990).

[115] P. N. Keating, Phys. Rev. **145**, 637 (1966).

[116] G. Kaczmarczyk, A. Kaschner, A. Hoffmann, and C. Thomsen, Phys. Rev. B **61**, 5353 (1999).

[117] J. L. Rouviere, M. Arlery, B. Daudin, G. Feuillet, and O. Briot, Mater. Sc. Engin. **B50**, 61 (1997).

[118] K. Kim, W. R. L. Lambrecht, and B. Segall, Phys. Rev. B **53**, 16310 (1996).

[119] F. Grosse and J. Neugebauer, Phys. Rev. B **63**, 085207 (2001).

[120] G. Wei, J. Zi, K. Zhang, and X. Xie, J. Appl. Phys. **82**, 622 (1997).

[121] J. Tersoff, Phys. Rev. Lett. **56**, 632 (1986).

[122] J. Tersoff, Phys. Rev. B **38**, 9902 (1988).

[123] J. Tersoff, Phys. Rev. Lett. **61**, 2879 (1988).

[124] J. Tersoff, Phys. Rev. B **39**, 5566 (1989).

[125] G. C. Abell, Phys. Rev. B **31**, 6184 (1985).

[126] J. H. Rose, J. R. Smith, and J. Ferrante, Phys. Rev. B **28**, 1835 (1983).

[127] T. Ito, Jpn. J. Appl. Phys. **37**, L574 (1998).

[128] Y. Kangawa, T. Ito, A. Mori, and A. Koukitu, J. Crystal Growth **220**, 401 (2000).

[129] T. Ito and Y. Kangawa, J. Crystal Growth **235**, 149 (2002).

[130] T. Suda, Y. Kangawa, K. Nakamura, and T. Ito, J. Crystal Growth **258**, 277 (2003).

[131] J. Nord, K. Albe, P. Erhart, and K. Nordlund, J. Phys.: Condens. Matter. **15**, 5649 (2003).

[132] W. H. Moon and H. J. Hwang, Phys. Lett. A **315**, 319 (2003).

[133] F. Benkabou, P. Becker, M. Certier, and H. Aourag, Phys. Status Solidi (b) **209**, 223 (1998).

[134] K. E. Khor and S. D. Sarma, Phys. Rev. B **38**, 3318 (1988).

[135] D. W. Brenner, Phys. Rev. Lett. **63**, 1022 (1989).

[136] K. Albe, K. Nordlund, J. Nord, and A. Kuronen, Phys. Rev. B **66**, 035205 (2002).

[137] F. H. Stillinger and T. A. Weber, Phys. Rev. B **31**, 5262 (1985).

[138] A. Béré, P. Vermaut, A. Hairie, E. Paumier, P. Ruterana, and G. Nouet, Mater. Sci. Forum. **294-296**, 223 (1999); P. Ruterana, B. Barbray, A. Béré, P. Vermaut, A. Hairie, E. Paumier, G. Nouet, A. Salvador, A. Botchkarev, and H. Morkoç, Phys. Rev. B **59**, 15917 (1999); P. Ruterana, A. Béré, and G. Nouet, MRS Internet J. Nitride Semicod. Res. **5S1**, W3.55(2000).

[139] N. Aïchoune, V. Potin, P. Ruterana, A. Hairie, G. Nouet, and E. Paumier, Comp. Mater. Sci. **17**, 380 (2000).

[140] J. Chen, P. Ruterana, and G. Nouet, Mater. Sci. Eng. **B82**, 117 (2001); J. Chen, P. Ruterana, and G. Nouet, Phys. Status Solidi (b) **228**, 411 (2001); A. Béré, J. Chen, P. Ruterana, A. Serra, and G. Nouet, Comput. Matre. Sci. **24**, 144 (2002); A. Béré, J. Chen, P. Ruterana, G. Nouet, and A. Serra, J. Phys.: Condens. Matter **14**, 12703 (2002); S. Kret, P. Dluzewski, G. Maciejewski, V. Potin, J. Chen, P. Ruterana, and G. Nouet, Diam. Relat. Mater. **11**, 910 (2002); J. Chen, P. Ruterana, and G. Nouet, Phys. Rev. B **67**, 205210 (2003).

[141] A. Béré and A. Serra, Interface Sci. **9**, 149 (2001); A. Béré and A. Serra, Phys. Rev. B **65**, 205323 (2002); A. Béré and A. Serra, Phys. Rev. B **66**, 085330 (2002); A. Béré and A. Serra, Phys. Rev. B **68**, 033305 (2003).

