Surfaces and Extended Defects in Wurtzite GaN

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vom

Fachbereich Physik
der Universität Gesamthochschule Paderborn

genehmigte

Dissertation

von Dipl. Math. Joachim Elsner

Eingereicht am 24. September 1998

Contents

Chapter 1

Introduction

1.1 Into the Blue: GaN

Semiconducting lasers on the basis of gallium arsenide (GaAs) and indium phosphide (InP) emit light in the red–infrared spectrum. These lasers have found many applications ranging from data storage on compact discs (CD) and data transmission via optical fibres to medical diagnostics and surgery.

Also blue and green lasers and laser diodes (LDs) are highly desirable. With their shorter wavelengths they would allow to reduce the storage space on CDs because the data can be "written" in a more compact form. Moreover, blue and green lasers are also expected to be employed in medical diagnostics: the red background colour of blood floated tissue makes the use of the available red laser light rather difficult whereas green or blue light would be much easier to recognise. In terms of market value, a very important by–product related to the development of semiconducting lasers are light emitting diodes (LEDs). Due to their high luminescence efficiency, quick response time and long lifetime LEDs are an attractive alternative to conventional sources of light: LEDs consume 10% of the energy of conventional light bulbs and have $\approx 10^3$ times longer lifetimes. As blue and green belong to the three primary colours (red, green, blue), blue and green LEDs are required to reproduce the full colour spectrum and achieve white light.

Having these applications in mind, many semiconductor companies started research for materials which can provide blue and green light. For a review see the textbook by Nakamura and Fasol [1]. Light emission in the blue spectrum requires band gaps of ≈ 3 eV. Only materials with direct band gaps are suitable to produce bright light because indirect band gaps require also phonons for optical transitions which reduce the luminescence efficiency. During the 80's, it was thought that lattice matching substrates were essential to grow materials in a stress–free manner. Fig. 1.1 shows the band gaps of compound semiconductors in dependence on the lattice constant.

With direct band gaps between ≈ 2.0 and ≈ 4.5 eV and lattice constants similar to that of GaAs, II–VI compounds such as ZnSSe were the favourite materials for a long time. However, although high quality II–VI materials were grown with densities of crystal defects below $10^4/\mathrm{cm}^2$, these state-of-the-art materials still show severe stability problems and degrade within hours thus making

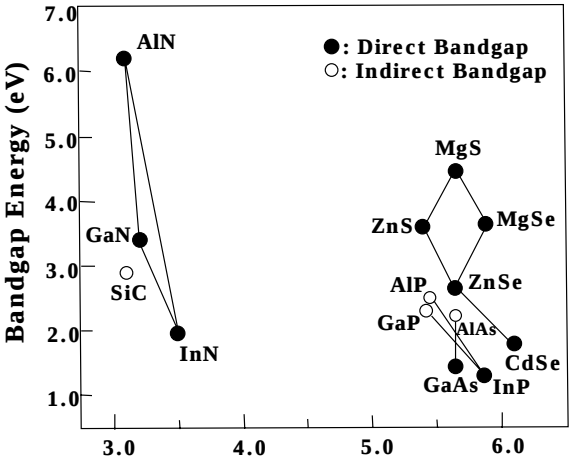

Figure 1.1: Band gap energy versus lattice constant of various materials for visible emission [1]. The lines indicate linear scaling of band gap energy and lattice constant for ternary compounds.

commercial applications impossible. It is generally thought that the rapid degradation is due to crystal defects because II–VI materials are very weakly bonded so that one defect can cause the propagation of other defects leading to failure of the devices even if the density of defects was low at the beginning. Another wide band material is SiC. However, SiC has an indirect band gap leading to very little brightness. Despite of their poor performance, 6H–SiC blue LEDs have been commercialised for a long time because no alternative existed [2, 3].

The remaining material group in Fig. 1.1 are group III–nitrides, in particular GaN which under standard growth conditions crystallises in the wurtzite phase (α–GaN). Since group III–nitrides possess direct band gaps ranging from $\approx 2.0 - 6.3$ eV, band gap engineering could lead to devices emitting bright light covering the entire spectrum from green to UV light.

However, until recently p–type doping of GaN could not be achieved so that efficiently working devices which require $p - n$ junctions could not be produced. This doping problem was overcome by Amano *et al.* [4] and Nakamura *et al.* [5] so that the first highly efficient light emitting devices based on GaN could be built and are commercially available now. p–type GaN films are obtained by Mg doping and annealing in an N_2 atmosphere [5].

The next major problem in the development of high–quality GaN based devices concerns the material quality. Sapphire is the most suitable substrate on which growth of α–GaN can be performed. However, sapphire has a lattice mismatch of 13% with respect to GaN. Therefore it is no surprise that attempts to grow GaN directly on sapphire resulted in a huge number of defects which penetrate the GaN epilayer (threading defects). The fabrication of light emitting devices based on these poor quality epilayers were obviously impossible. In part, this problem was overcome by Akasaki *et al.* [6] in 1989 who suggested that GaN growth on top of an AlN buffer layer (see Fig. 1.2) reduces significantly the defect density. This growth technique has been successfully adopted by Nakamura *et al.* [1] who found the epilayer quality to be still improved by replacing the AlN buffer layer by a GaN buffer layer. In addition, Nakamura *et al.* used a modified metal organic chemical vapour deposition (MOCVD) reactor where a subflow of N_2 and H_2 molecules provides a uniform distribution of the main flow gases to the substrate.

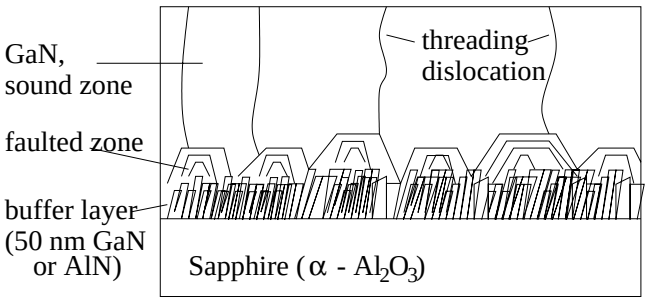

Figure 1.2: Using a GaN (or AlN) buffer layer, GaN can be grown with a sufficiently high quality on a sapphire substrate despite of the very large lattice mismatch. The buffer layer is normally grown at lower temperatures than the GaN device layers [6].

In spite of these improvements the density of threading defects (see Fig. 1.2) remains considerably high ($\approx 10^9/\mathrm{cm}^2$ in typical MOCVD grown epilayers) and gives rise to one of the most puzzling questions about GaN based devices [1]:

Why in general do GaN based devices work in spite of the large number of defects incorporated?

The answer is not yet known, instead it seems that most of the types of threading defects are harmless whereas the existence of some of them can seriously influence the material quality by inducing deep electronic states in the band gap. Deep states lead to parasitic components in the emission spectrum and, if occurring at high densities, render the sample useless for sophisticated optical applications, in particular for lasers [1].

The most commonly observed parasitic component in the GaN spectrum is the defect related broad band yellow luminescence (YL). The YL is generally associated with n–type material where it can have a strong intensity varying all over the material as shown in Fig. 1.3. The precise origin of the YL is not known and even the class of defect, e.g. point defect, line defect or planar defect, responsible for the YL remains unclear. A characterisation of the properties and possibly the origin of the most commonly observed defects in GaN is therefore a very important matter in order to get an idea about their influence on the material qualities and give growers hints on how to avoid "dangerous species". A variety of point defects and related defect complexes in GaN have already been investigated by theoretical groups [8, 9] and the results can be related to experimental work [10, 11]. On the other hand, most of the frequently occurring extended threading defects, in particular dislocations, have not been studied yet. This is mainly due to the fact that a theoretical investigation of extended defects is computationally very expensive since they require models which are much larger than those employed for studying point defects. Recently, numerically efficient theoretical methods have been implemented on parallel machines so that larger structures can now be investigated. The present thesis will try to fill the lack of information concerning the properties of the most commonly observed extended defects in wurtzite GaN. Also the interaction of these extended defects with stable point defects, impurities and related defect complexes will be explored.

Another important topic concerns the investigation of surfaces. A detailed knowledge of their properties is vital for high quality device fabrication, where in general electrically and chemically inactive surfaces are desired to form junctions and contacts. Therefore the stabilities of intrinsic GaN surfaces and their possible passivation by means of oxygen will be studied. These investigations are

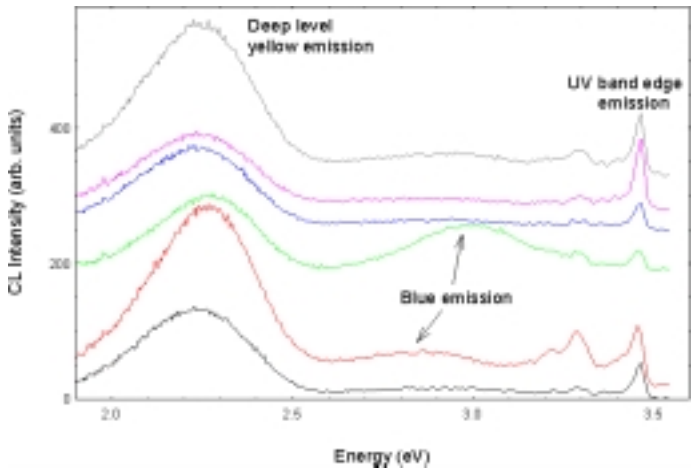

Figure 1.3: CL spectra from several different regions of a GaN film grown on a sapphire substrate. There are striking variations from point to point on the sample. *C. Traeger–Cowan et al.* [7].

related to the study of extended defects because some of the extended defects contain internal surfaces. Moreover, knowledge of the surface properties might help to determine growth conditions under which almost defect– and impurity–free material can be produced.

1.2 Outline of this thesis

In chapter 2, a brief summary of density functional theory which is a theoretical approach for determining total energies, structures, and to some extend electrical properties is given. The *AIMPRO* and *DFTB* methods used within this work are based on density functional theory and explained in the following chapters. Particular emphasis is given to the extension to periodic systems of the self-consistent charge *DFTB* method (*SCC–DFTB*) developed by M. Elstner for clusters. To this end in cooperation with D. Porezag and M. Haugk, the author developed a new functional expression suitable to account for the energy arising from charge fluctuations in extended and periodic systems. A benchmark will be given for a parallel code developed and implemented together with M. Haugk. Chapter 5 describes how total energies can be transformed into formation energies which will be used throughout the application chapters to comment on the stabilities of the structures.

The theory is applied in chapter 6 to characterise nonpolar wurtzite surfaces in view of the extended defects explored in the following. Chapter 7 and 8 constitute the main part of this thesis: threading line and planar defects are investigated in their pure form and with segregated impurities. Their possible implication for the yellow luminescence is discussed.

In chapter 9 polar surfaces corresponding to the main growth directions are investigated. A mechanism is suggested to identify the polarity during MBE growth in dependence on the surface reconstructions observed. Also the adsorption of oxygen will be discussed.

Finally, chapter 10 gives a summary of and an outlook on related topics in GaN.

Chapter 2

Density Functional Theory

Knowledge of the total energy of a structure is the key point in a theoretical investigation. Electronic structure calculations attempt to determine the total energy of a system of nuclei and their electrons. The structures considered in this work correspond to complex systems with a large number of degrees of freedom. It is therefore necessary to make a number of simplifications.

The motion of electrons and nuclei are described by the many–body Schrödinger equation. Within the Born–Oppenheimer approximation one separates the motion of the electrons from the motion of the nuclei so that for given nuclear positions in a first step one only needs to solve the many–electron Schrödinger equation. With this solution one can in a second step calculate the forces on the nuclei, optimise the nuclear positions with respect to the total energy and hence derive the equilibrium geometry.

The many–electron Schrödinger equation is usually solved using either wavefunction based methods such as the Hartree–Fock formalism or electron density based schemes such as density–functional theory (DFT). The latter approach is pursued within this work. It circumvents the computationally expensive aspects of theory by treating the *electron density* with only *three degrees of freedom* as the fundamental variable instead of the full wavefunction with $3 \times N_{\mathrm{el}}$ *degrees of freedom.*

All expressions used in the theory chapters are in atomic units, i.e. $\hbar, e, m_e$ and $4\pi\epsilon_0$ are set to unity. Therefore energies are given in Hartree (1 H = 27.21 eV), lengths are given in Bohr radii (1 $a_0 = 0.5292$ Å) and masses in multiples of the electron mass ($\mathrm{m}_e = 9.109 \times 10^{-31}$ kg). Furthermore, $\int d\mathbf{r}$ and $\int d\mathbf{r}\,'$ are denoted by $\int$ and $\int'$, respectively.

2.1 Reduction of the many–electron equation into effective one–electron (Kohn–Sham) equations

In 1964 Hohenberg and Kohn [12] established the basis of density–functional theory. DFT allows to determine the ground–state energy E_0 of an interacting electron gas in an external potential. Consider the many–electron problem:

$$H\ |\varphi_0(\{\mathbf{r_i}\})\rangle = E_0\ |\varphi_0(\{\mathbf{r_i}\})\rangle\ . \tag{2.1}$$

H is the Hamilton operator for N_{el} electrons in the field of N_{nuc} nuclei consisting of the following contributions: the kinetic energies of the electrons T_e, the Coulomb interaction between electrons V_{ee} and the Coulomb interaction between electrons and nuclei, denoted by a more general external potential V_{ext}:

$$H = T_e + V_{ee} + V_{ext} \ ,$$

where:

$$T_e = \sum_i^{N_{el}} -\frac{\nabla_i^2}{2} \ , \quad V_{ee} = \frac{1}{2} \sum_{\substack{i,j \\ i \neq j}}^{N_{el}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad \text{and} \quad V_{ext} = \sum_{i,I}^{N_{el},N_{nuc}} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \ .$$

The many–electron problem leads to a charge density n_0 for the ground–state:

$$n_0(\mathbf{r}) = \langle \varphi_0(\{\mathbf{r_i}\}) | \ \sum_{i=1}^{N_{el}} \delta(\mathbf{r} - \mathbf{r}_i) \ | \varphi_0(\{\mathbf{r_i}\}) \rangle \ .$$

Let us now define the wide class of densities

$$A_N = \{n(\mathbf{r}) | \text{n comes from an N–particle ground–state}\} \ .$$

Note that A_N contains all possible N–particle densities and does not refer to any specific external potential V_{ext}. The most important features of density–functional–theory can then be summarised within the following two theorems:

Theorem 1 : The ground–state energy E_0 of an electron gas is a functional of the ground–state charge density $n_0(\mathbf{r})$:

$$E_0 : A_N \to \mathbb{R}, \ n \to E_0[n] \ .$$

In words: for every possible ground–state charge density n, there is one and only one ground–state energy $E_0[n]$.

Based on this theorem the energy E_0 can be written as a functional of the charge–density n_0:

$$E_0[V_{ext}] = E_0[V_{ext}[n_0]] = \int n_0(\mathbf{r}) V_{ext}(\mathbf{r}) + F[n_0] \ .$$

Here $F[n]$ is defined on A_N by Eq. (2.1) as:

$$F[n] = E_0[n] - \int n(\mathbf{r}) V_{ext}(\mathbf{r}) \ .$$

Theorem 2 : The functional $E_0[V_{ext}, n]$ attains its minimum at $n = n_0$:

$$E_0[V_{ext}, n_0] \leq E_0[V_{ext}, n] \ .$$

In order to evaluate $E_0[V_{\text{ext}}, n]$ explicitly Kohn and Sham [13] proposed to write the functional $F[n]$ as:

$$F[n] = T_s[n] + E_H[n] + E_{\text{XC}}[n]. \tag{2.2}$$

Here $T_s[n]$ is the true kinetic energy of a non–interacting electron gas with density n. $E_H[n]$ is the so–called Hartree–energy:

$$E_H[n] = \frac{1}{2} \int\int' \frac{n(\mathbf{r}')n(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} \ .$$

The last term in Eq. (2.2) is called exchange and correlation energy (XC–energy). In addition to the actual exchange and correlation energy it contains contributions of the kinetic energy and corrections arising from the self–interaction of particles:

$$E_{\text{XC}}[n] = T_e[n] - T_s[n] + V_{ee}[n] - E_H[n] \ .$$

With the decomposition (2.2) the energy functional $E_0[V_{\text{ext}}, n]$ reads:

$$E_0[V_{\text{ext}}(n)] = T_s[n] + \frac{1}{2} \int\int' \frac{n(\mathbf{r}')\, n(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} + \int V_{\text{ext}}(\mathbf{r})\, n(\mathbf{r}) + E_{\text{XC}}[n] \ . \tag{2.3}$$

Applying the variational–principle (Theorem 2 guarantees a minimum in A_N) yields a condition for the ground–state charge density n_0:

$$\frac{\delta T_s}{\delta n} + V_{\text{ext}} + \int' \frac{n(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} + V_{\text{XC}}[n] - \mu \Big|_{n=n_0} = 0 \ . \tag{2.4}$$

μ is a Lagrange–multiplier, expressing particle conservation:

$$N_{\text{el}} = \int n(\mathbf{r}) \ . \tag{2.5}$$

V_{XC} is the functional derivative of E_{XC} and is called exchange– and correlation potential:

$$V_{\text{XC}} = \frac{\delta E_{\text{XC}}[n]}{\delta n} \ .$$

Eq. (2.3) and (2.4) determine the ground–state energy and the ground–state charge density. However the functionals $T_s[n]$, $E_{\text{XC}}[n]$ and $V_{\text{XC}}[n]$ in these equations are still unknown. The functionals $E_{\text{XC}}[n]$ and $V_{\text{XC}}[n]$ can only be approximated. The most commonly used approximation is described in section 2.2. On the other hand, $T_s[n]$ can be determined exactly assuming a potential $V_{\text{eff}}[n]$, for which n is the ground–state charge density of a system of non–interacting electrons. Although this potential V_{eff} does not always exist [14], it can be assumed for most applications. Let now $|\psi_i\rangle$ $i = 1, \ldots, N_{\text{el}}$ denote the one–particle wavefunctions of the non–interacting electron gas described with the effective potential V_{eff}. We then have:

$$(-\frac{\nabla^2}{2} + V_{\text{eff}}[n])|\psi_i\rangle = \varepsilon_i|\psi_i\rangle \tag{2.6}$$

and

$$n(\mathbf{r}) = \sum_i^{occ} n_i |\psi_i(\mathbf{r})\rangle\langle\psi_i(\mathbf{r})| \ . \tag{2.7}$$

The variables n_i are the occupation numbers of the one–particle states. Assuming vanishing temperature and the system to be in the ground–state the occupation numbers are given by

$$n_i = 2\,\Theta(\varepsilon_i - \mu)\ . \tag{2.8}$$

Here Θ is the Heaviside step function, i.e. $\Theta(x) = 0,\ x \leq 0$ and $\Theta(x) = 1,\ x > 0$, μ is the Lagrange–multiplier from Eq. (2.4) and corresponds to the electron chemical potential. μ is determined by equations (2.7) and (2.5).

Using the expansion of n in the one–particle states $|\psi_i\rangle$ the kinetic energy $T_s[n]$ can be written in the simple form:

$$T_s[n] = \sum_i^{occ} n_i \langle\psi_i| -\frac{\nabla^2}{2}|\psi_i\rangle\ . \tag{2.9}$$

Thus the energy functional $E_0[n, V_{\rm ext}]$ finally reads:

$$\begin{aligned} E_0[n, V_{\rm ext}] \;&=\; \sum_i^{occ} n_i \langle\psi_i| -\frac{\nabla^2}{2}|\psi_i\rangle + \frac{1}{2}\int\int{}' \frac{n(\mathbf{r}')\,n(\mathbf{r})}{|\mathbf{r}'-\mathbf{r}|} \\ &\quad + \int V_{\rm ext}(\mathbf{r})\,n(\mathbf{r}) \;+ E_{\rm XC}[n]\ . \end{aligned} \tag{2.10}$$

Of course, the one–particle states $|\psi_i\rangle$ are related to n via (2.7).

The ground–state density n_0 for which the functional (2.10) attains the minimum can be calculated using Hohenberg and Kohn's second theorem. To this end we consider a system of non–interacting electrons which in an external potential $V_{\rm eff}[n_0]$ has the same ground–state density n_0 as the system of interacting electrons. Following Hohenberg–Kohn's second theorem the functional of the non–interacting electron gas attains its minimum at n_0, too, i.e. the functional is stationary at n_0:

$$E_{s,0}[n, V_{\rm eff}[n_0]] = T_s[n] + \int V_{\rm eff}[n_0](\mathbf{r})\,n(\mathbf{r}) \tag{2.11}$$

$$\frac{\delta E_{s,0}[n, V_{\rm eff}[n_0]]}{\delta n} - \mu'\Big|_{n=n_0} = \frac{\delta T_s[n]}{\delta n} + V_{\rm eff}[n_0](\mathbf{r}) - \mu'\Big|_{n=n_0} = 0\ . \tag{2.12}$$

Comparing (2.12) and (2.4) we derive the following form for the effective one–particle potential $V_{\rm eff}[n_0]$:

$$V_{\rm eff}[n_0] = V_{\rm ext} + \int{}' \frac{n_0(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|} + V_{\rm XC}[n_0]\ . \tag{2.13}$$

Inserting this expression into Eq. (2.6), we obtain the eigenvalue-problem:

$$(-\frac{\nabla^2}{2} + \underbrace{V_{\rm ext} + \int{}' \frac{n_0(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|} + V_{\rm XC}[n_0])}_{V_{\rm eff}}\,|\psi_i\rangle = \varepsilon_i|\psi_i\rangle\ . \tag{2.14}$$

$$n_0(\mathbf{r}) = \sum_i^{occ} n_i |\psi_i(\mathbf{r})\rangle\langle\psi_i(\mathbf{r})| \tag{2.15}$$

These two equations are coupled via the construction of $\mathrm{V_{eff}}$ and together constitute the *Kohn–Sham equations*. They have to be solved self–consistently, i.e. the effective potential $\mathrm{V_{eff}}$ constructed from the charge density via (2.13) put into (2.14) must lead to the one–particle states out of which the

charge density has been constructed via 2.15. The self–consistent solution gives then the ground–state charge density of the system. Apart from the exchange and correlation potential $E_{\rm XC}$ all terms in (2.10) are known explicitly. We can thus write for the ground–state energy E_0 of an interacting electron gas in an external potential:

$$E_0 = \sum_i^{occ} n_i \langle \psi_i | -\frac{\nabla^2}{2} | \psi_i \rangle + \frac{1}{2} \int \int ' \frac{n_0(\mathbf{r}') n_0(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} + \int V_{\rm ext}(\mathbf{r})\, n_0(\mathbf{r}) + E_{\rm XC}[n_0] \; . \tag{2.16}$$

Multiplying (2.14) with $\langle \Psi_i^* |$, summing over the occupied states and using (2.7) and (2.9) we derive the identity:

$$T_s[n_0] = \sum_i^{occ} n_i \varepsilon_i - \int V_{\rm ext}(\mathbf{r}) n_0(\mathbf{r}) - \int \int ' \frac{n_0(\mathbf{r}) n_0(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} - \int V_{\rm XC}[n_0](\mathbf{r}) n_0(\mathbf{r}) \; .$$

Inserting it in (2.3), we obtain an alternative form for the total energy:

$$E_0 = \sum_i^{occ} n_i \varepsilon_i - \frac{1}{2} \int \int ' \frac{n_0(\mathbf{r}') n_0(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} - \int V_{\rm XC}[n_0](\mathbf{r}) n_0(\mathbf{r}) + E_{\rm XC}[n_0] \; . \tag{2.17}$$

The total energy is therefore the sum of the one–particle eigenvalues frequently called *band structure energy* corrected by terms usually referred to as *double counting* contributions plus the exchange–correlation energy $E_{\rm XC}$.

2.2 Approximating the exchange and correlation energy: the local density approximation (LDA)

The transformation of the many–electron problem (2.1) to a system of effective one–electron equations (2.14) has been exact so far, i.e. no approximations have been made up to this point. However, the one particle Hamilton operator (2.14) and the expressions for the total energies (2.16) and (2.17) contain the functionals $V_{\rm XC}[n]$ and $E_{\rm XC}[n]$. Although the existence of these functionals can be mathematically justified in most cases their explicit form is unknown. Using quantum Monte–Carlo calculations one could in principle determine numerically the exchange and correlation energy of a given system up to any required accuracy. However, quantum Monte–Carlo calculations are computationally too expensive to treat any but the simplest systems such as the uniform electron gas [15].

The most common approximation for the exchange and correlation energy and potential is the *local density approximation* (LDA). Within the LDA approximation one assumes that for any small region in the system, the exchange-correlation is the same as that for the uniform electron gas with the same electron density. This approximation applies to a *non–spinpolarized* system, and the exchange-correlation energy is approximated by:

$$E_{\rm xc} = \int n(\mathbf{r}) \epsilon_{\rm xc}(n) \; ,$$

where $\epsilon_{\rm xc}(n)$ is the exchange-correlation density for the homogeneous electron gas. For a spin polarised system, one simply applies the same assumptions using the exchange-correlation energy

density of the spin-polarised uniform electron-gas, $\epsilon_{\mathrm{xc}}(n_\uparrow, n_\downarrow)$. This is termed the local spin density approximation (LSDA) and implementing this within DFT is often called local spin density functional theory (LSDFT).

It is possible to go beyond such a *local* approximation and to consider further terms in a series representation of the charge density, termed the *generalised gradient approximation* (GGA) [16]. However, the merits of such an approach are not accepted universally. Therefore in this work the exchange and correlation energies and potentials are exclusively approximated by LDA.

Summary

It has been shown that DFT results in a simple yet powerful way of solving the many electron Schrödinger equation: the whole problem reduces to find the solutions of one–particle equations, the Kohn–Sham equations (2.14), in an effective potential. It should however be noted that in contrast to Hartree–Fock theory, which in principle gives meaningful one–particle states and eigenvalues that correspond to the true ionisation energies (Koopman's theorem) of the system the wavefunctions ψ_i derived in density–functional theory do not satisfy this condition. Instead they are related to occupation numbers. It is possible to augment the density functional theory with GW theory[1] which predicts quasi-particle energies with reasonable accuracy [18].

[1] A full discussion of GW theory is beyond the scope of this thesis, but the framework within which the GW approximation is formulated is that of a perturbation expansion of one–particle Greens functions $G(p, w)$. See for example reference [17]

Chapter 3

AIMPRO Methodology

In the previous chapter effective one–particle equations, the Kohn–Sham equations, have been deduced from the many–body Schrödinger equation. In this chapter the main features of a practical approach to the solution of the Kohn–Sham equations implemented in the *ab initio* modelling *pro*gram (AIMPRO) is described. The AIMPRO method uses pseudopotentials and a Gaussian basis set for the expansion of the one–particle wavefunctions. A full review of the methodology behind AIMPRO is given in Reference [19]. Currently on parallel machines AIMPRO can be applied routinely to clusters of sizes up to ~ 350 atoms. In this thesis AIMPRO will be used to determine the electronic properties of extended defects in GaN.

3.1 Pseudopotentials

DFT as described above would still prove computationally too difficult for system sizes useful for the modelling of surfaces, point and line defects in semiconductors. This is due to the fact that these systems, especially if they contain *non first row elements* usually have a large number of electrons $N_{\rm el}$ and according to equation (2.8) this number is correlated with the number of Kohn–Sham equations (2.14) $N_{\rm KS}$ to be solved by $N_{\rm KS} \sim N_{\rm el}/2$. However, not all the electrons need be considered. They are divided into two groups: core and valence electrons. The former are bound close to the ions and do not play a part in bonding. It is highly advantageous to remove these from the theory. This can be done through the use of pseudopotentials. These rely on the fact that only the valence electrons are involved in chemical bonding. Therefore it is possible to incorporate the core states into a bulk nuclear potential, or pseudo-potential, and only deal with the valence electrons separately.

For example, in Ga the $1s$, $2s$, $2p$, $3s$, $3p$ (and sometimes also the $3d$) electrons are regarded as part of the core while the $4s$ and $4p$ (and in the cases where they are not included in the core also the $3d$) electrons are part of the valence shell. In the same way, only N $2s$ and $2p$ electrons are considered as valence electrons whereas the N $1s$ electrons are regarded as part of the core. The four bonds associated with a chemical unit of GaN in the tetrahedrally bonded GaN bulk arise from these eight valence electrons.

By reducing the problem in a way that without modifying the result only the Kohn–Sham equations for these valence electrons need to be solved, a considerable simplification is achieved. The pseudopotential is constructed by insisting that it has exactly the same valence energy levels as the true atomic potential, e.g. the $4s$ and $4p$ pseudopotential levels in Ga are the same as the $4s$ and $4p$ levels in an all–electron calculation. Moreover, the corresponding pseudo-wavefunctions are exactly equal to the true wavefunctions outside a core whose size depends on the type of the atom. Inside the core, the pseudo-wavefunctions are not equal to the true valence ones but are very smooth nodeless functions which are easy to represent mathematically. On the other hand, the true $4s$ and $4p$ wavefunctions of Ga oscillate inside the core and are rather difficult to represent mathematically. As the pseudo-wavefunctions are nodeless inside the core the pseudopotentials have no core states at all. Indeed, the lowest bound state solutions are the valence energy eigenvalues and, by construction of the pseudopotential, these are the same as the true valence energy levels.

In addition to the advantages achieved by reducing the number of one–particle equations and by simplifying the evaluation of integrals in the core region pseudopotentials allow for the treatment of heavier atoms in which relativistic effects are important. Whereas for a description of the core electrons the Dirac equation is required, the valence electrons can be treated non-relativistically. Therefore, removal of the core electrons allows a non-relativistic approach to be maintained (some corrections must be included in the core electron pseudopotentials to account for relativistic effects).

The pseudopotentials used in *AIMPRO* come from work done by Bachelet, Hamann and Schlüter [20] who produced pseudopotentials for all elements up to Pu.

3.2 Numerical solution via a basis set expansion into Gaussian orbitals

Despite the great simplifications gained by the use of pseudopotentials the numerical solution of the interacting one–particle equations (2.14) in the effective potential $V_{\rm eff}$ remains a challenging task. For a practical solution the one–particle wavefunctions $|\psi_i(\mathbf{r})\rangle$ have to be expanded in terms of a basis $|\phi_\nu(\mathbf{r})\rangle$:

$$|\psi_i(\mathbf{r})\rangle = \sum_{\nu}^{M} c_{\nu i}|\phi_\nu(\mathbf{r})\rangle \ . \tag{3.1}$$

Two choices for the basis functions $|\phi_\nu\rangle$ are often made: plane waves and Cartesian Gaussian orbitals. Using the first is equivalent to making a Fourier transform of the wavefunction. This has problems with certain elements, e.g. first row, transition and rare-earth elements where the wavefunction varies rapidly and a great number of plane waves are required.

Within the *AIMPRO* methodology, Cartesian Gaussian orbitals are used as basis functions. Centred at the point R_ν they have the form:

$$\phi_\nu(\mathbf{r}) = (x - R_{\nu x})^{n_1}(y - R_{\nu y})^{n_2}(z - R_{\nu z})^{n_3} e^{-a_\nu(\mathbf{r}-\mathbf{R}_\nu)^2}, \quad n_i \in \ I\!N.$$

In practice, they are centred at the nuclei and sometimes at bond centred sites between the nuclei. From suitable combinations of n_i, one can construct the s, p, d, ... orbitals. For example, $n_i = 0$, $i = 1, 2, 3$ generates an s-Gaussian orbital.

The use of Gaussian orbitals as a basis set has considerable advantages:

1. In contrast to a plane wave basis, Gaussian orbitals are localised and therefore a significantly smaller basis set is needed to describe localised wavefunctions. In GaN this is particularly important since Ga $3d$ and N $2s$ wavefunctions are very localised.
2. In contrast to Slater type orbitals, which are also localised, matrix elements in Gaussian orbitals can be evaluated analytically leading to a reduction in the computational task.

Inserting the expansions (3.1) for the one–particle wavefunctions into (2.14) gives:

$$(-\frac{\nabla^2}{2} + V_{\text{eff}})|\sum_{\nu}^{M} c_{\nu i}\phi_\nu\rangle = \varepsilon_i|\sum_{\nu}^{M} c_{\nu i}\phi_\nu\rangle \ . \tag{3.2}$$

Multiplying this equation by $\langle\phi_\mu^*$ for $1 \leq \mu \leq M$ reduces the differential equation to a set of algebraic equations:

$$\sum_{\nu}^{M} c_{\nu i}(H_{\mu\nu} - \varepsilon_i S_{\mu\nu}) = 0 \ , \quad 1 \leq \mu, i \leq M \ ,$$

which in matrix notation read as a generalised eigenvalue problem. $H_{\mu\nu}$ and $S_{\mu\nu}$ are matrix elements of the Hamiltonian and the overlap matrix, respectively:

$$\begin{aligned} H_{\mu\nu} &= \langle\phi_\mu| - \frac{\nabla^2}{2} + V_{\text{eff}}|\phi_\nu\rangle \\ S_{\mu\nu} &= \langle\phi_\mu|\phi_\nu\rangle \ . \end{aligned}$$

Once this equation has been solved we can generate the *output* charge density:

$$n_{\text{out}}(\mathbf{r}) = \sum_{i}^{occ} n_i|\psi_i(\mathbf{r})\rangle\langle\psi_i(\mathbf{r})| \ .$$

In general, this is not equal to the *input* charge density used to generate the effective potential V_{eff} according to (2.13). Of course, this is an inconsistency and as indicated in the previous chapter we need to devise a scheme whereby the input charge density, which is used to construct V_{eff}, is equal to the output density. This is usually carried out by an iterative procedure and the process of achieving equality in the densities is called the self-consistent cycle. Methods which use a self–consistent cycle in this way are also referred to as self–consistent field (SCF) methods. With this self–consistent density the total energy can be calculated via (2.17). The atomic forces needed to perform a structural relaxation follow from the derivatives of the total energy with respect to the atomic coordinates via the application of the Hellman–Feynman theorem. Further details of the *AIMPRO* methodology, and in particular a description of the practical realisation of the implementation can be found in the review written by Jones and Briddon [19].

3.3 Application of AIMPRO to GaN

In this work *AIMPRO* is used to determine the structural and electrical properties of point defects and extended defects in GaN. All the structures investigated with *AIMPRO* are modelled in clusters

(see appendix D). In this work no absolute energies are calculated with the *AIMPRO* method. An explicit inclusion of the Ga $3d$ electrons as valence electrons, which hybridise with the N $2s$ electrons and thus might influence the absolute energy of a structure, is therefore not necessary. In this work *AIMPRO* is rather used to investigate the electrical properties which are well described within a converged basis of the Ga $4s$, $4p$ and N $2s$, $2p$ electrons. In the current application we use 5 (4) s and p-Gaussian orbitals with different exponents for each Ga (N) atom yielding a large real–space basis set of 20 (16) Gaussian orbitals on every Ga (N) atom. Applied to a 72 atom H-terminated stoichiometric cluster arranged as in wurtzite GaN, gave bond lengths within 4% of experiment and a maximum vibrational frequency of 729 cm^{-1} compared with 741 cm^{-1} experimentally found for the $E_1(LO)$ mode [21].

Summary

The *AIMPRO* method has been presented which solves the Kohn–Sham equations within a Gaussian basis. Since Gaussian basis functions are localised *AIMPRO* is a suitable program to describe the N $2s$ wavefunctions in GaN.

Chapter 4

The density functional based tight binding methodology

Density–functional theory realized with pseudopotentials and a Gaussian basis set as implemented in the *AIMPRO* methodology described in the previous chapter allows to treat accurately non–periodic structures within clusters up to ~ 350 atoms. However, some of the structures considered in this work require models which considerably exceed this size. Because of limited computer resources these structures cannot be investigated by any method based on a rigorous implementation of density functional theory, like *AIMPRO* but require more approximate schemes.

Empirically constructed potentials, derived from fitting parameters to equilibrium structures are very efficient and thus capable of dealing with extended systems. However, they suffer from a transferability problem: they generally only work well within the regime in which they were fitted and thus are not predictive in structural simulations. Therefore, more approximate schemes combining the advantage of the efficiency of the empirical potentials with the transferability and accuracy of the *SCF* methods are highly desirable. In this context, besides numerous traditional quantum chemical methods, tight-binding (TB) models have recently become very popular [22, 23, 24], providing one of the most accurate alternatives in the determination of the total energy and equilibrium geometry of various systems. In particular, two–centre–oriented schemes considering only two–centre integrals in the Hamiltonian give results that deviate only slightly from those of more sophisticated methods. In the standard TB-method eigenstates of a Hamiltonian are expanded in an orthogonalised basis of atomic-like orbitals. The exact many-body Hamilton operator is represented by a parameterised Hamiltonian matrix, where the matrix elements are fitted to the band structure of suitable reference systems. However, the choice of reference systems to which the matrix elements are fitted is rather arbitrary. The fitting is thus often complicated and does not guarantee a general transferability to all scale systems.

To circumvent this problem a density–functional based tight binding scheme *DFTB* has been developed [25, 26]. Using a minimal basis linear combination of atomic orbitals (LCAO) - framework [25] this scheme avoids difficulties arising from an empirical parametrisation. Instead, the two–centre Hamiltonian and overlap matrix elements are calculated from atom–centred valence–electron orbitals and atomic potentials are derived from *SCF* single–atom calculations within the local–density approximation. This scheme can be seen as an approximate DFT scheme. In contrast with the usual

parametrised TB schemes it has a well defined procedure for determining the desired matrix elements. Moreover, interactions extending beyond the first shell of neighbours are taken into account which is not the case in empirical TB methods. This is of crucial importance for GaN, because the second neighbour Ga-Ga distances (3.18 Å) are comparable to the distances between neighbouring atoms in bulk Ga which range from 2.45 Å to 2.71 Å.

In GaN systems charge transfer may play an important role, especially if bonds are broken as at surfaces and extended defects, since although GaN is a covalently bonded material, it has partial ionic character. Recently, we have extended the *DFTB* scheme by incorporating a self-consistency cycle for the atomic charges [27]. This self-consistent charge density–functional based tight binding (*SCC–DFTB*) scheme can be used to investigate systems where interatomic charge transfer plays an important role.

In the following, the *DFTB* scheme and its charge self-consistent extension *SCC–DFTB* will be deduced following Elstner *et al.* [27] from general density–functional theory described in chapter 2. Particular emphasis will be given to the extension of *SCC–DFTB* to periodic systems where within this thesis a well behaved functional is derived to describe the energy arising from charge transfer for extended systems [27]. The numerical efficiency will be outlined. It will be shown that as we have implemented the code on parallel machines [28] a high degree of parallelisation can be reached making *SCC–DFTB* a very useful scheme for treating the extended GaN systems examined within this thesis.