[142] J. Kioseoglou, G. P. Dimitrakopulos, H. M. Polatoglou, L. Lymperakis, G. Nouet, and P. Komninou, Diam. Relat. Mater. **11**, 905 (2002); J. Kioseoglou, H. M. Polatoglou, L. Lymperakis, G. Nouet, and P. Komninou, Comp. Mater. Sci. **27**, 43 (2003); J. Kioseoglou, P. Komninou, G. P. Dimitrakopulos, T. Kehagias, H. M. Polatoglou, G. Nouet, and T. Karakostas, Solid State Electron. **47**, 553 (2003).

[143] H. Balamane, T. Halicioglu, and W. A. Tiller, Phys. Rev. B **46**, 2250 (1992).

[144] A. Polian, M. Grimsditch, and I. Grzegory, J. Appl. Phys. **79**, 3343 (1996).

[145] J. E. Northrup, J. Neugebauer, and L. T. Romano, Phys. Rev. Lett. **77**, 103 (1996).

[146] C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1996).

[147] E. J. Tarsa, B. Heying, X. H. Wu, S. P. DenBaars, and J. S. Speck, J. Appl. Phys. **82**, 5472 (1997).

[148] G. B. Stephenson, J. A. Eastman, C. Thompson, O. Auciello, L. J. Thompson, A. Munkholm, P. Fini, S. P. DenBaars, and J. S. Speck, Appl. Phys. Lett. **74**, 3326 (1999).

[149] M. H. Xie, S. M. Seutter, L. X. Zheng, H. Wu, and S. Y. Tong, Phys. Rev. Lett. **82**, 2749 (1999).

[150] H. Marchand, J. P. Ibbetson, P. T. Fini, P. Kozodoy, S. Keller, S. DenBaars, J. S. Speck, and U. K. Mishra, MRS Internet J. Nitride Semicond. Res. **3**, 3 (1998).

[151] T. Nishida, T. Akasaka, and N. Kobayashi, Jpn. J. Appl. Phys. **37**, L459 (1998).

[152] T. Nishida, N. Maeda, T. Akasaka, and N. Kobayashi, Journal of Crystal Growth **195**, 41 (1998).

[153] L. X. Zheng, M. H. Xie, S. M. Seutter, S. H. Cheung, and S. Y. Tong, Phys. Rev. Lett. **85**, 2352 (2000).

[154] M. H. Xie, L. X. Zheng, X. Q. Dai, H. S. Wu, and S. Y. Tong, Surf. Sci. **558**, 195 (2004).

[155] C. M. Bertoni, G. Roma, and R. Di Felice in Ref. [258], pp. 274-278.

[156] E. Penev, Ph.D. thesis, FHI, Berlin (2002), available on line: `http://www.fhi-berlin.mpg.de/th/publications/PhD-penev-2002.pdf`.

[157] A. L. da Rosa, Ph.D. thesis, FHI, Berlin (2003), available on line: `http://edocs.tu-berlin.de/diss/2003/darosa_andreia.htm`, (English text).

[158] J. W. Gibbs, *The collected works of J. Willard Gibbs: : in two volumes, vol. 1: Thermodynamics* (Yale University Press, New Haven, 1948).

[159] N. Chetty and R. M. Martin, Phys. Rev. B **44**, 5568 (1991).

[160] N. Chetty and R. M. Martin, Phys. Rev. B **45**, 6074 (1992).

[161] N. Chetty and R. M. Martin, Phys. Rev. B **45**, 6089 (1992).

[162] N. Moll, A. Kley, E. Pehlke, and M. Scheffler, Phys. Rev. B **54**, 8844 (1996).

[163] K. Papcewicz, B. Chen, B. Yakobson, and J. Bernholc, Phys. Rev. B **57**, 7281 (1998).

[164] P. Kratzer, C. G. Morgan, E. Penev, A. L. Rosa, A. Schindlmayr, L. G. Wang, and T. Zywietz, ***FHI98MD*** *– Computer code for density-functional theory calculations for poly-atomic systems, User's Manual*, Berlin, 1999 Available on line: `http://www.fhi-berlin.mpg.de/th/fhi98md/doc/FHI98md.pdf`.

[165] J. Neugebauer, T. K. Zywietz, M. Scheffler, J. E. Northrup, H. Chen, and R. M. Feenstra, Phys. Rev. Lett. **90**, 056101 (2003).