4.1 Density-functional basis of TB-theory

According to DFT described in chapter 2 the total energy of a system of $N_{\rm el}$ electrons in the field of $N_{\rm nuc}$ nuclei at positions $\mathbf{R}_I$ may be written as a functional of a charge density $n(\mathbf{r})$ (see (2.10)):

$$E = \sum_i^{occ} n_i \langle \psi_i | -\frac{\nabla^2}{2} + V_{\rm ext} + \frac{1}{2}\int' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}|\psi_i\rangle + E_{\rm XC}[n(\mathbf{r})] + \frac{1}{2}\sum_{I,J}^{N_{\rm nuc}} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \ . \tag{4.1}$$

The first sum is over occupied *Kohn-Sham* eigenstates ψ_i with occupation numbers n_i, the second term is the exchange-correlation (XC) contribution and the last term which was not considered in (2.10) accounts for the ion-ion core repulsion, E_{ii}.

We now follow the approach of Foulkes and Haydock [29] and substitute the "true", usually self–consistently determined charge density $n = n(\mathbf{r})$ in (4.1) by a superposition of a reference or input density $n_0 = n_0(\mathbf{r})$ and a small fluctuation $\delta n = \delta n(\mathbf{r})$:

$$n = n_0 + \delta n \ .$$

Inserting this superposition into (4.1) gives for the total energy:

$$\begin{aligned} E \ = \ & \sum_i^{occ} n_i \langle \psi_i | -\frac{\nabla^2}{2} + V_{\rm ext} + \int' \frac{n_0'}{|\mathbf{r}-\mathbf{r}'|} + V_{\rm XC}[n_0]|\psi_i\rangle - \frac{1}{2}\iint' \frac{n_0'(n_0+\delta n)}{|\mathbf{r}-\mathbf{r}'|} \\ & - \int V_{\rm XC}[n_0](n_0+\delta n) + \frac{1}{2}\iint' \frac{\delta n'(n_0+\delta n)}{|\mathbf{r}-\mathbf{r}'|} + E_{\rm XC}[n_0+\delta n] + E_{ii} \ , \end{aligned}$$

where E_{ii} has been introduced as a shorthand for the ion–ion repulsion. The second term in this equation corrects for the double-counting of the new *Hartree*, the third term for the new XC contribution in the leading matrix element and the fourth term comes from dividing the full *Hartree* energy in (4.1) into a part related to n_0 and to δn.

Expanding $E_{\rm XC}$ at the reference density into a Taylors series, which we truncate after the second order terms, gives the total energy correct to second order in the density fluctuations. After a simple transformation this reads:

$$\begin{aligned} E = \sum_i^{occ} n_i \langle\psi_i|\hat{H}_0|\psi_i\rangle \quad &- \quad \frac{1}{2}\iint' \frac{n_0' n_0}{|\mathbf{r}-\mathbf{r}'|} + E_{\rm XC}[n_0] - \int V_{\rm XC}[n_0] n_0 + E_{ii} \\ &+ \quad \frac{1}{2}\iint' \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \left.\frac{\delta^2 E_{\rm XC}}{\delta n\, \delta n'}\right|_{n_0} \right) \delta n\, \delta n'. \end{aligned} \tag{4.2}$$

Note, that the terms linear in δn cancel each other at any arbitrary input density n_0. $\hat{H}_0$ denotes the Hamiltonian operator resulting from the input density n_0.

4.2 Zero-th order approximation: standard–DFTB

Within the *standard–DFTB* method one supposes that the initial charge density n_0 is very close to the real self–consistent density n. In this case δn_0 is sufficiently small so that all higher order terms, in particular the last term in the final equation for the total energy (4.2) can be neglected. A frozen-core approximation is applied to reduce the computational efforts by only considering the valence orbitals. A frozen core approximation differs from the pseudopotential approach described in chapter 3 mainly by the fact that since the core wave functions are not modified also the valence wavefunctions oscillate within the core region requiring a fine integration mesh. However, as will be seen later, within the *SCC-DFTB* scheme the evaluation of integrals plays only a minor role in the computational effort so that a frozen–core approach is suitable. The *Kohn-Sham* equations are then solved non-self-consistently and the second-order correction is neglected. The contributions in (4.2) that depend on the input density n_0 only and the core-core repulsion are taken to be a sum of one- and two-body potentials [29]. The latter, denoted by $E_{\rm rep}$, are strictly pairwise, repulsive and short-ranged. The total energy then reads:

$$E_0^{\rm TB} = \sum_i^{occ} n_i \langle\psi_i|\hat{H}_0|\psi_i\rangle + E_{\rm rep}\ . \tag{4.3}$$

Making a linear combination of atomic orbital (LCAO)-*ansatz* the single-particle wavefunctions ψ_i are expanded into a suitable set of localised atomic orbitals ϕ_ν:

$$|\psi_i(\mathbf{r})\rangle = \sum_\nu^M c_{\nu i} |\phi_\nu(\mathbf{r}-\mathbf{R}_I)\rangle\ . \tag{4.4}$$

We employ confined atomic orbitals ϕ_ν in a Slater-type representation. These are determined by solving a modified Schrödinger equation for a free neutral pseudoatom within SCF-LDA calculations [26]. For further details of the construction of these basis functions see also [30].

Within this expansion (4.3) transforms to:

$$E_0^{\mathrm{TB}} = \sum_i^{occ} n_i \sum_\mu^M \sum_\nu^M c_{\mu i}^* H_{\mu\nu}^0 c_{\nu i} + E_{\mathrm{rep}} \ , \tag{4.5}$$

where

$$H_{\mu\nu}^0 = \langle \phi_\mu | \hat{H}_0 | \phi_\nu \rangle$$

are the Hamilton matrix elements in the LCAO basis.

Denoting the overlap matrix elements with

$$S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle$$

and applying the Lagrange formalism with the condition that the number of electrons in the system remains constant, i.e. $\sum_i^{occ} n_i = N_{\mathrm{el}}$, to this zero-th order energy functional (4.5) gives:

$$\frac{\partial}{\partial c_{\mu i}^*} \Big[E_0^{\mathrm{TB}} - \epsilon_i \left(\sum_\delta^{occ} n_\delta \sum_\sigma^M \sum_\nu^M c_{\sigma\delta}^* S_{\sigma\nu} c_{\nu\delta} - N_{\mathrm{el}} \right) \Big] = 0 \ . \tag{4.6}$$

The derivatives are:

$$\frac{\partial E_0^{\mathrm{TB}}}{\partial c_{\mu i}^*} = n_i \sum_\nu^M H_{\mu\nu}^0 c_{\nu i}$$

and

$$\frac{\partial}{\partial c_{\mu i}^*} \varepsilon_i \left(\sum_\delta^{occ} n_\delta \sum_\sigma^M \sum_\nu^M c_{\sigma\delta}^* S_{\sigma\nu} c_{\nu\delta} - N_{\mathrm{el}} \right) = \epsilon_i n_i \sum_\nu^M S_{\mu\nu} c_{\nu i} \ .$$

Inserting these derivatives into (4.6) gives the *Kohn-Sham* equations of the zero-th order *DFTB* method as a set of algebraic equations:

$$\sum_\nu^M c_{\nu i} (H_{\mu\nu}^0 - \varepsilon_i S_{\mu\nu}) = 0 \ , \quad 1 \le \mu, i \le M \ . \tag{4.7}$$

Within the *DFTB* method the effective one-electron potential of the many-atom structure V_{eff} in

$$\hat{H}_0 = \hat{T} + V_{\mathrm{eff}}$$

is approximated as a sum of spherical *Kohn-Sham* potentials of neutral pseudoatoms due to their *confined* electron density. Furthermore, several terms in the Hamilton matrix elements, in particular multicentre contributions are neglected. This is consistent with the construction of the effective one-electron potential [25] and gives:

$$H_{\mu\nu}^0 = \begin{cases} \varepsilon_\mu^{\text{neutral free atom}} & \text{if } \mu = \nu \\ \langle \phi_\mu^I | \hat{T} + V_0^I + V_0^J | \phi_\nu^J \rangle & \text{if } I \neq J \\ 0 & \text{otherwise} \end{cases} \ . \tag{4.8}$$

Since indices I and J indicate the atoms on which the wavefunctions and potentials are centred, only two-centre Hamiltonian matrix elements are treated and explicitly evaluated in combination with the two-centre overlap matrix elements. As follows from (4.8), the eigenvalues of the free atom serve as diagonal elements of the Hamiltonian, thus guaranteeing the correct limit for isolated atoms.

By solving the general eigenvalue problem (4.7), the first term in (4.5) becomes a simple summation over the eigenvalues ε_i of all occupied Kohn-Sham orbitals:

$$E_0^{\mathrm{TB}} = \sum_i^{occ} n_i \varepsilon_i + E_{\mathrm{rep}} \ . \tag{4.9}$$

A transferable parametrised short range potential $E_{\mathrm{rep}} = \frac{1}{2}\sum_{I,J}^{I\neq J} \Phi(I,J)$ can easily be determined as a function of distance by taking the difference of the SCF-LDA cohesive and the corresponding TB band structure energy for a suitable reference system:

$$E_{\mathrm{rep}}(R) = \frac{1}{2}\sum_{I,J}^{I\neq J} \Phi(I,J) = \left\{ E_{LDA}^{SCF}(R) - \sum_i^{occ} n_i \varepsilon_i(R) \right\}\Bigg|_{\text{reference structure}} \ . \tag{4.10}$$

An analytic expression for the interatomic forces follows by taking the derivative of the final *DFTB* energy (4.9) with respect to the nuclear coordinates:

$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial E_0^{\mathrm{TB}}}{\partial \mathbf{R}_I} = -\sum_i^{occ} n_i \sum_\mu^M \sum_\nu^M c_{\mu i} c_{\nu i} \left[\frac{\partial H^0_{\mu\nu}}{\partial \mathbf{R}_I} - \varepsilon_i \frac{\partial S_{\mu\nu}}{\partial \mathbf{R}_I} \right] - \frac{\partial E_{\mathrm{rep}}}{\partial \mathbf{R}_I} \ .$$

Atomic charges, which will be of particular interest in the charge self-consistent second order extension of *DFTB* (see next section) can be approximately evaluated as Mulliken charges:

$$q_I = \frac{1}{2}\sum_i^{occ} n_i \sum_{\mu\in I} \sum_\nu^M \left(c^*_{\mu i}\, c_{\nu i}\, S_{\mu\nu} + c^*_{\nu i}\, c_{\mu i}\, S_{\nu\mu} \right) \ . \tag{4.11}$$

For many problems, like the calculation of surface energies or defect formation energies, it is very useful to know the energy contribution of single atoms, which for atom I is derived from (4.5) as:

$$E_I = \sum_i^{occ} n_i \sum_{\mu\in I} \sum_\nu^M c^*_{\mu i} H^0_{\mu\nu} c_{\nu i} + \sum_J^{J\neq I} \Phi(I,J) \ . \tag{4.12}$$

Summary

In this section the *non-SCC DFTB*-approach has been described, which in the following will also be called *standard–DFTB*. Provided a clever guess of the initial or input charge density of the system, the energies and forces are correct to second order of charge density fluctuations. Furthermore, the short-range two-particle repulsion (determined once using a proper reference system) operates transferably in very different bonding situations considering various scale systems. Indeed *standard–DFTB* has been successfully applied to systems of different materials with sizes ranging from small clusters [31, 32, 33] over fullerenes [34, 35] to surfaces and interfaces [36, 37] of a variety of semiconductors.

4.3 Second-order self-consistent charge extension, SCC-DFTB

The *standard–DFTB* scheme discussed above is suitable when the electron density of the many-atom structure may be represented as a sum of atomic-like densities in good approximation. However, since the *standard–DFTB*-variant neglects effects of charge redistribution due to Coulomb–interactions it cannot accurately describe systems with considerable charge transfer. This means that *standard–DFTB* will normally fail if the chemical bonding is controlled by a delicate balance of the interatomic charge transfer. In systems containing atoms having different electro-negativity, especially in polar semiconductors and in heteronuclear molecules this is often the case. Therefore, we have extended the approach in order to improve total energies, forces, and transferability in the presence of long-range Coulomb interactions.

We start from equation (4.2) and now explicitly consider the second order term in the density fluctuations.

In order to include associated effects in a simple and efficient TB concept, we first decompose $\delta n(\mathbf{r})$ into atom-centred contributions which decay quickly with increasing distance from the corresponding centre. The second-order term then reads:

$$E_{2nd} = \frac{1}{2}\sum_{I,J}^{N_{\rm nuc}} {\iint}' \Gamma[\mathbf{r},\mathbf{r}',n_0]\,\delta n_I(\mathbf{r})\,\delta n_J(\mathbf{r}'), \tag{4.13}$$

where we have used the functional Γ to denote the *Hartree* and XC coefficients. Secondly, the δn_I may be expanded in a series of radial and angular functions:

$$\delta n_I(\mathbf{r}) = \sum_{l,m} K_{ml} F^I_{ml}(|\mathbf{r}-\mathbf{R}_I|)\, Y_{lm}\left(\frac{\mathbf{r}-\mathbf{R}_I}{|\mathbf{r}-\mathbf{R}_I|}\right) \approx \Delta q_I F^I_{00}(|\mathbf{r}-\mathbf{R}_I|) Y_{00}\,, \tag{4.14}$$

where F^I_{ml} denotes the normalised radial dependence of the density fluctuation on atom I for the corresponding angular-momentum. While the angular deformation of the charge density, e.g. in covalently bonded systems, is usually described very well within the non-SCC approach, charge transfers between different atoms are not properly handled in many cases. Truncating the multipole expansion (4.14) after the monopole term accounts for the most important contributions of this kind while avoiding a substantial increase in the numerical complexity of the scheme. Also, it should be noted that higher-order interactions decay much more rapidly with increasing interatomic distance. Finally, expression (4.14) preserves the total charge in the system, i.e. $\sum_I \Delta q_I = \int \delta n(\mathbf{r})$. Substitution of (4.14) into (4.13) yields the simple final expression for the second-order energy term:

$$E_{2nd} = \frac{1}{2}\sum_{I,J}^{N_{\rm nuc}} \Delta q_I \Delta q_J\, \gamma_{IJ}\,, \qquad \text{where}$$

$$\gamma_{IJ} = {\iint}' \Gamma[\mathbf{r},\mathbf{r}',n_0] \frac{F^I_{00}(|\mathbf{r}-\mathbf{R}_I|)F^J_{00}(|\mathbf{r}\,'-\mathbf{R}_J|)}{4\pi}\,. \tag{4.15}$$

In the limit of large interatomic distances, the XC contribution vanishes within LDA and E_{2nd} may be viewed as a pure Coulomb interaction between two point charges Δq_I and Δq_J. In the opposite case, where the charges are located at one and the same atom, a rigorous evaluation of γ_{II} would require the knowledge of the actual charge distribution. This could be calculated by expanding the

charge density into an appropriate basis set of localised orbitals. In order to avoid this numerically expensive expansion we make an attempt to evaluate γ_{IJ} analytically. To this end in a first step the exchange and correlation contribution is neglected (the second order contribution of E_{XC} will be included a posteriori for short distances) and γ_{IJ} in (4.15) is evaluated analytically for the Coulomb contribution only:

$$\gamma_{IJ} = \iint' \frac{1}{|\mathbf{r}-\mathbf{r}'|} \times \frac{F_{00}^I(|\mathbf{r}-\mathbf{R}_I|)F_{00}^J(|\mathbf{r}'-\mathbf{R}_J|)}{4\pi} \ . \tag{4.16}$$

In accordance with the Slater-type orbitals used as a basis set to solve the Kohn-Sham equations [31], we assume an exponential decay of the monopole term of the density fluctuations (as will be seen, the values for the exponentials τ_I will follow from the evaluation of γ_{II} at $R = 0$):

$$\frac{F_{00}^I(\mathbf{r}-\mathbf{R_I})}{2\sqrt{\pi}} = \frac{\tau_I^3}{8\pi} e^{-\tau_I|\mathbf{r}-\mathbf{R}_I|} \ . \tag{4.17}$$

Inserting (4.17) into (4.16) gives:

$$\begin{aligned} \gamma_{IJ} &= \int\int' \frac{1}{|\mathbf{r}-\mathbf{r}'|} \cdot \frac{\tau_I^3}{8\pi} e^{-\tau_I|\mathbf{r}-\mathbf{R}_I|} \cdot \frac{\tau_J^3}{8\pi} e^{-\tau_J|\mathbf{r}'-\mathbf{R}_J|} \\ &= \int' \Phi(\mathbf{r}') \frac{\tau_J^3}{8\pi} e^{-\tau_J|\mathbf{r}'-\mathbf{R}_J|} \ , \end{aligned}$$

where

$$\Phi(\mathbf{r}') = \int \frac{1}{|\mathbf{r}-\mathbf{r}'|} \cdot \frac{\tau_I^3}{8\pi} e^{-\tau_I|\mathbf{r}-\mathbf{R}_I|} \ .$$

Via Poisson's equation we obtain for Φ:

$$\Phi(\mathbf{r}') = \left[\frac{1}{|\mathbf{r}'-\mathbf{R}_I|} - \left(\frac{\tau_I}{2} + \frac{1}{|\mathbf{r}'-\mathbf{R}_I|}\right) e^{-\tau_I|\mathbf{r}'-\mathbf{R}_I|}\right] \ .$$

Hence γ_{IJ} becomes:

$$\gamma_{IJ} = \int' \left[\frac{1}{|\mathbf{r}'-\mathbf{R}_I|} - \left(\frac{\tau_I}{2} + \frac{1}{|\mathbf{r}'-\mathbf{R}_I|}\right) e^{-\tau_I|\mathbf{r}'-\mathbf{R}_I|}\right] \cdot \frac{\tau_J^3}{8\pi} e^{-\tau_J|\mathbf{r}'-\mathbf{R}_J|} \ . \tag{4.18}$$

Setting $R = |R_I - R_J|$, after some coordinate transformations one gets for $R \neq 0$ (see appendix A.1):

$$\gamma_{IJ} = \frac{1}{R} - \begin{cases} e^{-\tau_I R}\left(\frac{\tau_J^4\tau_I}{2(\tau_I^2-\tau_J^2)^2} - \frac{\tau_J^6-3\tau_J^4\tau_I^2}{(\tau_I^2-\tau_J^2)^3 R}\right) - e^{-\tau_J R}\left(\frac{\tau_I^4\tau_J}{2(\tau_J^2-\tau_I^2)^2} - \frac{\tau_I^6-3\tau_I^4\tau_J^2}{(\tau_J^2-\tau_I^2)^3 R}\right) \\ \text{if } \tau_I \neq \tau_J \qquad \text{and} \\ -e^{-\tau_I R}\left(\frac{1}{R} + \frac{11\tau_I}{16} + \frac{3\tau_I^2 R}{16} + \frac{\tau_I^3 R^2}{48}\right) \\ \text{if } \tau_I = \tau_J \end{cases} \tag{4.19}$$

In the limit of large interatomic distances, $\gamma_{IJ} \to 1/R$ and thus represents the Coulomb interaction between two point charges Δq_I and Δq_J. This accounts for the fact that at large interatomic distances the exchange-correlation contribution vanishes within the local density approximation and only the Coulomb contribution remains important, as stated above.

Now we need to determine the values for τ_I. This is done by examining γ_{IJ} in the limit $R \to 0$ which corresponds to the "on–site" contribution of the second order energy in (4.13). Expanding the exponentials in γ_{IJ} we find (see Appendix A.1):

$$\gamma_{IJ} \overset{R\to 0}{=} \gamma_{II} = \frac{5}{16}\tau_I \ . \tag{4.20}$$

Parisers [38] suggested that γ_{II} can be approximated by the difference of the atomic ionisation potential I_I and the electron affinity A_I. This difference is related to the chemical hardness η_I, [39] or the Hubbard parameter U_I:

$$\gamma_{II} \approx I_I - A_I \approx 2\eta_I \approx U_I \ .$$

Inserting this approximation which is widely used in semi-empirical quantum chemistry methods into (4.20) one gets:

$$\tau_I = \frac{16}{5} U_I \ .$$

This result can be interpreted by noting that elements with a high chemical hardness tend to have localised wavefunctions which in turn implies a "localised fluctuation" of their charge densities. The Hubbard parameters U_I are constants which can be calculated for any atom type within LDA-DFT as the second derivative of the total energy of a single atom with respect to the occupation number of the highest occupied atomic orbital. This includes the influence of the second order contribution of E_{XC} in γ_{IJ} for small distances where it is important.

Since γ_{IJ} is explicitly known, the DFT total energy (4.2) can be transformed into a transparent TB-form:

$$E_2^{\mathrm{TB}} = \sum_i^{occ} n_i \langle \psi_i | \hat{H}_0 | \psi_i \rangle + \frac{1}{2} \sum_{I,J}^{N_{\mathrm{nuc}}} \gamma_{IJ} \Delta q_I \Delta q_J + E_{\mathrm{rep}} \ , \tag{4.21}$$

where $\gamma_{IJ} = \gamma_{IJ}(U_I, U_J, |\mathbf{R}_I - \mathbf{R}_J|)$. As discussed earlier, the contribution due to $\hat{H}_0$ depends only on n_0 and is therefore exactly the same as in the previous non-SCC studies [31]. However, since the atomic charges depend on the one-particle wave functions ψ_i, an iterative procedure is required to find the minimum energy in expression (4.21). To solve (4.21) explicitly we use again the pseudoatomic basis set expansion (4.4) for the wavefunctions ψ_i and hence obtain for (4.21):

$$E_2^{\mathrm{TB}} = \sum_i^{occ} n_i \sum_\mu^M c_{\mu\nu}^* H_{\mu\nu}^0 c_{\nu i} + \frac{1}{2} \sum_{I,J}^{N_{\mathrm{nuc}}} \gamma_{IJ} \Delta q_I \Delta q_J + E_{\mathrm{rep}} \ . \tag{4.22}$$

The charge fluctuations $\Delta q_I = q_I - q_I^0$ are estimated by means of Mulliken charges (4.11). Application of the Lagrange formalism to expression (4.22) gives:

$$\frac{\partial}{\partial c_{\mu i}^*} \Big[E_2^{\mathrm{TB}} - \epsilon_i \left(\sum_\delta^{occ} n_\delta \sum_\sigma^M \sum_\nu^M c_{\sigma\delta}^* S_{\sigma\nu} c_{\nu\delta} - N_{\mathrm{el}} \right) \Big] = 0 \ . \tag{4.23}$$

To this end we have to evaluate the derivatives $\frac{\partial E_2^{\mathrm{TB}}}{\partial c_{\mu i}^*}$:

$$\frac{\partial E_2^{\mathrm{TB}}}{\partial c_{\mu i}^*} = n_i \sum_\nu^M c_{\nu i} H_{\mu\nu}^0 + \sum_I^{N_{\mathrm{nuc}}} \sum_J^{N_{\mathrm{nuc}}} \frac{\partial q_I}{\partial c_{\mu i}^*} \Delta q_J \gamma_{IJ} \ .$$

Using (4.11) we get:

$$\frac{\partial q_I}{c_{\mu i}^*} = \frac{1}{2} n_i \Big[\sum_{\sigma \in I} S_{\mu\sigma} c_{\sigma i} + \delta_{I\mu} \sum_\delta^M S_{\sigma\delta} c_{\delta i} \Big], \ \delta_{\mu I} = \begin{cases} 1 \text{ if } \mu \in I \\ 0 \text{ otherwise} \end{cases}$$

and hence:

$$\begin{aligned}\frac{\partial E_2^{\mathrm{TB}}}{\partial c_{\mu i}^*} &= n_i \sum_{\nu}^{M} c_{\nu i} H_{\mu\nu}^0 + \frac{1}{2} \sum_{I}^{N_{\mathrm{nuc}}} \sum_{J}^{N_{\mathrm{nuc}}} (\sum_{\sigma \in I} S_{\mu\sigma} c_{\sigma i} + \delta_{I\mu} \sum_{\delta}^{M} S_{\mu\delta} c_{\delta i}) \Delta q_J \gamma_{IJ} \\ &= n_i \sum_{\nu}^{M} c_{\nu i} \Big[H_{\mu\nu}^0 + \frac{1}{2} S_{\mu\nu} \sum_{I}^{N_{\mathrm{nuc}}} \Delta q_I (\gamma_{JI} + \gamma_{KI}) \Big], \ \mu \in J, \ \nu \in K \ .\end{aligned}$$

Inserting these derivatives into (4.23) gives the Kohn–Sham equations (2.14) in matrix notation:

$$\sum_{\nu}^{M} c_{\nu i} (H_{\mu\nu} - \varepsilon_i S_{\mu\nu}) = 0, \qquad 1 \leq \mu, i \leq M, \tag{4.24}$$

with Hamilton and overlap matrix elements:

$$\begin{aligned}H_{\mu\nu} &= \langle \phi_\mu | \hat{H}_0 | \phi_\nu \rangle + \frac{1}{2} S_{\mu\nu} \sum_{K}^{N_{\mathrm{nuc}}} (\gamma_{IK} + \gamma_{JK}) \Delta q_K \\ &= H_{\mu\nu}^0 + H_{\mu\nu}^1 \ ; \quad S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle \ , \qquad \mu \epsilon I, \ \nu \epsilon J.\end{aligned} \tag{4.25}$$

As in the *standard–DFTB* approach, we make use of the two-centre approximation for the integrals of H^0 (4.8). It should however be noted that, since the overlap matrix elements $S_{\mu\nu}$ generally extend over a few nearest neighbour distances, also multiparticle interactions are incorporated in the *SCC–DFTB* scheme. The second-order correction due to charge fluctuations is now represented by the non-diagonal Mulliken charge dependent contribution $H_{\mu\nu}^1$ to the matrix elements $H_{\mu\nu}$.

In consistency with equation (4.10), we determine the short-range repulsive pair potential E_{rep} as a function of distance by taking the difference of the SCF-LDA cohesive energy and the corresponding *SCC-DFTB* electronic energy (first two terms in equation (4.21)) for a suitable reference structure. Since charge transfer effects are now considered explicitly, the transferability of E_{rep} is improved compared to the non-SCC approach.

Equation (4.25) has been derived for finite systems. A gamma point approximation to periodic systems is found in a straightforward manner by replacing the sum over the atoms K by the sum over all atoms K in the cell summed over all cells:

$$\begin{aligned}H_{\mu\nu} &= H_{\mu\nu}^0 + \frac{1}{2} S_{\mu\nu} \cdot \sum_{K}^{N_{\mathrm{nuc}}} \sum_{\mathbf{R}} \Delta q_K (\gamma_{IK}(\mathbf{R}) + \gamma_{JK}(\mathbf{R})) \\ &= H_{\mu\nu}^0 + \frac{1}{2} S_{\mu\nu} \cdot \sum_{K}^{N_{\mathrm{nuc}}} \Delta q_K \left(\sum_{\mathbf{R}} \gamma_{IK}(\mathbf{R}) + \sum_{\mathbf{R}} \gamma_{JK}(\mathbf{R}) \right), \ \forall \mu \in I, \ \nu \in J \ .\end{aligned}$$

Here $\gamma_{IK}(\mathbf{R})$ means that the function γ_{IK} in (4.19) is evaluated at $\mathbf{R}_I + \mathbf{R}$ and $\mathbf{R}_K$. We can now take the Fourier transform to obtain the formula for the Hamilton matrix elements in k–space:

$$H_{\mu\nu}(k) = H_{\mu\nu}^0(k) + \frac{1}{2} S_{\mu\nu}(k) \cdot \sum_{K}^{N_{\mathrm{nuc}}} \Delta q_K \left(\sum_{\mathbf{R}} \gamma_{IK}(\mathbf{R}) + \sum_{\mathbf{R}} \gamma_{JK}(\mathbf{R}) \right) \ .$$

where $H_{\mu\nu}(k)$ and $S_{\mu\nu}(k)$ are the Fourier transforms of $H_{\mu\nu}$ and $S_{\mu\nu}$, respectively. The same correction for $H_{\mu\nu}^0(k)$ can be rigorously derived by starting from the energy expression in a periodic

system given by:

$$\begin{aligned} E_2^{\mathrm{TB}} &= \sum_i^{occ} \sum_k W_{ik} \sum_\mu^M \sum_\nu^M c^*(k)_{\mu i} c_{\nu i}(k) H^0_{\mu\nu}(k) \\ &+ \frac{1}{2} \sum_I^{N_{\mathrm{nuc}}} \sum_J^{N_{\mathrm{nuc}}} \Delta q_I \Delta q_J \sum_{\mathbf{R}} \gamma_{IJ}(\mathbf{R}) \ . \end{aligned}$$

In periodic systems, the evaluation of $\overline{\gamma}_{IJ} = \sum_{\mathbf{R}} \gamma_{IJ}(\mathbf{R})$ demands the evaluation of an infinite sum which is not absolutely convergent, i.e. the value depends on the order of summation. This is by no means a trivial task. As can be seen from (4.19) $\overline{\gamma}_{IJ}$ consists of a long range part $\sum_{\mathbf{R}} \frac{1}{|\mathbf{R}|}$ and a short range part which is the sum over the terms following the curly bracket in (4.19). The long range part can be evaluated using the standard Ewald technique, whereas the short range part decays exponentially and can therefore be summed over a small number of unit cells (see Appendix A.2 for a detailed description of the numerical evaluation of $\overline{\gamma}_{IJ}$ for periodic systems). Hence the functional form for γ_{IJ} (4.19) derived in this work yields a well defined expression for extended and periodic systems. This is in contrast to common functional forms for γ_{IJ} frequently employed in semi–empirical schemes for molecules. Those functional forms are based on empirical studies and may cause severe numerical problems when applied to periodic systems since Coulomb-like behaviour is only accomplished for large interatomic distances. For example, for periodic systems the expression

$$\gamma_{IJ} = \frac{1}{\sqrt{(\mathbf{R}_I - \mathbf{R}_J)^2 + \frac{1}{4}\left(\frac{1}{U_I} + \frac{1}{U_J}\right)^2}}$$

used in MINDO [40] yields ill–conditioned energies with respect to the Hubbard parameters. This means that small changes in the Hubbard parameters may result in considerable variations for the value of $\overline{\gamma}_{IJ} = \sum_{\mathbf{R}} \gamma_{IJ}(\mathbf{R})$, i.e. the derivative of $\overline{\gamma}_{IJ}$ with respect to the Hubbard parameters has a large norm. Therefore expressions like this can not be used for systems where long range Coulomb interactions occur and thus limit considerably the applicability of the scheme.

Finally an analytic expression for the interatomic forces follows by taking the derivative of the final *SCC-DFTB* energy (4.21) with respect to the nuclear coordinates:

$$\begin{aligned} \mathbf{F}_I = &- \sum_i^{occ} n_i \sum_{\mu,\nu}^M c^*_{\mu i} c_{\nu i} \left(\left[\frac{\partial H^0_{\mu\nu}}{\partial \mathbf{R}_I} + \frac{1}{2} \frac{\partial S_{\mu\nu}}{\partial \mathbf{R}_I} \sum_L^{N_{\mathrm{nuc}}} (\gamma_{KL} + \gamma_{JL}) \Delta q_L \right] - \varepsilon_i \frac{\partial S_{\mu\nu}}{\partial \mathbf{R}_I} \right) \\ &- \Delta q_I \sum_L^{N_{\mathrm{nuc}}} \frac{\partial \gamma_{IL}}{\partial \mathbf{R}_I} \Delta q_L - \frac{\partial E_{\mathrm{rep}}}{\partial \mathbf{R}_I} \ , \quad \mu \in K \ , \quad \nu \in J \ . \end{aligned} \tag{4.26}$$

For periodic systems the derivative of γ_{IJ} is again evaluated by means of the Ewald technique (see Appendix A.2).

In analogy to (4.12) the atomic energies for the *SCC–DFTB* method are given by:

$$E_I = \sum_{\mu \in I} \sum_\nu^M \sum_i n_i c_{\mu i} c_{\nu i} H^0_{\mu\nu} + \frac{1}{2} \sum_J^N (q_I - q_I^0)(q_J - q_J^0) \gamma_{IJ} + \sum_J^{J \neq I} \Phi(I, J) \ . \tag{4.27}$$

Summary

In this section the *SCC–DFTB* method has been described. It has been derived from DFT by considering second order fluctuations in the charge density. To this end we have deduced an analytic functional to include the energy arising from charge transfer. This resulted in a charge self–consistent extension to *standard–DFTB*. In contrast to *standard–DFTB* the *SCC-DFTB* method also allows to treat systems where a correct description of the chemical bonding requires an accurate distribution of the electronic charges. Furthermore, the *SCC-DFTB* method can be applied to determine the energetic ordering of systems where electrostatic interactions play a crucial role. The best known example for the latter case are the well studied GaAs $\beta(2 \times 4)$ and $\beta 2(2 \times 4)$ surfaces. Whereas *standard–DFTB* as well as other TB methods [43] which neglect electrostatic interactions found $\beta(2 \times 4)$ and $\beta 2(2 \times 4)$ to have the same surface energies [36], *SCC–DFTB* favours the $\beta 2(2 \times 4)$ surface [27] in agreement with SCF–LDA calculations [44] and high resolution scanning tunnelling microscopy experiments [45].

4.4 Performance of standard–DFTB and SCC–DFTB

The computationally most expensive step in *standard–DFTB* and *SCC–DFTB* calculations is the solution of the generalised eigenvalue problem (4.7) and (4.24). While the *standard–DFTB* method requires this task to be solved only once for every step in the ionic relaxation process, the *SCC–DFTB* scheme is converged after ≈ 5 iterations in the charge self–consistency cycle, provided the structure is semiconducting. Therefore, *SCC–DFTB* is only slightly less efficient than *standard–DFTB*.

Moreover it can be shown that also all the other computationally expensive steps in *(SCC)–DFTB*, i.e. evaluation of Mulliken charges, atomic energies and calculation of forces, which all scale with M^3, where M is the total number of basis functions of the system, can be transformed into linear algebra operations and can thus easily and efficiently be parallelised [28].

As an illustrative example Fig. 4.1 shows the time scaling depending on the system size, i.e. the number of basis functions, for one iteration in the charge self-consistency cycle. The benchmarking has been performed on a typical workstation (HP 735/125) and on a parallel machine (*Cray T3E*) employing different numbers of processors. The discrete data points were fitted to cubic polynomials. In this work 8 (13) basis functions per GaN unit are used in the *standard–* (*SCC)–DFTB* method (see section below). As can be seen in Fig. 4.1 run on a parallel machine the *(SCC)–DFTB* method allows to treat large systems within a reasonable time making it a suitable tool for the investigation of surfaces and extended defects in GaN.

4.5 Application of standard–DFTB and SCC–DFTB to GaN

In this work *standard–DFTB* and *SCC–DFTB* are used to determine the structural properties and energetics of surfaces, point defects, extended defects and domain boundaries in GaN. According to the topology of the structures they were modelled either in clusters or in supercells whichever type of modelling is more appropriate for the problem under consideration (see appendix D).

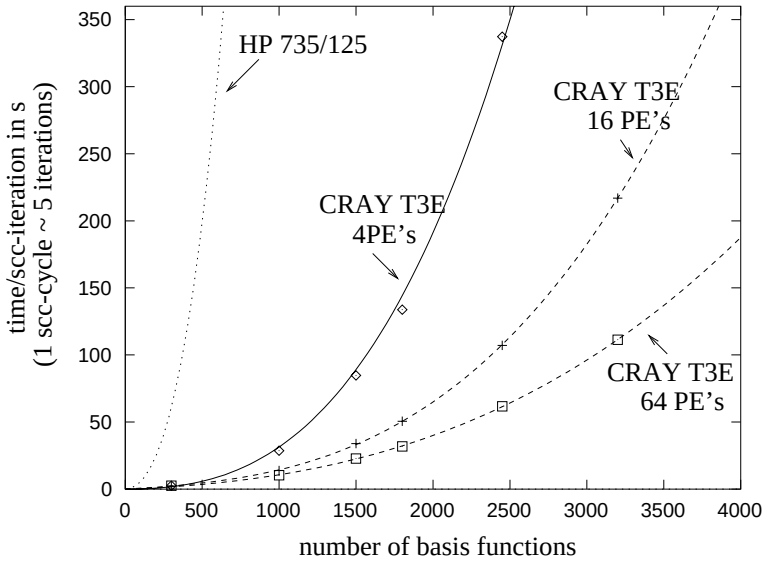

Figure 4.1: Time scaling of the *SCC–DFTB* method for a supercell with a GaAs (100) surface.

In most of the cases the geometries are already found to be well described within the *standard–DFTB* method. In order to provide geometry optimisations in a numerically efficient way in this work the *standard–DFTB* method uses a minimal basis set consisting of Ga $4s$, $4p$ and N $2s$, $2p$ wavefunctions. The Ga $3d$ electrons are not treated as valence electrons, an approximation which as in the case of *AIMPRO* (see previous chapter) is permitted if no energetics are considered. Indeed, test calculations show that an inclusion of the Ga $3d$ electrons into the valence band has almost no effect on the geometries.

In order to determine the sometimes complex energetic ordering of structures the use of the more accurate *SCC–DFTB* method is more appropriate. As mentioned above charge self–consistency is particularly important for the energetics of systems where electrostatic interactions play a crucial role. In the current applications we use the same minimal basis set of Ga $4s$, $4p$ and N $2s$, $2p$ wavefunctions as in the *standard–DFTB* method but with the Ga $3d$ electrons included in the valence band. As indicated by Northrup *et al.* [46], this is important to obtain accurate formation energies since the Ga $3d$ and N $2s$ levels hybridise.

A detailed description of the construction of the wavefunctions used for GaN in *standard–* and *SCC–DFTB* along with the resulting electronic band structures can be found in appendix B. In this appendix also the structures employed to generate the repulsive potentials E_{rep} for Ga–Ga, Ga–N and N–N interactions are given.

Chapter 5

Formation Energies of Surfaces and Defects in Thermodynamic Equilibrium

This chapter deals with the stability of surfaces and defects in a thermodynamical context. The energy necessary to create a structure (here and in the following structure stands for both, surfaces and defects) is called the formation energy. This energy is not constant but depends on specific growth conditions. In GaN the relative abundance of Ga, N and impurity atoms during the crystal growth determines the formation energy of the structure. In a thermodynamical context, these abundances are described by the chemical potentials associated with the reservoirs from which Ga, N and impurity atoms enter the growth process.

A simple expression for the formation energy will be given where specific growth conditions are described via chemical potentials. The derivation of this expression will be given for clean GaN structures. The result will then be generalised to formation energies of GaN structures with impurities.

Finally, from these calculated formation energies some experimentally interesting quantities will be deduced.

5.1 Formation energies of charge neutral GaN structures

The equilibrium structure is determined by minimising the grand canonical potential for the formation energy. Given a specific temperature and pressure the grand canonical potential for the formation energy E as a function of composition reads:

$$E = G(p, T, n_{\mathrm{Ga}}, n_{\mathrm{N}})_{\mathrm{structure}} - \mu_{\mathrm{Ga}}(p, T) \cdot n_{\mathrm{Ga}} - \mu_{\mathrm{N}}(p, T) \cdot n_{\mathrm{N}} \quad . \tag{5.1}$$

Here

$$G(p, T, n_{\mathrm{Ga}}, n_{\mathrm{N}})_{\mathrm{structure}} = E_{\mathrm{structure}} + PV - TS$$

is the Gibbs free energy of the structure. $E_{\text{structure}}$ is the total energy of the structure which is obtained from a calculation at $T = 0$. μ_{Ga} and μ_{N} denote the Ga and N chemical potentials and n_{Ga} and n_{N} the number of Ga and N atoms in the structure. In the following we assume the structure to be in equilibrium with its surroundings, i.e. with the perfect GaN bulk and with the gas phase Ga and N. The chemical potential $\mu_i = dG/dn_i$ of a given atomic species (i = Ga or N) is the same in each of the phases that are in contact in equilibrium. Therefore, each μ_i can be considered as the free energy per particle in each reservoir for particle i.

Equation (5.1) gives the formation energy of a charge neutral structure depending on the pressure and the temperature. As shown in [47] the pressure dependence can usually be neglected in condensed–state systems which are not easily compressible. Therefore the free energy may be approximated at $P = 0$. On the other hand, the free energy has a temperature dependent contribution arising from the vibrational entropy $-TS$. The energy $-TS$ is not necessarily small compared with the differences in formation energies for different structures. It comes from the fact that atoms at surfaces and defects are less strongly bound than in bulk material and thus can vibrate more easily. A quantitative evaluation would be computationally very demanding. However, for comparable structures, as e.g. different types of point defects, the vibrational entropy is usually very similar. If we are not interested in the precise values of the absolute formation energies but rather want to compare the stabilities of structures we can neglect the contribution coming from the vibrational entropy and may replace the free energy $G_{\text{structure}}$ of the structure by the total energy $E_{\text{structure}}$ of the structure. Note that this approximation means that $G_{\text{structure}} = E_{\text{structure}}$ which is exact at $T = 0$.

For the gaseous phase, the effect of T and P upon the chemical potential cannot be ignored, and any value of μ_{Ga} and μ_{N} may be experimentally attained. Qian, Martin and Chadi [48] showed, however, that there are limits on the allowable range of μ_{Ga} and μ_{N} under equilibrium conditions with all possible phases. Thermodynamic equilibrium implies that the chemical potential of the elementary compounds cannot be larger than the chemical potential of the lowest energy phase in which the elementary compound is involved. Otherwise this phase would be adopted.[1]

It is a good approximation to assume that Ga in its bulk structure and nitrogen in the form of N_2 molecules are the corresponding phases of lowest energy. This means for the upper limits of the elemental chemical potentials:

$$\mu_{\text{Ga}} \leq \mu_{\text{Ga(bulk)}}(p,T) \quad \text{and} \quad \mu_{\text{N}} \leq \mu_{\text{N}_2}(p,T) \; . \tag{5.2}$$

For $T = 0$ (corresponding to an upper limit) $\mu_{\text{Ga(bulk)}}$ and μ_{N_2} can be obtained from total energy calculations as $E_{\text{Ga(bulk)}}$ and $\frac{1}{2}E_{\text{N}_2}$, respectively. Thermodynamic equilibrium conditions also imply that the crystalline GaN structure can exchange pairs of Ga and N atoms with the reservoir. As stated above for a pair of Ga and N in the crystal the free energy μ_{GaN} is the total bulk energy per pair $E_{\text{GaN(bulk)}}$. An exchange of pairs of GaN units between structure and reservoirs implies that the sum of the chemical potentials μ_{Ga} and μ_{N} in the reservoirs equals the chemical potential of a crystal unit:

$$\begin{aligned} \mu_{\text{Ga}} + \mu_{\text{N}} &= \mu_{\text{GaN(bulk)}} \quad (\approx E_{\text{GaN(bulk)}}) \\ &= \mu_{\text{Ga(bulk)}} + \mu_{\text{N}_2} - \Delta H_f \; , \end{aligned} \tag{5.3}$$

where ΔH_f is the heat of formation of GaN. The heat of formation can be determined from enthalpy measurements and has been found to be 1.14 eV for GaN [49]. Of course, ΔH_f can also be derived

[1] This would then result in a dissociation of the GaN crystal, in contradiction to thermodynamic equilibrium.

from total energy calculations for GaN in the wurtzite phase $E_{\mathrm{GaN(bulk)}}$, Ga in the orthorhombic bulk phase $E_{\mathrm{Ga(bulk)}}$ and the N_2 molecule $\frac{1}{2}E_{\mathrm{N_2}}$. Combined with Eq. (5.2), Eq. (5.3) sets a lower limit on the elemental chemical potentials μ_{Ga} and μ_{N} so that the structure can be in equilibrium with its surroundings only if the chemical potentials satisfy the relations:

$$\mu_{\mathrm{Ga(bulk)}} - \Delta H_f \leq \mu_{\mathrm{Ga}} \quad \text{and} \quad \mu_{\mathrm{N_2}} - \Delta H_f \leq \mu_{\mathrm{N}} \ . \tag{5.4}$$

For a more detailed theoretical discussion about the validity of the limits for the chemical potentials see Kley [47].

Finally, we can use (5.3) to eliminate one of the elemental chemical potentials in (5.1). Here and in the following we choose to eliminate μ_{N}. With this choice the formation energies depend only on the gallium chemical potential.[2] This gives the following simple expression for the formation energy of a charge neutral structure:

$$E = E_{\mathrm{structure}} - E_{\mathrm{GaN(bulk)}} n_{\mathrm{N}} - \mu_{\mathrm{Ga}}(n_{\mathrm{Ga}} - n_{\mathrm{N}}) \ , \tag{5.5}$$

where according to (5.2) and (5.4) the allowed range of the gallium chemical potential is given by:

$$\mu_{\mathrm{Ga(bulk)}} - \Delta H_f \leq \mu_{\mathrm{Ga}} \leq \mu_{\mathrm{Ga(bulk)}} \ . \tag{5.6}$$

The way in which the limits for μ_{Ga} used in this work are determined for calculations with the *SCC–DFTB* method is described in Appendix C.

5.2 Formation energies of structures with impurities

Impurities, in particular H, C, O, Si and Mg play an important role in GaN. In this work the stabilities and electrical properties of hydrogen and oxygen as impurities and surface adsorbates will be examined.

The formalism for formation energies described above can be generalised in a straightforward manner. Denoting the impurities by I we have:

$$E = E_{\mathrm{structure}} - E_{\mathrm{GaN(bulk)}} n_{\mathrm{N}} - \mu_{\mathrm{Ga}}(n_{\mathrm{Ga}} - n_{\mathrm{N}}) - \sum_{\mathrm{I}} n_{\mathrm{I}} \mu_{\mathrm{I}} \ . \tag{5.7}$$

Assuming that the elemental chemical potentials of the impurities can be chosen independently from the gallium chemical potential, one would need to represent (5.7) as a multi–dimensional plot. However, as will be shown below for the hydrogen and oxygen impurities considered in this work this task can often be simplified by assigning the chemical potential of the impurities with a fixed value which corresponds to the growth conditions at which the process should be examined.

5.2.1 Hydrogen

From statistical mechanics [50] the chemical potential of hydrogen μ_{H} in the form of H_2 gas is given by:

$$2\mu_{\mathrm{H}} = E_{\mathrm{H_2}} + \tau ln(PV_Q/\tau) - \tau ln Z_{\mathrm{rot}} - \tau ln Z_{\mathrm{vib}} \ ,$$

[2] Of course, expressing the formation energies in terms of the nitrogen chemical potential μ_N would describe the same physics.

where E_{H_2} is the total energy of an H_2 molecule, P is the pressure, $V_Q = (h^2/2\pi mkT)^{3/2}$ is the quantum volume and $\tau = kT$ is the temperature. $Z_{\rm rot}$ and $Z_{\rm vib}$ are the rotational and vibrational contributions to the partition function describing the internal degrees of freedom of an H_2 molecule.

At an MOCVD growth temperature of $\approx 1300^{\rm o}$ K and with a nominal H_2 pressure of ≈ 1 atm, $\mu_{\rm H}$ is ≈ 1 eV lower than $E_{H_2}/2$. For further discussion and a temperature dependent plot of $\mu_{\rm H}$ see reference [51].

5.2.2 Oxygen

Unfortunately, in contrast to hydrogen, the oxygen pressure during MOCVD or MBE growth is not really known. It is therefore not possible to derive a value for the oxygen chemical potential in the same explicit way as for the H chemical potential.

In many cases, it is however satisfactory to get a rough idea about the stability of a defect. To this end one commonly assumes [52, 8, 9] that oxygen and gallium are in equilibrium with gallium oxide (Ga_2O_3):

$$2\mu_{\rm Ga} + 3\mu_{\rm O} = \mu_{\rm Ga_2O_3} \ . \tag{5.8}$$

This corresponds to the upper oxygen concentration. If $\mu_{\rm O}$ were larger than $\mu_{\rm Ga_2O_3}$ gallium oxide would form during growth under thermodynamic equilibrium. In order to get an absolute formation energy which is useful to determine the concentration of a defect (see below) one often furthermore assumes Ga rich growth conditions ($\mu_{\rm Ga} = \mu_{\rm Ga(bulk)}$). The value of $\mu_{\rm O}$ for the upper limit of the oxygen concentration is then fixed via (5.8):

$$\mu_{\rm O}({\rm upper}) = \frac{\mu_{\rm Ga_2O_3} - 2\mu_{\rm Ga(bulk)}}{3} \ .^{3} \tag{5.9}$$

The ideas outlined in the last paragraph are very plausible if effects of oxygen incorporation during GaN growth are examined. Here the growth of GaN is still the dominating process while the incorporation of oxygen plays a secondary role. This allows to set the oxygen chemical potential to a fixed value. On the other hand, if the chemisorption of oxygen at surfaces is considered it is more meaningful to express the surface formation energies via the oxygen chemical potential. Therefore in the following $\mu_{\rm O}$ will be chosen as the independent variable in (5.7). We assume that during the chemisorption of O the surface is in an environment where O is in the gas phase (O_2). Hence the upper limit of $mu_{\rm O}$ is given by equilibrium with O_2 molecules. The lower limit of $\mu_{\rm O}$ can be chosen somehow arbitrary. We assume that it should be not much less than $\mu_{\rm O}$ in equilibrium with gallium oxide under Ga–rich growth conditions. Otherwise the resulting oxygen configuration at the surface would be significantly more stable than bulk gallium oxide which is not very likely. The energy equation for gallium oxide is:

$$\mu_{\rm Ga_2O_3} = 2\mu_{\rm Ga(bulk)} + 3\mu_{O_2} - \Delta H_{f(\rm Ga_2O_3)} \ , \tag{5.10}$$

[3] It should be noted that the absolute formation energy obtained by inserting $\mu_{\rm O(upper)}$ and $\mu_{\rm Ga(bulk)}$ into equation (5.7) is also a good quantity for the upper limit of the oxygen concentration under N–rich growth conditions: oxygen adopts a nitrogen site in GaN and oxygen defects often have an equivalent stoichiometry to Ga_2O_3, as e.g. $V_{\rm Ga}$–$(O_{\rm N})_3$. This means that by moving towards a N–rich environment the formation energy gained by increasing the upper limit $\mu_{\rm O(upper)}$ in (5.9) (now $\mu_{\rm Ga} < \mu_{\rm Ga(bulk)}$) is compensated by the formation energy lost due to the fact that the structure has an N deficient stoichiometry.

where $\Delta H_{f(Ga_2O_3)}$ is the heat of formation for Ga_2O_3, which has been calculated from bulk Ga_2O_3, the orthorhombic Ga bulk phase and the O_2 molecule to be 12.05 eV (see C) in good agreement with the heat of formation of 11.3 eV determined from enthalpy measurements [49]. Hence, equilibrium with gallium oxide under Ga–rich growth conditions as a lower limit and with O_2 molecules as an upper limit implies:

$$\mu_{O_2} - \frac{1}{3}\Delta H_{f(Ga_2O_3)} \leq \mu_O \leq \mu_{O_2} \ , \tag{5.11}$$

extending from the Ga_2O_3 like ($\mu_O = \mu_{O_2} - \frac{1}{3}\Delta H_{f(Ga_2O_3)}$) to the O–rich ($\mu_O = \mu_{O_2}$) environment.

5.3 Formation energies of charged structures

In the previous section expressions for the formation energy of charge neutral structures were deduced. However, point defects frequently occur in a charged state, as e.g. in III–V semiconductors where the energetically low (high) lying anion (cation) derived electronic states become filled (emptied). A charged defect is stable if its formation energy is lower than the formation energy of the charge neutral defect. To obtain the formation energy of charged defects we need to include the energy of the reservoir from which the electrons come from (in the case of a negatively charged defect) or are transfered to (in the case of a positively charged defect). The energy of the electronic reservoir can be expressed via the position of the Fermi level E_F. Therefore if q denotes the charge of the defect, we can extend (5.7):

$$E = E(q) = E_{\text{structure}}(q) - E_{\text{GaN(bulk)}} n_N - \mu_{Ga}(n_{Ga} - n_N) - \sum_I n_I \mu_I - qE_F \ . \tag{5.12}$$

This formula allows an easy interpretation: often negatively charged defects are particularly stable in n–type GaN where electrons are abundant whereas in p–type material, where there is a lack of electrons, positively charged defects tend to have low energies.

5.4 Applications to surfaces and defects

In the following it will be shown how some experimentally interesting quantities can be derived from the formation energy of a structure.

Absolute Surface Energies

If the surfaces are modelled by slabs (see section D.1) the surface energy is obtained by dividing the formation energy of the slab by the area A of the slab surface:

$$\gamma = E/A \ .^4 \tag{5.13}$$

Frequently, the energy needed to create a surface from perfect bulk material, called the *absolute surface energy*, is of particular interest. Examples for processes which can be explained by the

[4] Of course, due to the approximations in the previous section γ does not include the entropy contribution and thus may contain a certain error.

knowledge of absolute surface energies are the adaption of an equilibrium crystal shape (ECS) [53], the difference in the growth rate of crystals in specific directions [54] and the formation of nanopipes in certain materials as SiC [55]. However, the surface energy derived by inserting the formation energy of the entire slab into (5.13) does not correspond to the absolute surface energy, but always contains the energy of another, often inequivalent surface which terminates the bottom of the slab.

For the (110), (100), (111) and $(\overline{111})$ surfaces of compound semiconductors in the zinc–blende phase it is possible to obtain absolute surface energies by dividing the slab into top and bottom ($\mathrm{slab_{top}}$ and a $\mathrm{slab_{bottom}}$) by a boundary consisting of crystal symmetry planes [56, 57]. Here $\mathrm{slab_{top}}$ is a well defined region below the surface which includes the surface and underlying layers that are affected by a relaxation due to the surface reconstruction. On the other hand, $\mathrm{slab_{bottom}}$ is the remaining region below $\mathrm{slab_{top}}$ containing only bulk–like GaN and the terminating pseudo–hydrogen layer (see section D.1).[5] Of course, only the energy contained in $\mathrm{slab_{top}}$ is related to the absolute surface energy, whereas the energy contained in $\mathrm{slab_{bottom}}$ includes terminating hydrogens and has no physical meaning. Therefore, if the formation energy E is evaluated for $\mathrm{slab_{top}}$ via equations (4.27) and (5.5) then γ represents the desired absolute surface energy.

The wurtzite structure has a lower crystalline symmetry than the zinc–blende structure. However, for the nonpolar $(10\bar{1}0)$ and $(11\bar{2}0)$ surfaces and for the polar $(10\bar{1}1)$ and $(10\bar{1}\bar{1})$ surfaces an accurate splitting into $\mathrm{slab_{top}}$ and $\mathrm{slab_{bottom}}$ can easily be found (see thesis of M. Haugk [59]). The splitting of slabs with zinc–blende (111) and $(\bar{1}\bar{1}\bar{1})$ surfaces into $\mathrm{slab_{top}}$ and $\mathrm{slab_{bottom}}$ implicitly makes use of the equivalence of the four tetrahedral bonds in the zinc–blende phase. It cannot therefore be strictly applied to (0001) and $(000\bar{1})$ surfaces in the wurtzite phase, where the non–ideal c/a ratio and u parameter lead to only three equivalent bonds. In GaN $c/a = 1.626$ and $u = 0.377$ are very close to the ideal values, however, thus yielding four nearly equivalent bonds in bulk material. We may therefore apply the same slab division as in zinc–blende material and expect it to give very reasonable absolute surface energies. For a rigorous mathematical justification see reference [58].

Concentration of Point Defects

For point defects it is useful to evaluate the absolute concentrations in thermodynamic equilibrium. To this end we also need to consider the configurational entropy and the energy contribution $-TS$ arising from the vibrational entropy. The latter has been neglected in all equations following (5.1). Estimates based on an Einstein model give values between 3 and $5k_B$ for the vibrational entropy S of native point defects in GaN [8]. The configurational entropy for point defects is simply given by the number of sites N_{sites} at which the defect can be created.[6] We can then write for the concentration in thermodynamic equilibrium:

$$c = N_{\mathrm{sites}} e^{S/k_B} e^{-E/k_B T} \quad . \tag{5.14}$$

Whether thermodynamic equilibrium conditions are reached depends on the mobility of the defect at the specific temperature. A high mobility implies that the assumption of thermodynamic equilibrium conditions should be valid so that the defect concentration will be given via (5.14). On the other hand, at a low mobility the defect concentration is rather likely to be controlled by surface kinetics.

[5] For a detailed description of the splitting of zinc–blende surface slabs into top and bottom see [36, 47, 58].

[6] N_{sites} is the concentration of possible sites. As an example a gallium vacancy can replace any gallium atom giving $N_{\mathrm{sites}} = 2.2 \times 10^{22}/cm^3$.

Line energy of one–dimensional defects

The energy per unit length of a one–dimensional defect is defined as:

$$E_{\text{line}} = E/L \ , \tag{5.15}$$

where L is the length of the line defect and E the formation energy in the structural model. Line energies can sometimes be used as a guide line to determine how frequently a specific type of line defect occurs.

In the case of dislocations E_{line} is mainly the core energy of the dislocation along with the elastic energy stored in the bulk–like region of the model. If the dislocations are modelled in a supercell by a dislocation dipole the elastic energy per dislocation is stored in a cylinder of radius R roughly equal to half the distance between the cores of the dislocation dipole.[7]

On the other hand, following linear elasticity theory, the energy per unit length within a cylinder of radius R, of a general straight dislocation with Burgers vector $\mathbf{b}$ is given by:

$$E_{\text{line}}(R) = \frac{\mu \mathbf{b}^2}{4\pi} \left(\cos^2 \beta + \frac{\sin^2 \beta}{1-\nu} \right) \ln \frac{\alpha R}{|\mathbf{b}|}. \tag{5.16}$$

Here ν is the Poisson's ratio and μ is the shear modulus of the medium. β is the angle between the dislocation line and the Burgers vector. The parameter α represents the core energy of the dislocation. Due to the heavily distorted bonds at the dislocation core the core energy cannot be derived from linear elasticity theory. Instead α could be obtained from a comparison of equations (5.15) and (5.16). To achieve a reasonable accuracy for the parameter α it would be necessary to evaluate E_{line} via (5.15) for different supercell sizes to check whether the stress filed in the region of the dislocation core and thus α is converged. This is a computationally very expensive task and not within the scope of this work.

Wall Energies of Domain Boundaries

The wall energy for a domain boundary is defined as:

$$E_{\text{wall}} = E/A \ , \tag{5.17}$$

where E and A are formation energy and area of the wall contained in the model. Often it is useful to compare domain wall energies with the energy of two surfaces which brought together form the domain boundary.

Summary

Assuming thermodynamic equilibrium it has been shown that the formation energy of a structure can be derived from total energy calculations. Moreover, it depends on the growth conditions and

[7] Note that according to 5.16 the elastic energy of a dislocation in an infinite crystal diverges, so that it makes more sense to define $E_{\text{line}}(R)$.

the position of the Fermi level described by the elemental chemical potentials and the electrochemical potential, respectively. Some examples of how experimental interesting quantities can be derived from the formation energy are given.

Chapter 6

Nonpolar GaN Surfaces

Nonpolar planes in compound semiconductors are characterised by an equal number of cations and anions. In wurtzite material these are the $\{10\bar{1}0\}$ and $\{11\bar{2}0\}$ planes. A nonpolar surface is a surface which lies in a nonpolar plane. Nonpolar surfaces have been investigated for a long time in III-V semiconductors [60, 61]. They can be obtained by cleavage and are important for the fabrication of lasers where they are employed as resonators. It might also be possible to use them as alternative growth directions for MBE or MOCVD. Finally in wurtzite GaN many of the extended defects threading in the [0001] direction exhibit internal nonpolar surfaces or contain atomic arrangements which are similar to them. Studies of the nonpolar surfaces may therefore provide information on the electronic properties of these defects. Oxygen is an impurity which is built in at high concentrations during MBE and MOCVD growth and may segregate to nonpolar surfaces and related extended defects. It is therefore of interest to explore how these impurities adsorb at the surfaces, influence the electrical properties and change the absolute surface energies.

In this chapter an extensive study of GaN $(10\overline{1}0)$ and $(11\overline{2}0)$ surfaces is presented. Atomic geometries, electrical properties and absolute surface energies are determined for stoichiometric as well as for Ga and N terminated surfaces. In addition the adsorption of O on $(10\overline{1}0)$ and $(11\overline{2}0)$ surfaces is investigated. We identify the gallium vacancy surrounded by three oxygen impurities (V_{Ga}–$(O_N)_3$) to be a particularly stable and electrically inert complex which might be involved during the formation of nanopipes (see chapter 7.2).

All stable surfaces obey a simple electron counting rule which has been derived from the nonpolar stoichiometric GaAs (110) surface [62] and is a useful tool for predicting the stabilities of surfaces and defects of III–V semiconductors.

6.1 Common relaxation mechanism for III–V semiconductor surfaces: the electron counting rule

Nonpolar surfaces have been investigated intensively for a variety of III–V semiconductors, in particular the (110) surface in zinc–blende GaAs has been subject to several detailed analysis [63, 64, 65]. A relaxation mechanism within the (1×1) surface unit cell was proposed by low–energy electron diffraction (LEED) studies [64]. The conclusion of this analysis was that the As atoms rotate out of the (110) surface plane and the Ga atoms inward from this plane by an angle $27^o \leq \omega \leq 31^o$. The bond lengths are nearly conserved.

Theoretical investigations confirmed these results and interpreted them by suggesting that the Ga derived dangling bond, which has a high energy in the band gap, transfers its charge to fill the lower lying As derived dangling bond. The resulting configuration has thus a lower energy. The charge transfer can also explain the surface geometry: the surface cation which has lost an electron favours an sp^2 like hybridisation. This makes the Ga atom at the stoichiometric GaAs (110) surface relax inward and form a more planar configuration. On the other hand the dangling bond of the anion is completely filled and s like. Therefore the anion forms bonds with the remaining three p orbitals explaining why the As atom relaxes outwards into a p^3 configuration. It should be noted that for many III–V semiconductors the empty p like cation orbital is pushed above the conduction band edge, whereas the full s like anion orbital is just below the valence band edge.

This phenomenon has since been observed at a variety of III–V semiconductor surfaces and can be summarised in the following *electron counting rule*:

A stable III-V semiconductor surface has all cation dangling bonds emptied and all anion dangling bonds filled resulting in a semiconducting surface.

The stabilities of a variety of III–V semiconductor surfaces could be explained by the electron counting model. In particular, all reconstructions occurring at GaAs (100) and (111) surfaces match this simple rule. Only at a few surfaces, growth under Ga–rich conditions sometimes results in metallic surfaces which do not obey the electron counting rule. Examples for this failure are the $(\sqrt{19} \times \sqrt{19}$ reconstruction of the GaAs $(\overline{111})$ surface [66] and the (1×1) reconstructions at the GaN (0001) and $(000\overline{1})$ surfaces (see chapter 9).

6.2 Stoichiometric $(10\bar{1}0)$ and $(11\bar{2}0)$ surfaces

Stoichiometric nonpolar compound semiconductor surfaces are terminated by an equal number of cation and anion atoms on the top surface layer.

The $(10\bar{1}0)$ surface

The $(10\overline{1}0)$ surface occurs at the walls of nanopipes (see chapter 7.2) and a similar atomic arrangement is found at the core of the threading edge dislocation (see chapter 7.3). Also domain boundaries of type DB–II terminate in $(10\overline{1}0)$ planes (see chapter 8).

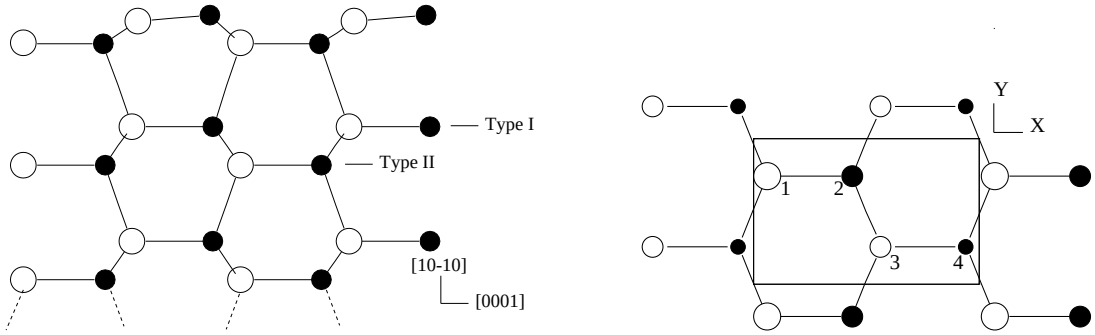

Figure 6.1: Side view along [11$\bar{2}$0] (*left*) and top view (*right*) of the GaN (10$\bar{1}$0) surface. Surfaces of type I have one dangling bond per atom and are therefore more stable than surfaces of type II which have two dangling bonds per atom (see Fig. 6.4). Ga (N) atoms rehybridise into sp^2 (p^3) resulting in an electrically inactive surface (see Fig. 6.2).

The (10$\bar{1}$0) surface exhibits two inequivalent surface types (type I and type II in Fig. 6.1). Surfaces of type I have Ga–N surface dimers and one dangling bond per surface atom. Surfaces of type II have two dangling bonds per atom.

Type I has been investigated theoretically by Pandey *et al.* [67] with a multiconfigurational Hartree-Fock scheme, by Northrup *et al.* [46] using SCF–LDA within a plane wave expansion and by us [68] within the *standard–DFTB* approximation. All these works found atomic geometries with the Ga atoms rehybridized into sp^2 and the N atoms into p^3 as predicted by the electron counting rule. However, as a surprising result the rotation angle was found to be $\sim 6^o$ which is significantly smaller than those known from other stoichiometric nonpolar III–V semiconductor surfaces. Also a contraction of the bond–lengths of the surface dimers of $\sim 6\%$ has been calculated in contrast to the bond–length conserving rotations reported for other III-V's.

Within *SCC–DFTB* the (10$\bar{1}$0) surface is modelled in a (3×2), i.e. 9.5 Å$\times$10.1 Å supercell. Table 6.2 gives details of the calculated geometrical structure along with the results of the first–principles calculations by Northrup and Neugebauer [46].

Table 6.1: Atomic displacements in Å for the top two layers of atoms at the GaN(10$\bar{1}$0) surface. Atom numbers refer to Fig. 6.1. Values in brackets are results of reference [46].

Atom	Δx	Δy	Δz
1 ($Ga_{3\times coord.}$)	-0.10 (-0.11)	0.00	-0.23 (-0.20)
2 ($N_{3\times coord.}$)	0.03 (0.01)	0.00	-0.01 (0.02)
3 ($Ga_{4\times coord.}$)	0.01 (0.05)	0.00	0.08 (0.05)
4 ($N_{4\times coord.}$)	0.04 (0.05)	0.00	0.07 (0.05)

The *SCC–DFTB* method gives an absolute surface energy of 121 meV/Å^2 which is in good agreement with the plane wave calculations (118 meV/Å^2). The stoichiometric surface of type II is found

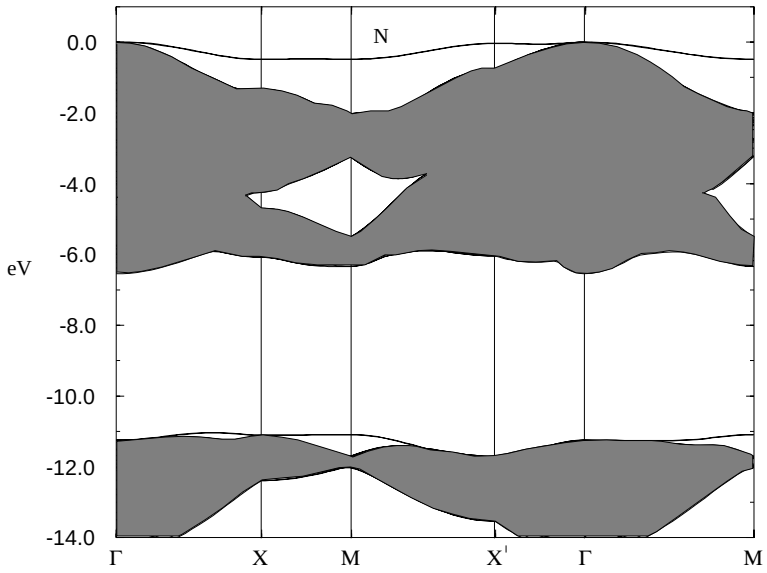

Figure 6.2: Valence band structure of the relaxed GaN $(10\bar{1}0)$ surface calculated within *SCC-DFTB*. The full line represents the N derived surface band. The shaded region corresponds to the bulk projected band structure. The Ga derived dangling bonds (not shown) are $\sim$ 4.1 eV (i.e. clearly more than 3.4 eV) above the valence band maximum. Therefore the surface is found to be electrically inactive.

to have a high surface energy of 152 meV/Å^2 and will therefore transform into a type I surface under equilibrium growth conditions.

Finally we calculated the electrical properties of the stoichiometric type I surface. The band structure is shown in Fig. 6.2. In agreement with Northrup and Neugebauer [46] we find that the relaxed structure has no deep gap states above the valence band maximum (the Ga derived dangling bonds lie $\sim$ 4.1 eV above VBM). Northrup and Neugebauer [46] conclude that the relaxed structure is electrically inactive.[1]

The $(11\bar{2}0)$ surface

The second non–polar surface in wurtzite GaN lies in $\{11\bar{2}0\}$ planes which are also the terminating planes of domain boundaries of type DB–I (see chapter 8).

As can be seen in Fig. 6.3 there is only one type of $(11\bar{2}0)$ surface with three–fold coordinated Ga and N surface atoms arranged in zickzag chains. It has been investigated by Northrup and Neugebauer [46] with SCF–LDA plane wave and in our previous work [68] with *standard–DFTB*. In analogy to the $(10\bar{1}0)$ surface the bond–length contraction of $\sim$ 5% is found to be larger than

[1] Due to the minimal basis set employed in *SCC–DFTB* we cannot describe the conduction band appropriately. Therefore here and in the following we suppose that any localised state which is clearly more than 3.4 eV above the valence band maximum is not a deep gap state. We will call a structure with no deep gap states electrically inactive. Within this approach also *SCC–DFTB* finds $(10\bar{1}0)$ to be electrically inactive.

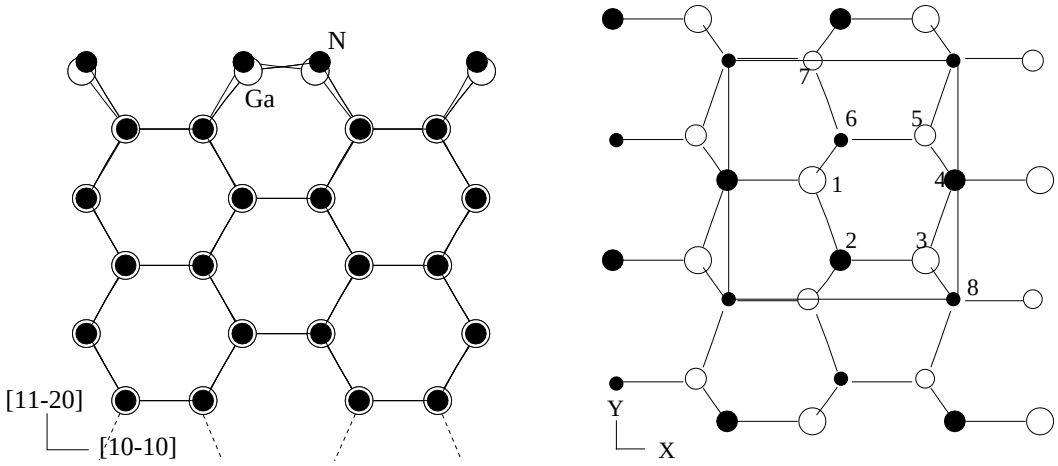

Figure 6.3: Side view along [0001] (*left*) and top view of the GaN (11$\bar{2}$0) surface. In analogy to (10$\bar{1}$0) surfaces Ga (N) atoms rehybridise into sp^2 (p^3) making the surface electrically inactive.

usually observed at nonpolar surfaces of III-V semiconductors, whereas the tilt angle of 6.5^o is again significantly smaller.

Within *SCC–DFTB* we model the (11$\bar{2}$0) surface in a (2×2), i.e. 11.0 Å× 10.3 Å supercell. Table 6.2 shows details of the calculated geometrical structure compared with the values of Northrup and Neugebauer [46].

Table 6.2: Atomic displacements in Å for the top two layers of atoms at the GaN(11$\bar{2}$0) surface. Atom numbers refer to Fig. 6.3. Values in brackets are results of reference [46].

Atom	Δx	Δy	Δz
1 (Ga)	-0.08 (-0.10)	-0.13 (-0.16)	-0.16 (-0.17)
2 (N)	0.04 (0.02)	-0.02 (-0.02)	0.05 (0.05)
3 (Ga)	-0.08 (-0.10)	0.13 (0.16)	-0.16 (-0.17)
4 (N)	0.04 (0.02)	-0.02 (0.02)	0.05 (0.05)
5 (Ga)	0.01 (0.02)	0.00 (0.00)	0.05 (0.05)
6 (N)	0.03 (0.01)	0.02 (0.01)	0.03 (0.02)
7 (Ga)	0.01 (0.02)	0.00 (0.00)	0.05 (0.05)
8 (N)	-0.01 (0.01)	-0.02 (-0.01)	0.03 (0.02)

The *SCC–DFTB* method gives an absolute surface energy of 127 meV/Å^2 (123 meV/Å^2 in reference [46]). We note that the energy difference between (10$\bar{1}$0) and (11$\bar{2}$0) surfaces, although small, might be the reason that GaN crystallites grow in hegaxons, i.e. they terminate in (10$\bar{1}$0) and not in the prismatic (11$\bar{2}$0) planes. Also the (11$\bar{2}$0) surface fulfils the electron counting rule. The *SCC-DFTB* method finds again that the N derived lone pairs are slightly below the valence band maximum and the Ga derived dangling bonds lie 4.0 eV above the valence band maximum. We thus claim that the surface is electrically inactive.

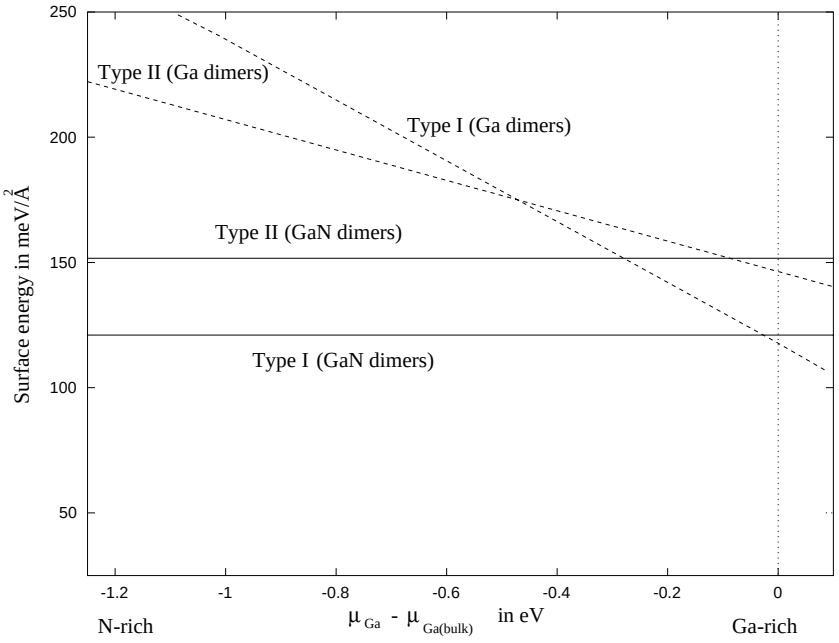

Figure 6.4: Absolute surface energies of different reconstructions of the GaN $(10\bar{1}0)$ surface depending on the Ga chemical potential μ_{Ga}. Only type I surfaces are stable.

6.3 Non–stoichiometric $(10\bar{1}0)$ and $(11\bar{2}0)$ surfaces

Under As-rich growth conditions the GaAs (110) surface is stable with As surface dimers [65]. It is therefore useful to consider also GaN $(10\bar{1}0)$ and $(11\bar{2}0)$ surfaces with modified terminations, i.e. with terminations which are different from the stoichiometric surface. In analogy to the GaAs (110) surface the most promising models are those where the top layer cations (anions) are substituted by the other species resulting in a complete anion (cation) coverage. Both surfaces obey the electron counting rule: the cation terminated surface has emptied cation dangling bonds, whereas the anion terminated surface has filled lone pairs.

The $(10\bar{1}0)$ surface

For the type I surface the full cation coverage is obtained if the N atom (No. 2 in Fig. 6.1) is replaced by Ga giving Ga–Ga dimers. The full anion coverage follows if the Ga atom (No. 1 in Fig. 6.1) is replaced by N yielding N–N dimers. The same can be done for the type II surface. The resulting surface energies are shown in Fig. 6.4.

Over a wide range of the gallium chemical potential μ_{Ga} the stoichiometric surface of type I is the most stable configuration. Ga dimers at the type I surface could be stable under Ga–rich growth conditions. Such a Ga–rich surface should have a higher sticking coefficient for N, and so it has been suggested [46] that it may be advantageous to employ such a surface as an intermediate stage in atomic layer epitaxy. The nitrogen terminated surfaces of type I and type II have very high surface

energies and are not drawn in the diagram. This comes from the fact that the N–N bonds in a nitrogen dimer are significantly shorter ($\sim$ 1.5 Å) than the Ga–N bonds (1.95 Å) resulting in a very strained configuration with a high surface energy. Surfaces of type II are never stable.