[166] O. Kubaschewski and C. B. Alcock, *Metallurgical Thermochemistry* (Pergamon, Oxford, 1979).

[167] M. E. Lin, S. Strite, A. Agarwal, A. Salvador, G. L. Zhou, N. Teraguchi, A. Rockett, and H. Morkoç, Appl. Phys. Lett. **62**, 702 (1993).

[168] W. C. Hughes, W. H. R. Jr, M. A. L. Johnson, S. Fujita, J. W. C. Jr, and J. F. Schetzina, J. Vac. Sci. Technol. B **13**, 1571 (1995).

[169] Q.-K. Xue, Q. Z. Xue, R. Z. Bakhtizin, Y. Hasegawa, I. S. T. Tsong, T. Sakurai, and T. Ohno, Phys. Rev. Lett. **82**, 3074 (1999).

[170] K. Rapcewicz, M. B. Nardelli, and J. Bernholc, Phys. Rev. B **56**, R12725 (1997).

[171] J. Elsner, M. Haugk, G. Jungnickel, and T. Frauenheim, Solid State Commun. **106**, 739 (1998).

[172] J. Fritsch, O. Sankey, K. Schmidt, and J. B. Page, Phys. Rev. B **57**, 15360 (1998).

[173] D. L. Abernathy, D. Gibbs, G. Grbel, K. G. Huang, S. G. J. Mochrie, A. R. Sandy, and D. M. Zehner, Surf. Sci. **283**, 260 (1993).

[174] A. Botchkarev, A. Salvador, B. Sverdlov, J. Myoung, and H. Morkoç, J. Appl. Phys. **77**, 4455 (1995).

[175] F. Widmann, B. Daudin, G. Feuillet, N. Pelekanos, and J. L. Rouvière, Appl. Phys. Lett. **73**, 2642 (1998).

[176] B. Heying, I. Smorchkova, C. Poblenz, C. Elsass, P. Fini, S. D. B. abd U. Mishra, and J. S. Speck, Appl. Phys. Lett. **77**, 2885 (2000).

[177] C. Kruse, S. Einfeldt, T. Böttcher, D. Hommel, D. Rudloff, and J. Christen, Appl. Phys. Lett. **78**, 3827 (2001).

[178] T. Zywietz, J. Neugebauer, and M. Scheffler, Appl. Phys. Lett. **73**, 487 (1998).

[179] J. Neugebauer, Phys. Stat. Sol. (b) **227**, 93 (2001).

[180] The spin polarized correction can be found under:
`http://physics.nist.gov/PhysRefData/DFTdata/Tables/ptable.html` .

[181] R. W. G. Wyckoff, *Crystal Structures*, vol. 1 (Wiley, New York, 1962), 2nd ed.

[182] A. Filippetti, V. Fiorentini, G. Cappellini, and A. Bosin, Phys. Rev. B **59**, 8026 (1999).

[183] J. E. Northrup and J. Neugebauer, Phys. Rev. B **53**, R10477 (1996).

[184] J. E. Jaffe, R. Pandey, and P. Zapol, Phys. Rev. B **53**, R4209 (1996).

[185] C. D. Lee, R. M. Feenstra, J. E. Northrup, L. Lymperakis, and J. Neugebauer, Appl. Phys Lett. **82**, 1793 (2003).

[186] C. D. Lee, R. M. Feenstra, J. E. Northrup, L. Lymperakis, and J. Neugebauer, Mat. Res. Soc. Symp. Proc. **743**, L4.1.1 (2003).

[187] M. Posternak, A. Baldereschi, A. Catellani, and R. Pesta, Phys. Rev. Lett. **64**, 1777 (1990).

[188] J. F. Nye, *Physical Properties of Crystals* (Oxford University Press, Oxford, 1985).

[189] F. Bernardini, V. Fiorentini, and D. Vandebilt, Phys. Rev. B **56**, R10024 (1997).

[190] O. Ambacher, J. Majewski, C. Miskys, M. Hermann, M. Eickhoff, M. Stutzmann, F. Bernardini, V. Fiorentini, V. Tilak, B. Schaff, and L. F. Eastman, J. Phys.: Condens. Matter. **14**, 3399 (2002).

[191] A. Bykhovski, B. Gelmont, and M. Shur, J. Appl. Phys. **74**, 6734 (1993).

[192] P. Waltereit, O. Brandt, M. Ramsteiner, A. Trampert, H. T. Grahn, J. Menniger, M. Reiche, R. Uecker, P. Reiche, and K. H. Ploog, Phys. Stat. Sol. (a) **180**, 133 (2000).

[193] P. Waltereit, O. Brandt, M. Ramsteiner, A. Trampert, H. T. Grahn, J. Menniger, M. Reiche, and K. H. Ploog, J. Crystal Growth **227-228**, 437 (2001).

[194] Y. J. Sun, O. Brandt, U. Jahn, T. Y. Liu, A. Trampert, S. Cronenberg, S. Dgar, and K. H. Ploog, J. Appl. Phys. **92**, 5741 (2002).

[195] Y. J. Sun, O. Brandt, and K. H. Ploog, J. Vac. Sci. Technol. B **21**, 1350 (2003).

[196] J. Pollmann and P. Krüger in Ref. [258], pp. 93-208.

[197] The pictures shown here are from the practical session *"Surface relaxation and band structures"* of the *"Workshop on Application of Density-Functional Theory in Condensed-Matter Physics, Surface Physics,Chemistry, Engineering, and Biology, 21-30 July 2003, Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany"*. For more information see: `http://www.fhi-berlin.mpg.de/th/Meetings/FHImd2003/`, and `http://w3.rz-berlin.mpg.de/ boeck/sfhingx/documentation/session-3.pdf`.

[198] J. L. A. Alves, J. Hebenstreit, and M. Scheffler, Phys. Rev. B **44**, 6188 (1991).

[199] M. Albrecht, L. Lymperakis, J. Neugebauer, J. E. Northrup, L. Kirtse, M. Leroux, I. Grzegory, S. Porowski, and H. P. Strunk, Phys. Rev. B **71**, 035314 (2005).

[200] R. H. Fowler and E. A. Guggenheim, *Statistical Thermodynamics* (University Press, Cambrige, 1949).

[201] R. H. Fowler, *Statistical Mechanics* (University Press, Cambrige, 1955).

[202] M. K. Behbehani, E. L. Piner, S. X. Liu, N. A. El-Masry, and S. M. Bedair, Appl. Phys. Lett. **75**, 2202 (1999).

[203] P. Ruterana, G. Nouet, W. V. der Stricht, I. Moerman, and L. Considine, Appl. Phys. Lett. **72**, 1742 (1998).

[204] J. R. Jenny, J. E. V. Nostrand, and R. Kaspi, Appl. Phys. Lett. **72**, 85 (1998).

[205] S. R. Lee, A. F. Wright, M. H. Crawford, G. A. Petersen, J. Han, and R. M. Biefeld, Appl. Phys. Lett. **74**, 3344 (1999), and references therein.

[206] M. Misra, D. Korakakis, H. M. Ng, and T. D. Moustakas, Appl. Phys. Lett. **74**, 2203 (1999).

[207] D. G. Elbing, L. Kirste, K. W. Benz, N. Teofilov, K. Thonke, and R. Sauer, J. Crystal Growth **227-228**, 453 (2001).

[208] M. Leroux, S. Dalmasso, F. Natali, S. Helin, C. Touzi, S. Laügt, M. Passerel, F. Omnes, F. Semond, J. Massies, and P. Gibart, phys. stat. sol. (b) **234**, 887 (2002).

[209] N. Teofilov, K. Thonke, R. Sauer, L. Kirste, D. G. Ebling, and K. W. Benz, Diam. Relat. Mater. **11**, 892 (2002).

[210] S.-H. Wei, L. H. Ferreira, J. E. Bernard, and A. Zunger, Phys. Rev. B **42**, 9622 (1990).

[211] K. Binder, Phys. Rev. Lett. **45**, 811 (1980).

[212] L. C. Davis and H. Holloway, Solid State Commun. **64**, 121 (1987).

[213] A. Zunger, S.-H. Wei, L. H. Ferreira, and J. E. Bernard, Phys. Rev. Rel. **65**, 353 (1990).

[214] M. Suzuki and T. Uenoyama in Ref. [4], pp. 168-171.

[215] S. D. Lester, F. A. Ponce, M. G. Craford, and D. A. Steigerwald, Appl. Phys. Lett. **66**, 1249 (1995).

[216] S. Nakamura, T. Mukai, and M. Senoh, Appl. Phys. Lett. **64**, 1687 (1994).

[217] J. S. Speck and S. J. Rosner, Physica B **273-274**, 24 (1999).

[218] F. A. Ponce, D. P. Bour, W. Götz, N. M. Johnson, H. I. Helava, I. Grzegory, J. Jun, and S. Porowski, Appl. Phys. Lett. **68**, 917 (1996).