The (11$\bar{2}$0) surface

In analogy to the Ga–N surface dimers at the (10$\bar{1}$0) surface, the Ga–N zickzag chains at the (11$\bar{2}$0) surface can be replaced by Ga–Ga or N–N chains. Ga–Ga (N–N) chains are achieved by substituting atoms No. 2 and 4 (No. 1 and 3) in Fig. 6.3 by Ga (N) atoms. In agreement with the plane wave calculations [46] *SCC–DFTB* shows that again the N–N chains are very strained and thus possess very high surface energies and also the Ga–Ga zickzag chains are unfavourable, even under Ga–rich growth conditions. The (11$\bar{2}$0) surface should therefore only occur in its stoichiometric form.

6.4 Oxygen at (10$\bar{1}$0) and (11$\bar{2}$0) surfaces

Oxygen is a very common impurity in GaN which unintentionally enters the material during growth. O has a size which is very similar to N and thus favours a nitrogen substitutional site O_N [8]. Therefore O is a single donor which might attract acceptors to reduce the Coulomb energy.

The stability of extended defects, in particular nanopipes which are surrounded by (10$\bar{1}$0) surfaces, has recently been linked to the O concentration in wurtzite GaN [69].

In the following we show that oxygen segregates to the (10$\bar{1}$0) and (11$\bar{2}$0) surfaces and discuss the stability at different adsorption sites.

O at (10$\bar{1}$0)

Placing O_N into a bulk–like position (six layers below the surface) gives an energy which is by $\sim$ 1.5 eV larger than the energy found for O_N at the surface where it can sit three–fold coordinated. This shows that there is a tendency for O to segregate to the surface. To some extend this tendency is off–set by the configurational entropy encouraging the defect to remain in the bulk.

We have then investigated O in a variety of positions at the (10$\bar{1}$0) surface including O_N, neighbouring O_N–O_{Ga} and O as an adatom where it sits as a bridge between the Ga dangling bond and the N lone pair, but find that if equilibrium with Ga_2O_3 is assumed all of them have higher energies than the ideal surface (see Fig. 6.6).

We next consider the V_{Ga}–$(O_N)_3$ defect (see Fig. 6.5) which is obtained by removing a surface gallium atom (No. 1 in Fig. 6.1) and replacing the surrounding nitrogen atoms by oxygen. A calculation showed that also this defect complex is more stable at the surface than in bulk material (2.2 eV). Furthermore, the defect is electrically inactive with the O atoms passivating the vacancy in the same way as the fully hydrogenated vacancy, VH_4, in Si. Two O neighbours of the surface vacancy are sub-surface and each bonded to three Ga neighbours (Ga–O bond length 1.78 Å to surface Ga, 1.81 Å to second layer Ga and 1.98 Å to third layer Ga atom), but the surface O is

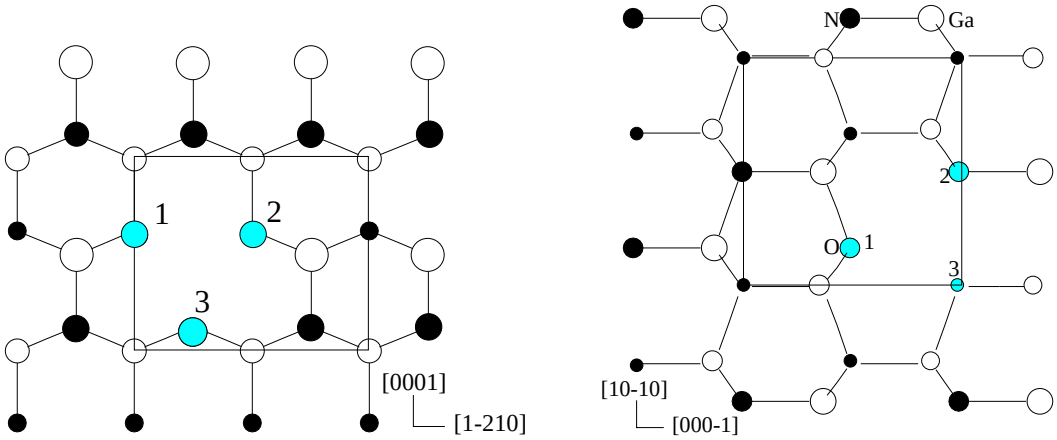

Figure 6.5: Schematic top view of the V_{Ga}–$(O_N)_3$ defect complex at the $(10\bar{1}0)$ surface (*left*) and at the $(11\bar{2}0)$ surface (*right*). White (black) circles represent Ga (N) atoms and large (small) circles top (second) layer atoms. At $(10\bar{1}0)$ atoms 1 and 2 are three–fold coordinated second layer O atoms each with one lone pair, atom 3 is a two–fold coordinated first layer O with two lone pairs. At $(11\bar{2}0)$ atoms 1 and 2 are two–fold coordinated surface layer O atoms each with two lone pairs, atom 3 is a three–fold coordinated second layer O with one lone pairs.

bonded to two subsurface Ga atoms at a normal oxygen bridge site (1.77 Å). With a formation energy of 1.7 eV per V_{Ga}–O_{3N} site (calculated for Ga–rich growth conditions and O in equilibrium with Ga_2O_3) below the defect free $(10\bar{1}0)$ surface, V_{Ga}–$(O_N)_3$ at the $(10\bar{1}0)$ surface is a very stable arrangement. Furthermore V_{Ga}–$(O_N)_3$ does not encourage overgrowth. Growth must proceed by adding a Ga atom to the vacant site but this leaves three electrons in shallow levels near the conduction band probably resulting in an unstable defect $(O_N)_3$ which even under Ga–rich growth conditions is by 1.5 eV higher than V_{Ga}–$(O_N)_3$. At the growth temperature, these O atoms will drift away diffusing to the new surface.

O at $(11\bar{2}0)$

The energetic order of the investigated oxygen configurations was exactly the same as at the $(10\bar{1}0)$ surface with V_{Ga}–$(O_N)_3$ being a very stable and electrically inert defect complex. Two oxygen atoms of V_{Ga}–$(O_N)_3$ at the $(11\bar{2}0)$ surface sit two–fold coordinated and only one oxygen is three–fold coordinated (see Fig. 6.5).

Comparison between the oxygen adsorption at $(10\bar{1}0)$ and $(11\bar{2}0)$ surfaces

Although the V_{Ga}–$(O_N)_3$ defect complex was found to be very stable at both surfaces, there is a fundamental difference for the formation of this complex during growth.

V_{Ga}–$(O_N)_3$ can form very easily on $(10\bar{1}0)$ surfaces. Suppose, one oxygen atom has emerged to the

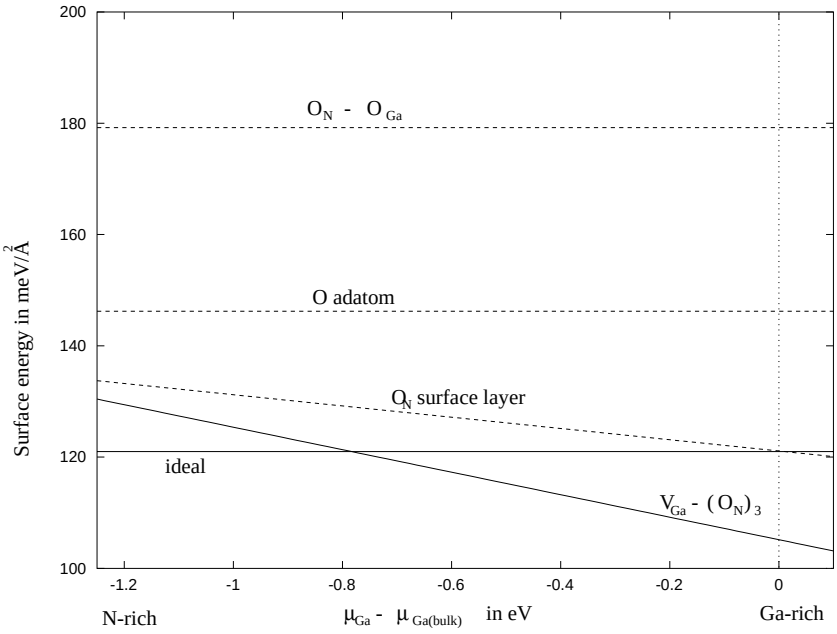

Figure 6.6: Absolute surface energies for the oxygen adsorption at the GaN (10$\bar{1}$0) surface depending on the Ga chemical potential μ_{Ga}. Oxygen is assumed to be in equilibrium with Ga_2O_3.

surface where it sits at a nitrogen site, i.e. atom No. 3 in Fig. 6.5 is the only oxygen. A gallium vacancy is very likely to be attracted reducing the Coulomb energy. Already this gallium vacancy oxygen complex is stable: the oxygen sits two fold coordinated in a bridge position whereas the two nitrogen atoms have three bonds. This V_{Ga}–O_N defect complex will then attract the other two oxygens to form a charge neutral V_{Ga}–$(O_N)_3$ complex.

This mechanism however does not exist at the (11$\bar{2}$0) surface. Assuming one oxygen at a surface nitrogen site, i.e. atom No. 1 in Fig. 6.5 is the only oxygen. A gallium vacancy would lead to a high energy configuration since there would be a two–fold coordinated surface nitrogen at pos. 2. A stable gallium vacancy at the surface would only be created if two surface oxygens sit on neighbouring nitrogen sites, i.e. at pos. 1 and pos. 2. This is not very likely to happen, since the oxygens are positively charged donors and thus repel each other.

As an illustrative example for the energy difference during growth we calculated the energy for a gallium vacancy surrounded by only one surface O. We found that due to the two–fold coordinated N at the (11$\bar{2}$0) surface the energy of V_{Ga}–O is indeed by 1.5 eV higher than at the (10$\bar{1}$0) surface. We therefore conclude that although V_{Ga}–$(O_N)_3$ defect complexes are stable at both types of nonpolar GaN surfaces they are only likely to occur at the (10$\bar{1}$0) surfaces.

Summary

Ga, N and O terminated nonpolar surfaces have been investigated. In agreement with SCF–LDA calculations the stoichiometric surfaces were found to be electrically inactive and to be stable

over a wide range of growth conditions. Only in a Ga–rich environment a Ga terminated $(10\bar{1}0)$ surface is energetically favourable. N terminated surfaces have very high formation energies in any environment. O segregation to the stoichiometric surfaces is likely to occur. In particular at the $(10\bar{1}0)$ surface this can lead to the formation of the energetically favourable V_{Ga}–$(O_N)_3$ defect complex which is electrically and chemically inactive.

Chapter 7

Line Defects: Threading Dislocations and Nanopipes

Frequently sapphire substrates are used to grow device quality wurtzite–(α) GaN with the metal-organic chemical vapour phase deposition (MOCVD) technique. In this case, growth proceeds along the c–axis. Figure 7.1 shows a cross–sectional TEM weak beam image of a typical sample. The

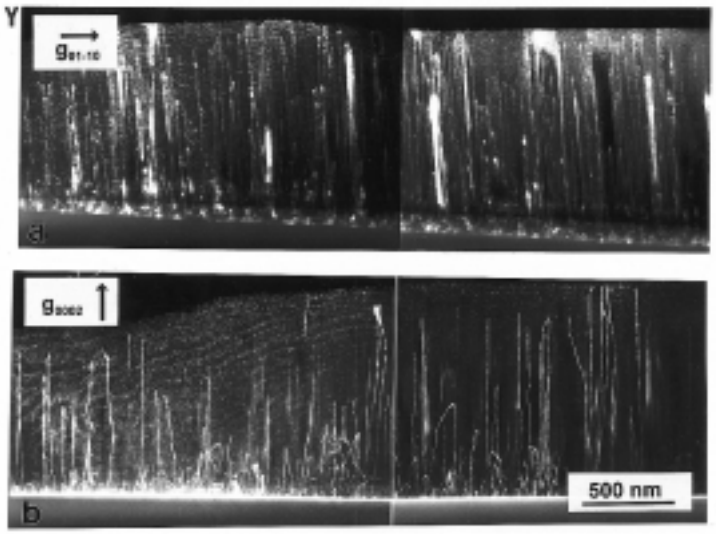

Figure 7.1: Dislocation arrangement in a GaN sample grown on sapphire by MOCVD: (a) cross-sectional TEM ($\mathbf{g}/3\mathbf{g}$) weak beam image, $\mathbf{g} = (01\bar{1}0)$. Screw dislocations with $\mathbf{b} = \pm[0001]$ are out of contrast. Dislocations with a $\mathbf{b}$-component in the interface, i.e. $\mathbf{b} = \pm\frac{1}{3}[\bar{1}2\bar{1}0]$ are visible. (b) cross-sectional TEM ($\mathbf{g}/3\mathbf{g}$) weak beam image, $\mathbf{g} = (0002)$. Screw dislocations with $\mathbf{b} = \pm[0001]$ are visible. Dislocations with a $\mathbf{b}$-component in the interface are out of contrast. S. Christiansen *et al.* [70].

large lattice misfit between GaN and the sapphire substrate of 13% results in dislocation tangles

near the interface. In addition to these geometric misfit dislocations which have dislocation lines in the basal plane, also isolated threading dislocations with dislocation lines parallel to **c** and Burgers vectors **c**, **a** and **c+a** persist beyond the interface [71, 72, 73]. Since many of them penetrate the entire epilayer from the substrate to the surface they are called threading dislocations.

An unexpected finding [74, 75] is that inspite of their high density (typically $\sim 10^9$ cm^{-2}) and the fact that they cross the active region of the devices (starting typically ~ 0.5 μm above interface) these threading dislocations in GaN do not lead to a pronounced reduction in the device-lifetime of the light-emitting diodes [76] or blue lasers [77]. This can be contrasted with GaAs where radiation enhanced dislocation motion [78] readily occurs and leads to an increase in non-radiative processes. It is therefore of considerable interest to understand the structural and electrical properties of threading dislocations in GaN and to compare them with those of dislocations in more traditional semiconductors.

Threading dislocations are often associated with the appearance of long nanopipes which are parallel to **c** and have hexagonal cross sections with constant diameters ranging from 20–250 Å [79, 80, 81, 69]. Nanopipes degrade the material quality. In particular, they can get filled by metal during the formation metal of contacts with GaN and have already caused short circuits in laser devices. Frank [82] predicted that a dislocation whose Burgers vector exceeds a critical value should have a hollow tube at the core. The equilibrium radius is achieved by balancing the elastic strain energy released by the formation of a hollow core against the energy of the resulting free surfaces. Liliental–Weber *et al.* [54] suggested another possible mechanism for the formation of nanopipes. They found the density of nanopipes to be increased with the impurity concentration and proposed that impurities poison the walls of the nanopipes which prevents the nanopipes from growing out.

Finally, the origin of defect–induced electronic states, which lie deep in the GaN band gap and can thus significantly alter the optical performance, is possibly related to threading dislocations. Especially in laser devices deep gap states are of concern since parasitic components in the emission spectrum are highly undesirable. The most commonly observed emission in unintentionally doped n–type GaN, the yellow luminescence (YL), is centred at 2.2-2.3 eV with a line width of ≈ 1 eV. Several models for the origin of the YL in GaN have been proposed. Most of them assume the transition to be between a shallow donor and a deep acceptor [83] or a deep donor and a shallow acceptor [84]. Recent work has however found evidence for the deep acceptor model [85]. Cathodoluminescence (CL) studies of the yellow luminescence have shown that the YL is spatially non-uniform (see Figure 1.3 in the introduction of this thesis). A possible reason for this non–uniform distribution of the YL could be related to threading dislocations, which are non–uniformly distributed throughout the epilayer and might be electrically active. Indeed, also atomic force microscopy (AFM) in combination with CL has led to the conclusion that threading dislocations act as non-radiative recombination centres and degrade the luminescence efficiency in the blue light spectrum of the epilayers [86]. However, the type of dislocations involved in the YL is not clear: Christiansen *et al.* [70] suggest that the YL arises from threading dislocations with a screw component whereas Ponce *et al.* [87] localise the YL at low angle grain boundaries which predominantly contain threading edge dislocations. Moreover, it is not clear whether dislocations in the pure, i.e. impurity free form or defects trapped in the stress field of dislocations are responsible for the YL.

In this chapter, the atomic geometries, electrical properties and line energies of threading screw and edge dislocations with full and open cores are investigated [88]. The results are interpreted by comparing elements of the dislocation cores with nonpolar $(10\bar{1}0)$ surfaces (see chapter 6). Possible

mechanisms for the formation of nanopipes are then examined [89]. Finally, we also explore the segregation of gallium vacancies and oxygen as well as related defect complexes to threading edge dislocations and discuss their implication for the yellow luminescence [91].

7.1 Screw dislocations

Threading screw dislocations in wurtzite material have a Burgers vector parallel to the dislocation line [0001]. The smallest screw dislocations have thus elementary Burgers vector $\pm$ **c**. Screw dislocations occur at a density $\sim 10^6$ cm^{-2} in α–GaN grown by MOCVD on (0001) sapphire. Since they nucleate in the early stages of growth at the sapphire interface and thread to the surface of the crystallites (see Fig. 7.1.b), screw dislocations are believed to arise from the collisions of islands during growth [72]. At a screw dislocation the surface is rough and has a high energy which favours the nucleation of islands. They are thus vital for the growth process. In GaN screw dislocations are unusual in often being associated with nanopipes [92]. However, full core screw dislocations [93] and screw dislocations with a very narrow opening of ~ 8 Å [94] are also reported.

7.1.1 Full core screw dislocations

Screw dislocations with a full core have been observed by Xin *et al.* [93] using high resolution Z–contrast imaging (see Fig. 7.2).

Within the *SCC–DFTB* method the dislocations are modelled in 210 atom clusters periodic along the dislocation line with periodicity **c** and in 576 atom $(12 \times 12 \times 1)$ supercells. Because of the large lateral extension of the supercell (12×12), only k–points for the sampling along the c–direction are necessary. Two k-points parallel to this direction were found sufficient to carry out the sum over the Brillouin zone: using four k–points gave only a difference of ≤ 0.02 eV/Å in the dislocation line energy. In the *AIMPRO* case, relaxations were carried out in 392 atom stoichiometric clusters. For further details concerning the modelling of dislocations see appendix D.3.

Both methods found heavily distorted bond lengths for the full core screw dislocation (see Fig. 7.3 and Table 7.1) yielding deep gap states ranging from 0.9–1.6 eV above the valence band maximum, VBM, and shallow gap states at ~ 0.2 eV below the conduction band minimum, CBM. An analysis of these gap states revealed that the states above the VBM are localised on N core atoms, whereas the states below CBM are localised on core atoms but have mixed Ga and N character. Therefore the full core screw dislocation is electrically active and could act as a non–radiative centre [88]. Similarly one could expect that dislocations of mixed type would also have deep states in the gap as a result of the distortion arising from their screw component. Indeed, CL experiments have related the yellow luminescence centred at 2.2 eV to screw dislocations [70]. In addition, atomic force microscopy in combination with CL imaging has shown that threading dislocations with a screw component act as nonradiative combination sites [86]. A calculation in a supercell containing a screw dipole consisting of two dislocations with $\mathbf{b} = [0001]$ and $-[0001]$, which are symmetrically equivalent, confirmed these results and gave a high line energy of 4.88 eV/Å. This is mainly the core energy of each screw dislocation together with the elastic energy stored in a cylinder of diameter roughly equal to the distance between the cores, 19.1 Å. See chapter 5 for a more detailed interpretation of dislocation line energies.

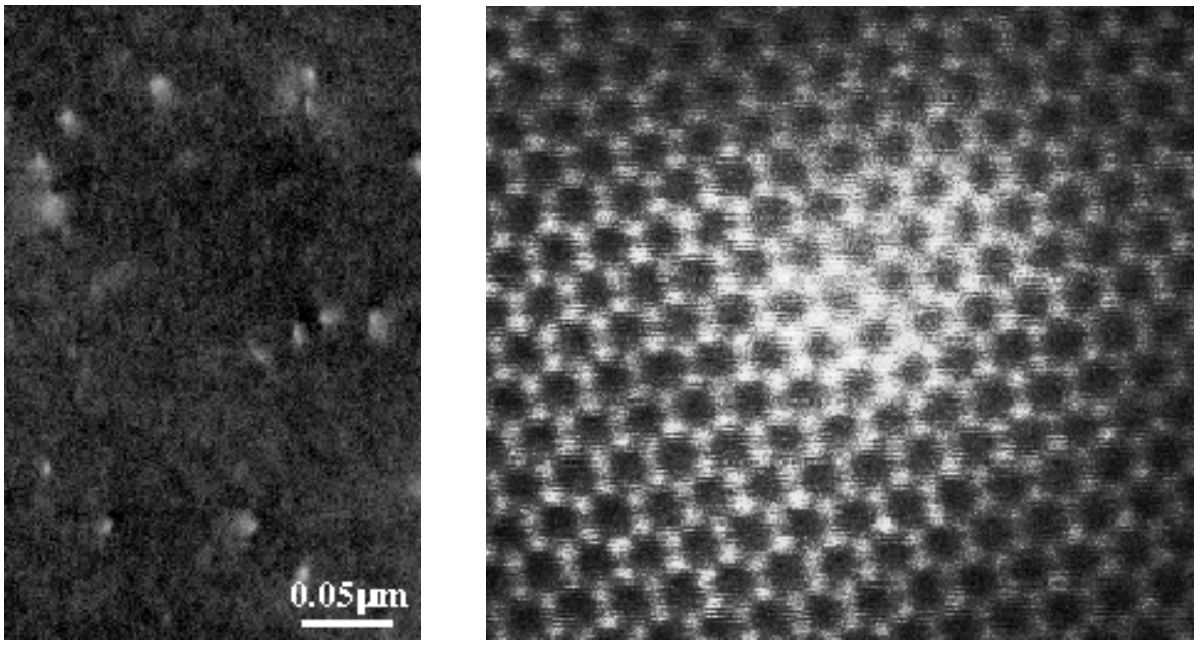

Figure 7.2: *Left:* Low magnification angular dark field image along [0001]. Threading dislocations show as bright dots due to their strain field. *Right:* High resolution Z–contrast image of an end–on pure screw dislocation showing a full core. Y. Xin *et al.* [93].

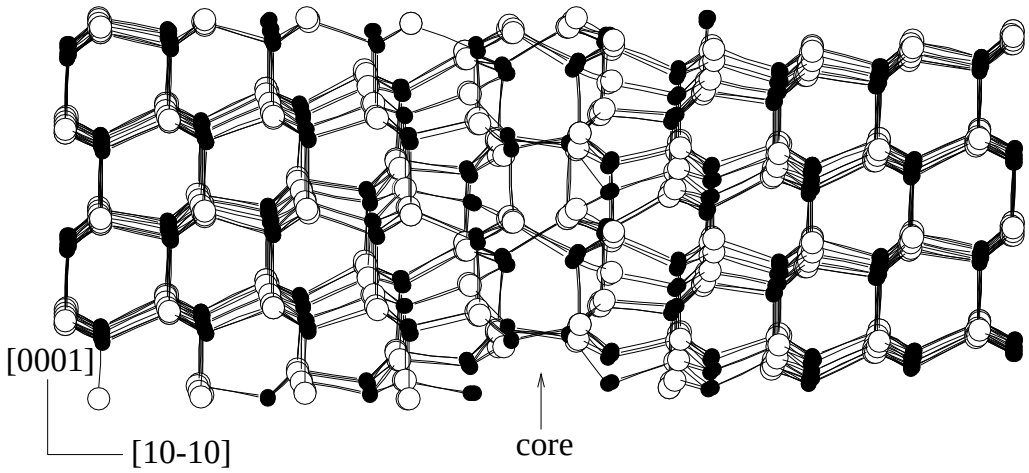

Figure 7.3: Side view (in $[11\bar{2}0]$) of the relaxed core of the full–core screw dislocation ($\mathbf{b} = [0001]$). The atoms at the dislocation core adopt heavily distorted configurations (see Table 7.1) yielding deep gap states.

Table 7.1: Bond lengths, min-max (average), in Å and bond angles (min-max) in ^{o} for the most distorted atoms at the core centre of the full–core screw dislocation ($\mathbf{b} = [0001]$).

Atom	bond lengths	bond angles
1 ($Ga_{4\times coord.}$)	1.85-2.28 (2.14)	68-137
2 ($N_{4\times coord.}$)	1.89-2.28 (2.13)	71-136

7.1.2 Screw dislocations with a narrow opening

We now investigate whether the line energy of the full core screw dislocation is reduced if material is taken from the core. Accordingly, calculations were then carried out using the same supercell as for the full core screw dislocations, but with the hexagonal core of each screw dislocation removed leading to a pair of open–core dislocations with diameters $d \approx 7.2$ Å. The relaxed structure (Fig. 7.4) preserved the hexagonal core character, demonstrating that the internal surfaces of the dislocation cores shown in Fig. 7.5 are similar to $\{10\bar{1}0\}$ type facets except for the topological singularity required by a Burgers circuit.

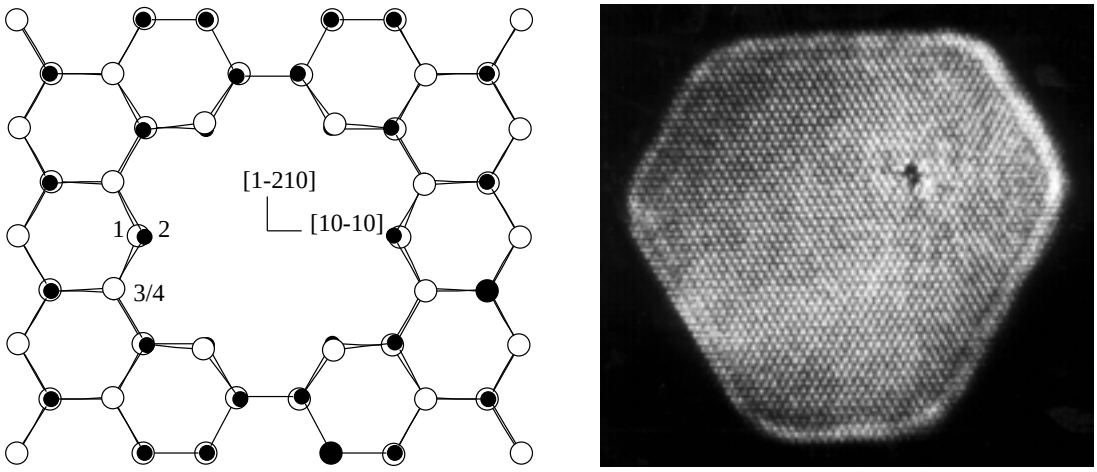

Figure 7.4: *Left:* Top view (in [0001]) of the relaxed core of the open–core screw dislocation ($\mathbf{b} = [0001]$). The three fold coordinated atoms 1 (Ga) and 2 (N) adopt a hybridisation similar to the $(10\bar{1}0)$ surface atoms. *Right:* TEM image of a nanopipe containing a dislocation with a screw component. During growth the nanopipe closes leaving the dislocation with an opening of three rows (≈ 8 Å) wide (see black arrangement within the nanopipe). Z. Liliental–Weber [94]

.

It is instructive to compare the distortions of the atoms situated at the wall of the open–core (Table 7.2) with the corresponding atoms at the $(10\bar{1}0)$ surface (Table 7.3). In both cases, the three fold coordinated Ga (N) atoms adopt an sp^2- (p^3)- like hybridisation which lowers the surface energy and cleans the band gap [46]. Indeed, we find that unlike the full–core screw dislocation, the gap is free from deep states [88].

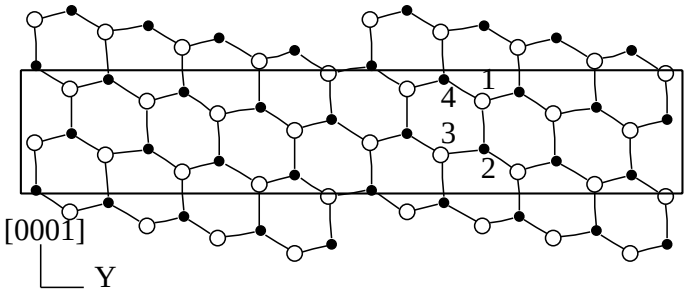

Figure 7.5: Projection of the wall of the open–core (d=7.2 Å) screw dislocation (**b** = [0001]). The three fold coordinated atoms 1 (Ga) and 2 (N) adopt a hybridisation similar to the $(10\overline{1}0)$ surface atoms.

Table 7.2: Bond lengths, min-max (average) in Å and bond angles, min-max (average) in ^{o} for the most distorted atoms at the wall of the open core screw dislocation (**b** = [0001]). Atom numbers refer to Fig. 7.4 and 7.5.

Atom	bond lengths	bond angles
1 ($Ga_{3\times coord.}$)	1.86-1.89 (1.88)	107-123 (117)
2 ($N_{3\times coord.}$)	1.88-2.05 (1.96)	102-111 (108)
3 ($Ga_{4\times coord.}$)	1.89-2.07 (1.96)	100-122
4 ($N_{4\times coord.}$)	1.93-2.03 (1.97)	98 -120

There are, however, in contrast to the $(10\overline{1}0)$ surface, energetically shallow gap states. Calculations were carried out for a distorted $(10\overline{1}0)$ surface, i.e. a $(10\overline{1}0)$ surface in a unit cell where the unit cell vectors were modified to give a distorted surface corresponding to that of the wall of the open–core screw dislocation with diameter $d = 7.2$ Å. We find that the distorted $(10\overline{1}0)$ surface has a spectrum with shallow states very similar to those of the open–core screw dislocation with $d = 7.2$ Å. We also calculated the spectrum for a nanopipe with $d = 7.2$ Å but without a dislocation core. This, like the undistorted $(10\overline{1}0)$ surface, possesses a gap free from deep states, although there are N (Ga) derived surface states lying slightly below (above) the VBM (CBM). These results indicate that the shallow states in the open–core screw dislocation with diameter $d = 7.2$ Å can be attributed to the distortion arising from the dislocation Burgers vector. Calculations for a series of different distortions of the $(10\overline{1}0)$ surface corresponding to open–core screw dislocations with different diameters also suggest that open–core screw dislocations with diameters greater than $\sim$ 20 Å should have no gap states at all. As can be seen in Table 7.2 the distortion in the open–core screw dislocation is significantly less than that in the full–core screw dislocation (see Table 7.1). It is therefore not surprising that the calculated line energy of 4.55 eV/Å is lower than the line energy of the full–core screw dislocation. The energy required to form the surface at the wall is compensated by the energy gained by reducing the strain. However, a further opening gave a higher line energy and we conclude that the equilibrium diameter is $\sim$ 7.2 Å. This opening has also been reported by Liliental–Weber *et al.* [94] who found some of the screw dislocations to have holes which are three atomic rows wide (see Fig. 7.4).

Table 7.3: Bond lengths, min-max (average) in Å and bond angles, min-max (average) in ^{o} for the top two layers of atoms at the GaN($10\bar{1}0$) surface. Atom numbers refer to Fig. 6.1.

Atom	bond lengths	bond angles
1 ($Ga_{3\times coord.}$)	1.83-1.88 (1.86)	116-117 (117)
2 ($N_{3\times coord.}$)	1.83-1.92 (1.89)	107-111 (108)
3 ($Ga_{4\times coord.}$)	1.91-2.02 (1.94)	107-112
4 ($N_{4\times coord.}$)	1.88-2.03 (1.93)	99-115

A theoretical approach to predict the opening of a screw dislocation was deduced by Frank [82]. By balancing the elastic dislocation strain energy released by the formation of a hollow core against the energy of the resulting free surfaces, he showed that, for isotropic linear elasticity and a cylindrical core, the equilibrium core radius r_{eq} is

$$r_{eq} = \frac{\mu \mathbf{b}^2}{8\pi^2\gamma} , \tag{7.1}$$

where γ is the surface energy, μ is the shear modulus and b is the Burgers vector. For a rough estimate of r_{eq}, we use the theoretical value for the surface energy of $\{10\bar{1}0\}$ facets which we found to be $\gamma = 121$ meV/Å^2 $= 1.9$ Jm^{-2} (see chapter 6). Taking $\mu = 8 \times 10^{10}$ Nm^2 as an upper limit and $\mathbf{b} = 0.5$ nm for the Burgers vector of an elementary screw dislocation yields $r_{eq} \approx 0.2$ nm. It is unlikely, that isotropic elasticity theory can describe the severely distorted full core dislocation which limits the usefulness of Frank's expression (7.1) concerning the precise quantitative value of the equilibrium diameter. Our calculated value of ~ 7.2 Å and Frank's value are reasonably close since the relatively small line energy difference found between full core and open core ($d \sim 7.2$ Å) screw dislocations suggests a shallow minimum which probably allows all intermediate structures to exist. In our calculations only structures constructed by removing entire hexagons, but not those obtained by removing single rows were considered. Calculating the latter ones, may lead to slightly lower energies.

In summary, it can be concluded from our calculations and from Frank's theorem that in GaN screw dislocations with an elementary Burgers vector $\mathbf{c}$ can exist with a full core and with a narrow opening up to ~ 7.2 Å. The full core screw dislocation is electrically active whereas the screw dislocation with a hexagonal opening has only shallow gap states. These states are induced by the distortion arising from the Burgers vector.

7.2 The formation of nanopipes

Nanopipes in α–GaN thread along the c–axis and have hexagonal cross sections, i.e. they are inclosed by $\{10\bar{1}0\}$ type walls (see Fig. 7.6). Nanopipes are commonly observed in MOCVD grown epilayers on sapphire [79, 80, 81]. However, they have also been reported in samples grown by MBE on SiC [69]. Nanopipes occur at a density up to $\sim 10^8$ cm^2 and have constant diameters ranging from 20–250 Å. The first suggestion was that they were the manifestation of screw dislocations with empty cores as discussed by Frank a long time ago [82]. However, as shown above *ab initio* calculations

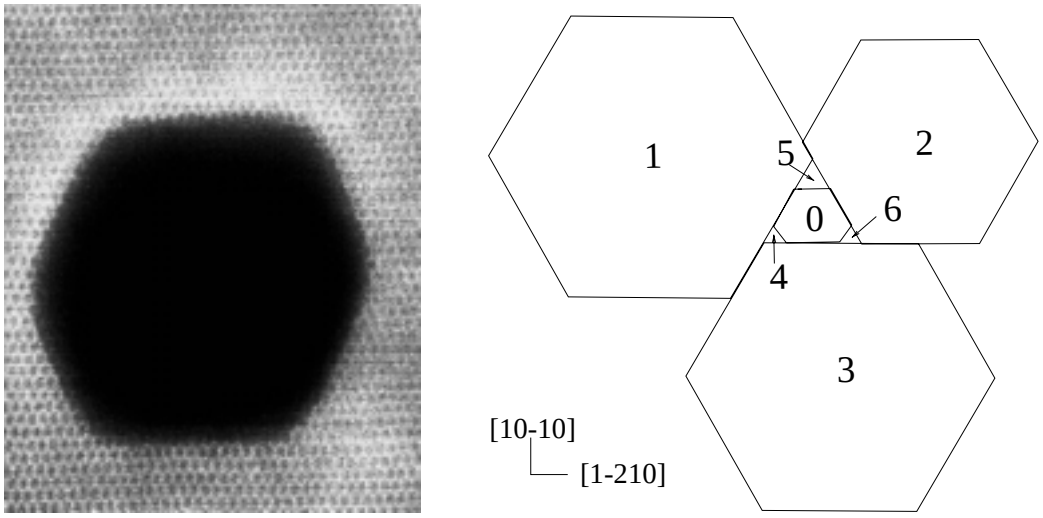

Figure 7.6: *Left:* High resolution Z–contrast image along [0001] of a nanopipe. Y. Xin unpublished. *Right:* Suggested mechanism for the formation of a nanopipe (area No. 0). Three hexagons (No. 1,2,3) are growing together. As the surface to bulk ratio at ledges (No. 4,5,6) is very large, they grow out quickly leaving a nanopipe (area No. 0) with $\{10\bar{1}0\}$ type facets.

as well as Frank's theorem do not support the idea that in GaN the core of a screw dislocation with Burgers vector equal to **c** is open with such a large diameter. Pirouz [55] has therefore argued that superscrew dislocations with Burgers vectors $n\mathbf{c}$ where $n > 1$ are formed during growth by the collision of islands. Clearly, if n were big enough, then the cores would be open. However, there is at present no microscopic evidence for such dislocations [95]. We also note that in a typical sample some nanopipes ($< 10\%$) were observed which could not be associated with a screw dislocation [95]. Another possibility is that the $\{10\bar{1}0\}$ type surface walls of the nanopipes are coated by hydrogen which is always present during MOCVD growth. This might result in a very low surface energy explaining the large diameter of the nanopipes via formula (7.1). The adsorption of H on (1010) surfaces has been investigated by Northrup and Neugebauer [51]. They found however, that at a usual MOCVD growth temperature of $\approx 1000^o$ C the energy of the $(10\bar{1}0)$ surface is not lowered by the adsorption of H and concluded that hydrogen is not responsible for the formation of nanopipes. This has been confirmed by the recent work of Liliental-Weber *et al.* [69] who detected nanopipes also in MBE grown material where the concentration of hydrogen is negligible. On the other hand, Liliental-Weber *et al.* [69] found the diameters and densities of nanotubes to be increased in the presence of impurities, e.g. O, Mg, In and Si, and argued that these impurities decorate the $\{10\bar{1}0\}$ walls of the nanotubes inhibiting overgrowth. O being the main source of unintentional doping in GaN, we will now discuss how O can cause the formation of nanopipes.

From Fig. 7.7 it can be seen that the surface walls of nanopipes are $(10\bar{1}0)$ surfaces which are predominantly of type I and flat, i.e. they usually have one dangling bond per atom and only little irregularities caused by surface steps. GaN samples usually contain a considerable concentration of gallium vacancies and oxygen which as our calculations show, have both a tendency to diffuse to $(10\bar{1}0)$ surfaces. It is therefore very likely that many gallium vacancies and oxygen atoms have segregated to the nanopipe walls where they can form V_{Ga}–$(O_N)_3$ defect complexes. In chapter 6 it

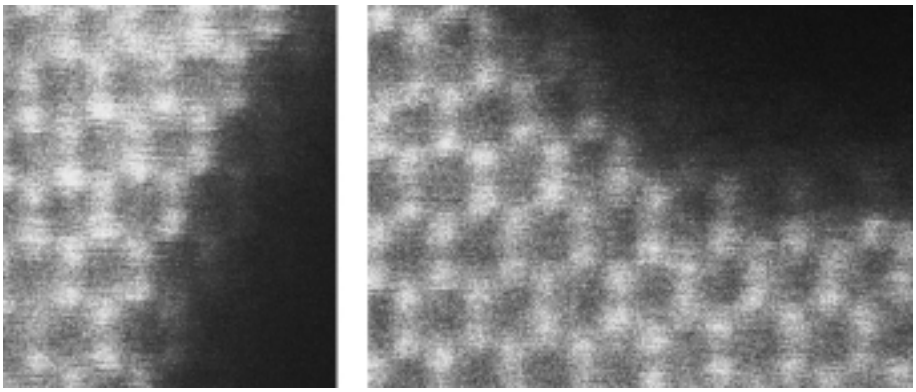

Figure 7.7: *Bottom:* Straight edge (*left)* and corner (*right*) at the $\{10\bar{1}0\}$ type wall. Y. Xin *et al.* [96].

was shown that V_{Ga}–$(O_N)_3$ are very stable defect complexes on $(10\bar{1}0)$ surfaces of type I. Moreover, overgrowth was determined to be difficult as oxygen atoms would drift away diffusing to the new surface. Since V_{Ga}–$(O_N)_3$ defect complexes do not lead to any noticeable change of the atomic positions at the surface they are consistent with the HRTEM image in Fig. 7.7. Unfortunately at present there seems to be no direct way for detecting V_{Ga}–$(O_N)_n$ at nanopipe walls by experiments.