[219] C. Kirchner, V. Schwegler, F. Eberhard, M. Kamp, K. J. Ebeling, K. Kornitzer, T. Ebner, K. Thonke, R. Sauer, P. Prystawko, M. Leszczynski, I. Grzegory, and S. Porowski, Appl. Phys. Lett. **75**, 1098 (1999).

[220] M. Leszczynski, B. Beaumont, E. Frayssinet, W. Knap, P. Prystawko, T. Suski, I. Grzegory, and S. Porowski, Appl. Phys. Lett. **75**, 1276 (1999).

[221] P. Gibart, B. Beaumont, and P. Vennégùes in Ref. [259], pp. 45-106.

[222] L. T. Romano in Ref. [4], pp. 221-225.

[223] S. K. Mathis, A. E. Romanov, , L. C. Chen, G. E. Beltz, W. Pompe, and J. S. Speck, phys. stat. sol. (a) **179**, 125 (2000).

[224] L. Lymperakis, J. Neugebauer, M. Albrecht, T. Remmele, and H. P. Strunk, Phys. Rev. Lett. **93**, 196401 (2004).

[225] Y. Xin, S. J. Pennycook, N. D. Browning, P. D. Nellist, S. Sivananthan, F. Omnès, B. Beaumont, J. P. Faurie, and P. Gibart, Appl. Phys. Lett. **72**, 2680 (1998).

[226] Y. Xin, E. M. James, I. Arslan, S. Sivananthan, N. D. Browning, S. J. Pennycook, F. Omnès, B. Beaumont, J. P. Faurie, and P. Gibart, Appl. Phys. Lett. **76**, 466 (2000).

[227] I. Arslan and N. D. Browning, Phys. Rev. B **65**, 075310 (2002).

[228] U. Bangert, A. Gutiérrez-Sosa, A. J. Harvey, C. J. Fall, and R. Jones, J. Appl. Phys. **93**, 2728 (2003).

[229] S. J. Rosner, E. C. Carr, M. J. Ludowise, G. Girolami, and H. I. Erikson, Appl. Phys. Lett. **70**, 420 (1997).

[230] T. Hino, S. Tomiya, T. Miyajima, K. Yanashima, S. Hashimoto, and M. Ikeda, Appl. Phys. Lett. **76**, 3421 (2000).

[231] T. Sugahara, H. Sato, M. Hao, Y. Naoi, S. Kurai, S. Tottori, K. Yamashita, K. Nishino, L. T. Romano, and S. Sakai, Jpn. J. Appl. Phys. **37**, L398 (1998).

[232] J. Y. Shi, L. P. Yu, Y. Z. Wang, G. Y. Zhang, and H. Zhang, Appl. Phys. Lett. **80**, 2293 (2002).

[233] S. Dassonneville, A. Amokrane, B. Sieber, J.-L. Farvacque, B. Beaumont, and P. Gibart, J. Appl. Phys. **89**, 3736 (2001).

[234] S. M. Lee, M. A. Belkhir, X. Y. Zhu, Y. H. Lee, Y. G. Hwang, and T. Frauenheim, Phys. Rev. B **61**, 16033 (2000).

[235] C. J. Fall, R. Jones, P. R. Briddon, A. T. Blumenau, T. Frauenheim, and M. I. Heggie, Phys. Rev. B **65**, 245304 (2002).

[236] C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1956), 2nd ed.

[237] F. Seitz, *The Modern Theory of Solids* (McGraw-Hill, New York, 1940), 1st ed.

[238] L. D. Landau and E. M. Lifschitz, *Elastizitätstheorie*, vol. III of *Lehrbuch der Theoretischen Physik* (Akademie-Verlag, Berlin, 1970).

[239] R. de Wit, J. Res. Nat. Bur. Stand. (USA) **77A**, 608 (1973).

[240] J. P. Hirth and J. Lothe, *Theory of Dislocations* (McGraw-Hill, New York, 1968).

[241] J. W. Steed, *Introduction to Anisotropic Elasticity Theory of Dislocations* (Oxford University Press, Oxford, 1973).

[242] X. Blase, K. Lin, A. Canning, S. G. Louie, and D. C. Chrzan, Phys. Rev. Lett. **84**, 5780 (2000).

[243] N. Lehto and S. Öberg, Phys. Rev. Lett. **80**, 5568 (1998).

[244] S. Hearne, E. Chason, J. Han, J. A. Floro, J. Figiel, J. Hunter, H. Amano, and I. S. T. Tsong, Appl. Phys. Lett. **74**, 356 (1999).