To explain the formation of nanopipes, we suppose [89, 90] that oxygen atoms constantly diffuse to the $\{10\bar{1}0\}$ type surfaces. Within the framework of Stranski-Krastanow growth, the internal $\{10\bar{1}0\}$ type surfaces between GaN islands are shrinking along with the spaces between colliding GaN islands (see Fig. 7.6). Therefore, the O coverage and density of V_{Ga}–$(O_N)_3$ defects is expected to increase. The maximum concentration of this defect would be reached if 50% (100%) of the first (second) layer N atoms were replaced by O (see Fig. 6.5). It is, however, likely that far lower concentrations are necessary to stabilise the surface and make further shrinkage of the inter-island spaces impossible thus leaving a nanopipe. Provided oxygen could diffuse to the surface fast enough, the diameter and density of the holes would be related to the density of oxygen atoms in the bulk. This has indeed been observed by Liliental–Weber *et al.* [54] who found that as the concentration of oxygen in the material changed by about an order of magnitude the number of nanopipes increased by a factor of ~ 3 and the diameter of the nanopipes changed from (3–10) nm to (6–12) nm. A more detailed prediction of the radii and density of nanopipes depending on the oxygen partial pressure would require thermodynamic equilibrium for the formation of the nanopipes. However, as can be seen in the large distribution of nanopipe radii, this is obviously not reached.

It is also necessary to explain why the tubes have $\{10\bar{1}0\}$ type surface walls. The other low index surface perpendicular to the growth direction which could become poisoned by O impurities and thus be responsible for the formation of nanopipes is the $(11\bar{2}0)$ surface. $\{11\bar{2}0\}$ type surfaces are not observed presumably because of their higher absolute surface energy (see chapter 6). Moreover we suggested that because of the different surface topologies V_{Ga}–$(O_N)_3$ is likely to form on $(10\bar{1}0)$ surfaces but not on $(11\bar{2}0)$ surfaces during growth.

Finally, we point out that our arguments are still valid if each nanopipe is associated with a screw dislocation since the walls of the tube with a dislocation are locally equivalent to a $(10\bar{1}0)$ surface which is distorted to form a helix (see 7.1.2). We therefore conclude that rather than being

responsible for the formation of nanopipes screw dislocation are attracted to nanopipes in order to reduce the elastic energy.

7.3 Threading edge dislocations

Pure edge dislocations lie on $\{10\bar{1}0\}$ planes and have a Burgers vector $\mathbf{b} = \mathbf{a} = [1\bar{2}10]/3$. They are a dominant species of dislocation, occurring at extremely high densities of $\sim 10^8 - 10^{11}$ cm^{-2} in α–GaN grown by MOCVD on (0001) sapphire (Fig. 7.1.a) and in analogy to screw dislocations are thought to arise from the collisions of islands during growth [72].

Within the *SCC–DFTB* method threading edge dislocations are modelled in 210 atom clusters periodic along the dislocation line with periodicity **c** and in 576 atom ($12 \times 12 \times 1$) supercells containing a dislocation dipole. In analogy to the models for the screw dislocations, two k-points parallel to **c** were used to carry out the sum over the Brillouin zone. In the *AIMPRO* case, relaxations were carried out in 286 atom stoichiometric clusters.

The relaxed core of the threading edge dislocation is shown in Fig. 7.8. The corresponding bond–lengths and bond angles of the most distorted atoms are given in Table 7.4. With respect to the perfect lattice the distance between columns (1/2) and (3/4) [and the equivalent on the right] are 9 % contracted while the distance between columns (9/10) and (7/8) [and the equivalent on the right] are 13 % stretched. This atomic geometry for the threading edge dislocation has recently been confirmed by Xin *et al.* [93] using atomic resolution Z–contrast imaging (see Fig. 7.9). Consistent with our calculation they determined a contraction (stretching) of 15 ± 10 % of the distances between the columns at the dislocation core. Our calculations show that in a manner identical to the $(10\bar{1}0)$ surface, the three–fold coordinated Ga (N) atoms (no. 1 and 2 in Fig. 7.8) relax towards sp^2 (p^3) leading to empty Ga dangling bonds pushed towards the CBM, and filled lone pairs on N atoms lying near the VBM. Thus we find threading edge dislocations to be electrically inactive [88].

From a supercell calculation, we obtain a line energy of 2.19 eV/Å for the threading edge dislocation. We note that this line energy is considerably lower than the one found for the screw dislocation with a narrow opening. This can be interpreted by noting that the edge dislocation has a smaller

Table 7.4: Bond lengths, min-max (average) in Å and bond angles, min-max (average) in ^{o} for the most distorted atoms at the core of the threading edge dislocation ($\mathbf{b} = \frac{1}{3}[1\bar{2}10]$). Atom numbers refer to Fig. 7.8.

Atom	bond lengths	bond angles
1 ($Ga_{3\times coord.}$)	1.85-1.86 (1.85)	112-118 (116)
2 ($N_{3\times coord.}$)	1.88-1.89 (1.86)	106-107 (106)
3/4 ($Ga/N_{4\times coord.}$)	1.86-1.95 (1.91)	97-119
5/6 ($Ga/N_{4\times coord.}$)	1.92-2.04 (1.97)	100-129
7/8 ($Ga/N_{4\times coord.}$)	1.94-2.21 (2.06)	94-125
9/10 ($Ga/N_{4\times coord.}$)	1.95-2.21 (2.11)	100-122

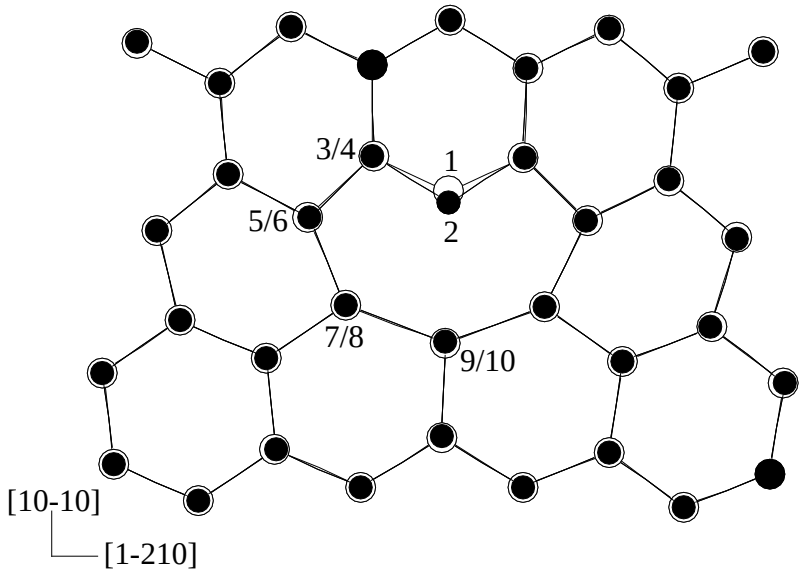

Figure 7.8: Top view (in [0001]) of the relaxed core of the threading edge dislocation ($\mathbf{b} = \frac{1}{3}[1\bar{2}10]$). The three fold coordinated atoms 1 (Ga) and 2 (N) adopt a hybridisation similar to the $(10\bar{1}0)$ surface atoms. The distance between columns (1/2) and (3/4) are by 9% contracted while the distance between columns (7/8) and (9/10) is by 13% stretched.

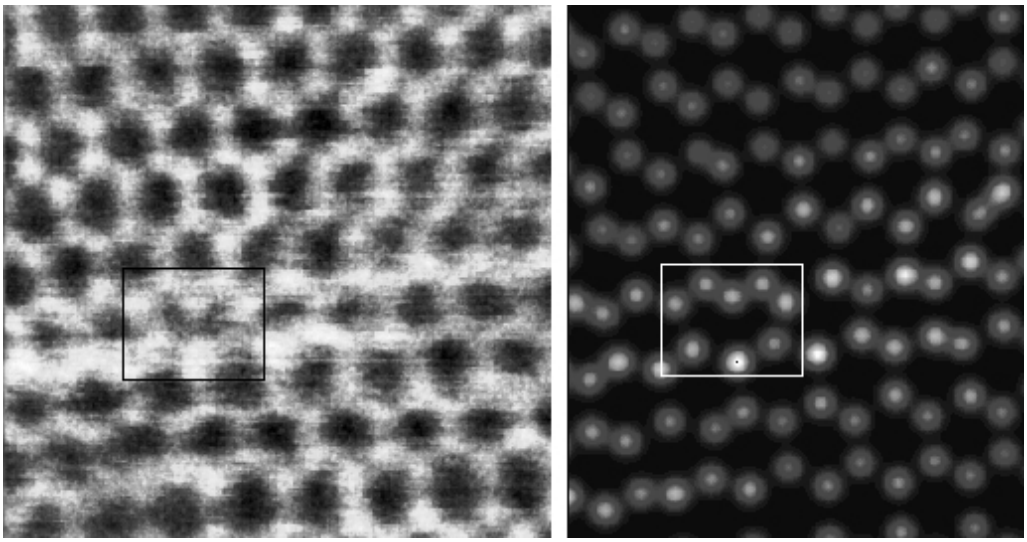

Figure 7.9: *Left*: High–resolution Z–contrast image of a threading edge dislocation looking down [0001]. The bright dots are atomic columns of alternating Ga and N atoms. The dislocation core is shown in the boxed region. *Right*: Maximum entropy image showing most probable column positions. The distance between the column of three–fold coordinated atoms and the columns on the left [and right] is found to be by 15 ± 10 % contracted. The distance between the column below the three–fold coordinated atoms and the neighbouring columns is found to be by 15 ± 10 % stretched. These results are consistent with our calculations. Y. Xin *et al.* [93].

number of three fold coordinated atoms than the open core screw dislocation as well as a smaller elastic strain energy arising from the smaller Burgers vector. This last energy is proportional to $k \cdot b^2$. Here, b is the magnitude of the Burgers vector and the constant k is equal to 1 for the screw dislocation, and $\frac{1}{1-\nu}$ for the edge dislocations, where ν is Poisson's ratio (0.37 for GaN [97]). Thus the ratio of the elastic energies is E_{screw}/E_{edge} which is approximately 1.66. Our calculations give the ratio of the line energies, which includes the core energies, to be 2.08. This could explain why threading edge dislocations occur at a higher density than threading screw dislocations.

In analogy to the open–core screw dislocations we have investigated whether the energy of the threading edge dislocation could be lowered by removing the most distorted core atoms (see Fig. 7.8). However, removal of either the columns of atoms (9,10), or the columns (1,2), (3,4), (5,6), (7,8) and their equivalents on the right, leads to considerably higher line energies. This implies that, in contrast with screw dislocations, which as discussed above can exist with a variety of cores, the threading edge dislocations should exist with a full core.

7.4 Deep acceptors trapped at threading edge dislocations: $\mathbf{V}_{\mathrm{Ga}}$ and $\mathbf{V}_{\mathrm{Ga}}$–$(\mathbf{O}_{\mathrm{N}})_n$

In the previous section we showed that in the defect free form the threading edge dislocation has a band gap free from deep lying states, hence implying that the pure dislocation cannot be responsible for the yellow luminescence detected in n–type GaN. However, as can be seen from Fig. 7.8 and table 7.4 the core atoms adopt a very particular geometry with atoms 1 and 2 being three–fold coordinated and atoms 9 and 10 having very stretched bonds with bond–lengths ranging from 2.0 to 2.2 Å. This geometry differs considerably from a position in bulk–like material and thus gives rise to a stress field which could act as a trap for intrinsic defects and impurities. Gallium vacancies ($\mathrm{V_{Ga}}$) have been detected by positron annihilation studies in bulk GaN and their concentration was found to be related to the intensity of the YL [10]. The relevant transition level in n–type GaN is at the centre of the YL spectrum ($E^{2-/3-} \approx 1.1$ eV referenced to the top of the valence band [52]). As a triple acceptor the gallium vacancy is three–fold negatively charged in n–type GaN and can attract up to three positively charged donors. Recent experimental [11, 98, 99] and theoretical [100] works suggest that oxygen at a nitrogen site (O_N) is the main cause of unintentional n-type conductivity in GaN. $\mathrm{V_{Ga}}$ forms defect complexes with O_N which sits as a next neighbour of $\mathrm{V_{Ga}}$ to reduces the Coulomb energy [52, 9]. $\mathrm{V_{Ga}}$ related defect complexes in GaN were found to have electrical properties dominated by the Ga vacancy [52], i.e. they are acceptors and exhibit gap states above the top of the valence band arising from the N dangling bonds surrounding $\mathrm{V_{Ga}}$. Furthermore, Youngman and Harris [101] studied the violet luminescence (VL) in AlN, which is believed to have essentially the same origin as the YL in GaN [9]. They found the VL in AlN to be correlated with the oxygen incorporation and extended defects which are also known to contain substantial amounts of oxygen [102]. Hence, in analogy to the VL in AlN it has been suggested that the YL in n–type GaN is caused by O related defect complexes.

7.4.1 Benchmark calculations for V_{Ga}, O_N and V_{Ga}–(O_N) in bulk material

In bulk material the $(V_{Ga}–O_N)^{2-}$ defect complex (see Fig. 7.10) as well as its constituents, V_{Ga}^{3-} and O_N^{+} have previously been investigated by Neugebauer *et al.* [52, 8] and Mattila *et al.* [9] using plane wave SCF–LDA methods. As a benchmark they are now investigated by the *SCC–DFTB*

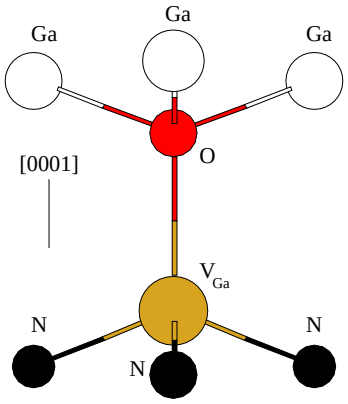

Figure 7.10: Schematic view of the V_{Ga}–O defect complex. Substituting further three–fold coordinated N by O leads to V_{Ga}–$(O_N)_2$ and V_{Ga}–$(O_N)_3$.

method where the defects are modelled in 128 atom wurtzite supercells using two k-points to sample the Brillouin zone (see appendix D). As in references [52, 8, 9], formation energies are evaluated assuming Ga–rich growth conditions, which are common in many growth techniques, O in equilibrium with Ga_2O_3, corresponding to an upper limit for the O concentration [8], and n–type material, i.e. the Fermi level is pinned close to the conduction band minimum. The atomic geometry of the triply charged Ga vacancy is characterised by a strong outward relaxation of the surrounding N atoms. The three equivalent N atoms relax by 10.2% (11.8% in ref. [52]) outwards and the remaining N atom moves 9.5% in [0001] (10.6% in ref. [52]). The formation energy is low (1.6 eV in this work, $\approx$ 1.3 eV in ref. [52] and $\approx$ 1.5 eV in ref. [9]). Oxygen on a nitrogen site has slightly larger Ga–O bonds than the Ga–N bond length in bulk GaN (1.95 Å). We obtain again a low formation energy of 1.7 eV ($\approx$ 1.7 eV in ref. [52] and $\approx$ 1.6 eV in ref. [9]). Bringing V_{Ga}^{3-} and O_N^{+} together, one gets $(V_{Ga}–O_N)^{2-}$. We find the distance between the vacancy core and the O (N) increased by 13.5% (8.9%) which is close to the values of 14.9% (9.8%) given by Neugebauer *et al.* [52]. Furthermore, we determined the energy ΔE for the reaction,

$$(V_{Ga} - O_N)^{2-} \rightarrow V_{Ga}^{3-} + O_N^{+},$$

to be 2.2 eV in good agreement with the plane wave methods (1.8 eV in ref. [52] and $\approx$ 2.1 eV in ref. [9]). We thus get an absolute formation energy of $\sim$ 1.1 eV which is again very close to the plane wave values ($\sim$ 1.1 eV in ref. [52] and $\sim$ 0.9 eV in ref. [9]) implying a high equilibrium concentration of $\sim 10^{18}/\mathrm{cm}^3$ [52, 9] at a usual MOCVD growth temperature of $\sim$ 1300 K.

The good agreement of the *SCC–DFTB* method for V_{Ga}^{3-}, O_N^{+} and $(V_{Ga}–O_N)^{2-}$ with SCF–LDA plane wave calculations suggests that *SCC–DFTB* allows a valid description of oxygen in GaN.

Table 7.5: Formation energies in eV of V_{Ga}^{3-}, O_N^+, $(V_{Ga}–O_N)^{2-}$, $(V_{Ga}–(O_N)_2)^{1-}$ and $V_{Ga}–(O_N)_3$ in a 128 atom bulk cell and at the threading edge dislocation (see Fig. 7.8 and D.2). Ga–rich growth conditions, O in equilibrium with Ga_2O_3 and n–type material are assumed.

position	$E(V_{Ga}^{3-})$	$E(O_N^+)$	$E(V_{Ga}–O_N)^{2-}$	$E(V_{Ga}–O_{2N})^{-}$	$E(V_{Ga}–O_{3N})$
bulk cell	1.8	1.7	1.1	0.7	0.8
pos. L (bulk–like)	1.7	1.5	1.0	0.9	0.7
pos. (1,2) (core)	-0.2	0.2	-2.3	-2.5	-3.0
pos. (5,6)	0.3	1.0	-1.0	-1.0	-0.8
pos. (9,10)	-0.3	1.3	-0.6	-0.3	-0.3

7.4.2 Properties of V_{Ga}, O_N and $V_{Ga}–(O_N)_n$ $(n = 1, 2, 3)$ in the stress field of the threading edge dislocation

In the following, the geometries, electrical properties and formation energies of V_{Ga}, O_N and $V_{Ga}–(O_N)_n$ $(n = 1, 2, 3)$ are investigated in the stress field of threading edge dislocations [91]. In the *SCC–DFTB* case the dislocations are modelled in a 312 atom supercell containing a dislocation dipole (see Fig. D.2). In order to reduce the interaction between the point defects we doubled the 312 atom supercell along the dislocation line, i.e. in [0001], to obtain a 624 atom supercell. In the AIMPRO case, we used 286 atom stoichiometric clusters with one dislocation.

Firstly we place the point defects into a bulk–like position, i.e. a position with a very small stress field, far away from the dislocation core in the supercell (position L in Fig. D.2). At position L in this cell we find the atomic geometries and formation energies of the point defects to be in good agreement with the values obtained in the 128 atom perfect lattice supercell (see first two lines in Table 7.5). We now put the defects at different positions (column (1/2), (5/6), (9/10)) in Fig. 7.8) in the dislocation stress field and evaluate the formation energies and electrical properties. As will be seen, some of the formation energies are negative suggesting that under equilibrium conditions the corresponding position would certainly be adopted by the defect. However, since gallium vacancies and oxygen are not necessarily in equilibrium with the dislocation stress field, the precise concentration of defect complexes in the dislocation stress field depends on the history of the sample.

V_{Ga}^{3-} is trapped in the dislocation stress field, in particular, at the dislocation core (pos. 1 in Fig. 7.8) and at pos. 9. Ga atoms in these positions would have high energies, caused by the under–coordination in pos. 1 or by the strongly strained bonds in pos. 9 (2.11 Å average bond length). This makes the formation of vacancies at these positions energetically favourable (see Table 7.5). It should be noted that at pos. 1 a Ga vacancy creates a two–fold coordinated N atom at pos. 2 which would result in a high energy. However, since Ga atoms at pos. 7 and its equivalent at the right are quite close to the N atom at pos. 2, this N atom forms a bond (2.00 Å) with one of these Ga atoms and thus achieves three–fold coordination. The new configuration has a distorted core and looks like a first step of a kink formation. This suggests that V_{Ga} play an important role in the dislocation motion.

Oxygen atoms sit preferentially two or three–fold coordinated. This explains why O_N^+ is by 1.3 eV more stable at the dislocation core (pos. 2) where it replaces a three–fold coordinated N atom than in a bulk–like region (pos. 0) where it is four–fold coordinated (see Table 7.5).

The high stabilities of V_{Ga}^{3-} and O_N^+ at the dislocation core imply also a very low formation energy for $(V_{Ga}–O_N)^{2-}$ (-3.3 eV below the energy for pos. L) and hence a high concentration. Here O sits two–fold coordinated in a bridge position with very strong Ga–O bonds (1.72 Å). Due to these strong bonds and the high complex binding energy of 2.3 eV at the dislocation core we expect $(V_{Ga}–O_N)^{2-}$ to be immobile. Finally, we investigated $(V_{Ga}–(O_N)_2)^-$ and $V_{Ga}–(O_N)_3$, which in analogy to $(V_{Ga}–O_N)^{2-}$ are found to be particularly stable at the core of the threading edge dislocations where they are likely to be immobile. See tables 7.5 for the detailed formation energies. All these results suggest that $(V_{Ga}–(O_N)_n)^{(3-n)-}$ $(n = 1, 2, 3)$ defect complexes increase the oxygen concentration near to threading edge dislocations and in particular at the dislocation core. Threading edge dislocations may therefore be used as a trap for undesired impurities. This has been suggested by Nakamura *et al.* [103] who proposed that during the initial stages of GaN growth threading edge dislocations should be permitted to clean the sample from impurities which emerge from the substrate. In a following step, a very thin SiO mask is then used to reduce the number of threading edge dislocations in the following region of the epilayer which will be used as the active region of the devices.

Concerning the electrical properties the *SCC-DFTB* calculations reveal that at bulk positions V_{Ga}^{3-}, $(V_{Ga}–O_N)^{2-}$, $(V_{Ga}–(O_N)_2)^-$ and $V_{Ga}–(O_N)_3$ defects are deep acceptors with gap states $\sim 1.0 - 1.2$ eV above VBM. In order to obtain information about the contribution of these defects to the YL we then calculated the difference of the formation energies depending on the charge states relevant to the transition in n–type material. The results referenced to VBM are given in Table 7.6. Subtracting them from the band gap ($\sim$ 3.4 eV) gives an estimate for the transition energies in n–type material. Since the energies for the different charge states are derived from total energies associated with fully relaxed atomic configuration, the calculated energy differences correspond to zero–phonon transitions. As can be seen, at a variety of positions the defects could contribute to the yellow luminescence. It is interesting to note that in a bulk–like position $V_{Ga}–(O_N)_3$ has a deep gap state ($\sim$ 1.0 eV above VBM) which comes from a three–fold coordinated nitrogen atom in a bulk position. At the dislocation core (col. (1/2)), however, for $V_{Ga}–(O_N)_3$ all three–fold coordinated nitrogens surrounding the Ga vacancy are replaced by oxygen. At this position $V_{Ga}–(O_N)_3$ adopts the same configuration as at the $(10\bar{1}0)$ surface (see chapter 6) and does not induce deep states in the band gap.

Table 7.6: Transition energies of V_{Ga}, $V_{Ga}–O_N$, $V_{Ga}–(O_N)_2$ and $V_{Ga}–(O_N)_3$ at the threading edge dislocation (see Fig. 7.8 and D.2) referenced to VBM.

position	$(V_{Ga})^{2-/3-}$	$(V_{Ga}–O_N)^{1-/2-}$	$(V_{Ga}–O_{2N})^{0/1-}$	$(V_{Ga}–O_{3N})^{1+/0}$
pos. L (bulk–like)	1.4	1.0	0.7	0.9
pos. (1/2) (core)	0.8	1.0	0.7	0.4
pos. (5/6)	0.8	1.4	1.0	0.9
pos. (9/10)	0.4	0.3	0.6	0.8

Summary

Line defects threading along the c–axis have been explored. We found full core screw dislocations to have a large distortion of the bonds at the dislocation core resulting in deep states in the band gap. Open core screw dislocations and in particular threading edge dislocations, which occur at very high densities, have a core structure similar to $(10\bar{1}0)$ surfaces and are therefore electrically inactive in their pure, i.e. impurity–free form. This and the fact that threading dislocations do not lie on the basal glide plane makes movement and the generation of large numbers of point defects difficult. In contrast, dislocations in GaAs glide and climb easily through recombination–enhanced mechanisms. This motion generates efficient radiative recombination centres which degrade optical emissions.

Oxygen–related defect complexes, some of which are electrically active, are found to possess very low formation energies at the core of threading edge dislocations. One specific oxygen related defect complex, V_{Ga}–$(O_N)_3$ is believed to be responsible for the formation of nanopipes.

Chapter 8

Domain Boundaries

In addition to the threading line defects discussed in the previous chapter, also planar defects thread along the c–axis in GaN and may influence the electrical properties of the devices. Fig. 8.1 shows a transmission electron microscopy image along [0001] of a typical sample grown by MBE on a GaP (111) substrate. Two kinds of threading planar defects, called "domain boundaries" (DB) can be distinguished [104, 105, 106, 107, 108, 109]. They lay on $\{11\bar{2}0\}$ and $\{10\bar{1}0\}$ planes and following Xin *et al.* [109] are denoted by DB–I and DB–II respectively. Domain boundaries are either described in terms of a double position boundary (DPB) [otherwise termed a stacking mismatch boundary (SMB)] consisting of a different stacking sequence across the boundary, or an inversion domain boundary (IDB) which is characterised by a polarity inversion across the boundary.

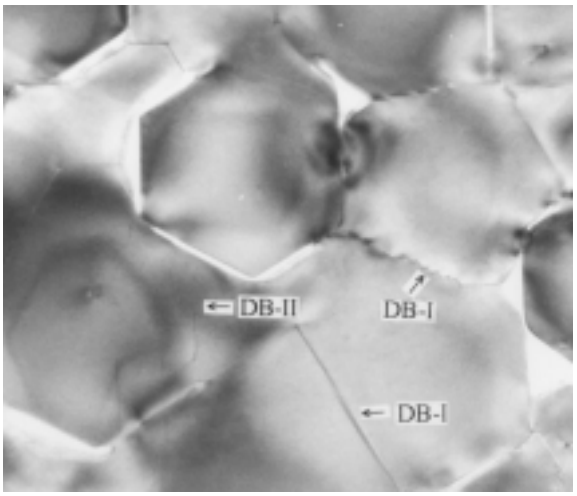

Figure 8.1: Plan view TEM bright field image of epitaxial wurtzite GaN of the region close to the growth surface. Domain boundaries on $\{1\bar{2}10\}$ planes (DB–I) and on $\{10\bar{1}0\}$ planes (DB–II) are visible. In the faceted DB–I boundary, the faceted segments are on $\{1\bar{2}10\}$. Y. Xin *et al.* [109].

Domain boundaries on $\{10\bar{1}0\}$ planes have been extensively explored experimentally and theoret-

ically. The results are summarised in the section below. For domain boundaries of the DB–I type structural models have been proposed based on high resolution transmission electron microscopy (HRTEM) studies by Xin *et al.* [109] and Rouvière *et al.* [108]. However, no theoretical investigations for the energetics and electrical properties of these models have been reported, presumably because of the larger supercells required to model domain boundaries terminating in $\{11\bar{2}0\}$ planes. To fill this gap we investigate DB–I domain boundaries with the *SCC–DFTB* method in this work.

8.1 Brief review of domain boundaries on $\{10\bar{1}0\}$ planes

Domain boundaries of type DB–II have been explored extensively using transmission electron microscopy (TEM) [104, 105, 106, 107].

Total energy calculations by Northrup *et al.* [110] show that an inversion domain boundary involving a $c/2$ translation along the $\langle 0001\rangle$ direction has a very low domain wall energy and is thus a suitable candidate for many of the vertical defects observed on $\{10\bar{1}0\}$ planes. At this shifted inversion domain boundary denoted by IDB*–II all atoms remain fourfold coordinated with Ga–N bonds across the boundary and therefore do not induce electronic states in the band gap. Domain boundaries of the IDB*–type originate at the substrate interface and thread along the whole epilayer since it would be energetically very expensive to terminate them by overgrowth.

Furthermore Northrup *et al.* [110] investigated a double position boundary (DPB–II). DPB–II could account for those domain boundaries on $\{10\bar{1}0\}$ planes for which no inversion of polarity across the boundary is observed [109]. Across the boundary DPB–II would have three–fold coordinated Ga and N atoms both in sp^2 hybridisations which gives rise to a deep acceptor state localised at the lone pair of the sp^2 hybridised N atoms.

8.2 Domain boundaries on $\{11\bar{2}0\}$ planes

In contrast to many of the DB–II type boundaries which originate at the epilayer substrate interface the DB–I type boundaries found in a GaN sample grown by MBE on GaP (see Fig. 8.1) extend only a short distance along the c–axis [109]. A high resolution Z–contrast image down [0001] reported by Xin *et al.* [96] shows clearly that DB–I has a *horizontal* displacement of $R_h = 1/2\langle 10\bar{1}0\rangle$ (see Fig. 8.2). This configuration which is called prismatic stacking fault is composed of four– and eight–fold rings along the fault.

In this work [111] DB–I domain boundaries are modelled within 256 atom supercells containing two boundaries and eight layers of atoms between the boundaries (see appendix D.4).

Assuming no additional displacement in the *vertical*, i.e. $\langle 0001\rangle$ direction gives a model for a double position boundary denoted by DPB–I. As can be seen in the side view in Fig. 8.3 DPB–I contains wrong, i.e. Ga–Ga and N–N bonds. Due to the very different bond lengths of both species (~ 2.7 Å in Ga bulk and ~ 1.5 Å in the N_2 molecule) wrong bonds give rise to a high energy and thus reduce the stability of the system. The lowest energy configuration is achieved for a spacing of 2.8 Å between the boundary planes (in the ideal lattice the corresponding distance would be ~ 1.6 Å)

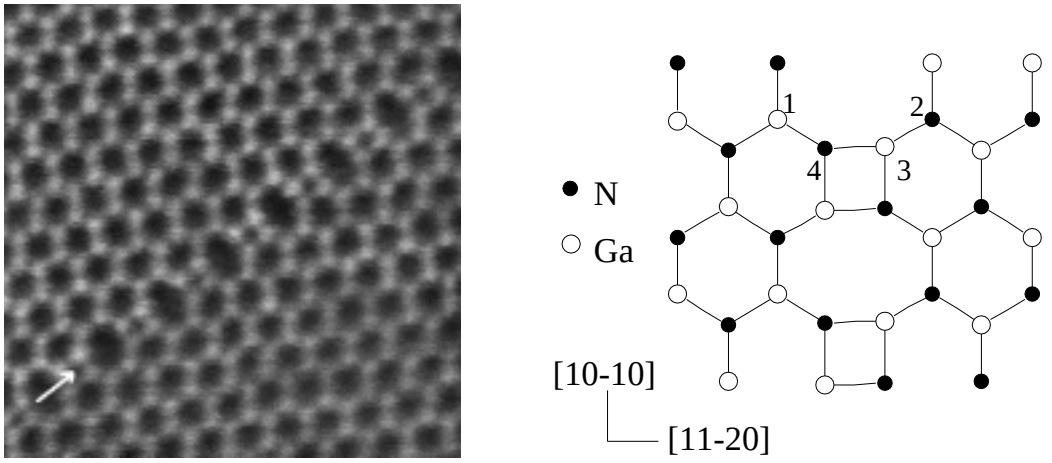

Figure 8.2: Top view along [0001] of a domain boundary of DB–I type, i.e. on $\{11\bar{2}0\}$ planes. From TEM experiments (*left*) the *horizontal* shift across the boundary is found to be $1/2\langle 10\bar{1}0\rangle$. *Right:* Structural model. All models discussed below (DPB–I, DPB*–I and IDB–I) agree with this top view. (Of course, in this figure the bonding across the boundary is arbitrary and varies with the different models). Here and in the following figures atom numbers 1 (2) refer to Ga (N) atoms in eight–fold rings close to the boundary, whereas atom numbers 3 (4) refer to Ga (N) atoms in four–fold rings with bonds across the boundary. Y. Xin *et al.* [96].

which is comparable with the bond length in bulk Ga. Our calculations find a high domain wall energy (see Eq. (5.17)) of $\sigma_{\rm wall} = 246$ meV/Å^2 which is only slightly less than the energy of two free surfaces (256 meV/Å^2). This suggests, that DPB–I should not occur frequently and if it occurs it should exist with different spacings. Indeed, we find that varying the spacing between the boundaries changes the wall energy only slightly since the wrong bonds across the boundary are very weak. We note that at the equilibrium distance of 2.8 Å the structure has shallow occupied N-derived states at ca. 0.2 eV above the valence band maximum (VBM) and unoccupied states at ca. 0.4 eV below the conduction band minimum (CBM). At larger distances the influence of the Ga–Ga bonds across the boundary should vanish so that the electrical properties correspond to free $(11\bar{2}0)$ surfaces which we found to be electrically inactive.

We now examine the structure with an additional vertical displacement of $1/2\langle 0001\rangle$ giving a total displacement of $1/2\langle 10\bar{1}1\rangle$ as derived from TEM by Xin *et al.* [109]. In this double position boundary denoted by DPB*–I all atoms along the boundary are four–fold coordinated and form Ga–N bonds across the boundary (see Fig. 8.3). Since Ga–N bonds are very strong DPB*–I has a clearly defined spacing of ~ 1.90 Å between the $\{11\bar{2}0\}$ planes at the boundary. The calculated domain wall energy of 99 meV/Å^2 is significantly lower than the energy of the unshifted DPB–I model suggesting that DPB*–I is a promising candidate for domain boundaries in $\{11\bar{2}0\}$ planes for which no polarity inversion across the boundary has been observed [109]. DPB*–I are thought to be associated with single growth faults in the basal plane [112, 109]: DPB*–I starts and ends with a basal plane stacking fault. Since these basal plane stacking faults have a low energy and thus are easily formed during growth, there are many possibilities for DPB*–I to nucleate but also to be overgrown. This explains why DPB*–I are observed throughout the whole epilayer but extend only over a short distance along the c–axis [109].

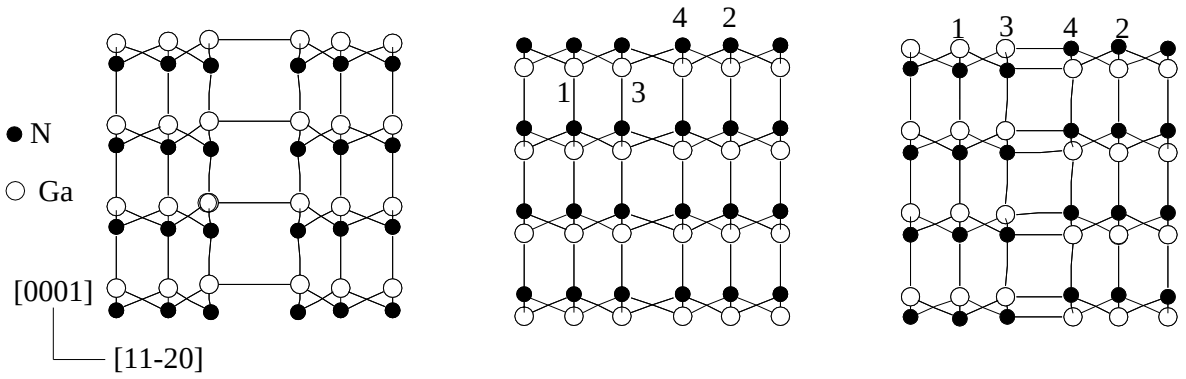

Figure 8.3: Side view along $[10\bar{1}0]$ of the DPB–I (*left*), DPB*–I (*middle*) and IDB–I (*right*) structures which have total displacements of $1/2\langle 10\bar{1}0\rangle$, $1/2\langle 10\bar{1}1\rangle$ and $1/2\langle 10\bar{1}0\rangle$ respectively. In the DPB–I structure wrong bonds yield a high energy which is only slightly less than that of two free $(11\bar{2}0)$ surfaces. In the DPB*–I and IDB–I structures all atoms are four–fold coordinated and exhibit strong Ga–N bonds across the boundary. DPB*–I has the lowest wall energy among all domain boundaries of type DB–I, the energy of IDB–I is slightly higher.

Table 8.1: Bond lengths in Å and bond angles in degree at the DPB*–I domain boundary. Atom numbers refer to Fig. 8.2 and 8.3.

Atom	Bond Lengths (min, max)	Bond Angles (min, max)
1 (Ga)	1.86, 1.95	107.0, 112.6
2 (N)	1.88, 1.96	106.1, 111.5
3 (Ga)	1.86, 2.11	80.6, 130.2
4 (N)	1.88, 2.11	86.3, 127.8

Details of the geometry of DPB*–I can be found in Table 8.1. As can be seen, some of the bonds are quite distorted which makes that DPB*–I induces shallow electronic states ~ 0.35 eV above VBM in the band gap. However, these states are not deep enough to be responsible for the yellow luminescence which is centred at ≈ 2.2 eV and observed in n–type GaN. On the other hand point defects may segregate to the DPB*–I boundary and change the electrical properties. As stated in section 7.4 gallium vacancies were experimentally found to be related to the intensity of the YL [10], a fact which is also supported by theoretical calculations [52, 9]. We therefore evaluated the formation energy of V_{Ga}^{3-} at the domain boundary and found it to be lower by 1.1 eV at pos. 3 with respect to a position in a bulk–like environment. The electronic properties of V_{Ga} at the DPB*–I were found to be similar to V_{Ga} at a perfect lattice position with deep acceptor states ≈ 1.1 eV above VBM and $E^{2-/3-} \approx 1.6$ eV with respect to VBM (in a bulk–like position we found $E^{2-/3-} \approx 1.4$ eV). This suggests that if Ga vacancies diffuse easily in GaN a lot of them will be trapped at DPB*–I where they would introduce deep acceptor states and can act as electron traps, in agreement with recent electron energy loss spectroscopy (EELS) measurements by Natusch *et al.* [113].

Table 8.2: Bond lengths in Å and bond angles in degree at the IDB–I domain boundary. Atom numbers refer to Fig. 8.2 and 8.3.

Atom	Bond Lengths (min, max)	Bond Angles (min, max)
1 (Ga)	1.88, 1.95	105.4, 112.4
2 (N)	1.87, 1.95	103.9, 111.7
3 (Ga)	1.87, 2.04	87.6, 142.3
4 (N)	1.87, 2.04	91.6, 141.0

A model for an inversion domain boundary on $\{11\bar{2}0\}$ planes (IDB–I) has been suggested by Rouvière *et al.* [108]. It has a total displacement of $1/2\langle 10\bar{1}0\rangle$ (see Fig. 8.3) and again four fold coordinated atoms with Ga–N bonds across the boundary yielding a spacing of ~ 2.0 Å between the boundary planes. Features of the geometry are listed in Table 8.2. The domain wall energy for IDB–I of 122 meV/Å^2 is slightly above the wall energy for DPB*–I. This can be understood by analysing the structural properties. At DPB*–I each of the boundary atoms (No. 3 and 4 in Fig. 8.2) has four bond angles near to the ideal sp^3 value of 109.3°. Only two angles at each atom deviate considerably ($\sim 80^o$ and $\sim 130^o$). At IDB–I only three angles at each boundary atom are near to the ideal value whereas each atom has two angles of $\sim 90^o$ and one angle as large as $\sim 140^o$. The bond angles are significantly more distorted at IDB–I compared to DPB*–I. This explains the higher domain wall energy found for IDB–I. In spite of the considerable distortion also IDB–I has only shallow gap states ~ 0.3 eV above VBM. It is worth noting that in contrast to DPB*–I which can be terminated by a low energy basal plane stacking fault, a mechanism to end IDB–I will be energetically much more costly. Therefore, domain boundaries of type IDB–I should thread over a long distance along the c–axis.

Summary

In summary, our calculations for structural models of domain boundaries in $\{11\bar{2}0\}$ planes reveal that in analogy to domain boundaries on $\{10\bar{1}0\}$ planes only structures which have Ga–N bonds across the boundary have low formation energies. The model with the lowest domain wall energy has a total displacement of $1/2\langle 10\bar{1}1\rangle$ which is in agreement with recent transmission electron experiments [109]. This boundary does not induce deep states in the band gap. However, gallium vacancies which are a common point defect in GaN could segregate to the domain boundary and adversely influence the electrical properties.

Chapter 9

Reconstructions of Ga, N, H and O terminated (0001)/(000$\bar{1}$) surfaces

The knowledge of the surface properties, in particular the type of reconstruction observed, can be important to produce high quality material. Indeed, the growth of material with specific properties is often related to the observation of a specific surface reconstruction pattern during the growth process. Also for an understanding of the growth mechanism it is essential to know the properties of the energetically favourable surface reconstructions. Upon these one can then simulate the diffusion of atoms (in particular Ga atoms which in GaN are the rate limiting species) and suggest improvements in the growth technique. Finally, many semiconducting devices depend crucially upon phenomena that occur at a surface or interface. Often an electrically and chemically inert surface is desired prior to device fabrication.

In this chapter the reconstructions of the main growth surfaces in wurtzite GaN, the (0001) and (000$\bar{1}$) surfaces, are explored. These surfaces are polar, i.e. they lie in planes which are characterised by an unequal number of cations and anions. The polar direction in which a crystal is grown is also called the polarity. We consider intrinsic, i.e. Ga and N terminated structures which account for surfaces observed during MBE growth. The stabilities of the reconstructions depending on the growth conditions expressed via the chemical potentials are discussed. From these results it is possible to determine the polarity of the material during growth. We also investigate the possible adsorption of H which could occur during the MOCVD growth process. Finally, we study the adsorption of oxygen, which with a size similar to that of nitrogen is a promising candidate for surface passivation, on top of the most common reconstructions.

9.1 Reconstructions of Ga, N and H terminated surfaces

The most common techniques for growing device quality wurtzite GaN are molecular beam epitaxy (MBE) or metal-organic chemical vapour phase deposition (MOCVD). In MBE, the growth temperature ranges from 600-800^o C and very little hydrogen is present during growth. MOCVD growth requires a temperature of $\approx 1000^o$ C and a substantial amount of H is present in the precursors. Thus the surface characteristics of GaN epilayers during growth may strongly depend on

the employed technique. The common growth direction is normal to the hexagonal {0001} basal plane which exhibits a polar configuration with two atomic sub–planes each consisting of either the cationic or the anionic element of the binary compound. Hence, the ideal GaN basal plane surface is either Ga– or N– terminated. Such polar surfaces are expected to have very different characteristics. Recently Ponce *et al.* [73] studied the polarity of MOCVD grown GaN and found that films grown in the (0001), Ga terminated, direction exhibit smooth facets, whereas rough facets also found correspond to (000$\bar{1}$), N terminated, planes. Also Sung *et al.* [114] report that MOCVD growth in the (000$\bar{1}$) direction results in rather rough surfaces with (1×1) periodicity.

Real–time monitoring has been very useful for the study of the crystal growth process. Usually, reflection high–energy electron diffraction (RHEED) is used in MBE for real–time monitoring of the growth process. Unfortunately in MOCVD more sophisticated techniques have to be developed since RHEED requires ultra high vacuum. RHEED experiments [115, 116, 117] observed transitions between (1×1), (2×2) and (4×4) reconstructions during MBE growth and cooling of GaN on sapphire. In these works the growth direction and thus surface polarity is not reported. Knowledge of the atomic positions and the corresponding absolute surface energies of the most stable surface reconstructions may help to establish a relation between the observed RHEED pattern and the growth direction and furthermore give a guidance for controlling the epitaxial growth process. Ga and N terminated surfaces have been investigated by Smith *et al.* experimentally using a variety of techniques, in particular STM [118] and theoretically with the SCF–LDA plane wave method [118]. In a recent work Smith *et al.* [119] related the surface periodicities observed by RHEED during MBE growth to the lattice polarity. As shown below the conclusions of this experimental work are in agreement with our *SCC–DFTB* calculations [120].

Schematic illustration of the surfaces

We examine the geometries and stabilities of ideal surfaces, models containing of a Ga mono– and a Ga bilayer and hydrogen terminated surfaces (see Fig. 9.1) to provide possible candidates for the observed (1×1) RHEED pattern.

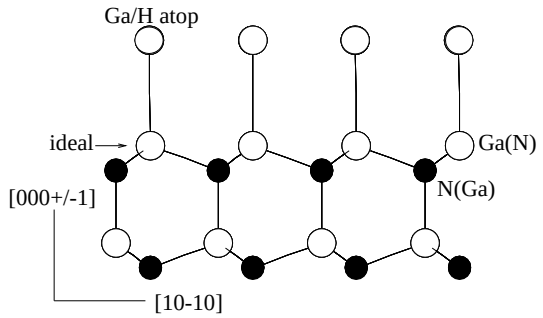

Figure 9.1: Side view of the Ga/H monolayer structure. Empty (filled) circles represent Ga (N) atoms for the (0001) surface and N (Ga) atoms for the (000$\bar{1}$). The (000$\bar{1}$) Ga bilayer structure is obtained by removing the Ga adlayer and changing the top layer N atoms into Ga atoms. In H terminated structures some of the adlayer Ga atoms are replaced by H.

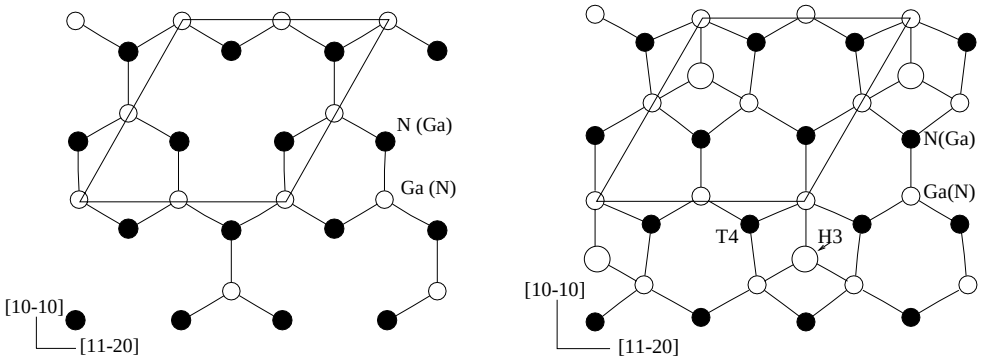

Figure 9.2: Top view of the (2×2) vacancy (*left*) and adatom (*right*) structures at the (0001) and $(000\overline{1})$ surfaces. Empty (filled) circles represent Ga (N) atoms for the (0001) surface and N (Ga) atoms for the $(000\overline{1})$ surface. The triangle structure can be obtained by replacing the adatom at the adatom structure by a trimer arranged in a unilateral triangle.

Following previous theoretical studies for (2×2) periodicities at AlN $(0001)/(000\overline{1})$ [121] and GaAs $(111)/(\bar{1}\bar{1}\bar{1})$ surfaces [53, 36] we have investigated the geometries and stabilities of the vacancy, adatom and trimer induced reconstructions (see Fig. 9.2) which satisfy the electron counting rule.

The (0001) Surface

MBE growth: Ga and N surface terminations

The absolute surface energies depending on the gallium chemical potential (see equation (5.5)) for the examined Ga and N covered reconstructions are shown in Fig. 9.3. Under Ga rich growth conditions the (1×1) reconstruction consisting of a Ga monolayer added in the atop configuration (Ga atop) has the lowest surface energy. This is in apparent contradiction with the recent works of Smith [118] who found the Ga atop structure to be unstable in any environment. We believe, however, that this difference can be explained by the different symmetry constraints employed in the calculations. By restricting the Ga atop structure to a small unit cell and therefore symmetry Smith *et al.* did not allow a Peierls distortion which often occurs at metallic surfaces. Such a distortion would reduce the symmetry (denoted $n \times m$), makes the surface semiconducting and thus lowers the energy. Since each Ga atom in the adlayer contributes 1.25 electrons to fill the bond with the Ga atom in the layer below, to achieve a semiconducting band structure the $(n \times m)$ reconstruction must satisfy the condition $nm = (8, 16, 24, ...)$. In our work we evaluated the surface energy within a (4×4) unitcell thus allowing a Peierls distortion which indeed reduces significantly the surface energy (0.31 eV/(1×1) cell $\sim$ 40 meV/Å^2) and therefore gives a stable surface under a Ga–rich environment [120]. In Fig. 9.3 the energy is therefore drawn for the distorted (1×1) Ga atop structure. Although the distortion breaks the (1×1) symmetry, the Peierls distortion does not exhibit any regularity within the (4×4) cell and is probably only weakly correlated between neighbouring (4×4) cells, i.e. it is likely to vary all over the surface. The Ga atop structure might therefore be a candidate for the (1×1) RHEED pattern observed in a Ga–rich environment [116, 117]. The disorder arising from the Peierls distortion would furthermore explain the fact that the RHEED images in Smith

et al [119] indicate that the reconstruction is not a clear (1×1) (for this reason Smith *et al.* denote it with "(1×1)"). Going further towards N–rich growth conditions we find that the Ga adatom model in a T4 position becomes favoured. In an extreme N–rich environment the stoichiometric (2×2) models, i.e. the Ga vacancy model and the N adatom model, could be stable having nearly degenerate energies. In the N adatom model due to its negative charge state the N adatom resides in the H3 position in order to reduce the electrostatic interaction with the third layer N atom. All other examined structures are unstable in any environment. Some of them have very high energies and are not drawn in Fig. 9.3. In particular, the N trimer structure has an extremely high surface energy, since the short N–N bonds in the trimer can only be achieved by a strong distortion of the underlying GaN surface layer which is energetically very expensive. This is in striking contrast to the corresponding GaAs (111) surface where the anion (As) trimer is very stable under As–rich growth conditions.

These results for the Ga and N covered (0001) surface suggest that if MBE growth proceeds in the (0001) direction, varying growth conditions should yield changes in the periodicity from (1×1) to (2×2) in the observed RHEED patterns. Indeed the diagram in Fig. 9.3 could explain the transitions observed during MBE growth [116, 117] where a (1×1) periodicity changes to a (2×2) periodicity if the Ga flux is lowered and/or the temperature is increased. The results are also in very good agreement with the recent results of Smith *et al.* [119] who find a "(1×1)" RHEED pattern in Ga–rich environment and a (2×2) pattern under an N–rich environment. Moreover they report (6×4) and (5×5) pattern in the intermediate range.

MOCVD growth: H adsorption at the surface

During MOCVD growth a large amount of H is present and may be adsorbed at the surface. Fig. 9.4 shows the energies of the ideal surface covered with 50% and 75% H together with the most stable Ga and N terminated surfaces from Fig. 9.3. We see that under typical MOCVD growth conditions the 75% and the 50% H–terminated surfaces have similar energies and are stable under N–rich environment. The 100% H coverage has a very high surface energy. We can therefore suggest that a 50-75% H coverage is bound to the surface under N–rich growth conditions. This agrees with the work of Rapcewicz *et al.* [122] who state that a considerable energy gain is achieved by H adsorption. However, no quantitative value is given in that work and furthermore the hydrogen chemical potential is fixed at $\mu_{\mathrm{H}} = \mu_{\mathrm{H}_2}$ which is not realistic for MOCVD growth temperatures [51]. The 75% H covered surface has one empty Ga dangling bond per (2×2) surface cell and is therefore semiconducting. If the 75% H configuration is ordered, it should be arranged in a (2×2) periodicity. If the sample is then transferred into vacuum this (2×2) periodicity could be distinguished by LEED and RHEED since all Ga atoms bound to H are sp^3 hybridised whereas the Ga atoms with the empty dangling bond adopt a planar sp^2 configuration. However, calculations within a (4×4) supercell show very little correlation between the positions of the H atoms giving an energy difference of less than 0.4 meV/Å^2 between the ordered and disordered positions. This suggests that "on average" a (1×1) periodicity should be observed.

From the energies shown in Fig. 9.4 we may suppose that MOCVD growth in the (0001) direction produces only surfaces with (1×1) periodicity no matter whether Ga or N–rich growth conditions are pursued. It should however be emphasised that, although the ideal surfaces terminated with hydrogen seem to be the most plausible H terminated surfaces, different models for H terminated surfaces, e.g. H on the Ga atop structure, exist which have not yet been examined.

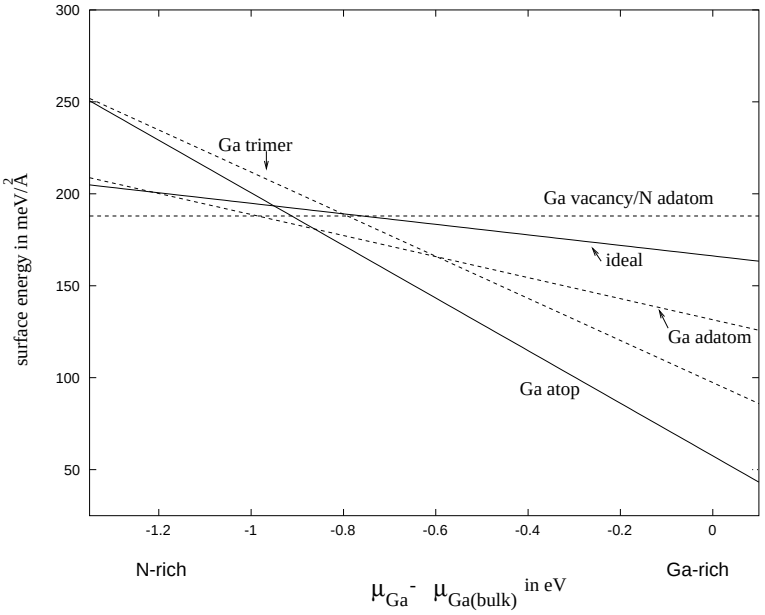

Figure 9.3: Surface energies in meV/Å^2 of Ga and N covered GaN (0001) surfaces plotted versus $\mu_{\mathrm{Ga}} - \mu_{\mathrm{Ga(bulk)}}$. The part on the left (right) of the diagram corresponds to N (Ga) rich growth conditions. This diagram might explain phase transitions observed during MBE growth.

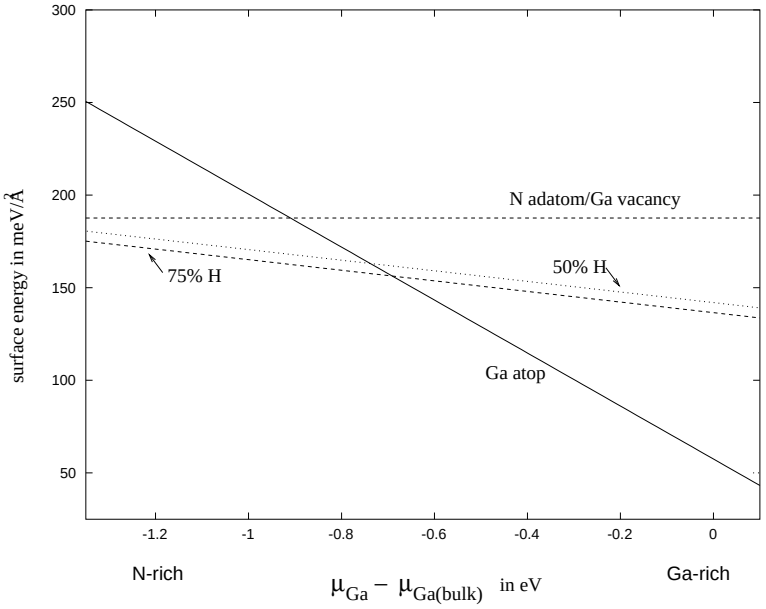

Figure 9.4: Surface energies in meV/Å^2 of the most stable Ga, N and H covered GaN (0001) surfaces plotted versus $\mu_{\mathrm{Ga}} - \mu_{\mathrm{Ga(bulk)}}$. The part on the left (right) of the diagram corresponds to N (Ga) rich growth conditions. This diagram might explain why MOCVD growth produces only surfaces with (1×1) periodicity.

Finally, the bonding configurations and electrical properties of the stable reconstructions at the (0001) surface are listed in Table 9.1. The eigenvalues of the highest (lowest) occupied (unoccupied) molecular orbitals HOMO (LUMO) are to be taken as a guidance for further experimental studies. The calculated *SCC–DFTB* band gap is 6.6 eV. The experimental band gap for bulk GaN is $\approx$ 3.4 eV. The values in brackets are "interpreted values", i.e. values the author believes to be the true ones if a larger basis set is used so that *SCC–DFTB* produces the correct band gap.[1]

Table 9.1: Bonding configuration and electrical properties of the most stable GaN (0001) surface reconstructions. In an $\approx sp^3$ configuration one of the four bonds is highly distorted. HOMO and LUMO are given in eV above the VBM. The calculated *SCC–DFTB* band gap is 6.6 eV. The experimental band gap for bulk GaN is $\approx$ 3.4 eV. In brackets are "interpreted values" (see text).

surface	1./2. layer	Bonding 1./2. layer	HOMO	LUMO
Ga monolayer	Ga/Ga	non–directional	1.0	1.1
Ga adatom	Ga/Ga	$p^3/\approx sp^3$	1.3	2.2
N adatom	N/Ga	$p^3/\approx sp^3$	1.0	3.9 (3.4)
Ga vacancy	Ga/N	sp^2/p^3	0.8	4.0 (3.4)
75% H	H/Ga	s/sp^3	0.0	4.0 (3.4)

The (000$\bar{1}$) Surface

MBE growth: Ga and N surface terminations

Fig. 9.5 shows the absolute surface energies of possible models for (1×1) and (2×2) reconstructions which are Ga and N covered. We see that except for extremely Ga–rich growth conditions where an accumulation of Ga resulting in a bilayer coverage appears to be stable, the Ga monolayer again arranged in the atop configuration is by far the most probable configuration. In analogy to the Ga atop reconstruction at the (0001) surface a calculation within a (4×4) cell shows that a possible distortion is unlikely to be ordered and will therefore result in a (1×1) RHEED pattern. It is worth noting that in the recent experimental work of Smith *et al.* [119] the (1×1) Ga atop structure is observed under N–rich growth conditions and cedes to (3×3), (6×6) and c(6×12) periodicities with even more Ga atoms at the surface in a Ga–rich environment. The fact that a (2×2) reconstruction is never observed agrees very well with our theoretical investigations but contradicts the theoretical studies by Smith *et al.* [118]. In analogy to the (0001) surface in the calculations of Smith *et al.* a Peierls distortion has not been allowed. *SCC–DFTB* calculations show that the energy gained by a Peierls distortion is considerable (0.10 eV/(1×1) cell $\sim$ 12 meV/Å^2). This might explain why the range where the Ga atop structure is stable was found to be larger in our work [120] than in the work of Smith *et al.* who suggest that reconstructions with (2×2) periodicity should be stable under N–rich growth conditions. In our work the ideal surface (not drawn in the diagram) and the examined (2×2) structures are unstable under typical growth

[1] The author believes that sp^2–type Ga derived states do not mix strongly with the conduction band states so that a basis extension will not change their positions significantly. Studies with an extended basis set would, however, be needed to justify this assumption.

conditions. Especially, the N terminated structures have very high surface energies. This is again in contrast to the corresponding GaAs $(\bar{1}\bar{1}\bar{1})$ surface where As terminated surfaces are stable over a wide range of the growth conditions and only in a Ga–rich environment cede to metallic surfaces which are terminated by both Ga and As atoms [66].

We can therefore suggest that if GaN is grown in the $(000\bar{1})$ direction by MBE then a (1×1) periodicity and following the experimental work by Smith *et al.* [119] (3×3), (6×6) and c(6×12) periodicities should be observed. Also from our calculations we exclude (2×2) periodicities at the $(000\bar{1})$ surface.

MOCVD growth: H adsorption at the surface

We finally examine some H terminated $(000\bar{1})$ surfaces. Fig. 9.6 shows the energies of the ideal surface covered with 50% and 75% H together with the most stable surfaces from Fig. 9.5. Under typical MOCVD growth conditions we see that in an N–rich environment a semiconducting 75% H coverage passivating three of the four N dangling bonds per (2×2) cell and leaving one N lone pair should be stable. This model has already been proposed to account for the (1×1) periodicity observed on MOCVD grown GaN $(000\bar{1})$ surfaces [114]. Indeed, a calculation in a (4×4) cell showed very little correlation between the H atoms (< 0.2 meV/Å^2). It is also worth noting that even if the arrangement were ordered the (2×2) pattern would probably not been seen because of the fact that nitrogen atoms which bound to hydrogen sit in sp^3 hybridised configurations and have a geometry very similar to nitrogens with a filled lone pair in a p^3 configuration. This is in contrast to the (0001) surface where a (2×2) pattern arising from 75% hydrogen coverage could in principle be distinguished with RHEED and LEED due to the sp^2 hybridised Ga atoms at the surface.

These results suggest that samples grown by MOCVD in the $(000\bar{1})$ direction effectively show a (1×1) periodicity. The stabilities of more complicated models, in particular with H on the Ga atop reconstruction, have not yet been examined though.

Finally, the bonding configurations and electrical properties of the stable structures at the $(000\bar{1})$ surface are summarised in Table 9.2.

Table 9.2: Bonding configuration and electrical properties of the stable reconstructions at the GaN $(000\bar{1})$ surface (see Table 9.1 and corresponding text).

surface	1./2. layer	Bonding 1./2. layer	HOMO	LUMO
Ga bilayer	Ga/Ga	non–directional	1.6	1.7
Ga monolayer	Ga/Ga	non–directional	0.6	1.2
75% H	H/N	s/sp^3	0.0	6.6 (3.4)

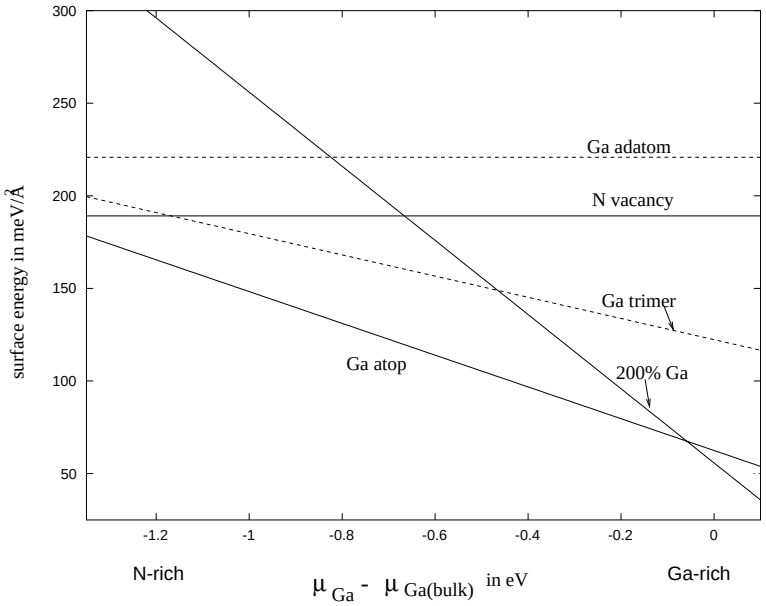

Figure 9.5: Surface energies in meV/Å^2 of Ga and N covered GaN (000$\bar{1}$) surfaces plotted versus $\mu_{\mathrm{Ga}} - \mu_{\mathrm{Ga(bulk)}}$. The part on the left (right) of the diagram corresponds to N (Ga)–rich growth conditions. In contrast to MBE growth in the (0001) direction no (2×2) periodicity seems to be stable.

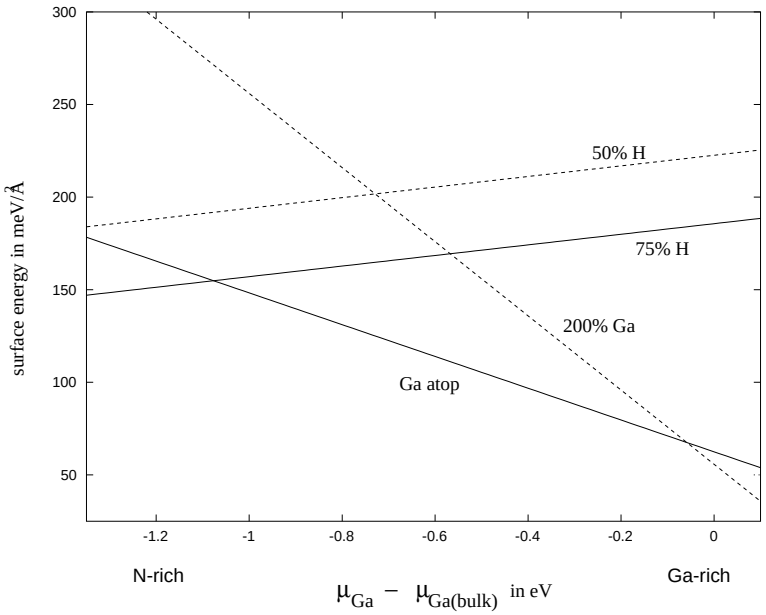

Figure 9.6: Surface energies in meV/Å^2 of the stable Ga, N and H covered GaN (000$\bar{1}$) surfaces plotted versus $\mu_{\mathrm{Ga}} - \mu_{\mathrm{Ga(bulk)}}$. The part on the left (right) of the diagram corresponds to N (Ga) rich growth conditions. In MOCVD growth in both the (0001) and (000$\bar{1}$) direction appears to exclusively result in surfaces with a (1×1) periodicity.

Summary: Polarity determination during MBE growth

In this section we have presented a theoretical study of possible models for the Ga, N and H terminated GaN (0001) and (000$\overline{1}$) surfaces. In particular, absolute surface energies and relative stabilities depending on the growth conditions have been examined. We suggest that GaN grown by MBE in the (0001) direction exhibits transitions between "(1×1)" and (2×2) periodicities. On the other hand, a (1×1) periodicity should be observed during MBE growth in the (000$\overline{1}$) direction over a wide range of growth conditions including the N–rich environment. In an extremely Ga–rich environment more complicated periodicities can occur which have not yet been examined theoretically.

9.2 The chemisorption of oxygen

A detailed knowledge of surface composition and electronic structure is important for understanding the contact formation which in turn is necessary for device fabrication. Often electrically and chemically inactive surfaces are desired. Since the intrinsic surface reconstructions determined in the previous section are Ga terminated and thus not chemically inert, it is of interest to know how common adsorbates react with the surface. In particular group VI elements have been successfully employed to passivate III–V surfaces, e.g. S at GaAs. For surfaces of group III–nitrides, O seems a promising candidate since its size is comparable to nitrogen. Experimentally it has been found that surface oxides are predominantly in the Ga_2O_3 form [123]. Cleaning mechanisms for GaN surfaces from surface oxides are described in [106].

A detailed experimental study of oxygen chemisorption on the GaN (0001) surface has been carried out by Bermudez [124]. Exposing the clean (1×1) surface to an excited O_2 flux he reports that chemisorption of O at the surface takes place. Although low energy electron diffraction (LEED) experiments indicate an ordered adsorbate layer, and x–ray photoemission spectroscopy (XPS) suggests a single chemically distinct adsorption site, a model for the chemisorbed layer could not be defined. In particular, the type of the observed bonding, i.e. Ga–O or N–O bonding could not be determined, and also the polarity of the underlying GaN surface, i.e. (0001) or (000$\overline{1}$) remained unclear.

In this work [125] we use *SCC–DFTB* to examine a variety of possible adsorption places for oxygen on top of the most stable reconstructions at the (0001) and (000$\overline{1}$) surfaces.

The (0001) Surface

The surface energies according to (5.7) of some of the examined models are shown in Fig. 9.7. Here we choose the oxygen chemical potential as the free variable in (5.7) and fix the gallium chemical potential at $\mu_{\mathrm{Ga}} = \mu_{\mathrm{Ga(bulk)}}$. The range of μ_{O} is then given by (5.11).

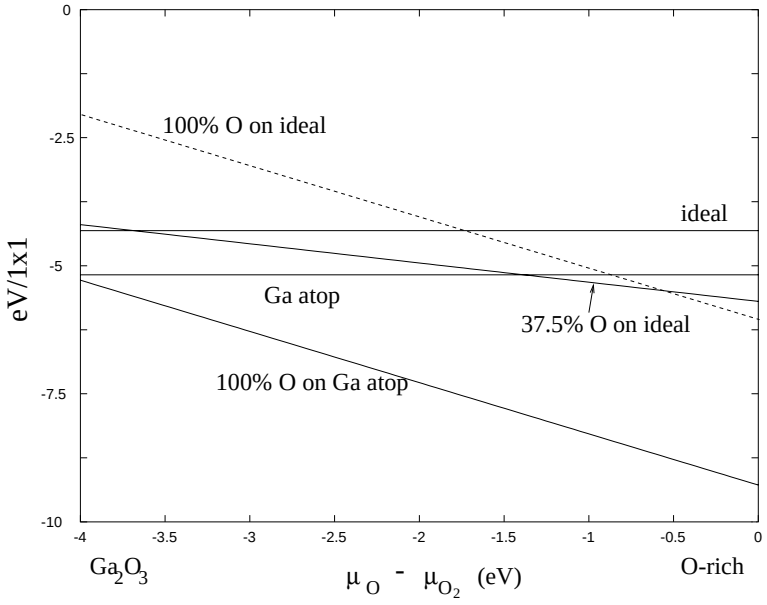

Figure 9.7: The relative energies calculated for possible models of O at the GaN (0001) surface are shown as a function of the O chemical potential (see (5.7)). All energies are evaluated for $\mu_{\mathrm{Ga}} = \mu_{\mathrm{Ga(bulk)}}$. The part on the left of the diagram corresponds to equilibrium with Ga_2O_3, the part on the right to an O–rich environment. The zero of energy in this figure is not related to the zero of energy in Fig. 9.9.

O on top of the ideal surface

We start with placing O on top of the ideal surface. The ideal surface (see Fig. 9.1) is never observed during GaN growth, no matter whatever growth conditions are chosen (see Fig. 9.3). However, with one free orbital per atom pointing along the surface normal the chemical behaviour of this surface with respect to the adsorption of O should be similar to the chemical behaviour of the Ga adatom, N adatom and Ga vacancy structures. As mentioned before these structures are believed to give rise to the experimentally observed (2×2) reconstructions [116, 118, 119, 120]. They have bonding configurations dominated by free Ga orbitals pointing along the surface normal. Therefore, the ideal surface is a good starting point for the investigation of O adsorption on Ga adatom, N adatom and Ga vacancy structures. Moreover, during the adsorption process, some diffusion of Ga and N atoms could occur changing the topology of Ga adatom, N adatom and Ga vacancy structures more and more to that of the ideal surface which clearly offers more free orbitals suitable for the oxygen adsorption than any of the (2×2) reconstructions.

Over a wide range of the oxygen chemical potential the 37.5% model (see Fig. 9.8) is stable with respect to the ideal surface and with respect to all of the (2×2) reconstructions which are not shown in the diagram. The model consists of the ideal surface covered by six oxygen adatoms per (4×4) cell. Four of the oxygens (No. 1–4 in Fig. 9.8) are three–fold coordinated in H3 positions and two oxygens (No. 5,6 in Fig. 9.8) sit two–fold coordinated in an asymmetric bridge position. The three–fold coordinated O have Ga–O bond lengths of 2.14 Å and the two–fold coordinated

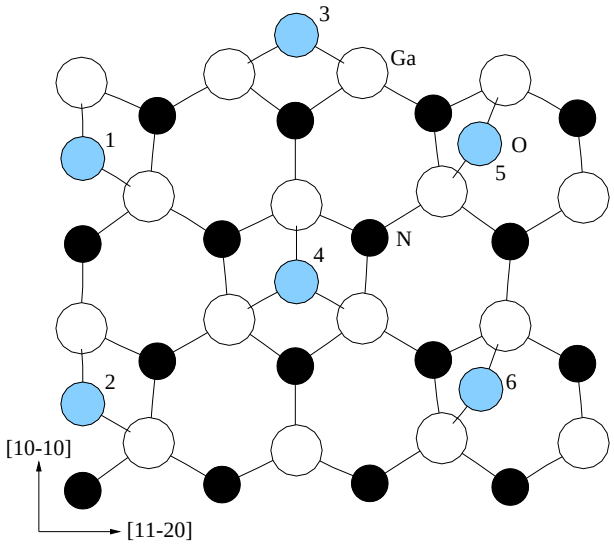

Figure 9.8: Top view of the 37.5% O model at the ideal (0001) surface. Empty (filled) circles represent Ga (N) atoms, grey circles O atoms. In this figure all three–fold coordinated O atoms sit in H3 positions. However, some of them might adopt T4 positions, i.e. they sit in top of the third layer N, yielding a disordered structure and therefore a (1×1) LEED pattern.

O in the asymmetrical bridges have bond lengths 1.92 Å and 2.07 Å. These lengths compare well with the bond lengths found in bulk Ga_2O_3 ranging from $\sim 1.8 - 2.1$ Å. Each of the underlying Ga atoms contributes 0.75 electrons per bond. Therefore each of the three–fold coordinated oxygens has a filled lone pair and 0.25 extra electrons. On the other hand, a two–fold coordinated O atom in a bridge position needs 0.5 electrons to fill its two lone pairs. Therefore, at a charge neutral surface charge transfer occurs from the three–fold to the two–fold coordinated O atoms yielding an emptied conduction band and filled O lone pairs. Hence, this model is semiconducting which explains the low surface energy. Placing some of the three–fold coordinated O from the H3 into a T4 position where they sit three–fold coordinated above the third–layer N atoms gives nearly degenerate energies which can be understood by the fact that the three–fold coordinated oxygens are nearly uncharged and thus have very little Coulomb interaction with the third layer N atoms. The structure is therefore expected to be disordered with some of the three–fold coordinated O in H3 and some in T4 positions and could give rise to a (1×1) LEED pattern.

Increasing the oxygen coverage to eight O atoms per (4×4) cell arranged in bridges results in a metallic surface and a significantly higher surface energy. Removing two O per (4×4) cell from the 37.5% model gives a structure with a 25% oxygen coverage which is only stable in a small range of the oxygen chemical potential near the Ga_2O_3–like environment. Another possibility would be to place 100% oxygen as an adlayer on top of the ideal surface, i.e. one oxygen per (1×1) cell in top of the gallium atom. The O form dimers so that each O has one Ga–O and one O–O bond. This full monolayer coverage is stable in a very O–rich environment (see Fig. 9.7). Here the different chemical behaviour between N and O becomes clear. N_2 has a very high binding energy and is chemically inert. Therefore N–N bonds in GaN are not formed. On the other hand, the O_2 molecule has not such a high binding energy so that O–O bonds can be formed at GaN surfaces.

We therefore conclude that over a wide range of the O chemical potential an oxygen coverage of $\Theta_O = \frac{6}{16}$ ML = 0.375 ML is energetically favourable at the ideal surface. Under O–rich growth conditions also a fully oxidised surface can be achieved.

O on top of the Ga atop reconstruction

As discussed in the previous section, the Ga atop surface (see Fig. 9.1) could be a promising candidate to explain the (1×1)–like RHEED pattern reported during and after growth in the (0001) direction [124, 115, 116, 117, 126, 119]. The experimentally observed surface might be slightly different from the ideal Ga atop surface in terms of possible Peierls distortions and irregularities in the arrangement of the top layer Ga atoms [119]. However, since the atoms in the top surface layer have exclusively Ga–Ga bonds the ideal Ga atop structure is very likely to have bonding properties similar to the Ga–rich (1×1)–like structures reported experimentally and serves thus as a good starting point.

As shown in Fig. 9.7 a 100% O coverage on top of the Ga adlayer is very stable over the entire range of the oxygen chemical potential. In this model all oxygens sit in three fold coordinated positions. Since the typical bond lengths in Ga_2O_3 are quite close to the bond length in GaN bulk (1.95 Å), no stress is induced in the surface layers. In addition, the ideal structure obeys nearly the electron counting rule: if one starts to count from the ideal GaN (0001) surface the 100% O on Ga atop surface has four bonds (one Ga–Ga and three Ga–O bonds) and one lone pair (situated on the O atom) per 1×1 unit cell. On the other hand, the surface contains 9.75 electrons per (1×1) unit cell ($0.75e^-$ from the Ga atom at the ideal surface, $3e^-$ from the Ga atom in top of it and $6e^-$ from the O atom). Therefore, to match the electron counting rule completely, 0.25 electrons per (1×1) unit cell are required. Our calculation within a (4×4) cell shows that two Ga–Ga bonds are broken thus giving the four electrons needed. Moreover, calculation shows that these three–fold coordinated Ga atoms adopt an sp^2 configuration thus lowering the energy and making the structure semiconducting.

It is also worth noting that even if O is assumed to be in equilibrium with Ga_2O_3 (left part in Fig. 9.7) the 100% O coverage has approximately the same energy as the Ga atop structure, showing that the oxygens adopt very low energy positions which are similar to those adopted in the lowest energy phase Ga_2O_3. This explains why our calculations find any other oxygen coverage on the Ga atop reconstruction to be unstable with respect to the 100% coverage model.

The (000$\bar{1}$) Surface

O on top of the Ga atop reconstruction

Fig. 9.9 shows the surface energies depending on the O chemical potential. According to section (9.1) the Ga atop structure seems a suitable reconstruction for the investigation of O on GaN (000$\bar{1}$) surfaces. We first consider a model which fulfils the electron counting rule. The 75% O + 25% N on Ga atop model (see Fig. 9.10) is energetically favourable over a wide range of μ_O. 12 O atoms per (4×4) cell have adsorbed on the Ga atop structure and the remaining four sites are occupied with N which could have diffused to the surface.