[245] T. Bötther, S. Einfeldt, S. Figge, R. Chierchia, H. Heinke, D. Hommel, and J. S. Speck, Appl. Phys. Lett. **78**, 1976 (2001).

[246] T. Böttcher, Ph.D. thesis, University of Bremen, Bremen, Germany (2002).

[247] R. Juza and H. Hahn, Zeitschr. Anorgan. Allgem. Chem. **239**, 282 (1938).

[248] M. Leszczynski, H. Teisseyre, T. Suski,I. Grzegory, M. Bockowski, J. Jun, S. Porowski, K. Pakula, J. M. Baranowski, C. T. Foxon, and T. S. Cheng, Appl. Phys. Lett. **69**, 73 (1996).

[249] R. G. Powell, N.-E. Lee, Y.-W. Kim, and J. E. Greene, J. Appl. Phys. **73**, 189 (1993).

[250] F. D. Murnaghan, Proc. Nat. Acad. Sci. USA **30**, 244 (1944).

[251] C. Stampfl and C. G. V. de Walle, Phys. Rev. B **59**, 5521 (1999).

[252] K. Kim, W. R. L. Lambrecht, and B. Segall, Phys. Rev. B **50**, 1502 (1994).

[253] K. Kim, W. R. L. Lambrecht, and B. Segall, Phys. Rev. B **53**, 16310 (1996).

[254] V. A. Savastenko and A. U. Sheleg, Phys. Stat. Sol. (a) **48**, K135 (1978).

[255] Y. Tagaki, M. Ahart, T. Sota, K. Suzuki, and S. Nakamura, Physica B **219&220**, 547 (1996).

[256] R. M. Martin, Phys. Rev. B **6**, 4546 (1972).

[257] B. Jogaia, J. Appl. Phys. **90**, 699 (2001).

[258] N. V. Richardson and S. Holloway, editors, K. Horn and M. Scheffler, volume editors, *Electronic Structure*, volume 2 of *Handbook of Surface Science*, 2000, ELSEVIER.

[259] P. Ruterana, M. Albrecht, and J. Neugebauer, eds., *Nitride Semiconductors, Handbook on Materials and Devices* (Wiley-VCH, Weinheim-Germany, 2003), 1st ed.

Curriculum Vitae

Name	Liverios Lymperakis
Birthdate	26.03.1975

July 2000 / December 2004	Ph.D student Fritz-Haber-Institut der Max-Planck-Gesellschaft Berlin, Germany
September 1997 / June 2000	Master Science degree in ”Physics of materials” Physics Department, Aristotle University of Thessaloniki Thessaloniki, Greece
September 1992 / June 1997	Bachelor degree in ”Physics” Physics Department, Aristotle University of Thessaloniki Thessaloniki, Greece

Publication List

- C. D. Lee, R. M. Feenstra, J. E. Northrup, L. Lymperakis, and J. Neugebauer, *Morphology and surface reconstructions of m-plane GaN*, Mat. Res. Soc. Symp. Proc. **743**, L.4.1.1 (2003).
- Ch. Adelmann, L. Lymperakis, J. Brault, G. Mula, J. Neugebauer, and B. Daudin, *Phase Transitions on GaN Surfaces*, Mat. Res. Soc. Symp. Proc. **743**, L.3.9 (2003).
- C. D. Lee, R. M. Feenstra, J. E. Northrup, L. Lymperakis, and J. Neugebauer, *Morphology and surface reconstructions of GaN(1100)*, Appl. Phys. Lett. **82**, 1793-1795 (2003).
- Ch. Adelmann, J. Brault, G. Mula, B. Daudin, L. Lymperakis, and J. Neugebauer, *Gallium adsorption on (0001) GaN surfaces*, Phys. Rev. B **67**, 165419 (2003).
- L. Lymperakis, J. Neugebauer, M. Albrecht, T. Remmele, and H.P. Strunk, *Strain induced metallization and deep electronic states around threading dislocations in GaN*, Phys. Rev. Lett. **93**, 196401 (2004).
- M. Albrecht, L. Lymperakis, J. Neugebauer, J. E. Northrup, L. Kirste, M. Leroux, I. Grzegory, S. Porowski, and H. P. Strunk, *Chemically ordered* $\mathrm{Al_xGa_{1-x}N}$ *alloys: Spontaneous formation of natural quantum wells*, Phys. Rev. B. **71**, 035314 (2005).