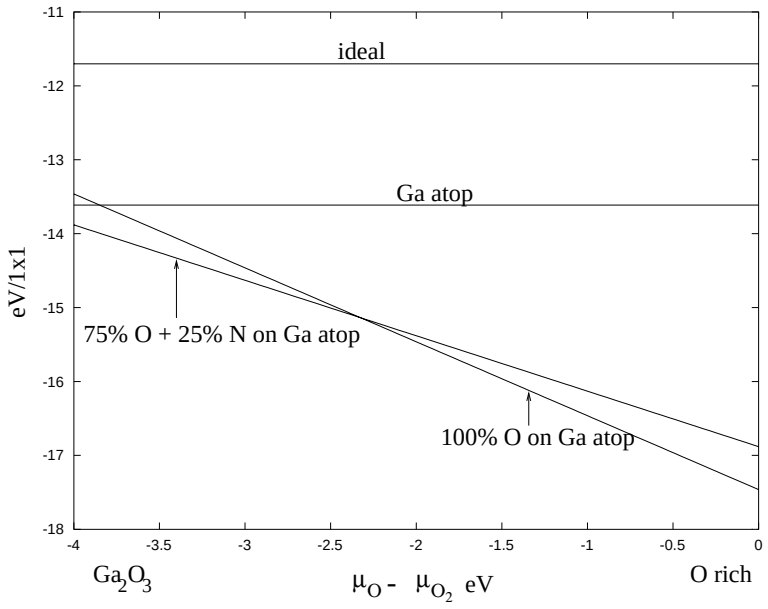

Figure 9.9: The relative energies calculated for possible models for O at the GaN $(000\overline{1})$ surface are shown as a function of the O chemical potential (see (5.7)). All energies are evaluated for $\mu_{\mathrm{Ga}} = \mu_{\mathrm{Ga(bulk)}}$. The part on the left of the diagram corresponds to equilibrium with Ga_2O_3, the part on the right to an O–rich environment. The zero of energy in this figure is not related to the zero of energy in Fig. 9.7.

In this model each O atom has one filled lone pair and 0.25 extra electrons. The N atoms have each 1.25 electrons in dangling bonds. Therefore, the extra electrons of three oxygens fill one N dangling bond. This results in an empty conduction band, one filled lone pair at all N and O atoms and thus in a semiconducting surface. The Ga–O bond lengths at the surface are very similar to the Ga–N bond lengths yielding very little stress at the surface which explains the very low surface energy.

Any related structure with a lower oxygen concentration was found to have a higher energy. We have also calculated structures where we replaced some of the Ga atoms in the adlayer by O so that N–O bonds could be formed. However, these structures turn out to have a very high formation energy suggesting that in any stable configuration only Ga–O bonds should exist.

Another possibility would be to place 100% O on top of the Ga atop reconstruction. This structure has four electrons per 4×4 cell too much, which in the case of the ideal structure, where all O sit threefold on N sites, would have to be placed in the conduction band. However, GaN has a wide band gap (~ 3.4 eV) which usually results in high energies for structures which possess electrons in the conduction band. Our calculations show that it is energetically more favourable to break two of the Ga–O bonds within the (4×4) cell giving rise to two additional O lone pairs and two Ga derived orbitals into which the four extra electrons can be placed. Since the Ga derived orbitals are occupied the Ga atoms *do not* rehybridise into sp^2 but remain in sp^3 positions. Of course, they introduce deep gap states which lie ~ 1.4 eV above VBM. This explains why the 100% O on Ga atop structure is less stable than the 75% O + 25% N on Ga atop structure with respect to Ga_2O_3 and becomes only stable under a more O–rich environment (see Fig. 9.9).

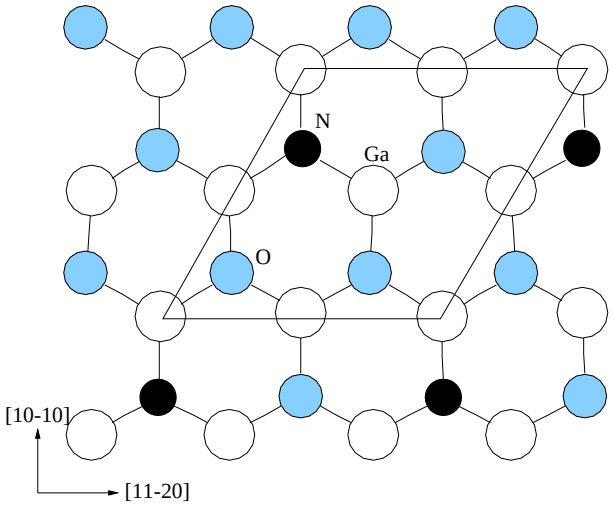

Figure 9.10: Top view of the 75% O + 25% N model at the (000$\bar{1}$) surface. Empty (filled) circles represent Ga (N) atoms, grey circles O. Although the structure can be arranged within a (2×2) periodicity (as shown in the figure) it is not likely to be ordered (cf. text). Replacing the N atoms by O gives the 100% O on Ga atop model.

Our results suggest that oxygen exposure of the (000$\bar{1}$) surface should yield a coverage of 0.75 ML $\leq \Theta_O \leq$ 1.0 ML depending on the O chemical potential and whether atomic N is available during the chemisorption of O at the surface to allow the 75% O + 25% N reconstruction. We finally note that both the 75% O + 25% N and the 100% O structure are not likely to exist in an ordered configuration but are rather disordered and thus give rise to a (1×1) RHEED pattern. Indeed, our calculations show that within a (4×4) cell the energy difference between ordered and disordered configurations is negligible.

Summary

In conclusion we have studied a variety of models for the O chemisorption at GaN (0001) and (000$\bar{1}$) surfaces. The results suggest that in an O–rich environment both surfaces can be fully oxidised. In all stable structures O is always bound to Ga. It is worth noting that at the (000$\bar{1}$) surface adsorbed O sits already in a nitrogen position. During growth this is likely to make the incorporation of O at this surface easier. Therefore, if a low oxygen concentration is desired, we suggest that the material should be grown in the (0001) direction.

Chapter 10

Conclusions

10.1 Summary

In this thesis the atomic structures, formation energies and electrical properties of surfaces and extended defects in GaN were investigated with methods based on density functional theory. To this end one of the methods, the *SCC–DFTB* scheme, has been extended to periodic systems so that with reasonable accuracy formation energies of heteropolar systems can be evaluated within supercells.

In view of the high density of threading defects in GaN the main objective of this thesis was to derive structural models for the most commonly observed line and planar defects and to discuss their implications on the luminescence properties. In particular we explore the relation–ship between extended defects and the frequently observed parasitic emission in the GaN spectrum, i.e. the yellow luminescence, which is believed to arise from a transition between a shallow donor and a deep acceptor state.

In a first step we have predicted structures for the edge and screw dislocations which were subsequently verified by experiment. We found full core screw dislocations to have very distorted bonds in the core region giving rise to deep gap states which possibly contribute to the YL. On the other hand, at the core structures of open core screw dislocations and pure edge dislocations Ga– and N–dangling bonds occur in pairs. In analogy to the nonpolar $(10\bar{1}0)$ surface this yields a rehybridisation of the threefold coordinated Ga and N atoms into sp^2 and p^3, respectively, resulting in a band gap free from deep states. We therefore conclude that extended defects which exhibit this reconstruction mechanism have no major impact on the luminescence properties. Our finding is consistent with experiments which show even material with a very high density of threading edge dislocations to luminesce in the blue spectrum. Subsequently, we investigated the behaviour of intrinsic point defects and impurities in the stress field of threading edge dislocations and found electrically active defect complexes consisting of gallium vacancies surrounded by oxygen to have low energies at the dislocation core. Since these point defects are believed to exist in high quantities in MBE and MOCVD grown epilayers we suggest that if they are mobile they will segregate to the dislocation core. This could explain the non–uniformly distributed YL frequently reported for unintentionally doped n–type GaN where O is believed to be the main source of free carriers.

Another topic which has often been related to threading dislocations concerns the formation of nanopipes. However, our calculations found their origin not to be caused by the elastic energy arising from threading dislocations. This is in agreement with Frank's theorem where superscrew dislocations would be required to account for the large diameters of the nanopipes. Instead we followed an experimental observation suggesting that impurities, in particular O, might diffuse to the inner walls of the nanopipes where they prevent further growth. Consequently we investigated the stability of a variety of oxygen related defect complexes at the nanopipe walls and identified a very stable V_{Ga}–$(O_N)_3$ configuration which is charge–neutral and passivates the surface impeding the growth process.

Also domain boundaries thread through GaN epilayers and might degrade the material properties. They were therefore of special interest in this work.

Finally, we focused on the main polar growth surfaces in order to help to understand the growth process aiming to improve the material quality. We evaluated the formation energies for a variety of Ga, N and H terminated surface reconstructions at the (0001) and $(000\bar{1})$ surfaces depending on the growth conditions. A surprising fact is that some stable polar GaN surfaces are purely Ga terminated and metallic with a (1×1) periodicity. This is in contrast to traditional III–V semiconductors which exhibit semiconducting surfaces with both cation and anion termination. We found a relation–ship between the surface periodicity observed by RHEED during MBE growth and the material polarity. This relation–ship is consistent with recently reported experimental results. Moreover, we explored the adsorption of O at stable surface reconstructions and suggest that oxygen can be adsorbed up to one monolayer if the sample is put into an oxygen rich environment.

10.2 Outlook

Within this thesis total energy calculations for given surfaces and defects in wurtzite GaN have been presented. In further investigations it will be interesting to see how the formation of these structures occurs during the growth process, especially if they involve the segregation of point defects at dislocations and nanopipes. A high diffusivity is necessary to achieve thermodynamic equilibrium which has been assumed within this work. It would therefore be desirable to determine the diffusion rate of common intrinsic point defects and impurities, in particular gallium vacancies and oxygen, in bulk material and at surfaces [127, 128]. This might also help growers to improve the material quality.

A related topic is the investigation of dislocation mobility based on the evaluation of kink pair formation energies. This might explain why dislocations in GaN are found to be far less mobile than those in GaAs and in II–VI materials and thus do not degrade the material properties significantly.

Finally, we believe that our work should stimulate experimentalists to carry out further investigations on the defect structures suggested. In particular, techniques which could determine the oxygen concentration at and next to extended defects with a high spatial resolution would be very desirable. Furthermore, in good quality material deep level transient spectroscopy (DLTS) experiments could analyse the electronic structure of defects and confirm the theoretical suggestions concerning the origin of the yellow luminescence.

Appendix A

Expressions for SCC–DFTB

A.1 Analytical evaluation of γ_{IJ}

The integral (4.18) consists of three parts:

$$\frac{\tau_J^3}{8\pi} \int \frac{1}{|\mathbf{r}-\mathbf{R}_I|} \cdot e^{-\tau_J |\mathbf{r}-\mathbf{R}_J|} \tag{A.1}$$

$$-\frac{\tau_I \tau_J^3}{16\pi} \int e^{-\tau_I |\mathbf{r}-\mathbf{R}_I|} \cdot e^{-\tau_J |\mathbf{r}-\mathbf{R}_J|} \tag{A.2}$$

$$-\frac{\tau_J^3}{8\pi} \int \frac{1}{|\mathbf{r}-\mathbf{R}_I|} e^{-\tau_I |\mathbf{r}-\mathbf{R}_I|} \cdot e^{-\tau_J |\mathbf{r}-\mathbf{R}_J|} \tag{A.3}$$

We therefore need to evaluate integrals of the form

$$\int |\mathbf{r}-\mathbf{R}_I|^n e^{-\tau_I |\mathbf{r}-\mathbf{R}_I|} e^{-\tau_J |\mathbf{r}-\mathbf{R}_J|} \qquad \text{for } n = -1, 0 \tag{A.4}$$

Setting $\mathbf{R}_I = 0$ and transforming to spherical coordinates with

$$r_a = |\mathbf{r}| \ , \quad r_b = |(\mathbf{R}_I - \mathbf{R}_J) + \mathbf{r}| \quad \text{and} \quad \xi_a = cos\theta_a \ ,$$

we get for (A.4):

$$\int_0^{2\pi} \int_{-1}^{1} \int_0^{\infty} |r_a|^n e^{-\tau_I r_a} e^{-\tau_J r_b(r_a,\xi_a)} r_a^2 dr_a d\xi_a d\phi$$

We now introduce spheroidal coordinates ($R = |\mathbf{R}_I - \mathbf{R}_J|$):

$$\xi = \frac{r_a + r_b}{R} \quad \eta = \frac{r_a - r_b}{R}$$

$$r_a = \frac{R}{2}(\xi + \eta) \quad \xi_a = \frac{1 + \xi\eta}{\xi + \eta} \text{ and } r_b = \frac{R}{2}(\xi - \eta)$$

with the functional determinant

$$\frac{\partial(r_a,\xi_a)}{\partial(\xi,\eta)} = \frac{R}{2}\frac{\xi^2-\eta^2}{(\xi+\eta)^2}$$

Setting $\alpha = \tau_I\frac{R}{2}$ and $\beta = \tau_J\frac{R}{2}$ the integral then becomes:

$$\frac{\pi}{4}R^3\left(\frac{R}{2}\right)^n \int_1^\infty d\xi \int_{-1}^1 d\eta(\xi+\eta)^{n+1}(\xi-\eta)e^{-(\alpha+\beta)\xi}\ e^{-(\alpha-\beta)\eta}$$

Setting $\gamma = \alpha+\beta$ and $\delta = \alpha-\beta$ we obtain by expanding the product $(\xi+\eta)^{n+1}$ into binomials:

$$\begin{aligned}\frac{\pi}{4}R^3\left(\frac{R}{2}\right)^n \quad &\cdot \quad \sum_{i=0}^{n+1}\binom{n+1}{i}\int_1^\infty d\xi e^{-\gamma\xi}\times \\ &\times \quad \left[\xi^{i+1}\int_{-1}^1 d\eta\ \eta^{n+1-i}e^{-\delta\eta} - \xi^i\int_{-1}^1 d\eta\ \eta^{n+2-i}e^{-\delta\eta}\right] \end{aligned} \tag{A.5}$$

The remaining integrals can be evaluated. For $\delta \neq 0$ one has:

$$\begin{aligned}\int_a^b x^l e^{-\delta x}\ dx &= -\frac{x^l}{\delta}e^{-\delta x}\Big|_a^b + \frac{l}{\delta}\int_a^b x^{l-1}e^{-\delta x}\ dx \\ &= -\frac{x^l}{\delta}e^{-\delta x}\Big|_a^b - \frac{lx^{l-1}}{\delta^2}e^{-\delta x}\Big|_a^b - \frac{l(l-1)x^{l-2}}{\delta^3}e^{-\delta x}\Big|_a^b \\ &= -\sum_{m=0}^{l}\frac{l!}{(l-m)!\delta^{m+1}}x^{l-m}e^{-\delta x}\Big|_a^b \end{aligned} \tag{A.6}$$

and for $\delta = 0$:

$$\int_a^b x^l\ dx = \frac{1}{l+1}x^{l+1}\Big|_a^b \tag{A.7}$$

Using (A.6) we get for (A.5) with $\delta \neq 0$:

$$\begin{aligned}&\frac{\pi}{4}R^3\left(\frac{R}{2}\right)^n\sum_{i=0}^{n+1}\binom{n+1}{i}\int_1^\infty d\xi e^{-\gamma\xi}\times \\ \times\quad &\left[\xi^{i+1}\sum_{m=0}^{n+1-i}\frac{(n+1-i)!}{(n+1-i-m)!\delta^{m+1}}\left((-1)^{n+1-i-m}e^\delta - e^{-\delta}\right)\right. \\ -\quad &\left.\xi^i\sum_{m=0}^{n+2-i}\frac{(n+2-i)!}{(n+2-i-m)!\delta^{m+1}}\left((-1)^{n+2-i-m}e^\delta - e^{-\delta}\right)\right]\end{aligned}$$

Using (A.6) once again we obtain the final result for $\delta \neq 0$:

$$\begin{aligned}&\frac{\pi}{4}R^3\left(\frac{R}{2}\right)^n\sum_{i=0}^{n+1}\binom{n+1}{i}e^{-\gamma}\times \\ \times\quad &\left[\left(\sum_{m=0}^{i+1}\frac{(i+1)!}{(i+1-m)!\gamma^{m+1}}\right)\sum_{m=0}^{n+1-i}\frac{(n+1-i)!}{(n+1-i-m)!\delta^{m+1}}\left((-1)^{n+1-i-m}e^\delta - e^{-\delta}\right)\right. \\ -\quad &\left.\left(\sum_{m=0}^{i}\frac{i!}{(i-m)!\gamma^{m+1}}\right)\sum_{m=0}^{n+2-i}\frac{(n+2-i)!}{(n+2-i-m)!\delta^{m+1}}\left((-1)^{n+2-i-m}e^\delta - e^{-\delta}\right)\right]\end{aligned} \tag{A.8}$$

For $\delta = 0$ (A.5) becomes:

$$\frac{\pi}{4}R^3\left(\frac{R}{2}\right)^n\sum_{i=0}^{n+1}\left(\begin{array}{c} n+1 \\ i \end{array}\right)e^{-\gamma} \times \left[\left(\sum_{m=0}^{i+1}\frac{(i+1)!}{(i+1-m)!\gamma^{m+1}}\right)\frac{1}{n+2-i}(1-(-1)^{n+2-i}) - \left(\sum_{m=0}^{i}\frac{i!}{(i-m)!\gamma^{m+1}}\right)\frac{1}{n+3-i}(1-(-1)^{n+3-i})\right]$$

We can now evaluate expressions (A.1-A.3).

In (A.1) $n=-1$ and $\tau_I = 0$ giving $\gamma = \tau_J\frac{R}{2}$ and $\delta = -\tau_J\frac{R}{2}$:

$$\frac{\tau_J^3}{8\pi}\int\frac{1}{|\mathbf{r}-\mathbf{R}_I|}e^{-\tau_J|\mathbf{r}-\mathbf{R}_J|} = \frac{\tau_J^3}{8\pi}\frac{\pi}{2}R^2\frac{e^{-\gamma}}{-\gamma^2}\left[\frac{2}{\gamma}\left(e^{-\gamma}-e^{\gamma}\right)+2e^{-\gamma}\right] = \frac{1}{R}\left(1-e^{-\tau_J R}(1+\frac{\tau_J R}{2})\right)$$

In (A.2) $n=0$, $\gamma = \frac{\tau_I+\tau_J}{2}R$ and $\delta = \frac{\tau_I-\tau_J}{2}R$. This gives for $\delta \neq 0$ from (A.8):

$$-\frac{\tau_I\tau_J^3}{16\pi}\int e^{-\tau_I|\mathbf{r}-\mathbf{R}_I|}\cdot e^{-\tau_J|\mathbf{r}-\mathbf{R}_J|} = -\frac{\tau_J^3\tau_I}{16\pi}\frac{\pi}{4}R^3\cdot 2\frac{e^{-\gamma}}{\gamma\delta}\left[\left(\frac{1}{\gamma^2}+\frac{1}{\gamma}-\frac{1}{\delta^2}\right)\left(e^{\delta}-e^{-\delta}\right)+\frac{1}{\delta}(e^{\delta}-e^{-\delta})\right]$$
$$= \frac{\tau_J^3\tau_I}{2R(\tau_I^2-\tau_J^2)^3}\left[(\tau_J^2-\tau_I^2)(\tau_I\cdot e^{-\tau_J R}-\tau_J\cdot e^{-\tau_I R})R+4\tau_I\tau_J(e^{-\tau_J R}-e^{-\tau_I R})\right]$$

For $\tau_I = \tau_J$ one has $\delta = 0$. $n=0$ and $\gamma = \frac{\tau_I+\tau_J}{2}R$ gives then for (A.2):

$$-\frac{\tau_I\tau_J^3}{16\pi}\int e^{-\tau_I|\mathbf{r}-\mathbf{R}_I|}\cdot e^{-\tau_I|\mathbf{r}-\mathbf{R}_J|} = -\frac{\tau_I^4}{16\pi}\frac{\pi}{4}R^3\cdot\left(\frac{4}{3\gamma}+\frac{4}{\gamma^2}+\frac{4}{\gamma^3}\right) = -\frac{1}{16}e^{-\tau_I}\left(\frac{\tau_I^3R^2}{3}+\tau_I^2R+\tau_I\right)$$

In (A.3) $n=-1$, $\gamma = \frac{\tau_I+\tau_J}{2}R$ and $\delta = \frac{\tau_I-\tau_J}{2}R$. This gives for $\delta \neq 0$ from (A.8):

$$-\frac{\tau_J^3}{8\pi}\int\frac{1}{|\mathbf{r}-\mathbf{R}_I|}e^{-\tau_I|\mathbf{r}-\mathbf{R}_I|}\cdot e^{-\tau_J|\mathbf{r}-\mathbf{R}_J|} = -\frac{\tau_J^3}{8\pi}\frac{\pi}{4}R^3\frac{2}{R}\left[\frac{e^{-\gamma}}{\gamma\delta}\left(\frac{1}{\gamma}-\frac{1}{\delta}\right)\left(e^{\delta}-e^{-\delta}\right)+2e^{\delta}\right]$$
$$= \frac{\tau_J^3}{(\tau_I^2-\tau_J^2)^2R}\left[\tau_J(e^{-\tau_J R}-e^{-\tau_I R})-\frac{1}{2}(\tau_I^2-\tau_J^2)Re^{-\tau_J R}\right]$$

For $\tau_I = \tau_J$ one has $\delta = 0$. $n=-1$ and $\gamma = \frac{\tau_I+\tau_J}{2}R$ gives then for (A.3):

$$-\frac{\tau_I^3}{8\pi}\int\frac{1}{|\mathbf{r}-\mathbf{R}_I|}e^{-\tau_I|\mathbf{r}-\mathbf{R}_I|}\cdot e^{-\tau_J|\mathbf{r}-\mathbf{R}_J|} = -\frac{\tau_I^3}{8\pi}\frac{\pi}{4}R^3\frac{2}{R}\left[e^{-\gamma}\left(\frac{2}{\gamma}+\frac{2}{\gamma^2}\right)\right] = -\frac{1}{8}e^{-\tau_I R}\left(\tau_I^2R+\tau_I\right)$$

Adding these contributions finally gives for γ_{IJ} in (4.18) for $R \neq 0$:

$$\gamma_{IJ} = \frac{1}{R} - \left\{\begin{array}{l} e^{-\tau_I R}\left(\frac{\tau_J^4\tau_I}{2(\tau_I^2-\tau_J^2)^2}-\frac{\tau_J^6-3\tau_J^4\tau_I^2}{(\tau_I^2-\tau_J^2)^3R}\right)-e^{-\tau_J R}\left(\frac{\tau_I^4\tau_J}{2(\tau_J^2-\tau_I^2)^2}-\frac{\tau_I^6-3\tau_I^4\tau_J^2}{(\tau_J^2-\tau_I^2)^3R}\right) \\ \quad \text{if } \tau_I \neq \tau_J \qquad \text{and} \\ -e^{-\tau_I R}\left(\frac{1}{R}+\frac{11\tau_I}{16}+\frac{3\tau_I^2R}{16}+\frac{\tau_I^3R^2}{48}\right) \\ \quad \text{if } \tau_I = \tau_J \end{array}\right. \tag{A.9}$$

The expression for γ_{IJ} in the case $\tau_I = \tau_J$ could have also been derived by setting $\tau_J = \tau_I + \epsilon$ in γ_{IJ} ($\tau_I \neq \tau_J$), expanding the exponentials and then taking the limit $\epsilon \to 0$. This shows in a mathematically rigorous way that γ_{IJ} is continuous at $\tau_I = \tau_J$; a fact which is clear on physical grounds since no discontinuity should arise if the Coulomb energy is evaluated for two identical charge distributions.

We now consider the limit for $R \to 0$. To this end the exponentials of the terms after curly brackets in (A.9) are expanded. The first term gives:

$$\begin{aligned} & (1 + \frac{-\tau_I R}{1!} + O(R^2)) \left(\frac{\tau_J^4 \tau_I}{2(\tau_I^2 - \tau_J^2)^2} - \frac{\tau_J^6 - 3\tau_J^4 \tau_I^2}{(\tau_I^2 - \tau_J^2)^3 R} \right) \\ = \; & \frac{\tau_J^4 \tau_I}{2(\tau_I^2 - \tau_J^2)^2} - \frac{\tau_J^6 - 3\tau_J^4 \tau_I^2}{(\tau_I^2 - \tau_J^2)^3 R} + \frac{\tau_I \tau_J^6 - 3\tau_J^4 \tau_I^3}{(\tau_I^2 - \tau_J^2)^3} + O(R) \end{aligned}$$

The second term is obtained by interchanging τ_I and τ_J. Adding both gives:

$$\begin{aligned} & -\frac{1}{R} + \frac{\tau_J^4 \tau_I + \tau_I^4 \tau_J}{2(\tau_I^2 - \tau_J^2)^2} + \frac{\tau_I \tau_J^6 - 3\tau_J^4 \tau_I^3 - \tau_J \tau_I^6 + 3\tau_I^4 \tau_J^3}{(\tau_I^2 - \tau_J^2)^3} + O(R) \\ = \; & -\frac{1}{R} + \frac{\tau_I \tau_J}{2(\tau_I + \tau_J)^3 (\tau_I - \tau_J)^3} \left[-\tau_I^5 - 5\tau_I^2 \tau_J^3 + 5\tau_J^2 \tau_I^3 + \tau_J^5 \right] + O(R) \\ = \; & -\frac{1}{R} - \frac{\tau_I \tau_J}{2(\tau_I + \tau_J)^3 (\tau_I - \tau_J)^3} \left[\left((\tau_I + \tau_J)^2 + \tau_I \tau_J \right) (\tau_I - \tau_J)^3 \right] + O(R) \\ = \; & -\frac{1}{R} - \frac{\tau_I \tau_J}{2(\tau_I + \tau_J)^3} \left[(\tau_I + \tau_J)^2 + \tau_I \tau_J \right] + O(R) \end{aligned}$$

And hence:

$$\gamma_{IJ} = \frac{\tau_I \tau_J}{2(\tau_I + \tau_J)^3} \left[(\tau_I + \tau_J)^2 + \tau_I \tau_J \right] + O(R)$$

At $R = 0$ we are at the same atom ($I = J$) and thus have $\tau_I = \tau_J$ from which relation (4.20) follows:

$$\gamma_{II} = \frac{5}{16} \tau_I$$

Of course, the latter relation could also have been obtained by expanding the exponential in the expression for $\tau_I = \tau_J$ in (A.9) and taking the limit $R \to 0$. This shows that the function γ_{IJ} is continuous at $R = 0$.

A.2 Numerical evaluation of $\overline{\gamma}_{IJ}$ and its first derivative for periodic systems via the Ewald summation technique

Evaluation of $\overline{\gamma}_{IJ}$

To compute the charge dependent correction of the matrix elements, we need to evaluate terms of the kind

$$\sum_I \Delta q_I \gamma_{JI}$$

This expression can be implemented in a straightforward manner for finite systems. In periodic systems the sum over the atoms I is replaced by the sum over all atoms I in the cell summed over all cells:

$$\sum_{\mathbf{R}}\sum_{I}\Delta q_I\gamma_{JI}=\sum_{I}\Delta q_I\sum_{\mathbf{R}}\gamma_{JI}=\sum_{I}\Delta q_I\overline{\gamma}_{JI}$$

$\overline{\gamma}_{IJ}$ can be split into a short range and a long range part. The short range part consists of the sum of the terms following the curly bracket in (4.19). The terms in this sum decay exponentially. The short range part is therefore absolutely convergent. Numerically this short range part is evaluated as a sum over a small number of unit cells.

The long range part representing the Coulomb interactions is not absolutely convergent, i.e. the value changes depending on the order of summation. However, as the long range part can be considered as the potential corresponding to a charge distribution, there is only one physically reasonable value to which it should converge. Ewald's method is used to compute this value:

$$\sum_{\mathbf{R}}^{\mathbf{R}\neq\mathbf{R}_J-\mathbf{R}_I}\frac{1}{|\mathbf{R}_J-(\mathbf{R}+\mathbf{R}_I)|}=\Phi(\mathbf{R}_J-\mathbf{R}_I) \tag{A.10}$$

Here Φ is the Ewald potential defined by [129, 130]:

$$\Phi(\mathbf{r})=\begin{cases}\frac{4\pi}{\Omega}\sum_{\mathbf{G}\neq 0}\frac{e^{-\frac{\mathbf{G}^2}{4\epsilon^2}}e^{i\mathbf{G}\mathbf{r}}}{|\mathbf{G}^2|}+\sum_{\mathbf{R}}\frac{1-erf(\epsilon|\mathbf{R}-\mathbf{r}|)}{|\mathbf{R}-\mathbf{r}|}-\frac{\pi}{\Omega\epsilon^2} & \text{for } \mathbf{r}\neq 0\\ \frac{4\pi}{\Omega}\sum_{\mathbf{G}\neq 0}\frac{e^{-\frac{\mathbf{G}^2}{4\epsilon^2}}e^{i\mathbf{G}\mathbf{r}}}{|\mathbf{G}^2|}+\sum_{\mathbf{R}}\frac{1-erf(\epsilon|\mathbf{R}-\mathbf{r}|)}{|\mathbf{R}-\mathbf{r}|}-\frac{\pi}{\Omega\epsilon^2}-\frac{2\epsilon}{\sqrt{\pi}} & \text{for } \mathbf{r}=0\end{cases} \tag{A.11}$$

where: Ω is the volume of the unit cell, $\mathbf{G}$ is a reciprocal lattice vector, $\mathbf{R}$ is a lattice vector, ϵ is a parameter determining the convergence (large ϵ allows the neglect of the real space term) and erf is the error function defined by

$$erf(x)=\frac{2}{\pi}\int_0^x\exp(-t^2)dt\ .$$

The potential Φ represents the periodic charge distribution of positive unit charge located at 0 with uniform negative background charge, so that the average potential is zero.

As we have $\sum_I\Delta q_I=0$ for a neutral periodic structure, the background charges cancel leaving exactly the potential describing the charge distribution in (A.10).

If a charged structure is considered, $\sum_I\Delta q_I\neq 0$. In this case, the background charges compensate the ionic charges leaving a charge neutral supercell which can be periodically repeated. This approximation of a compensating uniform background charge is often used to model charged defects in supercells (see section D.2).

Evaluation of $\frac{\partial}{\partial\mathbf{R}_I}\overline{\gamma}_{IJ}$

In a finite system the derivative of γ_{IJ} can be evaluated in a straight forward manner. For a periodic system the derivative of $\overline{\gamma}_{IJ}=\sum_{\mathbf{R}}\gamma_{IJ}$ is again split into short range and long range part. While

the short range part consisting of the derivatives of the terms after curly brackets in (A.9) can be differentiated explicitly, the long range term is again evaluated by means of Ewald's technique:

$$\frac{\partial}{\partial \mathbf{R}_I} \sum_{\mathbf{R}}^{\mathbf{R}\neq\mathbf{R}_I-\mathbf{R}_J} \frac{1}{|\mathbf{R}_I-(\mathbf{R}+\mathbf{R}_J)|} = \frac{d}{d\mathbf{R}_I}\Phi(\mathbf{R}_I-\mathbf{R}_J)$$

where Φ is the Ewald potential defined in (A.11). The first derivative reads for $\mathbf{r}\neq 0$:

$$\frac{\partial\Phi(\mathbf{r})}{\partial\mathbf{r}} = \frac{4\pi}{\Omega}\sum_{\mathbf{G}\neq 0}\frac{\partial}{\partial\mathbf{r}}\frac{e^{-\frac{\mathbf{G}^2}{4\epsilon^2}}e^{i\mathbf{G}\mathbf{r}}}{|\mathbf{G}^2|} + \frac{\partial}{\partial\mathbf{r}}\sum_{\mathbf{R}}^{\mathbf{R}\neq\mathbf{r}}\frac{1-erf(\epsilon|\mathbf{R}-\mathbf{r}|)}{|\mathbf{R}-\mathbf{r}|}$$

Let us first evaluate the reciprocal space term:

$$\frac{\partial}{\partial\mathbf{r}}REZ = \frac{\partial}{\partial\mathbf{r}}\frac{4\pi}{\Omega}\sum_{\mathbf{G}\neq 0}\frac{e^{-\frac{\mathbf{G}^2}{4\epsilon^2}}e^{i\mathbf{G}\mathbf{r}}}{|\mathbf{G}^2|} = \frac{4\pi}{\Omega}\sum_{\mathbf{G}\neq 0}\frac{e^{-\frac{\mathbf{G}^2}{4\epsilon^2}}(-\mathbf{G})\sin(\mathbf{G}\mathbf{r})}{|\mathbf{G}^2|}$$

As the sum is absolutely convergent, we were allowed to change the order of differentiation and summation. In a similar way we get for the real space term:

$$\begin{aligned}\frac{\partial}{\partial\mathbf{r}}REAL &= \frac{\partial}{\partial\mathbf{r}}\sum_{\mathbf{R}}^{\mathbf{R}\neq\mathbf{r}}\frac{1-erf(\epsilon|\mathbf{R}-\mathbf{r}|)}{|\mathbf{R}-\mathbf{r}|}\\ &= \sum_{\mathbf{R}}^{\mathbf{R}\neq\mathbf{r}}(\mathbf{r}-\mathbf{R})\frac{-\frac{2}{\sqrt{\pi}}e^{-\epsilon^2|\mathbf{R}-\mathbf{r}|^2}\cdot\epsilon\cdot|\mathbf{R}-\mathbf{r}|-1+erf(\epsilon|\mathbf{R}-\mathbf{r}|)}{|\mathbf{R}-\mathbf{r}|^3}\end{aligned}$$

Fortunately, no derivative needs to be evaluated for the case $\mathbf{r}=0$, where Φ is discontinuous, since $\mathbf{r}=0$ means $\mathbf{R}_I=\mathbf{R}_J$.

Appendix B

(SCC)-DFTB Parameters

Table B.1 shows the values chosen to create the basis functions used for *(SCC)–DFTB* following Porezag *et al.* [31] and Elstner *et al.* [27]. The on–site energies are those of a free atom calculated within SCF–LDA. The Hubbard parameters were derived from SCF–LDA calculations as the second derivative of the total energy of a free atom with respect to the occupation number of the highest occupied atomic orbital. According to Porezag *et al.* [31] confined atomic wavefunctions are used for *(SCC)–DFTB* as a minimal basis because they are suitable to describe the charge density in molecules and solid state systems which usually is more contracted than the atomic charge density. Confined atomic basis functions are created by adding a potential $\left(\frac{r}{r_0}\right)^2$ to the effective potential V_{eff} in the Kohn–Sham equations for the single atom [31]. The confining radius r_0 is chosen $\approx 2r_{\text{cov}}$, where r_{cov} is the covalent radius of the atom. Table B.1 shows the precise values of the r_0 used in this work. Inserted into the *DFTB* and *SCC–DFTB* methods the confined atomic basis functions

Table B.1: On–site energies ε_i, Hubbard parameters τ_I and confining radii r_0 of the minimal basis set used within the *(SCC)–DFTB* method. All quantities are given in a.u. See text and references [31, 27] for further explanations.

	Ga 3d	Ga 4s	Ga 4p	N 2s	N 2p	O 2s	O 2p	H 1s
ε_i	-0.7360	-0.3280	-0.1017	-0.6760	-0.2662	-0.8712	-0.3383	-0.2335
τ_I	0.2084	0.2084	0.2084	0.4303	0.4303	0.4946	0.4946	0.4065
r_0	4.55	4.55	4.55	2.71	2.71	2.60	2.60	1.30

gave the electronic band structures shown in Fig. B.1. In order to calculate formation energies within the *SCC–DFTB* method the Ga $3d$ electrons are included into the valence band since they hybridise with the N $2p$ electrons. Note that due to the minimal basis set employed the band gap calculated with *(SCC)–DFTB* is by far too large.

In Table B.2 we list the reference structures used to generate the repulsive potentials E_{rep} according to Eq. 4.10. For the Ga–N interaction bulk GaN in the zinc–blende (β) phase was chosen to be the reference structure. The energy versus the lattice constant was evaluated via the Murnaghan

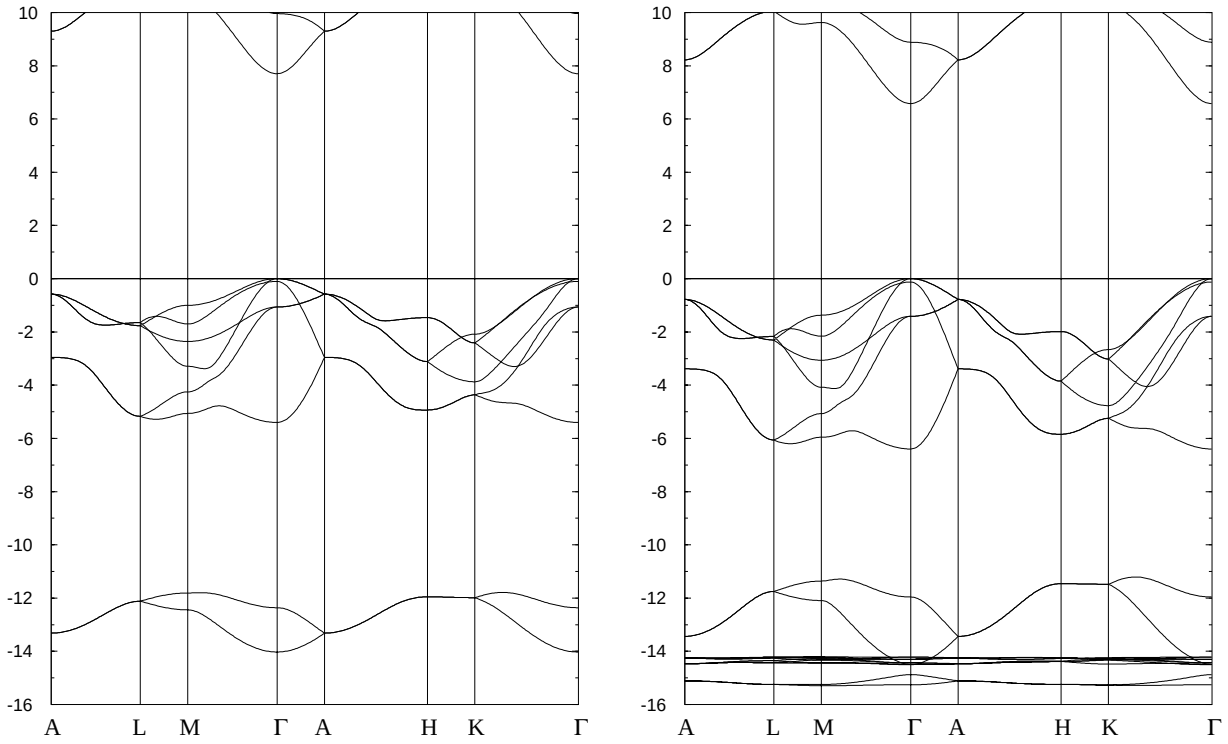

Figure B.1: Electronic band structure for α–GaN calculated within the *DFTB* (*left*) and *SCC–DFTB* (*right*) approximation. In order to obtain more accurate formation energies, the Ga $3d$ orbitals are included within *SCC–DFTB* since they hybridise with the N $2s$ levels. Note that the valence bands in the Fig. on the right show a much larger dispersion.

equation of state [131] using the bulk modulus and its derivative with respect to pressure.[1] All other repulsive interactions were derived from fully saturated molecules. The SCF–LDA reference energies were calculated with the all–electron *CLUSTER* code [132]. The advantage of using fully saturated molecules as reference structures is that they usually possess large band gaps so that a crossing of occupied and unoccupied electronic levels for varying distances ("level crossing") is very unlikely. This results in constant occupation numbers n_i so that Eq. 4.10 can easily be used. On the other hand, unsaturated molecules, in particular dimers, have occupation numbers which are discontinuous for varying distances. These level crossings produce kinks in the energy–versus–distance curve making a determination of E_{rep} rather difficult.

Table B.2: Reference structures used to determine the repulsive potential E_{rep} within the *(SCC)–DFTB* method. See text and references [31, 27] for further explanations.

Ga–Ga	Ga–N	Ga–O	Ga–H	N–N	N–O	N–H	O–O	O–H	H–H
Ga_2H_4	β–GaN	GaH_2–OH	GaH_3	N_2	N–O	NH_3	O_2	H_2O	H_2

[1] For β–GaN the values used in this work are: $a_0 = 4.50$ Å, $B_0 = 1.84$ Mbar, $B_0' = 4.0$.

Appendix C

Range of the Chemical Potentials within SCC–DFTB

C.1 The elemental chemical potentials

GaN

If SCF–LDA total energy calculations are used to investigate the formation energies of surfaces and defects [46, 8] one usually calculates the heat of formation ΔH_f of GaN to fix the range of the chemical potential μ_{Ga} given by Eq. (5.6). According to Eq. (5.3) ΔH_f can be determined from the orthorhombic bulk phase of Ga, the wurtzite lattice phase of GaN and the N_2 molecule. The SCF–LDA value for ΔH_f is usually in reasonable agreement with the experimentally derived value of 1.14 eV [49]. For example, Northrup *et al.* find a value of 0.90 eV [46]. However, within *SCC–DFTB* we make some approximations and, in particular, use a minimal basis set which is only suitable for structures consisting of sp^3, sp^2 and p^3–like bonding configurations. Therefore, we cannot expect to obtain precise values for the total energies of structures containing more complicated bonding configurations. In particular, energies of molecules which have triple bonds such as N_2 are not converged within a minimal basis. Therefore, the *DFTB* energy of N_2 is considerably too high resulting in an increased value of ΔH_f for GaN (see Table C.1). To circumvent this problem, we plot the formation energy depending on the Ga chemical potential within a range given by $\mu_{\mathrm{Ga(bulk)}}$ and the experimental heat of formation:

$$\mu_{\mathrm{Ga(bulk)}} - 1.14 \text{ eV} \leq \mu_{\mathrm{Ga}} \leq \mu_{\mathrm{Ga(bulk)}} \ .$$

The value of $\mu_{\mathrm{Ga(bulk)}}$ has been lowered by 0.15 eV (see Table C.1) since also the orthorhombic bulk phase of Ga lies slightly too high in energy and therefore had to be slightly corrected. This comes from the fact that the extended wavefunctions of metals can not in general be well described within our minimal basis of localised orbitals. The range of the chemical potential μ_{Ga} determined in this way has been successfully applied in chapters 6 and 9 to determine the formation energies of non–stoichiometric nonpolar and polar surfaces giving results very similar to those of Northrup *et al.* [46] and Smith *et al.* [118]. Also in GaAs we defined the range of the gallium chemical potential by taking the calculated value of $\mu_{\mathrm{Ga(bulk)}}$ lowered by 0.15 eV for the upper limit and used the experimental

value for ΔH_f to fix the lower limit of μ_{Ga}. Within this range we determined the formation energies of surfaces [36]. The results are in very good agreement with SCF–LDA calculations [53] showing that this procedure is transferable.

Table C.1: *SCC–DFTB* total energies $\mathrm{E_{tot}}$, correction of total energies (Corr) and resulting heat of formations ΔH_f for the elements and compounds used in this work. All quantities are given in eV.

	Ga (bulk)	N (N_2)	O (O_2)	H (H_2)	GaN (bulk)	Ga_2O_3 (bulk)
$\mathrm{E_{tot}}$	-224.15	-67.32	-89.04	-9.70	-294.26	-729.3
Corr	-0.15	not known	-0.51	-	-	-
ΔH_f*(SCC-DFTB)*	-	-	-	-	2.64	12.05
ΔH_f(exp) [49]	-	-	-	-	1.14	11.3

Ga_2O_3

The heat of formation for Ga_2O_3 has been determined by enthalpy measurements from Ga in the orthorhombic bulk phase and O_2 to be 11.3 eV [49]. In contrast to N_2 the O_2 molecule has no triple bond and is therefore better described within a minimal basis. We may thus attempt to calculate ΔH_f for Ga_2O_3 within *SCC–DFTB*. However, the O_2 molecule adopts a paramagnetic triplet structure in the ground–state whereas *SCC–DFTB* does not include spin polarisation effects, i.e. it gives the energy for the singlet state as the ground–state energy. We therefore correct the *SCC–DFTB* energy for the singlet state by the energy difference between singlet and triplet state which we obtained from an SCF–LDA calculation to be 1.02 eV for O_2. Using this corrected value we obtain the heat of formation as 12.05 eV (see Table C.1), in good agreement with the experimental value.

C.2 The electro-chemical potential

In a semiconductor the electro-chemical potential (*Fermi level*) can vary from the valence band maximum (VBM) in p–type material to the conduction band minimum (CBM) in n–type material. Defects usually lead to a broadening of the valence band. However, the Fermi level in Eq. (5.7) is to be taken with respect to the valence band maximum of a perfect crystal. Therefore, one should use the VBM calculated within a perfect bulk cell, denoted by VBM(bulk cell). If charged defects are treated within periodic cells a compensating uniform background charge is introduced via the Ewald–summation technique (see Eq. (A.11)) and gives rise to an artificial potential. This results in a rigid shift of the spectrum so that the suitable VBM is not any more the same as VBM(bulk cell). In order to get the value by which VBM(bulk cell) needs to be shifted one can use a characteristic bulk level in the spectrum of both, the charged defect cell and the uncharged perfect bulk cell. Via this level one can determine the shift which added to VBM(bulk cell) gives the appropriate VBM:

$$\mathrm{VBM(defect\ cell)} = \mathrm{VBM(bulk\ cell)} + \varepsilon_{\mathrm{char}}(\mathrm{defect\ cell}) - \varepsilon_{\mathrm{char}}(\mathrm{bulk\ cell})\ .$$

Appendix D

Structural Modelling of Surfaces and Defects

D.1 Surfaces

In this work *SCC–DFTB* is used to investigate surfaces. The surfaces were modelled by ten mono-layer thick slabs with periodic boundary conditions in two dimensions. The first six mono-layers were allowed to relax, while the remaining atoms were fixed to preserve the bulk lattice spacing. In order to prevent artificial charge transfer between the bottom of the slab and the surface, we follow the approach of Shiraishi [133] and saturate the dangling bonds on the bottom with pseudo–hydrogen. By demanding that the charge distribution in our slab model should not depend on whether we terminate our slab with an N or a Ga monolayer, one can derive an equation for the charge contribution of the pseudo–hydrogen atoms, yielding 1.25 and 0.75 electrons per H atom, for the replacement of an N and a Ga atom respectively [59]. These charges correspond to the charge per bond contributed from a tetrahedrally bound N or Ga atom. For an illustration see Fig. D.1.

The vacuum region separating the slabs in the direction of the surface normal is chosen to be 25 Å. This is sufficient to reduce spurious fields in the vacuum region which can arise from a repeated configuration with different polarities and might artificially influence the surface reconstructions. Numerical tests for polar surfaces have shown that the change in the surface energies is smaller than 2 meV/Å^2 if another monolayer is added and smaller than 0.25 meV/Å^2 if the vacuum region is doubled thus showing that our model is converged with respect to the slab thickness.

For calculations of the formation energies the lateral dimensions of the surfaces are chosen sufficiently large (Fig. D.1) to justify the gamma point approximation implemented in the *SCC–DFTB* method. The specific lateral dimensions chosen depend on the surface orientation and the periodicity to be examined and will be given in the corresponding application chapters.

For a selected set of surfaces band structure calculations are performed. In contrast to geometry optimisations and calculations of the surface formation energy the smallest lateral real–space unit cell possible for the specific reconstruction (for an example see Fig. D.1) is used in order to obtain a large Brillouin zone in reciprocal–space.

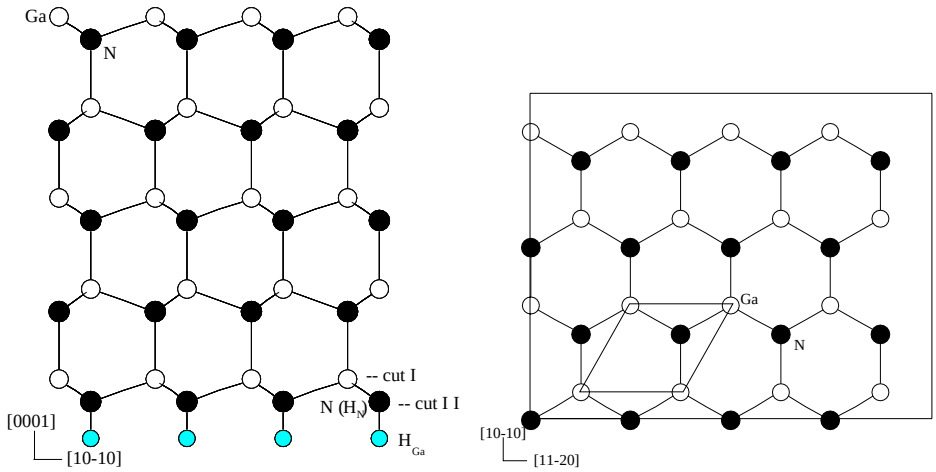

Figure D.1: Side view (along $[11\bar{2}0]$) and top view of the GaN (0001) surface. The bottom of the slab is saturated with pseudo–hydrogen of charge 3/4. Formation energies are evaluated within the gamma–point approximation in a large supercell in order to compare different reconstructions. Band structure calculations are performed within the smallest possible real space supercell so that a large reciprocal space Brillouin zone is obtained.

D.2 Point defects

Within the *SCC–DFTB* method we use supercells to model point defects. This allows to calculate formation energies and compared with modelling in a cluster avoids the interaction between the defect and the cluster surface. The inconvenient of a supercell modelling is that one does not describe an isolated point defect but a repeated sequence of defects. This could give rise to a non–negligible dispersion of the defect levels due to defect–defect interaction. The simplest way to avoid this would be to increase the size of the supercells and thus the distance between the point defects. Computationally expensive methods, as e.g. plane wave schemes, make use of the fact that an average of the defect band is very close to the level of the isolated defect. Employing special k–points one performs an integration over the Brillouin zone and thus achieves the desired averaging even within a small 32 atom supercell [8]. *SCC-DFTB* offers the possibility to treat large systems. We therefore take the approach of using large supercells of 128 atoms to overcome the defect–defect interaction problem. Another critical problem is caused by the Coulomb interaction arising if charged defects are modelled in supercells which are periodically repeated. This is usually solved by introducing a compensating background charge. As explained in chapter 4 *SCC-DFTB* does this automatically by making use of the Ewald–summation technique for the evaluation of the Coulomb sums.

Within the *AIMPRO* method clusters are used to describe point defects. The advantage of modelling point defects in clusters over a modelling in periodic systems is that in clusters the defect–defect interaction mentioned above is avoided. Unfortunately, this advantage is in general compensated by a possible interaction of the defect with the cluster surface if the cluster is not chosen very large. Of course, as in the case for the bottom of the surface slabs also the surfaces of the clusters should be saturated with pseudo–hydrogens.

D.3 Line defects

Line defects are defects which are periodic in only one dimension. The most interesting type of line defects are dislocations which represent the boundary of a region where slip between adjacent atomic planes has taken place. Thus, a single dislocation must either be a closed loop within the crystal, or terminate on the surface at both ends. The displacement is given by the Burgers vector **b**. Away from the dislocation line the crystal is locally only negligibly different from the perfect crystal, and near the line the atomic positions are substantially different from the original crystalline sites. The resulting strain pattern is that of the dislocation characterised jointly by its path through the crystal and the Burgers vector **b**. In a crystal the Burgers vector must generally be equal to a lattice vector in order to maintain the crystallinity of the material. Such dislocations are called perfect.

Dislocations are characterised by the angle β between the dislocation line and the Burgers vector, see figure D.2. Special cases are edge dislocations where the Burgers vector **b** is perpendicular to the dislocation line ($\beta = 90°$), and screw dislocations where **b** is parallel to the line ($\beta = 0°$). In wurtzite GaN the most interesting dislocations have dislocation lines parallel to the growth direction **c** and are therefore called *threading dislocations*. See Fig. D.2 for an example.

In this work the initial positions of atoms are determined via linear elasticity theory. If the atoms are embedded in an infinite, continuous, and isotropic medium, then a straight dislocation will displace an atom at (x, y, z) by

$$\begin{aligned} u_x &= \frac{b_e}{2\pi}\left(\tan^{-1}\frac{y}{x} + \frac{xy}{2(1-\nu)(x^2+y^2)}\right) \\ u_y &= -\frac{b_e}{2\pi}\left(\frac{1-2\nu}{4(1-\nu)}\ln(x^2+y^2) + \frac{x^2-y^2}{4(1-\nu)(x^2+y^2)}\right) \\ u_z &= \frac{b_s}{2\pi}\tan^{-1}\frac{y}{x}, \end{aligned} \tag{D.1}$$

where ν is the Poisson's ratio of the medium, and b_s, b_e are the screw and edge components of the Burgers vector, respectively. The structures are then optimised using *AIMPRO* or *SCC–DFTB* in the conjugate gradient technique.

In order to evaluate line energies (see equation (5.15)) it is suitable to use *SCC-DFTB* in the supercell approach. To this end two dislocations with opposite Burgers vector are put together. This results in a vanishing Burgers vector and thus perfect crystalline periodicity. The arrangement can therefore be modelled within a supercell (for an example see Fig. D.2). The dimensions of the supercells used in this work will be given in the corresponding application chapters.

A modelling of dislocations in a cluster like configuration is especially useful if one wants to determine the geometries at the dislocation core and wants to exclude any artificial distortion arising from the interaction of the dislocations in the supercell. Within *SCC–DFTB* a very convenient model can be constructed by placing the dislocation into a cluster which is periodic along the dislocation line, i.e. along [0001] for threading dislocations in wurtzite GaN. The dangling bonds at the sides of these one dimensionally periodic clusters are saturated with pseudo–hydrogen as described in section D.1. Since the current version of *AIMPRO* does not support periodic boundary conditions full clusters have to be used to represent dislocations.

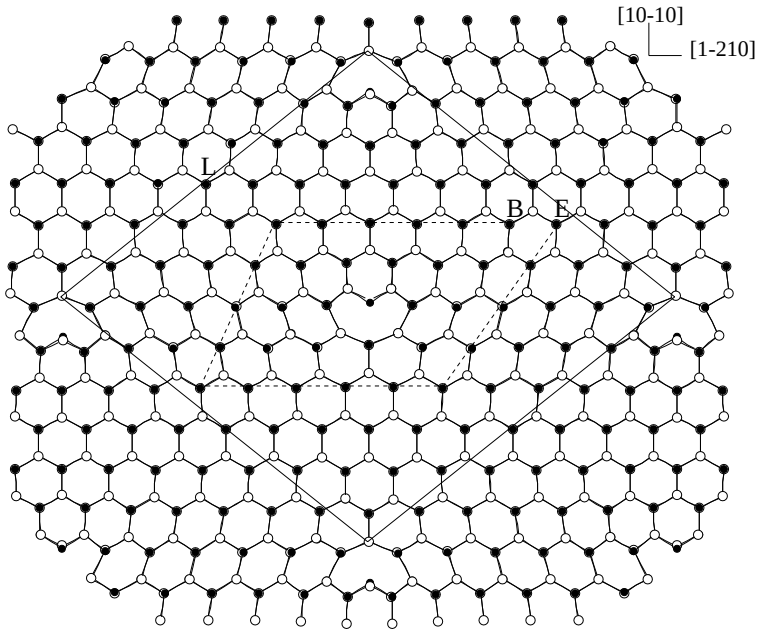

Figure D.2: View along the dislocation line ([0001]) of a wurtzite supercell containing a dipole of threading edge dislocations with Burgers vector $\mathbf{b} = \mathbf{BE} = \pm\frac{1}{3}[1\bar{2}10]$. The cell contains 312 atoms and has a periodicity of [0001] along the dislocation line. Cutting out one dislocation and saturating the dangling bonds with pseudo–hydrogens gives a cluster periodic along the dislocation line (see text). Position L is used as a bulk reference position to model point defects (see section 7.4).

D.4 Domain boundaries

In this thesis domain boundaries are investigated by the *SCC-DFTB* method using supercells with two domain boundaries per cell. The supercells were chosen sufficiently large to reduce the interaction between the domain boundaries and to justify the gamma point approximation (see Fig. D.3).

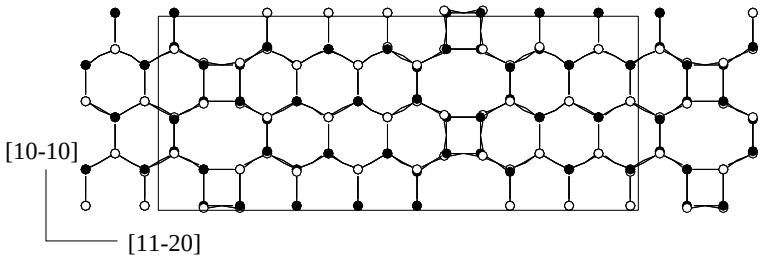

Figure D.3: Top view (in [0001]) of a supercell modelling a dipole of domain boundaries.

Bibliography

[1] S. Nakamura and G. Fasol, *The Blue Laser Diode*, Springer Verlag, Berlin 1997.

[2] K. Koga and T. Yamaguchi, Prog. Cryst. Growth and Charact. **23**, 127 (1991).

[3] J. Edmond, H. Kong and V. Dmitrieve, Inst. Phys. Conf. Series **137**, 515 (1994).

[4] H. Amano, M. Kito, K. Hiramatsu and I. Akasaki, Inst. Phys. Conf. Series **106**, 725 (1989).

[5] S. Nakamura, T. Mukai, M. Senoh and N. Iwasa, Jpn. J. of Appl. Phys. **28**, L2112 (1992).

[6] I. Akasaki, H. Amano, Y. Koide, K. Hiramatsu and N. Sawaki, J. Cryst. Growth **98**, 209 (1989).

[7] C. Trager–Cowan, P.G. Middleton and K.P. O'Donell, MRS Internet J. Nitride Semic. Res. **1** (1996).

[8] J. Neugebauer and C.G. Van de Walle, Festkörperprobleme **35**, 25 (1996).

[9] T. Mattila and R.M. Nieminen, Phys. Rev. B **55**, 9571 (1997).

[10] K. Saarinen, T. Laine, S. Kuisma, J. Nissilä, P. Hautojärvi, L. Dobrzynski, J.M. Baranowski, K. Pakula, R. Stepniewski, M. Wojdak, A. Wysmolek, T. Suski, M. Leszcynski, I. Grzegoryy and S. Porowski, Phys. Rev. Lett. **79**, 3030 (1997).

[11] C. Wetzel, T. Suski, J. W. Ager, E. R. Weber, E. E. Haller, S. Fischer, B. K. Meyer, R. J. Molnar and P. Perlin, Phys. Rev. Lett. **78**, 3923 (1997).

[12] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[13] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[14] R. M. Dreizler and E. K. U. Gross, *Density Functional Theory*, Springer Verlag, Berlin 1990.

[15] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[16] See for example J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).

[17] L. Hedin, Phys. Rev. **139**, A796 (1965).

[18] R. Delsole, L. Reining and R. W. Godby, Phys. Rev. B **49**, 8024 (1994).

[19] R. Jones and P.R. Briddon, in *The ab initio cluster method and the dynamics of defects in semiconductors*, Semiconductors and Semimetals: Identification of Defects in Semiconductors, edited by M. Stavola, Academic Press, New York 1997.

[20] G. B. Bachelet, D. R. Hamann and M. Schlüter, Phys. Rev., **B26**, 4199 (1982).

[21] D. Behr, J. Wagner, J. Schneider, H. Amano and I. Akasaki, Appl. Phys. Lett. **68**, 2404 (1996).

[22] K. Laasonen and R. M. Nieminen, J. Phys. C Condens. Matter **2**, 1509 (1990).

[23] C. H. Xu, C. Z. Wang, C. T. Chan and K. M. Ho, J. Phys. C Condens. Matter **4**, 6047 (1992).

[24] M. Menon and K. R. Subbaswamy, Phys. Rev. B **47**, 12754 (1993).

[25] G. Seifert, H. Eschrig and W. Bieger, Z. Phys. Chemie *(Leipzig)* **267**, 529 (1986).

[26] H. Eschrig, *Optimized LCAO Method and the Electronic Structure of Extended Systems*, Akademie-Verlag, Berlin 1988.

[27] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Shuhai and G. Seifert, Phys. Rev. B **58**, 7260 (1998).

[28] M. Haugk, J. Elsner, Th. Frauenheim, Th. Heine and G. Seiffert, J. Comp. Material Science in press (1998).

[29] W. Foulkes and R. Haydock, Phys. Rev. B **39**, 12520 (1989).

[30] D. Porezag, *Development of ab initio and approximate density functional methods and their application to complex fullerende systems*, PhD thesis at Fakultät für Naturwissenschaften at Technische Universität Chemnitz–Zwickau 1997.

[31] D. Porezag, Th. Frauenheim, Th. Köhler, R. Kaschner and G. Seifert, Phys. Rev. B **51**, 12947 (1995).

[32] J. Elsner, M. Haugk, G. Jungnickel and Th. Frauenheim, J. Mater. Chem. **6**, 1649 (1996).

[33] A. Sieck, D. Porezag, Th. Frauenheim, M. R. Pederson and K. Jackson, Phys. Rev. A **56**, 4890 (1997).

[34] D. Porezag, M. R. Pederson, Th. Frauenheim and T. Köhler, Phys. Rev. B **52**, 14963 (1997).

[35] D. Porezag, G. Jungnickel, Th. Frauenheim, G. Seifert, A. Ayuela and M. R. Pederson, Appl. Phys. A **64**, 321 (1997).

[36] M. Haugk, J. Elsner and Th. Frauenheim, J. Phys. C Condens. Matter **9**, 7305 (1997).

[37] M. Sternberg, W. R. L. Lambrecht, Th. Frauenheim, Phys. Rev. B **56**, 1568 (1997).

[38] R. Pariser, J. Chem. Phys. **24**, 125 (1956)

[39] R. G. Parr and R. G. Pearson, J. Am. Chem. Soc. **105**, 1503 (1983).

[40] M. J. S. Dewar and N. L. Hojvat, J. Chem. Phys. **34**, 1232 (1961).

[41] G. Klopman, J. Am. Chem. Soc. **28**, 4550 (1964).

[42] N. Mataga and K. Nishimoto, Z. Phys. Chemie (Frankfurt) **13**, 140 (1957).

[43] D.J. Chadi, J. Vac. Sci. Technol. A **5**, 834 (1987).

[44] J.E. Northrup and S. Froyen, Phys. Rev. B **50**, 2015 (1994).

[45] T. Hashizume, Q. K. Xue, J. M. Zhou, A. Ichimiya and T. Sakurai, Phys. Rev. Lett. **73**, 2208 (1994).

[46] J. E. Northrup and J. Neugebauer, Phys. Rev. B **53**, 10477 (1996).

[47] A. Kley, *Theoretische Untersuchungen zur Adatomdiffusion auf niederindizierten Oberflächen von GaAs*, PhD thesis at Fachbereich Physik at Technische Universität Berlin 1997.

[48] G.-X. Qian, R. M. Martin and D. J. Chadi, Phys. Rev. B **38**, 7649 (1988).

[49] *CRC Handbook of Chemistry and Physics*, 67 ed., edited by R.C. West, CRC, Boca Raton FL 1986.

[50] C. Kittel, *Thermal Physics*, Wiley, New York 1969.

[51] J.E. Northrup, R. Di Felice and J. Neugebauer, Phys. Rev. B **56**, R4325 (1997).

[52] J. Neugebauer and C. Van de Walle, Appl. Phys. Lett. **69**, 503 (1996).

[53] N. Moll, A. Kley, E. Pehlke and M. Scheffler, Phys. Rev. B **54**, 8844 (1996).

[54] Z. Liliental–Weber, Y. Chen, S. Ruvimov and J. Washburn, Mat. Science Forum **258–263**, 1659 (1997).

[55] P. Pirouz, Phil. Mag. A **78**, 727 (1998).

[56] N. Chetty and R. M. Martin, Phys. Rev. B **45**, 6074 (1992).

[57] N. Chetty and R. M. Martin, Phys. Rev. B **45**, 6089 (1992).

[58] K. Rapcewicz, B. Chen, B. Yakobson and J. Bernholc, Phys. Rev. B **57**, 7281 (1998).

[59] M. Haugk, *Theoretische Untersuchung von komplexen Strukturen in GaAs* to be submitted as PhD thesis at Fakultät für Physik at Universität–Gesamthochschule Paderborn 1999.

[60] C. Messmer and J.C. Bilello, J. Appl. Phys. **52**, 4623 (1981).

[61] G.-X. Qian, R. M. Martin and D. J. Chadi, Phys. Rev. B **37** 1303 (1988).

[62] S. Y. Tong, G. Xu and W. N. Mei, Phys. Rev. Lett. **52**, 1693 (1984).

[63] D. J. Chadi, Phys. Rev. B **19**, 2074 (1978).

[64] S. Y. Tong, A. R. Lubinsky, B. J. Mrstik and M. A. Van Hove, Phys. Rev. B **17**, 3303 (1978).

[65] J.E. Northrup, Phys. Rev. B **44**, 1349 (1991).

[66] M. Haugk, J. Elsner, M. Sternberg and Th. Frauenheim, J. Phys. C Condens. Matter **10**, 4523 (1998).

[67] J. E. Jaffe, R. Pandey and P. Zapol, Phys. Rev. B **53**, R4209 (1996).

[68] J. Elsner, M. Haugk and Th. Frauenheim, MRS Conference Proc. **423**, 421 (1996).

[69] Z. Liliental–Weber, Y. Chen, S. Ruvimov and J. Washburn, Phys. Rev. Lett. **79**, 2835 (1997).

[70] S. Christiansen, M. Albrecht, W. Dorsch, H.P. Strunk, C. Zanotti-Fregonara, G. Salviati, A. Pelzmann, M. Mayer, M. Kamp and K. J. Ebeling, MRS Internet J. of Nitride Res. **1** (1996).

[71] X. H. Wu, L. M. Brown, D. Kapolnek, S. Keller, B. Keller, S. P. DenBaars and S. J. Speck, J. Appl. Phys. **80**, 3228 (1996).

[72] X.J. Ning, F.R. Chien and P. Pirouz, J. Mater. Res. **11**, 580 (1996).

[73] F.A. Ponce, D. Cherns, W.T. Young and J.W. Steeds, Appl. Phys. Lett. **69**, 770 (1996).

[74] S. D Lester, F. A. Ponce, M. G. Cranford and D. A. Steigerwald, Appl. Phys. Lett. **66**, 1249 (1996).

[75] L. Sugiura, J. Appl. Phys. **81**, 1633 (1997).

[76] S. Nakamura, T. Mukai, and M. Senoh, J. Appl. Phys. **76**, 8189 (1994).

[77] S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyohu and Y. Sguimoto, Jpn. J. Appl. Phys. **35**, L74 (1996).

[78] K. Maeda and S. Takeuchi, J. Physique **44**, C4-375 (1983).

[79] W. Qian, G.S. Rohrer, M. Skowronski, K. Doverspike, L. B. Rowland and D. K. Gaskill, Appl. Phys. Lett. **67**, 2284 (1995).

[80] W. Qian, M. Skowronski, K. Doverspike, L. B. Rowland and D. K. Gaskill, J. Cryst. Growth **151**, 396 (1995).

[81] P. Vennégues, B. Beaumont, M. Vaille and P. Gibart, Appl. Phys. Lett. **70**, 2434 (1997).

[82] F. C. Frank, Acta. Crys. **4**, 497 (1951).

[83] T. Ogino and M. Aoki, Jpn. J. Appl. Phys. **19**, 2395 (1980)..

[84] E. R. Glaser *et al.*, Phys. Rev. B **51**, 13326 (1995).

[85] M. Godlewski *et al.*, Mat. Sci. Forum **258–263**, 1149 (1997).

[86] S. J. Rosner, E. C. Carr, M. J. Ludowise, G. Girlami and H. I. Erikson, Appl. Phys. Lett. **70**, 420 (1997).

[87] F. A. Ponce, D. B. Bour, W. Götz and P. J. Wright, Appl. Phys. Lett. **68**, 57 (1996). Recent combined cathodoluminescence and atomic force microscopy points to threading dislocations as as a source for these centres. [86]

[88] J. Elsner, R. Jones, P. K. Sitch, V. D. Porezag, M. Elstner, Th. Frauenheim, M. I. Heggie, S. Öberg and P.R. Briddon, Phys. Rev. Lett. **79**, 3672 (1997).

[89] J. Elsner, R. Jones, M. I. Heggie, M. Haugk, Th. Frauenheim, S. Öberg and P. R. Briddon, Phil. Mag. Lett. in press (1998).

[90] J. Elsner, R. Jones, M. Haugk, R. Gutierrez, M. Heggie, Th. Frauenheim, S. Öberg and P. R. Briddon, Appl. Phys. Lett. **73**, 3530 (1998).

[91] J. Elsner, R. Jones, M. I. Heggie, P. K. Sitch, M. Haugk, Th. Frauenheim, S. Öberg and P. R. Briddon, Phys. Rev. B **58**, 12571 (1998).

[92] D. Cherns, W. T. Young, J. W. Steeds, F. A. Ponce and S. Nakamura, J. Cryst. Growth **178**, 201 (1997).

[93] Y. Xin, S. J. Pennycook, N. D. Browning, P. D. Nelist, S. Sivananthan, F. Omnès, B. Beaumont, J.–P. Faurie and P. Gibart, Appl. Phys. Lett. **72**, 2680 (1998).

[94] Z. Liliental–Weber, private communications at the EDS 1998.

[95] D. Cherns, private Communications, Bristol 1997.

[96] Y. Xin, S. J. Pennycook, N. D. Browning, P. D. Nellist, S. Sivananthan, J.-P. Faurie and P. Gibart, *Nitride Semicond.* **482** 781, edited by F.A. Ponce, S.P. Den Baars, B.K. Meyer, S. Nakamura, S. Strite, Mat. Res. Soc., Pennsylvania (1998).

[97] V. A. Savastenko and A. U. Shelog, Phys. Status Solidi A **48**, K135 (1978).

[98] C. Wetzel *et al.*, Proc. ICPS 23 (World Scientific, Singapore), 2929 (1996).

[99] P. Perlin *et al.*, Proc. Mater. Res. Soc. Symp. **449**, 519 (1997).

[100] C. G. Van de Walle and J. Neugebauer, Mater. Sci. Forum **258–263**, 19 (1997).

[101] R. A. Youngman and J. H. Harris, J. Am. Ceram. Soc. **73**, 3238 (1990).

[102] A. D. Westwood *et al.*, J. Mater. Res. **10**, 1270 (1995).

[103] S. Nakamura, *Nitride Semiconductors* **482** 1145, edited by F.A. Ponce, S.P. Den Baars, B.K. Meyer, S. Nakamura, S. Strite, Mat. Res. Soc., Pennsylvania (1998).

[104] Z. Sitar, M. J. Paisley, B. Yan, and R. F. Davis, Mater. Res. Soc. Symp. Proc. **162**, 537 (1990).

[105] S. Tanaka, R. Scott Kern, and R. F. Davis, Appl. Phys. Lett. **66**, 37 (1995).

[106] D. J. Smith, D. Chandrasekhar, B. Sverdlov, A. Botchkarev, A. Salvador, and H. Morkoc, Appl. Phys. Lett. **67**, 1830 (1995).

[107] B. N. Sverdlov, G. A. Martin, H. Morkoc, and D. J. Smith, Appl. Phys. Lett. **67**, 2063 (1995).

[108] J.–L. Rouvière, M. Arlery, A. Bourret, R. Niebuhr, and K. Bachem, Inst. Phys. Conf. Series **146**, 285 (1995).

[109] Y. Xin, P. D. Brown and C.J. Humphreys, Appl. Phys. Lett. **70**, 1308 (1997).

[110] J. E. Northrup, J. Neugebauer and L. T. Romano, Phys. Rev. Lett. **77**, 103 (1996).

[111] J. Elsner, M. Kaukonen, M. I. Heggie, M. Haugk, Th. Frauenheim and R. Jones, Phys. Rev. B **58**, 15347 (1998).

[112] C.M. Drum, Phil. Mag. A **11**, 313 (1964).

[113] M.K.H. Natusch, G.A. Botton, R.F. Broom, P.D. Brown, D.M. Tricker and C.J. Humphreys, *Nitride Semiconductors* **482** 763, edited by F.A. Ponce, S.P. Den Baars, B.K. Meyer, S. Nakamura, S. Strite, Mat. Res. Soc., Pennsylvania (1998).

[114] M. M. Sung, J. Ahn, V. Bykov, J. W. Raballais, D. D. Koleske, and A. E. Wickenden, Phys. Rev. B **54**, 14652 (1996).

[115] K. Iwata, H. Asahi, S. Jae Yu, K. Asami, H. Fujita, M. Fushida and S. Gonda, Jpn. J. Appl. Phys. **35**, L289 (1996).

[116] P. Hacke, G. Feuillet, H. Okumura and S. Yoshida, Appl. Phys. Lett. **69**, 2507 (1996).

[117] P. Hacke, G. Feuillet, H. Okumura and S. Yoshida, J. Crystal Growth **175**, 94 (1997).

[118] A. R. Smith, R. M. Feenstra, D. W. Greve, J. Neugebauer and J. E. Northrup, Phys. Rev. Lett. **79**, 3934 (1997).

[119] 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).

[120] J. Elsner, M. Haugk, G. Jungnickel and Th. Frauenheim, Sol. State Comm. **106**, 739 (1998).

[121] J.E. Northrup, R. Di Felice and J. Neugebauer, Phys. Rev. B **55**, 13878 (1997).

[122] K. Rapcewicz, M. Buongiorno Nardelli and J. Bernholc, Phys. Rev. B **56**, 12725 (1997).

[123] K. Prabhakaran, T. G. Andersson and K. Nozawa, Appl. Phys. Lett. **69**, 3212 (1996).

[124] V. M. Bermudez, J. Appl. Phys. **80**, 1190 (1996).

[125] J. Elsner, R. Gutierrez, B. Hourahine, M. Haugk, Th. Frauenheim and R. Jones, Sol. State Comm. **108**, 953 (1998).

[126] V. M. Bermudez, T. M. Jung, K. Doverspike and A. E. Wickenden, J. Appl. Phys. **79**, 110 (1996).

[127] Diffusion processes in bulk GaN are currently being investigated in our group (Prof. Th. Frauenheim, Uni Paderborn).

[128] Diffusion processes on GaN surfaces are currently being investigated by T. Zywietz and J. Neugebauer (FHI der MPG Berlin).

[129] J. M. Ziman, *Prinzipien der Festkörpertheorie*, Verlag Harri Deutsch, Zürich und Frankfurt am Main 1992.

[130] J. C. Slater, *Insulators, Semiconductors and Metals*, Mc Graw Hill, New York 1967.

[131] F. D. Murnaghan, Proc. Natl. Acad. Sci. USA **30**, 244 (1944).

[132] All electron code developed by M. Pederson, K. Jackson, D. Porezag and A. Sieck.

[133] K. Shiraishi, J. Phys. Soc. Jap. **59**, 3455 (1990).

Danksagung

An dieser Stelle möchte ich all jenen danken, die mir bei der Arbeit an dieser Dissertation mit Rat und Tat zur Seite standen. Mein Dank gilt insbesondere Prof. Dr. Thomas Frauenheim, Dr. Bob Jones und Dr. Malcolm Heggie, die sich für meine Arbeit sehr interessiert und sie mit Inspiration betreut haben. Intensive Unterstützung erfuhr ich auch von der fruchtbaren und freundschaftlichen Zusammenarbeit mit Michael Haugk. Natürlich dürfen die Hilfe und Diskussionen mit allen Mitgliedern der DFTB Gruppe in Paderborn und Chemnitz nicht vergessen werden. Insbesondere nennen möchte ich Paul Sitch, Alexander Sieck, Gerd Jungnickel, Dirk Porezag, Rafael Gutierrez, Silke Uhlmann, Markus Elstner, Zoltan Hajnal und Jörg Widany. Die engagierten Diskussionen mit den genannten haben dieser Arbeit gut getan. Michael Sternberg sei an dieser Stelle für seinen grossen Einsatz in der Systemverwaltung gedankt. Prof. Dr. M. Schreiber danke ich für die wohlwollende Förderung dieser Arbeit an seinem Lehrstuhl und die Bereitstellung von Arbeitsmitteln.

Während meines Aufenthalts in Exeter möchte ich der AIMPRO Gruppe, Paul Leary (PL abroad), Ben Hourahine, Chris Latham, Chris Ewels, Antonio Resende, Patrick Briddon, Jonathan Goss und James Buttler sowie allen weiteren des Department of Physics für die fachlichen Ratschläge und die gute Unterhaltung danken. Stellvertretend seien hier J. P. Srivastava, Simon Gay, Mehmet Kaknak, Julia Gasson, Caroline Hoad, Mauro Boero und Hussein genannt.

Auch Mitglieder anderer Arbeitsgruppen haben sehr zu dieser Arbeit beigetragen. Insbesondere seien hier Michael Natusch, Alexander Kley, Markus Kaukonen, Yan Xin, Zusanna Liliental–Weber, Tosja Zywietz, Jörg Neugebauer, Sven Öberg und Niclas Lehto erwähnt.

Schliesslich möchte ich Stefan Matzig und meinem Vater für das Lesen meiner Dissertation und Thomas Köhler für die Einarbeitung zu Beginn danken.